

1 ***Vibrio cholerae* pathogenicity island 2 encodes two distinct types of restriction**  
2 **systems**

3

4 Grazia Vizzarro<sup>1</sup>, Alexandre Lemopoulos<sup>1</sup>, David William Adams<sup>1#</sup>, Melanie  
5 Blokesch<sup>1#</sup>

6

7 <sup>1</sup>Laboratory of Molecular Microbiology, Global Health Institute, School of Life  
8 Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne,  
9 Switzerland

10

11 Running Title: VPI-2 encodes two restriction systems

12

13 #Address correspondence to David W. Adams ([david.adams@epfl.ch](mailto:david.adams@epfl.ch)) and Melanie  
14 Blokesch ([melanie.blokesch@epfl.ch](mailto:melanie.blokesch@epfl.ch))

15

16 Keywords: *V. cholerae*, pathogenicity island, restriction systems, DNA modification

17

## 18 **Abstract**

19 In response to predation by bacteriophages and invasion by other mobile genetic  
20 elements such as plasmids, bacteria have evolved specialised defence systems that  
21 are often clustered together on genomic islands. The O1 El Tor strains of *Vibrio*  
22 *cholerae* responsible for the ongoing seventh cholera pandemic (7PET) contain a  
23 characteristic set of genomic islands involved in host colonisation and disease, many  
24 of which contain defence systems. Notably, *Vibrio* pathogenicity island 2 contains  
25 several characterised defence systems as well as a putative Type I restriction-  
26 modification system (T1RM), which, interestingly, is interrupted by two genes of  
27 unknown function. Here, we demonstrate that the T1RM system is active, methylates  
28 the host genomes of a representative set of 7PET strains, and identify a specific  
29 recognition sequence that targets non-methylated plasmids for restriction. We go on  
30 to show that the two genes embedded within the T1RM system encode a novel two-  
31 protein modification-dependent restriction system related to the GmrSD family of  
32 Type IV restriction enzymes. Indeed, we show that this system has potent anti-phage  
33 activity against diverse members of the *Tevenvirinae*, a subfamily of bacteriophages  
34 with hypermodified genomes. Taken together these results expand our  
35 understanding of how this highly conserved genomic island contributes to the  
36 defence of pandemic *V. cholerae* against foreign DNA.

37

## 38 **IMPORTANCE**

39 Bacterial defence systems are specialised immunity systems that allow bacteria to  
40 counter the threat posed by bacterial viruses (bacteriophages) and other invasive  
41 mobile genetic elements such as plasmids. Although these systems are numerous  
42 and highly diverse, the most common types are restriction enzymes that can

43 specifically recognise and degrade non-self DNA. In this work we show that the  
44 *Vibrio* pathogenicity island 2, present in the human pathogen *Vibrio cholerae*,  
45 encodes two types of restriction systems that use distinct mechanisms to sense non-  
46 self DNA. The first is a classical Type I restriction-modification system that  
47 recognises specific DNA sequences, which are protected in the host genome by  
48 methylation. The second, is a novel modification-dependent Type IV restriction  
49 system that recognises hypermodified cytosines present in certain bacteriophage  
50 genomes. Curiously, these systems are embedded one within the other, forming a  
51 single cluster, suggesting that the systems collaborate to create a multi-layered  
52 defence system.

## 53 **Introduction**

54 Mobile genetic elements (MGEs) such as plasmids, transposons and integrative-  
55 conjugative elements can confer significant fitness advantages by facilitating the  
56 transfer of beneficial traits to the host bacterium, including key virulence factors or  
57 antibiotic-resistant genes (1). However, their maintenance and or activity can also  
58 impose a metabolic burden on the host cell, while elements that integrate on the  
59 chromosome have the potential to disrupt important genomic features (2).  
60 Furthermore, the replication of some MGEs results in the death of the host cell (3).  
61 Indeed, predation by lytic bacteriophages (phages), which are ubiquitous bacterial  
62 viruses, has imposed a strong evolutionary pressure to develop multiple lines of  
63 defence against these elements, including a vast array of specialised defence  
64 systems (4, 5).

65       Upon recognising an infection, these systems can either respond directly by  
66 degrading the invading non-self-DNA and thus provide individual level protection, or  
67 alternatively can sacrifice the host cell prior to phage induced lysis to protect the  
68 surrounding population (abortive infection, Abi) (6). The most common and best-  
69 studied defence systems are Restriction-modification (RM) systems, which use  
70 restriction enzymes to directly degrade non-self DNA (5). Type I-III RM systems are  
71 modification-blocked enzymes that recognise specific DNA sequences and only cut  
72 DNA when it is unmodified, while the corresponding sequences in the host genome  
73 are protected by epigenetic modification with a cognate methylase (7–9). In contrast,  
74 Type IV systems are modification-dependent enzymes that can recognise and  
75 degrade invading DNA with specific modifications, which are used by some phages  
76 to avoid restriction by modification-blocked systems (9, 10).

77           Diverse defence systems, including RM systems, tend to cluster together  
78 within genomic islands known as "defence islands" (11, 12). This pattern also applies  
79 to the defence systems identified so far in *Vibrio cholerae*, the causative agent of  
80 cholera. This bacterium features specialized islands crucial to its pathogenic  
81 evolution. Indeed, only certain *V. cholerae* strains, referred to as toxigenic isolates,  
82 can cause cholera. This ability is due to the presence of two key  
83 virulence/colonization factors: the cholera toxin (CT) and toxin-coregulated pilus  
84 (TCP), encoded on the filamentous phage CTX $\Phi$  and the *Vibrio* pathogenicity island  
85 1 (VPI-1), respectively (13–16). The ongoing seventh cholera pandemic is caused by  
86 the O1 El Tor *V. cholerae* lineage (7PET), which uniquely carries the *Vibrio* seventh  
87 pandemic islands I and II (VSP-I and VSP-II), characteristic of the 7PET strains (17,  
88 18). These genomic islands are implicated in defence, as they encode for instance  
89 CBASS and AvcD systems (VSP-I) and the Lamassu system DdmABC on VSP-II  
90 (19–23). Additionally, toxigenic *V. cholerae* strains carry the *Vibrio* pathogenicity  
91 island 2 (VPI-2), which is believed to enhance pathogenicity by giving the pathogen a  
92 competitive advantage in using sialic acid as a carbon source during gut colonization  
93 (24–26). This capability is encoded within the island's *nan-nag* genomic region (24–  
94 26). Moreover, the island houses several genes believed to protect against MGEs,  
95 including (i) a predicted Zorya Type I system, a phage defence system identified  
96 across a wide range of bacterial genomes and experimentally studied primarily  
97 through *Escherichia coli* homologs (27, 28); (ii) the DNA defence module DdmDE  
98 that targets and degrades small multicopy plasmids (23); and (iii) a gene  
99 cluster/operon predicted to encode a Type I restriction-modification (T1RM) system  
100 (24). The presence of both predicted and established defence systems encoded  
101 within VPI-2 suggests that it may serve as a genuine defence island.

102           In this study we set out to characterize the predicted T1RM operon within VPI-  
103 2. We show that the T1RM system promotes methylation of the genomes of 7PET *V.*  
104 *cholerae* strains, and identify a specific recognition sequence that can target non-self  
105 derived plasmids for restriction. Furthermore, we identify two genes embedded within  
106 the T1RM operon that form a novel modification-dependent restriction system related  
107 to the GmrSD family of Type IV restriction enzymes, which we term TgvAB. When  
108 produced in *E. coli* this system has potent anti-phage activity against phages with  
109 hypermodified genomes. Collectively, these findings enhance our understanding of  
110 how this highly conserved genomic island contributes to the defence of pandemic *V.*  
111 *cholerae* against foreign DNA.

112

## 113 **Results and Discussion**

### 114 ***In silico* analysis of VPI-2 and the T1RM cluster**

115 Although VPI-2 was discovered over 20 years ago (24), the genes it carries have not  
116 yet been fully characterized. To begin bridging this knowledge gap, we started by re-  
117 evaluating the conservation of the island. Consistent with earlier findings (24), this  
118 revealed that VPI-2 is highly conserved among a set of 7PET O1 strains isolated  
119 between 1975 and 2011. Our analysis confirmed the island's modular structure as  
120 described by Jermyn and Boyd (24, 29), including a predicted Zorya system (27)  
121 encoded by genes *VC1761-64* (as per reference strain N16961; (30)), a predicted  
122 T1RM system (*VC1765-69*), the DdmDE defence module (*VC1770-71*) (23), the *nan-*  
123 *nag* sialic acid utilisation cluster (*VC1773-1784*), and a region with phage-like  
124 properties (*VC1791-1809*) (24) (Fig. 1a). Notably, O139-serogroup strains such as  
125 MO10 carry a highly truncated version of VPI-2 that retains only the phage-like  
126 region (24, 31) (Fig. 1a).

127           Given the observed conservation of VPI-2 and the presence of established  
128 defence systems, we explored the possibility that the putative T1RM system was  
129 also actively involved in restricting foreign DNA. Interestingly, the previously  
130 annotated T1RM region sits within a five-gene operon, of which three genes encode  
131 homologs of the known T1RM components (Fig. 1b). These host-specificity  
132 determinant (*hsd*) genes encode: the specificity subunit HsdS, which recognises a  
133 specific DNA recognition sequence; the methylase subunit HsdM, which methylates  
134 (and therefore protects) the recognition sequences in the host genome; and HsdR,  
135 the restriction enzyme subunit, which upon encountering foreign DNA with an  
136 unmethylated recognition sequence translocates the flanking DNA and cleaves at  
137 variable distances from the recognition site (7, 32–34). These components function  
138 together as multi-subunit complexes capable of both methylating and restricting  
139 DNA. Importantly, restriction requires a pentameric complex of 2HsdR + 2HsdM +  
140 HsdS, and although HsdR is dispensable for methylation, HsdS is required for both  
141 activities (7, 32). Interestingly, two genes of unknown function are embedded within  
142 the T1RM cluster (*VC1766-67*; Fig. 1b), which we characterize in the subsequent  
143 sections below.

144

### 145 **Deciphering the recognition motif of VPI-2's T1RM system**

146 If the T1RM system is active in *V. cholerae* then we predicted that we should be able  
147 to detect a specific methylation signature that is absent in strains lacking this system.  
148 To test this hypothesis, we used SMRT PacBio whole-genome sequencing, which  
149 can detect the presence of various DNA modifications including methylation, to  
150 determine the methylomes of a selection of 7PET O1 serogroup strains (strains as in  
151 Fig. 1a), as well as those of control strains lacking the T1RM system (see methods)

152 (35–37). As shown in Figure 2a, this analysis revealed a unique 13-nucleotide motif  
153 with methylation marks located on the second nucleotide within the sequence  
154 GATGNNNNNNCTT (m6A: GATGNNNNNNCTT:2). Upon further examination, we  
155 discovered that this DNA motif is present in over 600 copies throughout the genome  
156 of each strain and is modified in nearly 100% of cases in all O1 serogroup strains,  
157 except DRC193A (Fig. 2a). This phenotype is likely explained by the interruption of  
158 *hsdS* in this strain by a IS256-like transposase gene (38) (Fig. 1b). Finally, and as  
159 expected, the O139 serogroup strain MO10, which is missing the T1RM-encoding  
160 region of VPI-2 (Fig. 1a), and both a VPI-2 and a VC1765-69-deficient deletion strain  
161 (Table S1) all lacked this particular methylation mark (Fig. 2a).

162

### 163 **The T1RM impairs plasmid acquisition**

164 Having identified the methylated recognition motif, we next tested the ability of this  
165 motif to target plasmids for restriction by the VPI-2 T1RM system. Serendipitously,  
166 we realised that the recognition sequence is present within the widely used  
167 gentamicin resistance cassette *aacC1*. We therefore created plasmid derivatives  
168 carrying *aacC1* either with the putative recognition sequence intact ( $P_{\text{motif}+}$ ) or with  
169 silent mutations that disrupt the nucleotide recognition sequence while preserving  
170 the protein coding sequence ( $P_{\text{motif}-}$ ) (Fig. 2b). We then purified these plasmids from  
171 *E. coli* and used them as substrates in an electroporation-based transformation  
172 assay to compare their transformation frequencies in various backgrounds. As  
173 shown in Figure 2b, transformation with plasmid  $P_{\text{motif}+}$  was below the detection limit  
174 in the WT background (strain A1552) even though transformants could readily be  
175 obtained with plasmid  $P_{\text{motif}-}$ . Furthermore, this disparity between the acquisition of



176 the two plasmids became even stronger in the absence of the DdmABC system (23)  
177 (Fig. 2b), which is known to target derivatives of this high-copy number plasmid (39).

178 To determine if the plasmid restriction was mediated by the T1RM system, we  
179 removed the corresponding operon from the *A1552ΔddmABC* background and then  
180 assessed the plasmid transformability of the resulting strain. As shown in Figure 2b,  
181 deletion of the five-gene restriction cluster on VPI-2 indeed led to the recovery of  
182  $P_{\text{motif+}}$  transformants. Moreover, the transformation difference between the  $P_{\text{motif+}}$  and  
183  $P_{\text{motif-}}$  plasmids was now no longer statistically significant. Consequently, we  
184 conclude that the T1RM system is active, that it methylates a specific recognition  
185 sequence, and that when this sequence is present on non-self DNA the acquisition  
186 of this non-methylated DNA is restricted.

187

### 188 **Genes embedded in the T1RM cluster protect against phages with modified** 189 **genomes.**

190 Type I restriction-modification systems are recognized for their important role in  
191 defending the cell against phage infection (40). Therefore, we aimed to investigate  
192 the ability of the entire RM cluster, including the two embedded genes, to protect  
193 against viral infections. However, given that commonly used *Vibrio* phages, such as  
194 ICP1, ICP2, and ICP3, are typically isolated using VPI-2-carrying 7PET strains as  
195 the host (for example, strain E7946 and its derivatives (41)), it is unlikely that any  
196 defence system encoded on VPI-2 would provide protection against these phages.  
197 Therefore, we engineered the *E. coli* strain MG1655 to carry an arabinose-inducible  
198 version of the entire five-gene RM cluster (*VC1765-69*), which was integrated into its  
199 chromosome. Utilizing this strain and a strain without the cluster as a control, we  
200 screened for protection against the BASEL collection, a recently established phage

201 collection that represents the natural diversity of *E. coli* phages (42). As shown in  
202 Figure 3a (and Fig. S1 for the data on the entire screen), we noted a reduction in the  
203 efficiency of plaquing of at least 1000-fold compared to the non-defence control upon  
204 infection with members of the *Tevenvirinae* subfamily. The *Tevenvirinae* subfamily is  
205 characterized by their unique cytosine modifications, which play a crucial role in their  
206 defence against RM systems like the T1RM (42). Specifically, *Tequatrovirus* group  
207 phages feature cytosines that are hydroxymethyl-glucosylated, while *Mosigviruses*  
208 possess cytosines that are hydroxymethyl-arabinosylated (10, 43).

209 To determine which part of the RM cluster was responsible for this protection,  
210 we created *E. coli* strains that either carried the T1RM cluster or just the embedded  
211 two-gene cluster independently. Strikingly, this revealed that the two-gene cluster  
212 alone was responsible for this protection (Fig. 3a and b). Furthermore, the two genes  
213 did not provide protection when expressed individually, indicating a necessity for  
214 their combined action to achieve the observed anti-phage activity (Fig. 3a). For  
215 reasons explained below, we named these two genes as **T**ype I-embedded **G**mrSD-  
216 like system of **V**PI-2, *tgvA* (VC1767) & *tgvB* (VC1766).

217 To dissect the underlying mechanism of anti-phage defence by TgvAB we  
218 monitored the growth kinetics of *E. coli* strains infected with increasing multiplicities  
219 of infection (MOI) for both *Tequatrovirus* (Fig. 3c) and *Mosigvirus* phages (Fig. 3d).  
220 As expected, cultures of the no system control strain grew and then lysed in an MOI-  
221 dependent manner (Fig. 3c-d). In contrast, TgvAB producing cultures infected with  
222 the *Tequatrovirus* Bas35 continued to grow at rates indistinguishable from those of  
223 the no phage control up to and including MOI 5, before being partially overcome at  
224 MOI 10 (Fig. 3c). This phenotype is consistent with TgvAB acting directly to target  
225 the invading phage. However, TgvAB producing cultures infected with either the

226 *Tequatrovirus* Bas40 or the *Mosigviruses* Bas46 and 47 all showed more variable  
227 levels of protection (Fig. 3c-d). Indeed, while protection was robust at MOI 0.2, at  
228 higher MOIs we observed growth inhibition and even partial lysis. Nevertheless,  
229 given that the cultures mostly continued to grow past the point at which they lysed in  
230 the no system control, together with the direct protection observed against Bas35 at  
231 all tested MOIs, we conclude that TgvAB likely also acts directly against these  
232 phages, but that they are better able to overwhelm the system at high MOI.

233

### 234 **The TgvAB defence system is a member of the GmrSD family of Type IV** 235 **restriction enzymes**

236 Bioinformatic analysis of the TgvAB system revealed that TgvB (VC1766) possesses  
237 two domains of unknown function (DUF), an N-terminal DUF262 domain and a C-  
238 terminal DUF1524 domain. In contrast, TgvA (VC1767) is predicted to carry only an  
239 N-terminal DUF262 domain (Fig. 4a-b). Interestingly, previous work by Machnicka *et*  
240 *al.* found that GmrS and GmrD proteins contain the DUF262 and DUF1524 domains,  
241 respectively, typically coming together to form GmrSD fusion proteins (44). Notably,  
242 the TgvB homolog from classical biotype *V. cholerae* (VC0395\_A1364) was also  
243 identified as a GmrSD homolog in this study (44). These double domain forms of  
244 GmrSD function as modification-dependent Type IV restriction enzymes, and are  
245 known to specifically recognise and cleave DNA containing sugar-modified hydroxy-  
246 methylcytosine. However, they exhibit no activity against unmodified DNA (44–47).  
247 Given that such modifications are typical of the *Tevenvirinae* (10) and the specific  
248 protective effect we observed against them (Fig. 3), this suggests that TgvAB may  
249 function in a similar manner. Importantly, and in contrast to classical single protein  
250 GmrSD such as Eco94GmrSD (Fig. 4a) (45), our phage infection assay revealed that

251 TgvA and TgvB cannot function independently, and that both proteins are required  
252 for anti-phage activity.

253 Machnicka *et al.*, showed that the predominant form of GmrSD is as a single  
254 multi-domain protein containing an N-terminal DUF262(GmrS) domain and a C-  
255 terminal DUF1524(GmrD) domain, separated by an alpha helical linker region (44).  
256 This domain organisation was subsequently confirmed by crystal structures of the  
257 related GmrSD family members BrxU, which also recognises and degrades DNA  
258 containing modified cytosines, and the phosphorothioate modification sensing  
259 enzyme SspE (48–50). Furthermore, biochemical experiments with these enzymes  
260 have shown that the N-terminal DUF262 likely functions as DNA modification sensor,  
261 and uses nucleotide binding and hydrolysis to regulate the activity of the C-terminal  
262 DUF1524, which functions as a nuclease to degrade non-self-DNA (48, 50).  
263 Strikingly, structural modelling of Eco94GmrSD and TgvAB using AlphaFold (51),  
264 revealed that TgvB is predicted to share a similar domain architecture, although in  
265 the case of TgvA, this similarity is limited to the N-terminal DUF262 domain (Fig. 4a-  
266 b). Moreover, the top hits in structural alignments of the TgvAB models were SspE  
267 and BrxU, reinforcing the idea that these proteins are related.

268 Next, to further investigate the relative contributions of the DUF262 and  
269 DUF1524 domains to TgvAB function, we used the structural modelling and  
270 alignments to identify key residues in each domain. For both TgvA and B, the three  
271 highly conserved motifs characteristically associated with the DUF262 domain (i.e. (i)  
272 QR, (ii) DGQQR and (iii) FxxxN) were readily identifiable (Fig. 4a-d) (44). Notably,  
273 the DGQQR motif is thought to form part of a nucleotide binding pocket and to be  
274 required for nucleotide hydrolysis. Indeed, site-directed mutants of either TgvA or  
275 TgvB encoding substitutions in this motif previously shown to disrupt NTPase activity

276 (48–50), all resulted in a total loss of anti-phage activity (Fig. 4c-e). In contrast, the  
277 DUF1524 domain contains a highly conserved H...N...H/N motif, which belongs to  
278 the His-Me finger nuclease superfamily and that assumes a characteristic  $\beta\beta\alpha$  fold  
279 (52, 53). Such a motif was readily apparent in C-terminal domain of the predicted  
280 TgvB structure, and consistent with previous findings (45, 48–50), substitutions  
281 designed to disrupt either the catalytic histidine (TgvB[H571A]) or the metal-binding  
282 asparagine (TgvB[N602A]) were sufficient to abolish anti-phage activity (Fig. 4d-e).

283 Overall, our results suggest that the TgvAB system senses phages with  
284 hypermodified cytosines in a manner that requires the DUF262 domains of both  
285 TgvA and B, and that the His-Me nuclease domain of TgvB likely functions as the  
286 effector against phage DNA. Nevertheless, why TgvB alone is not sufficient for  
287 phage protection remains unclear. One possibility is that TgvA is required to  
288 overcome a phage encoded inhibitor. For example, some GmrSD family enzymes  
289 such as Eco94GmrSD are inhibited by the protein IPI\*, which is co-injected into the  
290 host cell with the T4 genome (45, 46, 54). However, TgvA could equally also play a  
291 regulatory or structural role and further work will therefore be needed to clarify these  
292 possibilities.

293

### 294 **Occurrence of the *tgvAB* system within and outside T1RM clusters**

295 To investigate the prevalence of *tgvAB* homologs within the T1RM cluster, we  
296 examined the distribution of the specific five-gene operon within 81,172 bacterial  
297 genomes (see methods for details). This *in silico* analysis revealed that the gene  
298 architecture found in VPI-2 of *V. cholerae* is also present in a variety of other  
299 bacterial genera (Fig. 5a) with 102 hits within this genome database, including  
300 several *Shewanella*, *Acinetobacter*, and *Pseudoalteromonas* species (see Table S1

301 for species-level details). This wider distribution indicates the potential functional  
302 conservation of these gene arrangements across different gram-negative bacteria.  
303 However, the genus *Vibrio* was still most prominently featured in these findings with  
304 55 hits (Fig. 5a). Precisely, apart from *V. cholerae*, species such as *Vibrio vulnificus*,  
305 *Vibrio antiquarius*, *Vibrio nigripulchritudo*, *Vibrio parahaemolyticus*, *Vibrio pelagius*,  
306 and the unclassified *Vibrio* strain B1ASS3 (*Vibrio* sp.) were identified to carry similar  
307 gene clusters (Fig. 5a). Despite the presence of these diverse *Vibrio* species, *V.*  
308 *cholerae* 7PET strains were the most commonly identified with 37 hits (Fig. 5a), likely  
309 reflecting their prominent representation in the NCBI database.

310 Subsequent analysis focused on the independent occurrences of the T1RM  
311 and TgvAB systems. As expected, the T1RM system was widespread (3886 hits)  
312 across numerous bacterial orders (Fig. 5b and Table S2 for species-level details).  
313 Homologs of the *tgxAB* operon alone were slightly less common with 1744 hits  
314 (Table S3 for species-level details), yet 17-times more prevalent than the instances  
315 of the five-gene operon described above. Indeed, as shown in Figure 5c, the  
316 occurrence of TgvAB homologs spans a wide array of bacterial orders, with species  
317 found in the human gut, like *Bacteroides fragilis* (Bacteroidales), to organisms  
318 isolated from permafrost, such as *Psychrobacter cryohalolentis* (Moraxellales).

319

## 320 **Conclusion**

321 In this study, we aimed to characterize the predicted restriction gene cluster of VPI-2.  
322 We showed that the T1RM system actively methylates the genomes of 7PET *V.*  
323 *cholerae* strains, while restricting unmethylated foreign DNA. Additionally, we  
324 identified a novel two-protein modification-dependent restriction system, TgvAB,  
325 which is embedded within the T1RM cluster. Interestingly, Picton *et al.* demonstrated

326 that the TgvB homolog BrxU, along with the Bacteriophage Exclusion (BREX)  
327 system (55), work in concert to offer complementary resistance against both  
328 modified and non-modified phages (44, 48). Therefore, it is tempting to speculate  
329 that the embedding of the *tgxAB* operon within the T1RM cluster serves a similar  
330 complementary role in *V. cholerae*. Supporting this notion, Machnicka *et al.* noted  
331 that GmrSD homologs are frequently encoded within Type I RM loci. An example  
332 includes the gene encoding the DUF262 domain-containing protein RloF of  
333 *Campylobacter jejuni*, which is situated between *hsdR* and *hsdS* of a T1RM operon  
334 (56), similar to the positioning of *tgxAB* described in this study. That defence system  
335 tend to cluster together within defence islands has been established over several  
336 years (11, 27, 57). However, this concept was recently extended by Payne and  
337 colleagues by identifying specific genes embedded within multi-gene defence  
338 clusters, highlighting the complex organization and integration of these systems  
339 within bacterial genomes (58). Notably, their research found GmrSD-like genes  
340 embedded within Hma (Helicase, Methylase, ATPase) defence gene clusters.  
341 However, unlike the HEC-05 (=BrxU) and HEC-06 GmrSD-like proteins identified in  
342 their work, which function independently (52), our findings indicate that the TgvAB  
343 defence operates as a two-protein system, underscoring the diversity in bacterial  
344 defence strategies.

345

## 346 **Material and Methods**

### 347 **Bacterial strains, plasmids, and culture conditions**

348 The bacterial strains and the plasmids used in this study are listed in Table S4.  
349 pUC18-mini-Tn7T-Gm-*lacZ* was a gift from Herbert Schweizer via Addgene plasmid  
350 #63120 (59). The primary *V. cholerae* strain used, A1552, is a fully sequenced

351 toxigenic O1 El Tor Inaba strain, representing the ongoing 7<sup>th</sup> cholera pandemic (60,  
352 61). Unless stated otherwise, bacteria were aerobically cultured in Lysogeny broth  
353 (LB; 1% tryptone, 0.5% yeast extract, 1% sodium chloride; Carl Roth, Switzerland)  
354 with shaking at 180 r.p.m., or on LB agar plates at either 30°C or 37°C. When  
355 required, antibiotic selection was applied using ampicillin (100 µg/ml), kanamycin (75  
356 µg/ml), and gentamicin (25 or 50 µg/ml). For natural transformation, chitin powder  
357 (Alfa Aesar via Thermo Fisher, USA) was combined with half-concentrated Instant  
358 Ocean medium (Aquarium Systems) and sterilized by autoclaving prior to adding the  
359 bacterial cultures.

360 Conjugation with MFDpir (62) was used to introduce the mini-Tn7 transposon  
361 derivatives into *E. coli* strain MG1655 on agar plates containing 0.3 mM  
362 diaminopimelic acid (DAP; Sigma-Aldrich). To induce expression from the  $P_{BAD}$   
363 promoter, cultures were grown in media containing 0.2% L-arabinose. For  
364 bacteriophages experiments, LB medium was supplemented with 5 mM CaCl<sub>2</sub> + 20  
365 mM MgSO<sub>4</sub>. Double-layer LB plates were prepared by adding 0.5% agar for semi-  
366 solid agar and 1.5% agar for the solid base.

367

### 368 **Genetic engineering of strains and plasmids**

369 Standard molecular cloning techniques were utilized for the cloning process (63)  
370 using the following enzymes: Pwo polymerase (Roche), Q5 High fidelity polymerase  
371 (New England Biolabs), GoTaq polymerase (Promega), restriction enzymes (New  
372 England Biolabs), and T4 DNA ligase (New England Biolabs). Enzymes were used  
373 according to the manufacturer's instructions. All constructs were verified through  
374 PCR and/or Sanger or Nanopore sequencing (performed by Microsynth AG,  
375 Switzerland) and analysed using SnapGene version 4.3.11.



376 *V. cholerae* strains were created through natural transformation and FLP  
377 recombination (TransFLP) (64–66) or through allelic exchange using derivatives of  
378 the suicide plasmid pGP704-Sac28 (67) and SacB-based counter-selection on NaCl-  
379 free LB plates with 10% sucrose. Mini-Tn7 transposons, containing *araC* and the  
380 gene(s) of interest regulated by the arabinose-inducible promoter  $P_{BAD}$ , were  
381 inserted in *E. coli* downstream of *glmS* via triparental mating, following established  
382 protocols (68). Site-directed mutations in these constructs were introduced by  
383 inverse PCR prior to their transposition into the *E. coli* chromosome.

384

### 385 **PacBio (SMRT) sequencing**

386 Genomic DNA was purified from overnight cultures using Qiagen's Genomic-tip  
387 procedure combined with the Genomic DNA buffer set (Qiagen, Switzerland),  
388 following the manufacturer's instructions. Sample processing, PacBio Single  
389 Molecule, Real-Time (SMRT) sequencing, and *de novo* genome assembly were  
390 performed at the University of Lausanne's Genomic Technology Facility, as  
391 previously described (35). All SMRT sequencing raw data have been made available  
392 on Zenodo (three datasets: 10.5281/zenodo.10838595; 10.5281/zenodo.10839511;  
393 10.5281/zenodo.10839547). Note that the assembled genomes of strains A1552,  
394 C6706, C6709, P27459, E7946, DRC193A, and MO10 have been previously  
395 reported without analysis of their epigenetic modifications (35, 36, 61).

396

### 397 **Electroporation-mediated transformation of *V. cholerae* using plasmids**

398 To explore the T1RM system's efficiency in restricting DNA with specific recognition  
399 sequences, we compared the uptake frequency of a plasmid harboring the putative  
400 recognition motif ( $P_{\text{motif+}}$ ) to that of a variant plasmid with silent mutations in *aacC1*

401 ( $P_{\text{motif}}$ ) altering its sequence while maintaining the encoded aminoglycoside-3-O-  
402 acetyltransferase-I protein. Transformation frequencies were assessed through  
403 electroporation. *V. cholerae* competent cells were prepared by standard protocols  
404 (63), involving 1:100 dilution of overnight cultures, growth for 2h and 30min at 37°C  
405 ( $OD_{600} \sim 1.0$ ), and washing steps with cold 2 mM  $\text{CaCl}_2$  and 10% glycerol before  
406 shock-freezing. After 2h at -80°C, electroporation with 100 ng plasmid was  
407 performed at 1.6 kV followed by recovery in 2xYT-rich medium at 30 °C for 2 h. Cells  
408 were plated on LB agar with and without kanamycin and incubated at 37°C  
409 overnight. Transformation frequencies were calculated as the ratio of kanamycin-  
410 resistant transformants to the total number of bacteria.

411

#### 412 **Bacteriophage handling and culturing**

413 The *E. coli* BASEL phage collection (42) was used in this study. To generate phage  
414 stocks, an *E. coli* MG1655 $\Delta$ araCBAD (69) overnight culture was diluted and grown to  
415 the exponential phase in LB medium supplemented with 5 mM  $\text{CaCl}_2$  and 20 mM  
416  $\text{MgSO}_4$ . Subsequently, the culture was 1:10 diluted in prewarmed medium, infected  
417 with  $10^4$  plaque forming units (PFU)/mL, and incubated under shaking conditions at  
418 37°C for 5 h. Following incubation, centrifugation and filtration were used to clear the  
419 lysate, which was then treated with 1% chloroform and stored at +4°C. Phage titers  
420 were determined using plaque assays on the propagation strain.

421

#### 422 **Bacteriophage plaque assays**

423 For plaque assays, *E. coli* MG1655 $\Delta$ araCBAD, either with the candidate defence  
424 system or the empty miniTn7 transposon control, was grown in LB medium.  
425 Overnight cultures were diluted 1:100 in LB medium supplemented with 0.2%

426 arabinose, 5 mM CaCl<sub>2</sub>, and 20 mM MgSO<sub>4</sub> and grown at 37°C with shaking for 2 h.  
427 Once reaching the exponential phase, the cultures were diluted 1:40 in 0.5% LB agar  
428 containing 5 mM CaCl<sub>2</sub>, 20 mM MgSO<sub>4</sub>, and 0.2% arabinose, then overlaid on 1.5%  
429 LB agar. Phage samples were serially diluted in LB medium with 5 mM CaCl<sub>2</sub> and 20  
430 mM MgSO<sub>4</sub> and spotted onto the bacterial overlays. After overnight incubation at  
431 37°C, plaques were counted to assess the defense system's effectiveness compared  
432 to the miniTn7-carrying control strain (= fold protection).

433

### 434 **Infection kinetics**

435 The infection kinetics assay of *Tequatrovirus* (Bas35, Bas40) and *Mosigviruses*  
436 (Bas46, Bas47) was conducted as follows: Overnight cultures of *E. coli* strains were  
437 diluted 1:100 in LB medium supplemented with 5 mM CaCl<sub>2</sub>, 20 mM MgSO<sub>4</sub> and  
438 0.2% arabinose. Bacterial cultures were then incubated at 37°C with shaking for 2 h.  
439 Subsequently, 20 µl of phage per well at multiplicities of infection (MOI) of 0, 0.2, 5,  
440 or 10 were added in technical triplicate to a 96-well plate. The cultures were further  
441 diluted 1:10 in the same LB condition, and 180 µl of each diluted culture was then  
442 added to the wells. The SpectraMax i3x plate reader from Molecular Devices was  
443 utilized to assess bacterial growth at 37°C, with measurements taken at 6-minute  
444 intervals over a total of 49 cycles. To calculate the MOI, cultures of strains  
445 MG1655Δ*araCBAD*-Tn-empty and MG1655Δ*araCBAD*-TnTgvAB were cultured  
446 following the protocol outlined in the Bacteriophage Plaque Assays section. Colony-  
447 forming units (per ml) were quantified by spotting serially diluted cultures onto LB  
448 plates. The calculated values represent the average of three technical replicates.

449

### 450 **Bioinformatics analyses**

451 The VPI-2 genomic region of 7PET O1 strains and one O139 serogroup strain  
452 (MO10) was compared and visualized using Clinker software (v 0.0.25, default  
453 parameters) (70) after reannotation of the genome sequence of strain A1552 using  
454 the Prokaryotic Genome Annotation pipeline version 2023-10-03.build7061 (71) to  
455 unify the annotation method. For sequence similarity, NCBI's blastp was utilized  
456 (default parameters, non-redundant protein database; accession August 2023), while  
457 structural modeling was conducted with ColabFold (1.5.2) (72) based on AlphaFold2  
458 (51) using default settings. DALI was employed for structural similarity predictions  
459 against the Protein Data Bank (PDB) (73, 74).

460 The distribution of the specific five-gene operon (*VC1765-69*) across bacterial  
461 species was examined with MacSyFinder (v.2.1) (75), using a comprehensive  
462 database of all sequenced and fully assembled bacterial genomes (taxid:2) from the  
463 NCBI database (accession date 26.01.2024). This analysis therefore covered a  
464 dataset comprising 81,172 bacterial genomes, which altogether contained over 304  
465 million protein sequences.

466 To build HMM profiles for each target CDS within the *VC1765-69* operon,  
467 homologous protein sequences were identified via PSI-blast searches in the NCBI  
468 database (3 iterations), using the non-redundant protein sequence database  
469 (accessed in February 2024) with a cutoff e-value of 1e-10.

470 After identifying homologous sequences for each CDS through PSI-blast, the  
471 sequences were aligned using MAFFT (v7.508, --maxiterate 1000 --localpair  
472 parameters for higher accuracy alignments) (76). From these multiple alignments,  
473 HMM profiles were generated with HMMER (v3.3.2, using hmmbuild with default  
474 parameters) (77), forming the basis for constructing different models in  
475 MacSyFinder. These models were used to search for the occurrence of the CDS in

476 various combinations encompassing the T1RM and/or TgvAB system genes. The  
477 constructed models were then applied in a search across the bacterial protein  
478 database mentioned above.

479

#### 480 **Statistics and reproducibility**

481 Results are derived from biologically independent experiments, as specified in the  
482 figure legends. Statistical analyses were conducted using Prism software (v10.2.1;  
483 GraphPad).

484

#### 485 **Acknowledgements**

486 We thank members of the Blokesch lab and especially Nicolas Flaugnatti and Alexis  
487 Proutière for fruitful discussions and Sandrine Stutzmann, Laurie Righi, and Loriane  
488 Bader for technical assistance. We acknowledge the staff of the Lausanne Genomic  
489 Technologies Facility and Christian Iseli & Nicolas Guex from the EPFL/UNIL  
490 Bioinformatics Competence Center for sample processing, sequencing, genome  
491 assembly, and SMRT sequencing analysis. We are also grateful to Alexander Harms  
492 for sharing the BASEL phage collection and for valuable discussions. Our  
493 appreciation goes to the Waters lab for coordinating the timing of their submission  
494 with ours. This work was supported by the Swiss National Science Foundation  
495 (310030\_185022) and an International Research Scholarship by the Howard Hughes  
496 Medical Institute (HHMI) (grant 55008726) awarded to M.B.

497

498 *Author contributions:* MB secured funding; MB & DWA conceived and supervised the  
499 study; GV, DWA, and MB designed the experiments and analysed the data. GV,  
500 DWA, and MB designed and constructed strains/plasmids. GV performed the

501 experiments. MB initiated the SMRT sequencing. AL performed comparative  
502 genomic and conservation analyses. MB, DWA, and GV wrote the manuscript. All  
503 authors approved the final version of the paper.

504

#### 505 **Authors' ORCIDs:**

506 Grazia Vizzarro: <https://orcid.org/0000-0002-2466-9679>

507 Alexandre Lemopoulos: <https://orcid.org/0000-0002-7997-5526>

508 David William Adams: <https://orcid.org/0000-0002-9644-5607>

509 Melanie Blokesch: <https://orcid.org/0000-0002-7024-1489>

510

#### 511 **References**

- 512 1. Rocha EPC, Bikard D. 2022. Microbial defenses against mobile genetic  
513 elements and viruses: Who defends whom from what? *PLOS Biol* 20:e3001514.
- 514 2. Baltrus DA. 2013. Exploring the costs of horizontal gene transfer. *Trends Ecol*  
515 *Evol* 28:489–495.
- 516 3. Stern A, Sorek R. 2011. The phage-host arms race: Shaping the evolution of  
517 microbes. *BioEssays* 33:43–51.
- 518 4. Dy RL, Richter C, Salmond GPC, Fineran PC. 2014. Remarkable Mechanisms  
519 in Microbes to Resist Phage Infections. *Annu Rev Virol* 1:307–331.
- 520 5. Georjon H, Bernheim A. 2023. The highly diverse antiphage defence systems of  
521 bacteria. *Nat Rev Microbiol* 21:686–700.
- 522 6. Lopatina A, Tal N, Sorek R. 2020. Abortive Infection: Bacterial Suicide as an  
523 Antiviral Immune Strategy. *Annu Rev Virol* 7:371–384.
- 524 7. Murray NE. 2000. Type I Restriction Systems: Sophisticated Molecular  
525 Machines (a Legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64:412–434.

- 526 8. Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. *Curr*  
527 *Opin Microbiol* 8:466–472.
- 528 9. Loenen WAM, Raleigh EA. 2014. The other face of restriction: modification-  
529 dependent enzymes. *Nucleic Acids Res* 42:56–69.
- 530 10. Weigele P, Raleigh EA. 2016. Biosynthesis and Function of Modified Bases in  
531 Bacteria and Their Viruses. *Chem Rev* 116:12655–12687.
- 532 11. Makarova KS, Wolf YI, Snir S, Koonin EV. 2011. Defense Islands in Bacterial  
533 and Archaeal Genomes and Prediction of Novel Defense Systems. *J Bacteriol*  
534 193:6039–6056.
- 535 12. Raleigh EA. 1992. Organization and function of the *mcrBC* genes of *Escherichia*  
536 *coli* K-12. *Mol Microbiol* 6:1079–1086.
- 537 13. Kovach ME, Shaffer MD, Peterson KM. 1996. A putative integrase gene defines  
538 the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio*  
539 *cholerae*. *Microbiology* 142:2165–2174.
- 540 14. Waldor MK, Mekalanos JJ. 1996. Lysogenic Conversion by a Filamentous  
541 Phage Encoding Cholera Toxin. *Science* 272:1910–1914.
- 542 15. Taylor RK, Miller VL, Furlong DB, Mekalanos JJ. 1987. Use of *phoA* gene  
543 fusions to identify a pilus colonization factor coordinately regulated with cholera  
544 toxin. *Proc Natl Acad Sci* 84:2833–2837.
- 545 16. Karaolis DKR, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR.  
546 1998. A *Vibrio cholerae* pathogenicity island associated with epidemic and  
547 pandemic strains. *Proc Natl Acad Sci* 95:3134–3139.
- 548 17. Balasubramanian D, López-Pérez M, Almagro-Moreno S. 2023. Cholera  
549 Dynamics and the Emergence of Pandemic *Vibrio cholerae*, p. 127–147. *In*

- 550 Almagro-Moreno, S, Pukatzki, S (eds.), *Vibrio* spp. Infections. Springer  
551 International Publishing, Cham.
- 552 18. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. 2002.  
553 Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with  
554 cholera endemic and pandemic disease. *Proc Natl Acad Sci* 99:1556–1561.
- 555 19. Davies BW, Bogard RW, Young TS, Mekalanos JJ. 2012. Coordinated  
556 Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for  
557 *V. cholerae* Virulence. *Cell* 149:358–370.
- 558 20. Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger A-K,  
559 Khataokar A, O'Hara BJ, Behrmann LV, Neiditch MB, Benning C, Waters CM,  
560 Ng W-L. 2018. Direct activation of a phospholipase by cyclic GMP-AMP in El  
561 Tor *Vibrio cholerae*. *Proc Natl Acad Sci* 115:E6048–E6055.
- 562 21. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen  
563 A, Doron S, Amitai G, Sorek R. 2019. Cyclic GMP-AMP signalling protects  
564 bacteria against viral infection. *Nature* 574:691–695.
- 565 22. Hsueh BY, Severin GB, Elg CA, Waldron EJ, Kant A, Wessel AJ, Dover JA,  
566 Rhoades CR, Ridenhour BJ, Parent KN, Neiditch MB, Ravi J, Top EM, Waters  
567 CM. 2022. Phage defence by deaminase-mediated depletion of  
568 deoxynucleotides in bacteria. *Nat Microbiol* 7:1210–1220.
- 569 23. Jaskólska M, Adams DW, Blokesch M. 2022. Two defence systems eliminate  
570 plasmids from seventh pandemic *Vibrio cholerae*. *Nature* 604:323–329.
- 571 24. Jermyn WS, Boyd EF. 2002. Characterization of a novel *Vibrio* pathogenicity  
572 island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae*  
573 isolates. *Microbiology* 148:3681–3693.



- 574 25. Almagro-Moreno S, Boyd EF. 2009. Sialic Acid Catabolism Confers a  
575 Competitive Advantage to Pathogenic *Vibrio cholerae* in the Mouse Intestine.  
576 Infect Immun 77:3807–3816.
- 577 26. Yan J, Liu Q, Xue X, Li J, Li Y, Su Y, Cao B. 2023. The Response Regulator  
578 VC1795 of *Vibrio* Pathogenicity Island-2 Contributes to Intestinal Colonization  
579 by *Vibrio cholerae*. Int J Mol Sci 24:13523.
- 580 27. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R.  
581 2018. Systematic discovery of antiphage defense systems in the microbial  
582 pangenome. Science 359:eaar4120.
- 583 28. Haidai Hu, Thomas C.D. Hughes, Philipp F. Popp, Aritz Roa-Eguiara, Freddie  
584 J.O. Martin, Nicole R. Rutbeek, Ivo Alexander Hendriks, Leighton J. Payne,  
585 Yumeng Yan, Victor Klein de Sousa, Yong Wang, Michael Lund Nielsen,  
586 Richard M. Berry, Marc Erhardt, Simon A. Jackson, Nicholas M.I. Taylor. 2023.  
587 Structure and mechanism of Zorya anti-phage defense system. bioRxiv  
588 2023.12.18.572097.
- 589 29. Jermyn WS, Boyd EF. 2005. Molecular evolution of *Vibrio* pathogenicity island-2  
590 (VPI-2): mosaic structure among *Vibrio cholerae* and *Vibrio mimicus* natural  
591 isolates. Microbiology 151:311–322.
- 592 30. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft  
593 DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin  
594 H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I,  
595 Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O,  
596 Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM.  
597 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
598 *cholerae*. Nature 406:477–483.

- 599 31. Dorman MJ, Domman D, Uddin MI, Sharmin S, Afrad MH, Begum YA, Qadri F,  
600 Thomson NR. 2019. High quality reference genomes for toxigenic and non-  
601 toxigenic *Vibrio cholerae* serogroup O139. *Sci Rep* 9:5865.
- 602 32. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG. 2014. Type I restriction  
603 enzymes and their relatives. *Nucleic Acids Res* 42:20–44.
- 604 33. Dussoix D, Arber W. 1962. Host specificity of DNA produced by *Escherichia*  
605 *coli*. *J Mol Biol* 5:37–49.
- 606 34. Arber W. 1965. Host-Controlled Modification of Bacteriophage. *Annu Rev*  
607 *Microbiol* 19:365–378.
- 608 35. Stutzmann S, Blokesch M. 2020. Comparison of chitin-induced natural  
609 transformation in pandemic *Vibrio cholerae* O1 El Tor strains. *Environ Microbiol*  
610 22:4149–4166.
- 611 36. Lemopoulos A, Miwanda B, Drebes Dörr NC, Stutzmann S, Bompangue D,  
612 Muyembe-Tamfum J-J, Blokesch M. 2024. Genome sequences of *Vibrio*  
613 *cholerae* strains isolated in the DRC between 2009 and 2012. *Microbiol Resour*  
614 *Announc* 13:e00827-23.
- 615 37. Beaulaurier J, Schadt EE, Fang G. 2019. Deciphering bacterial epigenomes  
616 using modern sequencing technologies. *Nat Rev Genet* 20:157–172.
- 617 38. Hennig S, Ziebuhr W. 2010. Characterization of the Transposase Encoded by IS  
618 256, the Prototype of a Major Family of Bacterial Insertion Sequence Elements.  
619 *J Bacteriol* 192:4153–4163.
- 620 39. William P Robins, Bradley T Meader, Jonida Toska, John J Mekalanos. 2022.  
621 Cell density-dependent death triggered by viral palindromic DNA sequences.  
622 *bioRxiv* 2022.11.18.517080.

- 623 40. Bickle TA, Krüger DH. 1993. Biology of DNA restriction. *Microbiol Rev* 57:434–  
624 450.
- 625 41. Seed KD, Bodi KL, Kropinski AM, Ackermann H-W, Calderwood SB, Qadri F,  
626 Camilli A. 2011. Evidence of a Dominant Lineage of *Vibrio cholerae*-Specific  
627 Lytic Bacteriophages Shed by Cholera Patients over a 10-Year Period in Dhaka,  
628 Bangladesh. *mBio* 2:e00334-10.
- 629 42. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, Sauer P,  
630 Willi L, Michaelis S, Hilbi H, Thaler DS, Harms A. 2021. Systematic exploration  
631 of *Escherichia coli* phage–host interactions with the BASEL phage collection.  
632 *PLOS Biol* 19:e3001424.
- 633 43. Thomas JA, Orwenyo J, Wang L-X, Black LW. 2018. The Odd “RB” Phage-  
634 Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-  
635 Like Phage RB69. *Viruses* 10:313.
- 636 44. Machnicka MA, Kaminska KH, Dunin-Horkawicz S, Bujnicki JM. 2015.  
637 Phylogenomics and sequence-structure-function relationships in the GmrSD  
638 family of Type IV restriction enzymes. *BMC Bioinformatics* 16:336.
- 639 45. He X, Hull V, Thomas JA, Fu X, Gidwani S, Gupta YK, Black LW, Xu S. 2015.  
640 Expression and purification of a single-chain Type IV restriction enzyme  
641 Eco94GmrSD and determination of its substrate preference. *Sci Rep* 5:9747.
- 642 46. Bair CL, Rifat D, Black LW. 2007. Exclusion of Glucosyl-Hydroxymethylcytosine  
643 DNA Containing Bacteriophages Is Overcome by the Injected Protein Inhibitor  
644 IPI\*. *J Mol Biol* 366:779–789.
- 645 47. Bair CL, Black LW. 2007. A Type IV Modification Dependent Restriction  
646 Nuclease that Targets Glucosylated Hydroxymethyl Cytosine Modified DNAs. *J*  
647 *Mol Biol* 366:768–778.

- 648 48. Picton DM, Luyten YA, Morgan RD, Nelson A, Smith DL, Dryden DTF, Hinton  
649 JCD, Blower TR. 2021. The phage defence island of a multidrug resistant  
650 plasmid uses both BREX and type IV restriction for complementary protection  
651 from viruses. *Nucleic Acids Res* 49:11257–11273.
- 652 49. Xiong X, Wu G, Wei Y, Liu L, Zhang Y, Su R, Jiang X, Li M, Gao H, Tian X,  
653 Zhang Y, Hu L, Chen S, Tang Y, Jiang S, Huang R, Li Z, Wang Y, Deng Z,  
654 Wang J, Dedon PC, Chen S, Wang L. 2020. SspABCD–SspE is a  
655 phosphorothioation-sensing bacterial defence system with broad anti-phage  
656 activities. *Nat Microbiol* 5:917–928.
- 657 50. Gao H, Gong X, Zhou J, Zhang Y, Duan J, Wei Y, Chen L, Deng Z, Wang J,  
658 Chen S, Wu G, Wang L. 2022. Nicking mechanism underlying the DNA  
659 phosphorothioate-sensing antiphage defense by SspE. *Nat Commun* 13:6773.
- 660 51. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O,  
661 Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C,  
662 Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J,  
663 Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska  
664 M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu  
665 K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with  
666 AlphaFold. *Nature* 596:583–589.
- 667 52. Jablonska J, Matelska D, Steczkiewicz K, Ginalski K. 2017. Systematic  
668 classification of the His-Me finger superfamily. *Nucleic Acids Res* 45:11479–  
669 11494.
- 670 53. Wu C-C, Lin JLJ, Yuan HS. 2020. Structures, Mechanisms, and Functions of  
671 His-Me Finger Nucleases. *Trends Biochem Sci* 45:935–946.

- 672 54. Rifat D, Wright NT, Varney KM, Weber DJ, Black LW. 2008. Restriction  
673 Endonuclease Inhibitor IPI\* of Bacteriophage T4: A Novel Structure for a  
674 Dedicated Target. *J Mol Biol* 375:720–734.
- 675 55. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y,  
676 Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system  
677 widespread in microbial genomes. *EMBO J* 34:169–183.
- 678 56. Miller WG, Pearson BM, Wells JM, Parker CT, Kapitonov VV, Mandrell RE.  
679 2005. Diversity within the *Campylobacter jejuni* type I restriction–modification  
680 loci. *Microbiology* 151:337–351.
- 681 57. Hochhauser D, Millman A, Sorek R. 2023. The defense island repertoire of the  
682 *Escherichia coli* pan-genome. *PLOS Genet* 19:e1010694.
- 683 58. Leighton J. Payne, Tom C. D. Hughes, Peter C. Fineran, Simon A. Jackson.  
684 2024. New antiviral defences are genetically embedded within prokaryotic  
685 immune systems. *bioRxiv* 2024.01.29.577857.
- 686 59. Choi K-H, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR,  
687 Schweizer HP. 2005. A Tn7-based broad-range bacterial cloning and  
688 expression system. *Nat Methods* 2:443–448.
- 689 60. Yildiz FH, Schoolnik GK. 1998. Role of *rpoS* in Stress Survival and Virulence of  
690 *Vibrio cholerae*. *J Bacteriol* 180:773–784.
- 691 61. Matthey N, Drebes Dörr NC, Blokesch M. 2018. Long-Read-Based Genome  
692 Sequences of Pandemic and Environmental *Vibrio cholerae* Strains. *Microbiol*  
693 *Resour Announc* 7:e01574-18.
- 694 62. Ferrières L, Hémerly G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo J-M.  
695 2010. Silent Mischief: Bacteriophage Mu Insertions Contaminate Products of  
696 *Escherichia coli* Random Mutagenesis Performed Using Suicidal Transposon

- 697 Delivery Plasmids Mobilized by Broad-Host-Range RP4 Conjugative Machinery.  
698 J Bacteriol 192:6418–6427.
- 699 63. Green MR, Sambrook J, Sambrook J. 2012. Molecular cloning: a laboratory  
700 manual.4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- 701 64. Marvig RL, Blokesch M. 2010. Natural transformation of *Vibrio cholerae* as a  
702 tool - Optimizing the procedure. BMC Microbiol 10:155.
- 703 65. De Souza Silva O, Blokesch M. 2010. Genetic manipulation of *Vibrio cholerae*  
704 by combining natural transformation with FLP recombination. Plasmid 64:186–  
705 195.
- 706 66. Blokesch M. 2012. TransFLP—A Method to Genetically Modify *Vibrio cholerae*  
707 Based on Natural Transformation and FLP-recombination. J Vis Exp 68:e3761.
- 708 67. Meibom KL, Li XB, Nielsen AT, Wu C-Y, Roseman S, Schoolnik GK. 2004. The  
709 *Vibrio cholerae* chitin utilization program. Proc Natl Acad Sci 101:2524–2529.
- 710 68. Bao Y, Lies DP, Fu H, Roberts GP. 1991. An improved Tn7-based system for  
711 the single-copy insertion of cloned genes into chromosomes of gram-negative  
712 bacteria. Gene 109:167–168.
- 713 69. Aoki SK, Lillacci G, Gupta A, Baumschlager A, Schweingruber D, Khammash M.  
714 2019. A universal biomolecular integral feedback controller for robust perfect  
715 adaptation. Nature 570:533–537.
- 716 70. Gilchrist CLM, Chooi Y-H. 2021. clinker & clustermap.js: automatic generation of  
717 gene cluster comparison figures. Bioinformatics 37:2473–2475.
- 718 71. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L,  
719 Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic  
720 genome annotation pipeline. Nucleic Acids Res 44:6614–6624.

- 721 72. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022.  
722 ColabFold: making protein folding accessible to all. *Nat Methods* 19:679–682.
- 723 73. Holm L. 2020. Using Dali for Protein Structure Comparison, p. 29–42. *In*  
724 Gáspári, Z (ed.), *Structural Bioinformatics*. Springer US, New York, NY.
- 725 74. Holm L. 2022. Dali server: structural unification of protein families. *Nucleic Acids*  
726 *Res* 50:W210–W215.
- 727 75. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSyFinder: A  
728 Program to Mine Genomes for Molecular Systems with an Application to  
729 CRISPR-Cas Systems. *PLoS ONE* 9:e110726.
- 730 76. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software  
731 Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30:772–  
732 780.
- 733 77. Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol*  
734 7:e1002195.
- 735

736 **Figure legends**

737 **Figure 1. VPI-2 exhibits high conservation across 7PET *V. cholerae* strains. a)**

738 Comparative genome alignment of the *Vibrio* pathogenicity island 2 (VPI-2) across a  
739 selection of 7PET O1 and O139 strains, isolated between 1975 and 2011. The  
740 genomes of these strains are displayed alongside their designated strain name.  
741 Coding sequences within the genomes are represented by arrows, with grey bars  
742 connecting them to indicate amino acid identity percentages at or above a threshold  
743 of 0.93. Instances of lower identity are highlighted in black boxes. Gene locus tags  
744 are derived from the reference genome of strain N16961. Predicted or established  
745 functions are labelled above each cluster. **b)** Close-up examination of the VC1765-  
746 69 gene cluster in strains A1552 and DRC193A reveals three genes responsible for  
747 the components of the putative T1RM system (*hsdR*, *hsdS*, *hsdM*). A comparative  
748 alignment highlights the disruption of the T1RM cluster in strain DRC193A, caused  
749 by an IS256 transposon insertion within *hsdS*.

750

751 **Figure 2. Type I RM system's role in chromosomal methylation and plasmid**

752 **restriction. a)** SMRT sequencing uncovers a distinctive modified DNA motif across  
753 various 7PET *V. cholerae* strains. The grey bars show the number of the DNA motif  
754 (GATGNNNNNCTT) in each genome, while the blue bars denote the percentage of  
755 this motif methylated in each strain's genome (see secondary Y-axis on the right). **b)**  
756 The T1RM system hinders plasmid acquisition. Transformation assays compare the  
757 uptake of a plasmid containing the recognition motif ( $P_{\text{motif+}}$ ) against a derivative  
758 plasmid with silent mutations altering the nucleotide sequence ( $P_{\text{motif-}}$ ). To the left,  
759 diagrams of the plasmids are depicted. Statistical differences were calculated on log-  
760 transformed data using a two-way ANOVA corrected for multiple comparisons with



761 Šidák's method. \*  $P < 0.05$ ; \*\*\*  $P < 0.001$ ; ns, not significant. < d.l., below detection  
762 limit.

763

764 **Figure 3. Protection against *Tevenvirinae* by Tgv proteins encoded by the**  
765 **T1RM-embedded genes. a)** Observed defence activity against then BASEL phage  
766 collection. Protection levels (fold-protection, as shown by the colour code on the  
767 right) were determined by comparing plaque formation in strains with the system to  
768 those without, using tenfold serial dilution assays. On the left, gene organization of  
769 the tested strains. Data represent the average of two replicates. The collective  
770 results of all 77 phage infections are detailed in Fig. S1. **b)** Phage plaque assays on  
771 *E. coli* strains harboring an empty transposon (control, Ctrl) or the two T1RM-  
772 embedded genes (*tgxAB*), using a tenfold serial dilution. **c, d)** Growth curves of *E.*  
773 *coli* cultures carrying an empty transposon (no system) or TntgvAB (+ *tgxAB*),  
774 without (NO phage) or with exposure to phages, initiated at time 0 with various MOIs  
775 (0.2, 2, 5, or 10). **c)** *Tequatroviruses* and **d)** *Mosigviruses* were used for infection.  
776 The presented data are the average of three independent experiments ( $\pm$ SD,  
777 illustrated with error bars).

778

779 **Figure 4. The two-protein TgvAB defence system is a member of the GmrSD**  
780 **family of Type IV restriction enzymes. a)** Structural models of Eco94GmrSD of *E.*  
781 *coli* STEC\_94C and TgvA (VC1767) & TgvB (VC1766) of *V. cholerae* 7PET strains.  
782 The models, produced via AlphaFold (ColabFold), portray the domains with  
783 corresponding colours, while also highlighting the residues characteristic to the  
784 DUF262 and DUF1524 domains. Images were generated using ChimeraX 1.7.1 **b)**  
785 Schematics displaying conserved domains identified in the TgvA and TgvB proteins.

786 **c, d)** Zoomed view of the conserved **(c)** DGQQR motif found in the DUF262 region  
787 of TgvA & TgvB and **(d)** of the His-Me finger motif within the DUF1524 of TgvB,  
788 highlighting the catalytic histidine (H) situated at the terminus of the  $\beta$ 1 strand, the  
789 Asparagine (N) residue positioned in the loop region and the final N residue within  
790 the  $\alpha$ -helix. **e)** Site-directed mutagenesis removed the antiviral effect. The level of  
791 protection was evaluated as described in Fig. 3. Mutagenesis aimed at disrupting  
792 NTPase or endonuclease functions exerted by DUF262 and DUF1524, respectively.  
793 The data are averages from three independent experiments ( $\pm$ SD, as shown by the  
794 error bars).

795

796 **Figure 5. Phylogenetic distribution of the restriction systems.** **a)** The presence  
797 of the 5-gene cluster (*VC1765-69*) was assessed across 81,172 bacterial genomes.  
798 The results revealed its distribution beyond *V. cholerae*, which represented 54% of  
799 all hits. \**V. cholerae* O37 serogroup strains are known to be closely related to  
800 classical O1 strains with highly similar chromosomal backbones. **b, c)** Exploration of  
801 the **(b)** T1RM system (*VC1769-68-65*) and **(c)** TgvAB system (*VC1766-67*) across  
802 the bacterial genomes demonstrates their assessment at the order level of  
803 taxonomy. Orders represented in less than 1% of instances were consolidated into a  
804 singular category labeled "Others" for the visualization. For details at the species  
805 level see Tables S1-S3.

806 **Supplemental Material**

807 **Table S1.** Summary of matching hits for the T1RM + TgvAB (*VC1765-69*) model,  
808 detected by MacSyFinder v.2.

809 **Table S2.** Summary of matching hits for the T1RM (*VC1769-68-65*) model, detected  
810 by MacSyFinder v.2.

811 **Table S3.** Summary of matching hits for the TgvAB (*VC1766-67*) model, detected by  
812 MacSyFinder v.2.

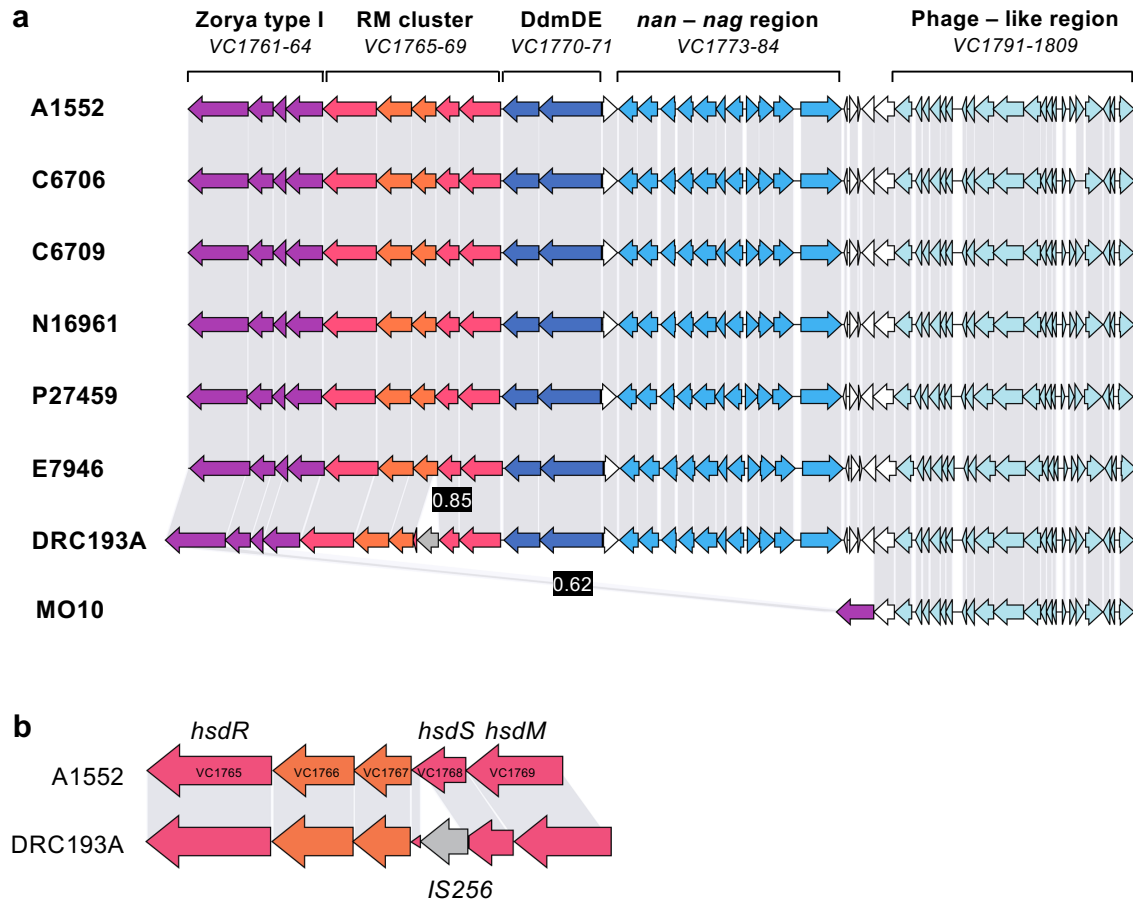
813 **Table S4.** Bacterial strains and plasmids used in this study.

814

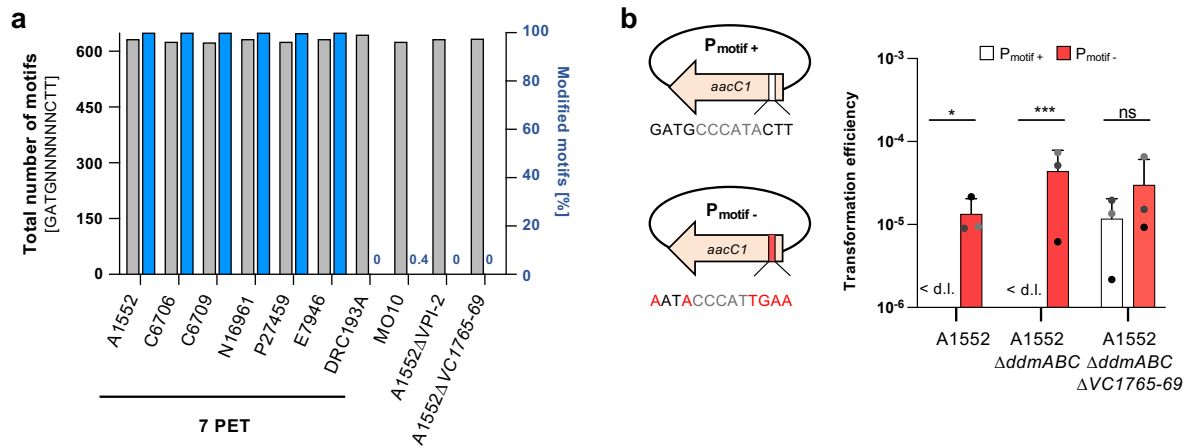
815 **Supplementary Figure**

816 **Figure S1. Observed defence activity against the BASEL phage collection.**

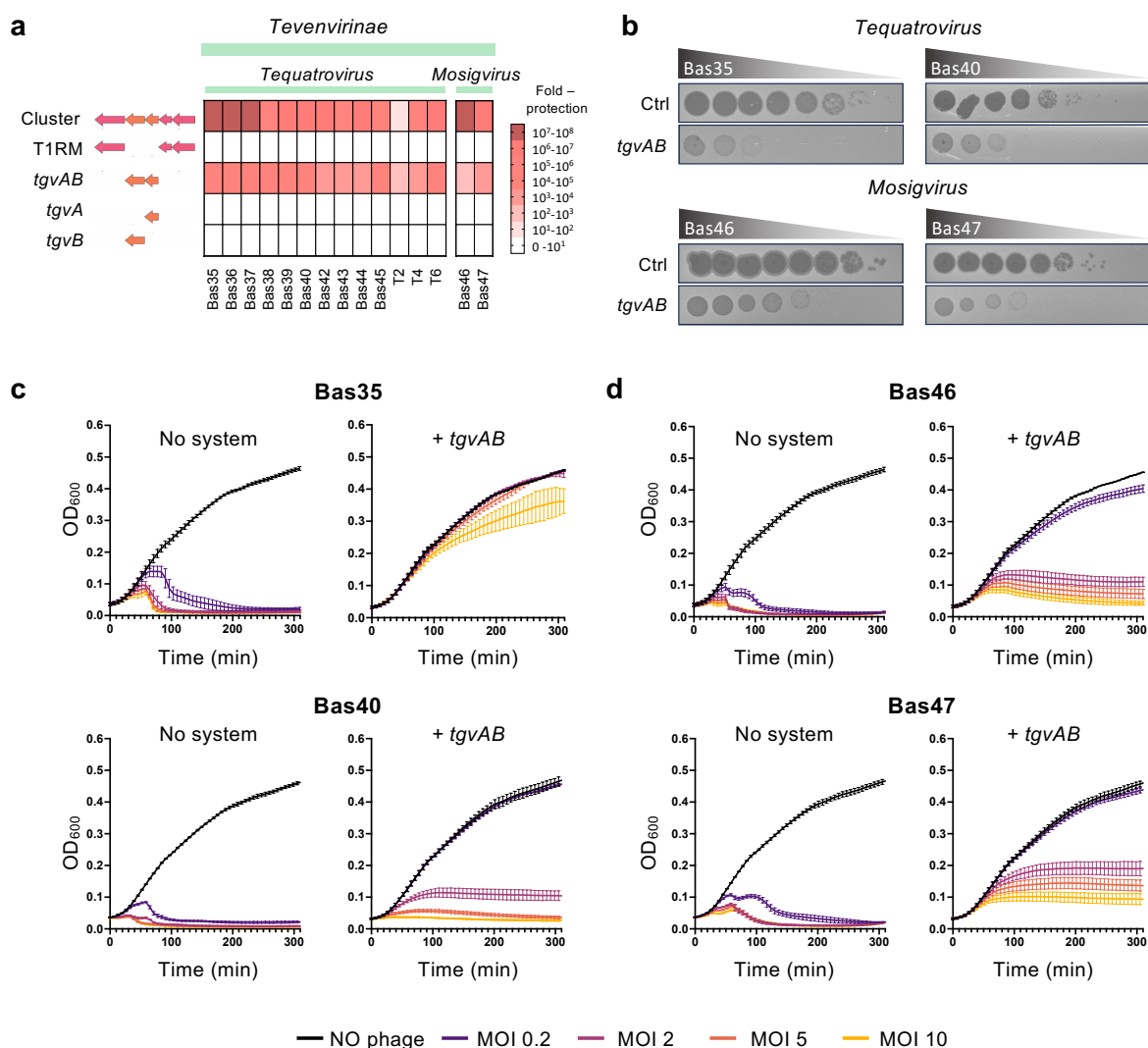
817 Protection levels (fold-protection, as shown by the colour code on the right) were  
818 determined by comparing plaque formation in strains with the system to those  
819 without, using tenfold serial dilution assays. Data represent the average of two  
820 replicates. Details as in Fig. 3a.



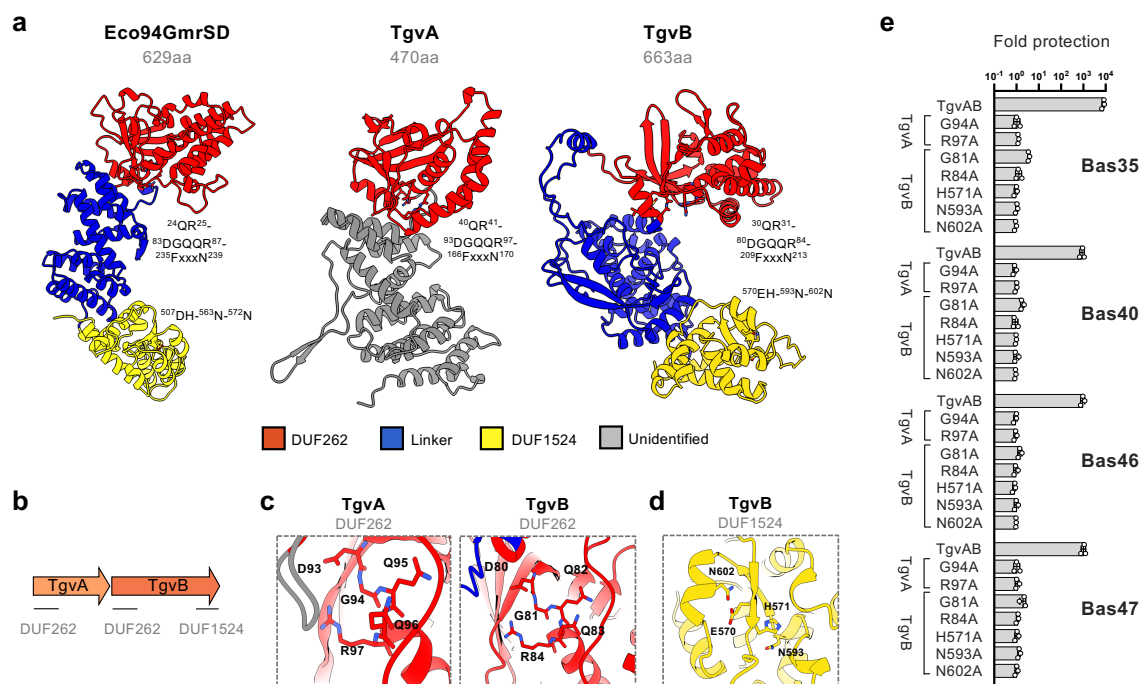
**Figure 1. VPI-2 exhibits high conservation across 7PET *V. cholerae* strains. a)** Comparative genome alignment of the *Vibrio* pathogenicity island 2 (VPI-2) across a selection of 7PET O1 and O139 strains, isolated between 1975 and 2011. The genomes of these strains are displayed alongside their designated strain name. Coding sequences within the genomes are represented by arrows, with grey bars connecting them to indicate amino acid identity percentages at or above a threshold of 0.93. Instances of lower identity are highlighted in black boxes. Gene locus tags are derived from the reference genome of strain N16961. Predicted or established functions are labelled above each cluster. **b)** Close-up examination of the VC1765-69 gene cluster in strains A1552 and DRC193A reveals three genes responsible for the components of the putative T1RM system (*hsdR*, *hsdS*, *hsdM*). A comparative alignment highlights the disruption of the T1RM cluster in strain DRC193A, caused by an IS256 transposon insertion within *hsdS*.



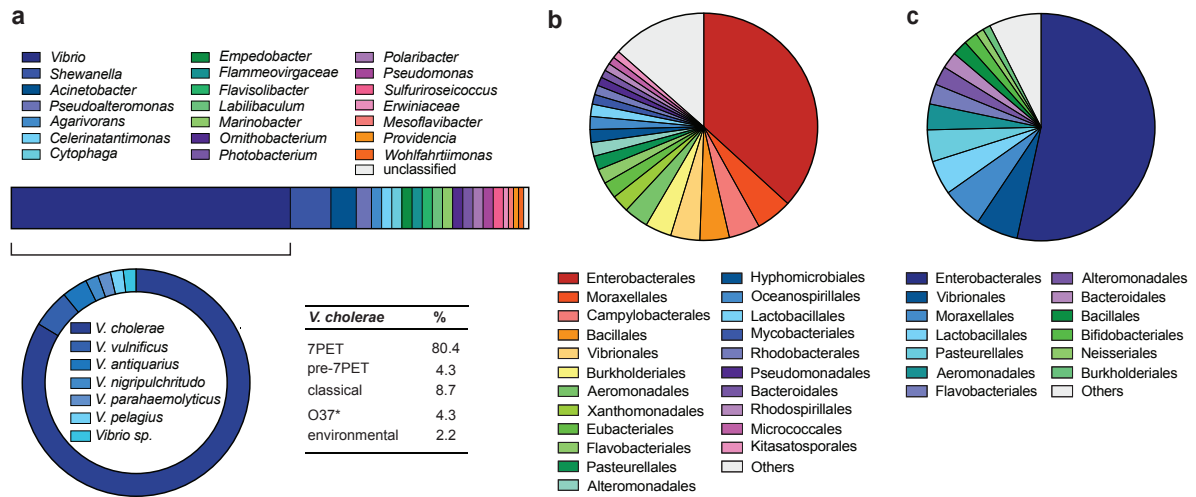
**Figure 2. Type I RM system's role in chromosomal methylation and plasmid restriction.** **a)** SMRT sequencing uncovers a distinctive modified DNA motif across various 7PET *V. cholerae* strains. The grey bars show the number of the DNA motif (GATGNNNNNCTT) in each genome, while the blue bars denote the percentage of this motif methylated in each strain's genome (see secondary Y-axis on the right). **b)** The T1RM system hinders plasmid acquisition. Transformation assays compare the uptake of a plasmid containing the recognition motif (P<sub>motif+</sub>) against a derivative plasmid with silent mutations altering the nucleotide sequence (P<sub>motif-</sub>). To the left, diagrams of the plasmids are depicted. Statistical differences were calculated on log-transformed data using a two-way ANOVA corrected for multiple comparisons with Šidák's method. \*  $P < 0.05$ ; \*\*\*  $P < 0.001$ ; ns, not significant. < d.l., below detection limit.



**Figure 3. Protection against *Tevenvirinae* by Tgv proteins encoded by the T1RM-embedded genes.** **a)** Observed defence activity against the BASEL phage collection. Protection levels (fold-protection, as shown by the colour code on the right) were determined by comparing plaque formation in strains with the system to those without, using tenfold serial dilution assays. On the left, gene organization of the tested strains. Data represent the average of two replicates. The collective results of all 77 phage infections are detailed in Fig. S1. **b)** Phage plaque assays on *E. coli* strains harboring an empty transposon (control, Ctrl) or the two T1RM-embedded genes (*tgvAB*), using a tenfold serial dilution. **c, d)** Growth curves of *E. coli* cultures carrying an empty transposon (no system) or TntgvAB (+ *tgvAB*), without (NO phage) or with exposure to phages, initiated at time 0 with various MOIs (0.2, 2, 5, or 10). **c)** *Tequatroviruses* and **d)** *Mosigviruses* were used for infection. The presented data are the average of three independent experiments ( $\pm$ SD, illustrated with error bars).



**Figure 4. The two-protein TgvAB defence system is a member of the GmrSD family of Type IV restriction enzymes. a)** Structural models of Eco94GmrSD of *E. coli* STEC\_94C and TgvA (VC1767) & TgvB (VC1766) of *V. cholerae* 7PET strains. The models, produced via AlphaFold (ColabFold), portray the domains with corresponding colours, while also highlighting the residues characteristic to the DUF262 and DUF1524 domains. Images were generated using ChimeraX 1.7.1 **b)** Schematics displaying conserved domains identified in the TgvA and TgvB proteins. **c, d)** Zoomed view of the conserved **(c)** DGQQR motif found in the DUF262 region of TgvA & TgvB and **(d)** of the His-Me finger motif within the DUF1524 of TgvB, highlighting the catalytic histidine (H) situated at the terminus of the  $\beta$ 1 strand, the Asparagine (N) residue positioned in the loop region and the final N residue within the  $\alpha$ -helix. **e)** Site-directed mutagenesis removed the antiviral effect. The level of protection was evaluated as described in Fig. 3. Mutagenesis aimed at disrupting NTPase or endonuclease functions exerted by DUF262 and DUF1524, respectively. The data are averages from three independent experiments ( $\pm$ SD, as shown by the error bars).



**Figure 5. Phylogenetic distribution of the restriction systems. a)** The presence of the 5-gene cluster (*VC1765-69*) was assessed across 81,172 bacterial genomes. The results revealed its distribution beyond *V. cholerae*, which represented 54% of all hits. \**V. cholerae* O37 serogroup strains are known to be closely related to classical O1 strains with highly similar chromosomal backbones. **b, c)** Exploration of the **(b)** T1RM system (*VC1769-68-65*) and **(c)** TgvAB system (*VC1766-67*) across the bacterial genomes demonstrates their assessment at the order level of taxonomy. Orders represented in less than 1% of instances were consolidated into a singular category labeled "Others" for the visualization. For details at the species level see Tables S1-S3.