

A Broadly Conserved Deoxycytidine Deaminase Protects Bacteria from Phage Infection

Geoffrey B. Severin^{1#®}, Brian Y. Hsueh^{2#}, Clinton A. Elg³, John A. Dover¹, Christopher R. Rhoades², Alex J. Wessel², Benjamin J. Ridenhour⁴, Eva M. Top³, Janani Ravi⁵, Kristin N. Parent¹, and Christopher M. Waters^{2*}

¹Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA, 48824

²Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA, 48824

³Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies,

Bioinformatics and Computational Biology Program, University of Idaho, Moscow, Idaho, USA,
83844

⁴Department of Mathematics and Statistical Sciences, University of Idaho, Moscow, Idaho, USA, 83844

⁵Department of Pathobiology and Diagnostic Investigation, Michigan State University, East Lansing, Michigan, USA, 48824

19 Running Title: A Broadly Conserved Cytidine Deaminase Protects Bacteria from Phage Infection

²⁰ #These authors contributed equally to this work.

21 @Current address: Department of Microbiology and Immunology, University of Michigan, Ann
22 Arbor, MI 48109

23 *Corresponding Author:

24 5180 Biomedical and Physical Sciences

25 567 Wilson Road

26 East Lansing, MI 48824

27 Telephone 517-884-5360

28 E-mail: watersc3@msu.edu

29 Keywords: cytidine deaminase, phage, genomic island, toxin-antitoxin, thymineless death,

30 APOBEC

31

32 **SUMMARY**

33 The El Tor biotype of *Vibrio cholerae* is responsible for perpetuating the longest cholera
34 pandemic in recorded history (1961-current). The genomic islands VSP-1 and -2 are two
35 understudied genetic features that distinguish El Tor from previous pandemics. To understand
36 their utility, we calculated the co-occurrence of VSP genes across bacterial genomes. This
37 analysis predicted the previously uncharacterized *vc0175*, herein renamed **deoxycytidylate**
38 **deaminase Vibrio** (*dcdV*), is in a gene network with *dncV*, a cyclic GMP-AMP synthase involved
39 in phage defense. *DcdV* consists of two domains, a P-loop kinase and a deoxycytidylate
40 deaminase, that are required for the deamination of dCTP and dCMP, inhibiting phage
41 predation by corrupting cellular nucleotide concentrations. Additionally, *DcdV* is post-
42 translationally inhibited by a unique noncoding RNA encoded 5' of the *dcdV* locus. *DcdV*
43 homologs are conserved in bacteria and eukaryotes and our results identify *V. cholerae* *DcdV*
44 as the founding member of a previously undescribed bacterial phage defense system.

45

46 **INTRODUCTION**

47 *Vibrio cholerae*, the etiological agent responsible for the diarrheal disease cholera, is a
48 motile, Gram-negative bacterium found ubiquitously in marine environments [1]. There
49 have been seven recorded pandemics of cholera, beginning in 1817, and the fifth and sixth
50 pandemics were caused by strains of the classical biotype. The seventh pandemic, which began
51 in 1961 and continues today, was initiated and perpetuated by circulating strains of the El Tor
52 biotype. Numerous phenotypic and genetic characteristics are used to distinguish the classical

53 and El Tor biotypes [2]. It is hypothesized that El Tor's acquisition of two unique genomic
54 islands of unknown origins, named the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and 2)
55 [3], played a pivotal role in El Tor's evolution to pandemicity and the displacement of the classic
56 biotype in modern cholera disease [4].

57 Combined, VSP-1 and VSP-2 encode ~36 putative open reading frames (ORFs) within
58 ~39 kb (Figs. 1A and S1B) [3, 5–7]. While the majority of the genes in these two islands remain
59 to be studied, it is hypothesized that the biological functions they encode may contribute to
60 environmental persistence [8] and/or the pathogenicity [9] of the El Tor biotype. In support of
61 this idea, VSP-1 encodes a phage defense system encompassing the genes *dncV*, *capV*,
62 *vc0180* and *vc0181* called the cyclic-oligonucleotide-based antiphage signaling system
63 (CBASS) [10] (Fig. 1A). CBASS limits phage invasion of bacterial populations via a process
64 termed abortive replication whereby upon phage infection cyclic GMP-AMP (cGAMP) synthesis
65 by DncV activates cell lysis by stimulating the phospholipase activity of CapV [10, 11]. During
66 our search for VSP-1 and 2 gene networks, we determined that the gene *vc0175*, renamed
67 herein as **deoxycytidylate deaminase in *Vibrio* (*dcdV*)**, cooccurs in bacterial genomes with
68 *dncV*, suggesting a common function.

69 We show that *dcdV*, exhibits deoxycytidylate deaminase (DCD) activity, catalyzing the
70 deamination of free deoxycytidine monophosphate (dCMP) substrates to form deoxyuridine
71 monophosphate (dUMP) and is part of the broader Zn-dependent cytosine deaminase (CDA)
72 family of enzymes [12–14]. The activity of DCD enzymes play a vital role in the de novo
73 synthesis of deoxythymidine triphosphate (dTTP) by supplying the dUMP required by
74 thymidylate synthase (TS) to form deoxythymidine monophosphate (dTDP) [12]. CDA enzymes
75 belonging to the APOBEC (Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like)
76 family also play an important role in viral immunity in higher organisms where their catalytic
77 activity is utilized for the deamination of nucleic acids rather than free nucleotide substrates to
78 restrict several types of viruses, such as retroviruses, and retroelements [15–19].

79 A primary challenge faced by lytic phage is to rapidly replicate many copies of its
80 genome, which requires sufficient nucleotide substrates [20]. During DNA phage infection, total
81 DNA within a bacterium can increase 5-10 fold, illustrating the vast amount of DNA replication
82 that must occur in a short window of time [21, 22]. To accomplish this feat, invading DNA phage
83 often corrupt the delicate balance of enzymatic activity across a host's deoxynucleotide
84 biosynthetic pathways by deploying their own DCD, dUTPase, TS, and ribonucleotide reductase
85 to ensure the appropriate ratio and abundance of deoxyribonucleotides [23–27].

86 Here we show that DcdV is a dual domain protein consisting of a putative N-terminal P-
87 loop kinase (PLK) and C-terminal DCD domain, and this novel domain architecture is present
88 across the tree of life. Overexpression of DcdV promotes cell filamentation, which has hallmarks
89 of nucleotide starvation resembling thymine-less death (TLD) toxicity [28–31]. Our results
90 demonstrate that ectopic expression of DcdV indeed corrupts the intracellular concentrations of
91 deoxynucleotides and this activity protects bacteria from phage infection. Moreover, we
92 demonstrate that DcdV activity is negatively regulated by a non-coding RNA encoded 5' of the
93 *dcdV* locus [renamed herein as **D**cdV **i**nsensitivity **f**actor in *Vibrio (difV)*]. Furthermore, *dcdV-*
94 *difV* systems are widely encoded in bacteria and we show that a subset of them function
95 similarly, establishing cytidine deaminase enzymes as antiphage defense systems in bacteria.

96

97 **RESULTS**

98 ***dncV* and *dcdV* co-occur in bacterial genomes**

99 To help identify functional interactions within the largely unclassified VSP-1 & 2 genes,
100 VSP island genes were classified into putative “gene networks” or sets of genes that form a
101 functional pathway to accomplish a biological task. Since gene networks often share deep
102 evolutionary history among diverse taxa, we hypothesized that the set of genes in a gene
103 network would co-occur together in the genomes of diverse taxa at a higher frequency than

104 chance alone would predict. Our software package was named 'Correology' inspired by [32] and
105 is described in detail in the materials and methods.

106 We calculated a Pearson correlation followed by a partial correlation correction between
107 each of the VSP island genes from the same island across the sequenced bacterial domain.
108 This resulting partial correlation correction " w_{ij} " has an output normalized to a range of -1 to 1,
109 with a w_{ij} of -1 revealing homologs of genes i and j never occur in the same species as
110 opposed to a value of 1 in which homologues of genes i and j always co-occur in the same
111 species. Previous research using well-classified *Escherichia coli* gene networks showed that
112 partial correlation values $w_{ij} > 0.045$ were highly correlated with shared biological functions [32].
113 Using the above-mentioned approach, we calculated a partial correlation value w_{ij} for all genes
114 i to j in VSP-1 (Supplemental File 1) and VSP-2 (Supplemental File 2). We generated a
115 visualization of the Maximum Relatedness Subnetworks (MRS) showing the single highest w_{ij}
116 value for each VSP gene (Figs. 1B, S1A, 1B).

117 One of our VSP-1 gene networks centered on *dncV* and identifies the experimentally
118 validated CBASS anti-phage system (Fig. 1B) [10]. Curiously, the putative deoxycytidylate
119 deaminase encoded by *vc0175*, which we renamed *dcdV*, was also found to co-occur with *dncV*
120 ($w_{ij} = 0.147$) but not with any of the other CBASS members ($w_{ij} < 0.045$) (Fig. 1B). Recognizing
121 that co-occurrence of *dncV* with *dcdV* may indicate a shared or common biological function, we
122 sought to understand the biological activity of *dcdV*.

123

124 **Ectopic expression of *dcdV* induces cell filamentation in the absence of VSP-1**

125 To assess the function of *DcdV*, we generated growth curves in both wild type (WT) *V.*
126 *cholerae* and a double VSP island deletion strain (Δ VSP-1/2) over-expressing *dcdV* (pDcdV) or
127 vector control (pVector). *DcdV* overexpression did not impact WT growth but did reduce growth

128 yield in the Δ VSP-1/2 background (Fig. 1C). We evaluated the cellular morphology of WT and
129 Δ VSP-1/2 strains after overexpression of DcdV and observed expression from pDcdV in the
130 Δ VSP-1/2 background yielded filamentous cell morphologies, suggesting these cells have a
131 defect in cell division that manifests in a reduced growth yield (Fig. 1D). We performed the same
132 image analysis in single island mutants (Δ VSP-1 and Δ VSP-2) and found that the mean cell
133 length increased significantly upon DcdV overexpression only in cells lacking VSP-1 (Fig. 1E).
134 Likewise, overexpression of pDcdV in a laboratory strain of *E. coli* also induced cell
135 filamentation that was inhibited by provision of a single copy cosmid containing VSP-1 (pCCD7)
136 but not the vector cosmid control (pLA7R) (Figs. S2A and S2B). The spiral nature of *V. cholerae*
137 filaments (Fig. 1D) is due to the natural curvature of *V. cholerae* mediated by *crvA* [33, 34].
138 Taken together, these results indicated that DcdV overexpression severely impacts cell
139 physiology in the absence of VSP-1.

140

141 **DifV is encoded immediately 5' of the *dcdV* locus in VSP-1**

142 To identify the negative regulator of DcdV activity encoded in VSP-1, we generated
143 partial VSP-1 island deletions and quantified cell filamentation following DcdV expression. Three
144 sections of VSP-1; *dcdV-vc0176*, *vspR-vc0181*, and *vc0182-vc0185*, were individually deleted.
145 Of the three partial VSP-1 deletion strains, expression of pDcdV only induced filamentation in
146 the Δ *dcdV-vc0176* mutant (Fig. 2A). Individual gene deletion mutants of *dcdV* and *vc0176*
147 maintained WT cell morphology following expression of DcdV (Fig. 2B), suggesting the 504 nt
148 intergenic region between *dcdV* and *vc0176* is the source of DcdV inhibition. We identified a
149 222 nucleotide (nt) open reading frame we named *ig²²²* encoded in the same orientation
150 immediately 5' of *dcdV* as a possible candidate for the DcdV regulation (Fig. 1A).
151 Overexpression of DcdV in the Δ *ig²²²* mutant led to cell filamentation (Fig. 2B). Furthermore,
152 complementation of *ig²²²* co-expressed from a second plasmid in the Δ *ig²²²* strain prevented
153 DcdV induced filamentation (Fig. 2B). We conclude that *ig²²²* contains the necessary genetic

154 components for inhibiting DcdV activity and refer to this negative regulator as DifV for **DcdV**
155 **insensitivity factor in *Vibrios*.**

156 As *dcdV* and *dncV* cooccur in a gene network (Fig. 1B), we hypothesized that the role of
157 DncV was to inactivate DifV, leading to the liberation of DcdV activity. However, co-expression
158 of both DncV and DcdV did not liberate DcdV activity as these cells were not filamentous (Fig.
159 S3). The Δig^{222} mutant is not filamentous in the absence of pDcdV expression which is likely
160 due to a polar effect originating from the deletion of *ig²²²*. Indeed, *dcdV* expression was reduced
161 at all growth phases in the Δig^{222} mutant (Fig. S4).

162

163 **DifV is an sRNA that post-translationally regulates the activity of DcdV**

164 The fact that DifV inhibits DcdV expressed from a plasmid with exogenous transcription
165 and translation start sites suggests DifV regulates DcdV at a post-translational level. To test this
166 hypothesis, we expressed a *dcdV* C-terminal 6x histidine tagged construct (DcdV^{6xHIS}) in WT
167 and Δig^{222} *V. cholerae* and probed for the cellular abundance of DcdV^{6xHIS} using Western blot
168 (Fig. 2C). When this tagged DcdV is expressed, Δig^{222} manifest a filamentation phenotype while
169 the WT strain does not, indicating the 6x histidine tag does not change the activity of DcdV nor
170 does it inhibit the ability of DifV to regulate DcdV (Fig. S5). Despite the lack of filamentation in
171 the WT strain, the cellular abundance of DcdV^{6xHIS} was slightly greater than Δig^{222} with an
172 average signal intensity ratio WT: Δig^{222} of 1.5 ± 0.3 across three biological replicates, although
173 this difference was not statistically significant. This result connotes that DifV limits DcdV activity
174 after it has been translated and not by reducing the abundance of DcdV.

175 Given that DifV regulates the activity of DcdV at the post-translational level, we
176 wondered if DifV was a small peptide or an untranslated small regulatory RNA (sRNA). Mutation
177 of the *ig²²²* rare CTG start codon to a TAG stop codon (222 nt^{STOP}) did not abrogate the ability of
178 this construct to inhibit DcdV activity in trans when co-expressed in the Δig^{222} strain (Fig. 2D).
179 We then examined a 174 nt ORF completely encoded within *ig²²²* (174 nt) and found it was also

180 sufficient to prevent DcdV induced filamentation (Fig. 2D). Additionally, expression of this 174 nt
181 ORF from constructs either lacking a ribosome binding site (174 nt^{RBS}) or where the native ATG
182 start codon was mutated to a TAA stop codon (174 nt^{STOP}) each retained the ability to inhibit
183 DcdV activity (Fig. 2D). We also identified an ATG start codon on the interior of the 174 nt ORF
184 corresponding to an alternative reading frame and mutation of this interior start codon to a TAA
185 stop codon (174 nt^{InteriorSTOP}) also failed to abrogate DifV inhibition of DcdV activity (Fig. 2D).
186 Together, these results suggest that translation of a gene product originating from within *ig*²²² is
187 not necessary for DifV activity.

188 To identify the minimum functional size of *difV* we further truncated this 174 nt segment
189 from both the 5' and 3' ends and found that removal of either 18 bp from the 5' end or 4 bp from
190 the 3' end was sufficient to abolish DifV activity (Fig. 2D). Additionally, expression in trans of
191 npcR_3991 [35], a 104 nt non-coding RNA of unknown function contained within *ig*²²², was also
192 unable to inhibit DcdV filamentation (Fig. 2D). Collectively, these results suggest that DifV is a
193 regulatory RNA that is between 152 to 174 nt long encoded 5' of the *dcdV* locus, and we will
194 therefore refer to the 174 nt locus as *difV* for the remainder of these experiments.

195

196 **DifV and DcdV constitute a two gene operon that resembles a Toxin-Antitoxin System**

197 The genomic orientation and proximity of *difV* to *dcdV* suggest they may constitute an
198 operon and two previous genome-wide transcriptional start site (TSS) analyses previously
199 identified a common putative TSS 5' of *difV* [36, 37]. To test if *difV* and *dcdV* are indeed
200 expressed as an operon, we performed diagnostic PCR with primers located within *difV* and
201 *dcdV* on cDNA generated from both WT and Δig^{222} RNA (Fig. 3A). As expected, *dcdV* was
202 detected in the cDNA generated from each strain while *difV* was only amplified using the WT
203 cDNA template (Fig. 3B). The presence of an 839 nt PCR product amplified using primers
204 spanning *difV* to *dcdV* from the WT cDNA template, that was not present with Δig^{222} cDNA,
205 confirmed that both genes are present on a shared transcript (Fig. 3B). Additionally, we

206 quantified the relative abundance of *difV* and *dcdV* RNA using qRT-PCR and found the *difV*
207 locus was approximately 40-, 20-, and 60-fold more abundant than *dcdV* at early exponential,
208 late exponential, and stationary phases, respectively (Fig. 3C). While having several unique
209 features, the co-transcription of *difV* and *dcdV* and the post-translational regulation of DcdV
210 activity by the abundant sRNA DifV resembles Type III Toxin-Antitoxin systems [38].

211

212 **DcdV induced filamentation requires conserved features of both the PLK and the CDA
213 domains**

214 DcdV is a 532 amino acid polypeptide composed of two putative domains: an
215 unannotated N-terminal domain and a DCD-like C-terminus (Figs. 4A, 4B). Analysis of the N-
216 terminal domain using Pfam did not reveal any conserved domains. However, Phyre2 [39] and
217 PSI-BLAST searches combined with InterProScan [40, 41] analyses revealed that the N-
218 terminus contained features of the P-loop containing nucleoside triphosphate hydrolase (IPR ID:
219 IPR027417) aka P-loop kinase (PLK) enzyme family (Figs. 4A,B and S6). PLKs catalyze the
220 reversible phosphotransfer of the γ -phosphate from a nucleotide triphosphate donor to a diverse
221 group of substrates, depending on the enzyme class, including deoxynucleotide
222 monophosphates. Three structural features commonly found in these enzymes include a P-
223 loop/Walker A motif {GxxxxGK[ST]}, a two helical LID module that stabilizes the donor
224 nucleotide triphosphates, and a Walker B motif {hhhh[D/E], where "h" represents a hydrophobic
225 residue} that is partly involved in coordinating Mg^{2+} [42, 43]. Interrogation of the Phyre2 DcdV
226 model (Fig. 4A), InterProScan predictions, and PSI-BLAST primary sequence alignments (Fig.
227 S6) revealed these three features are likely present in the N-terminal domain, suggesting the N-
228 terminus of DcdV is a PLK domain involved in binding nucleotide substrates and performing a
229 phosphotransfer reaction. The C-terminal DCD domain contains a highly conserved zinc-
230 dependent CDA active site motif ([HAE]X₂₈[PCXXC]) (Figs. 4A,B and S6). The constellation of
231 residues that make up the Zn^{2+} binding pocket is composed of three critical amino acids in

232 DcdV; H382, C411, and C414. Zn²⁺ is required for the catalytic deprotonation of water by a
233 conserved glutamate residue (E384 in DcdV) for the hydrolytic deamination of a cytosine base
234 to uridine.

235 Hypothesizing that one of the two domains present in DcdV is responsible for cell
236 filamentation in the absence of *difV*, we made site-specific mutations in the conserved residues
237 predicted to be essential for activity in both the PLK and DCD domains. Two variant constructs
238 were generated in the PLK domain targeting the Walker A motif (DcdV^{S52K}) and the Walker B
239 motif (DcdV^{D162A + Q163A}) (Fig. 4B). Two variants were constructed in the DCD active site; a
240 double substitution of both C411A and C414A (DcdV^{C411A + C414A}) to abrogate Zn²⁺ binding and
241 an E384A substitution (DcdV^{E384A}) to inhibit the deprotonation of water required for the hydrolytic
242 deamination of cytosine (Fig. 4B). Unlike WT DcdV (DcdV^{WT}), all four of the variants failed to
243 induce filamentation when ectopically expressed in *E. coli* (Fig. 4C). The cellular abundance of
244 these variants is comparable to WT DcdV (Fig. S7). This result shows both DcdV domains are
245 necessary for induction of filamentation.

246 We performed a genetic screen to identify DcdV variants whose activity was no longer
247 inhibited by DifV by expressing a random library of *dcdV* mutants in a $\Delta dcdV$ mutant strain
248 where *difV* remains intact. Ectopic expression of WT *dcdV* in a $\Delta dcdV$ mutant does not induce
249 filamentation (Fig. 4D) or produce small, wrinkled colonies on solid agar due to the genomic
250 copy of *difV*. However, *dcdV* mutants that are insensitive to *difV* exhibit a small colony
251 phenotype. Screening ~ 15,000 potential mutants, we identified five unique *dcdV* mutations that
252 encoded single amino acid substitutions (E123K, A126T, K201R, K511E, and Q514R) located in
253 both the PLD and DCD domains that rendered DcdV insensitive to DifV inhibition (Figs. 4B, 4D).
254 Based on the Phyre2 DcdV structural model, all five residues are located on the exterior of the
255 protein (Fig. 4A) suggesting they may be involved in mediating molecular interactions between
256 DifV and DcdV. The only mutation found within a conserved domain feature was the seemingly

257 innocuous K201R substitution, which is modeled to lie between the two helices of the PLK LID
258 module (Fig. 4A).

259

260 **DcdV induced filamentation is due to impaired genome replication**

261 Filamentation is a phenotype often associated with thymineless death (TLD) [28] due to
262 nucleotide starvation. A hallmark of TLD is an increased genomic origin to terminus (*ori/ter*) ratio
263 resulting from repeated attempts to initiate replication from the origin that ultimately fail to reach
264 the terminus due to a lack of dTTP substrate [44]. Hypothesizing that DcdV induced
265 filamentation may be a consequence of replication inefficiency, analogous to TLD, we measured
266 the *ori/ter* ratio of *V. cholerae* chromosome 1 from WT and Δig^{222} *V. cholerae* grown to
267 stationary phase overexpressing WT DcdV or a vector control. There was no significant
268 difference in the *ori/ter* ratios following ectopic expression of WT DcdV in WT *V. cholerae* (Fig.
269 4E), consistent with the observation that these strains do not filament (Figs. 1D and 1E).
270 However, ectopic expression of WT DcdV in the Δig^{222} mutant, which lacks *difV*, resulted in an
271 *ori/ter* ratio ~ 3 times greater than the vector control (Fig. 4F), consistent with cell filamentation
272 (Fig. 2B). We also measured the *ori/ter* ratio of the Δig^{222} mutant expressing *dcdV* with
273 mutations in the PLK or DCD domain. In agreement with an inability to induce filamentation (Fig.
274 4C), the *ori/ter* ratio of these variants was not significantly different from the empty vector control
275 (Fig. 4F). Therefore, DcdV corruption of DNA replication is dependent upon both the PLK and
276 DCD domains.

277

278 **DcdV catalyzes the deamination of both dCMP and dCTP**

279 Based on the TLD-like genome instability driven by DcdV, we hypothesized this enzyme
280 deaminates free nucleic acid substrates. Though we determined DcdV and DcdV variants were
281 readily retained in *E. coli* lysates (Fig. S7), numerous attempts to purify active DcdV were
282 unsuccessful. This suggested that an unknown cofactor or cellular condition may contribute to

283 the activity of DcdV that was missing in our purification conditions. Soluble lysates from *E. coli*
284 ectopically expressing DcdV or the DCD active site variant DcdV^{E384A} were supplemented with
285 amine containing nucleotides and monitored for the evolution of NH₄⁺, a product of nucleotide
286 deamination. Lysates containing DcdV evolved significantly more ammonium when incubated
287 with dCMP and dCTP, which was not detected in lysates containing the DCD active site variant
288 DcdV^{E384A} (Fig. 5A).

289 DCD enzymes are unique among the CDAs for their allosteric regulation by both dCTP
290 and dTTP which activate and repress the catalytic deamination of dCMP, respectively, through
291 a G[Y/W]NG allosteric site motif [45, 46]. Such allosteric regulation ensures that nucleotide
292 homeostasis is maintained even if DCD enzymes are present. The allosteric site found in DcdV
293 is composed of a divergent GCND motif suggesting allosteric regulation by dNTPs may not be
294 preserved. In support of this, the deamination of both dCMP and dCTP by soluble lysates
295 containing DcdV were not inhibited by the addition of equimolar dTTP (Fig. S8).

296 To further understand the catalytic activity of DcdV we spiked 1 μM dCTP into soluble
297 lysates collected from *E. coli* ectopically expressing either WT DcdV or a vector control and
298 quantified the concentrations of dUTP and dUMP over 30 minutes using UPLC-MS/MS.
299 Following addition of 1 μM dCTP the concentrations of both dUTP (Fig. 5B) and dUMP (Fig. 5C)
300 increased in lysates containing DcdV within the first minute while those found in vector control
301 lysates did not dramatically change over the course of the entire experiment. The concentration
302 of dUTP in DcdV containing lysates peaked after five minutes and slowly receded over time
303 (Fig. 5B) while the concentration of dUMP in these lysates continued to increase to a final
304 concentration of ~ 1 μM after 30 minutes (Fig. 5C). Importantly, the equimolar stoichiometry of
305 the 1 μM dCTP substrate spike and the 1 μM dUMP detected at the conclusion of the
306 experiment demonstrates that DcdV does not modify nucleotides in a unique manner which
307 would alter their mass. Together these experiments indicate that DcdV deaminates both dCTP

308 and dCMP substrates and DcdV containing lysates ultimately funnel dCTP to dUMP, indicating
309 DcdV is likely to have profound effects on intracellular nucleotide metabolism.

310

311 **DcdV decreases intracellular dCTP, dCMP, and dUTP in *E. coli***

312 Our genetic and in vitro evidence suggested that DcdV catalyzes the deamination of
313 both dCMP and dCTP to the detriment of DNA replication. To quantify the impact of DcdV
314 activity on the intracellular concentrations of deoxyribonucleotide species, we overproduced
315 DcdV, DcdV^{S52K}, DcdV^{E384A}, and an empty vector control in *E. coli* and measured the abundance
316 of these molecules by UPLC-MS/MS. While all strains contained similar levels of dATP, dGTP,
317 dTTP, and dUMP, the intracellular abundance of dCTP and dCMP were significantly reduced in
318 *E. coli* expressing WT DcdV (Figs. 5D, S9). No dUTP was found following expression of WT
319 DcdV while trace amounts of dUTP were detected in the vector and the two DcdV variant strains
320 (Figs. 5D, S9). Unlike the results observed with the in vitro DcdV lysates (Fig. 5C), no increase
321 in intracellular dUMP concentrations were observed when DcdV was expressed. We speculate
322 the difference between dUMP detected in lysate versus in vivo extractions are due to
323 compensatory metabolic pathways active in live cells which are lost in the lysates. Similar
324 results were obtained when a DcdV homolog derived from enterotoxigenic *E. coli* (DcdV^{ETEC}),
325 discussed later in this study, was overexpressed in the same heterologous *E. coli* host (Figs.
326 5D, S9). Importantly, inactivating amino acid substitutions in conserved features of the PLK
327 (DcdV^{S52K}) or DCD (DcdV^{E384A}) domains blocked DcdV activity, indicating both domains are
328 necessary for the DcdV dependent depletion of intracellular dC pools (Figs. 5D, S9).

329

330 **Conservation and evolution of DcdV**

331 To identify if DcdV is widely conserved, we used six DcdV homologs as starting points
332 from *V. cholerae*, *Vibrio parahaemolyticus*, *E. coli*, *Proteus mirabilis*, *Aeromonas veronii*, and
333 *Enterobacter cloacae* to perform homology searches across the tree of life (see Methods). We

334 used a combination of protein domain and orthology databases, homology searches, and
335 multiple sequence alignment for detecting domains, signal peptides, and transmembrane
336 regions to reconstruct the domain architectures of the query DcdV proteins (Fig. S6). In
337 agreement with the Phyre2 model of *V. cholerae* DcdV (Fig. 4A), we identified two distinct
338 domains in all six DcdV homologs, the N-terminal PLK domain and the C-terminal DCD domain
339 (Fig. S6).

340 We identified numerous homologs containing the core PLK+DCD architecture as well as
341 other variations, which included multiple PLK domain fusions in proteobacteria (e.g., *Klebsiella*,
342 *Vibrio*) and a nucleic acid binding domain (e.g., *Mannheimia*, *Bibersteinia*) (Table S4).
343 Homologs of DcdV were identified in multiple bacterial phyla including Proteobacteria,
344 Actinobacteria, Bacteroidetes, and Firmicutes (Figs. 6A, a few dominant clusters of homologs
345 are labeled). Interestingly, we found DcdV-like proteins in Archaea (e.g., Thaumarchaeota) and
346 Eukaryota (e.g., Ascomycota) (Figs. 6A, Table S4). While the percentage similarity is ~50% and
347 <30% for archaeal and eukaryotic homologs, respectively, we note these contain comparable
348 domain architectures to the query proteins (Table S4).

349

350 **Identification and evaluation of Gram-negative DcdV-DifV system homologs**

351 To evaluate the conservation of enzymatic activity we selected three of the core DcdV
352 homologs used in the initial homolog search; *V. parahaemolyticus* O1:Kuk FDA_R31, *P.*
353 *mirabilis* AR379, and *E. coli* H10407 ETEC (Figs. 6A and S10). Expression of all three DcdV
354 homologs in *E. coli* resulted in filamentous cells analogous to *V. cholerae* DcdV (Fig. 6B). These
355 *dcdV* homologs are encoded 3' of a small ORF, annotated as a hypothetical protein, in an
356 orientation, size, and proximity consistent with *V. cholerae* *difV*. While there was no strong
357 amino acid or nucleotide sequence similarity among the small ORFs 5' of the *dcdV* homologs
358 (Figs. S11 and S12) we hypothesized these could encode cognate *difV* negative regulators.
359 Consistent with the inhibition of DcdV activity by DifV from *V. cholerae*, co-expression of the

360 corresponding DifV with its DcdV partner suppressed the cell filamentation phenotype (Fig. 6B).
361 Additionally, overexpression of DcdV^{ETEC} in a heterologous *E. coli* host also decreased the
362 intracellular concentrations of dCMP, dCTP, and dUTP (Fig. 5D), indicating the catalytic activity
363 of these DcdV homologs are analogous to *V. cholerae* DcdV.

364 To determine if DifV and the three ORFs encoded upstream of *dcdV* homologs could
365 provide cross-species inhibition of DcdV, we challenged each of the four *dcdVs* with each of the
366 four *difVs* in *E. coli* and looked for DcdV dependent filamentation. Cross-species inhibition of
367 DcdV induced filamentation was observed between *V. parahaemolyticus* and *V. cholerae* when
368 each species' *difV* was expressed in trans (Fig. 6C). However, *difV* from *P. mirabilis* and *E. coli*
369 ETEC were only able to inhibit the activity of their own cognate DcdV (Fig. 6C). These data
370 suggest that while the general mechanism of DifV inhibition of DcdV activity is conserved the
371 specific molecular interactions that mediate this process are not.

372

373 **Ectopic expression of DcdV reduces phage titers and slows predation**

374 We initiated studies of *dcdV* based on our discovery that this gene co-occurs in bacterial
375 genomes with *dncV*, a critical member of the CBASS antiphage abortive infection system [10,
376 47]. Additionally, cytidine deaminases are conserved anti-viral defense mechanisms in
377 eukaryotes [15, 17, 48]. These connections led us to hypothesize that DcdV can also provide
378 phage defense by manipulating cellular nucleotide concentrations. To test this hypothesis, we
379 challenged *V. cholerae* WT and $\Delta dcdV$ with two *V. cholerae* lytic phage with dsDNA genomes,
380 ICP1 and ICP3 [49, 50]. However, we observed no differences in the ability of these phages to
381 kill *V. cholerae* in these conditions (Figs S13A and S13B).

382 Because ICP1 and ICP3 have coevolved with El Tor *V. cholerae*, it is likely that these
383 phages have evolved mechanisms to counteract *dcdV*. Such resistance to other *V. cholerae*
384 phage defense mechanisms by ICP-1 has been previously demonstrated [51–53]. Therefore,
385 we selected the heterologous host *Shigella flexneri*, a Gram-negative human pathogen, and its

386 bacteriophage Sf6, a dsDNA phage from the *Podoviridae* family [54, 55], as a naïve host-phage
387 pair to test the antiphage activity of DcdV and its homologs. Ectopic expression of *dcdV* or its
388 homologs did not impact the growth of *S. flexneri* before the onset of phage killing at ~110
389 minutes (Figs. 7A-D). *S. flexneri* strains ectopically expressing *dcdV* or its homologs delayed the
390 onset of population collapse caused by Sf6 predation, although the impact of the *V. cholerae*
391 DcdV was more modest than the other three homologs (Figs. 7A-D). Additionally, induction of all
392 four DcdV homologs significantly reduced Sf6 progeny following infection compared to the
393 control strains lacking induction of DcdV (Fig. 7E). Together, these data indicate that DcdV
394 enzymes confer defense against phage infection by delaying population collapse and reducing
395 the proliferation of viable phage progeny.

396

397 **DISCUSSION**

398 Uncovering the contributions to bacterial fitness of the ~36 genes encoded within the El
399 Tor *V. cholerae* VSP-1 and 2 genomic islands may help elucidate the longevity and persistence
400 of the seventh cholera pandemic. Our bioinformatic approach using Correlogeny accurately
401 identified a gene network composed of the VSP-1 antiphage CBASS system (*capV-dncV-*
402 *vc0180-vc0181*). Interestingly, this also revealed *dncV* is frequently found in genomes with the
403 previously uncharacterized gene *dcdV*. The only function previously ascribed to *dcdV* was an
404 undefined involvement in quorum sensing controlled *V. cholerae* aggregate formation [56].

405 We showed that DcdV contains a functional DCD domain that catalyzes the deamination
406 of deoxycytidine nucleotides and a putative PLK-like domain of unknown function. We further
407 demonstrate that homologs of this protein are present across the tree of life. Collectively, both
408 domains are required for DcdV to disrupt deoxynucleotide pool homeostasis, which impairs
409 DNA replication and manifests in a filamentous cell morphology. DcdV activity is post-
410 translationally regulated by DifV, a sRNA encoded immediately 5' of the *dcdV* locus in VSP-1,
411 though the details of this inhibition remain to be fully elucidated. Finally, we demonstrate that

412 DcdV and a set of homologs from other Gram-negative bacteria confer phage resistant
413 properties when expressed in a heterologous host.

414 Cell filamentation is a hallmark of TLD, observed in bacteria and eukaryotes, which
415 arises from a sudden loss of thymine during robust cellular growth [31]. Interestingly, this
416 phenomenon is not limited to dTTP as dGTP starvation elicits a similar response in *E. coli* and is
417 also hypothesized to occur when other deoxynucleotide substrates become disproportionately
418 scarce [29]. In the case of DcdV, it is conceivable the observed filamentation phenotype is a
419 consequence of a TLD-like reduction in dCTP pools that can be termed '*cytosineless death*'.
420 However, while DcdV activity also reduces the intracellular dC pool, it did not significantly
421 increase the intracellular concentrations of dTTP or dUMP in vivo, suggesting a cellular
422 compensatory pathway to combat DcdV activity is at work in intact cells. We speculate that the
423 DCD and PLK domains of DcdV are responsible for this conversion of dC nucleotides to dUMP
424 observed in the bacterial lysates, but we cannot rule out the contribution of other unknown
425 cellular factors. The deamination of dCTP is canonically performed by non-zinc dependent
426 enzymes [57] making the dual substrate repertoire of dCMP and dCTP in DcdV a rare trait.

427 The delicate balance of enzymatic activity across the pyrimidine biosynthesis pathway
428 can be corrupted by viruses that deploy their own DCD, dUTPase, and TS enzymes to hijack
429 host nucleotide biosynthesis to ensure the appropriate ratio and quantities of
430 deoxyribonucleotide precursors for replicating their own genomes [23, 24, 26, 27]. For example,
431 biDCD from chlorovirus PBCV-1, the only DCD previously reported to deaminate both dCMP
432 and dCTP substrates, rapidly catalyzes the conversion of host dC nucleic acids into dTTP thus
433 aiding replication of the A+T rich viral genome [23]. biDCD is allosterically regulated by dCTP
434 and dTTP to activate and inactivate the deaminase, respectively. This regulation provides a
435 means to fine-tune the pool of available dNTPs by preventing the enzyme from deaminating all
436 available dC substrates. Interestingly, DcdV does not appear to have maintained the allosteric
437 nucleotide binding site nor does excess dTTP added to cell lysates alter the catalytic activity of

438 DcdV towards dCMP or dCTP (Fig. S8), and we propose these differences in enzyme activity
439 are consistent with the function of DcdV as a phage defense mechanism that inhibits phage
440 replication by corrupting cellular nucleotide pools (graphical abstract). Altering pools of available
441 nucleotides has been shown to fend off biological attacks. For example, prokaryotic viperins
442 protect against T7 phage infection by producing modified ribonucleotides that ultimately inhibit
443 phage polymerase-dependent transcription [58]. The SAMHD1 phosphohydrolase enzyme in
444 eukaryotes also inhibits viral infections by depleting cellular nucleotide pools, although its
445 structure and activity are different than DcdV [59–61].

446 In lieu of a conserved deoxynucleotide allosteric site, DcdV is regulated post-
447 translationally by the DifV untranslated RNA, which is unique among the CDA-family. The
448 spacing, orientation, and relationship of *difV* and *dcdV* may have adapted to perform functions
449 in a manner analogous to Type 2 and Type 3 Toxin-Antitoxin (TA) systems found across the
450 bacterial phyla of which some are involved in antiphage defense and bacterial stress response
451 [62]. While the RNA antitoxin of Type 3 TA systems encode nucleotide repeats [62] no repeat
452 sequences are obvious in DifV indicating that DcdV/DifV may constitute a new TA class. We
453 hypothesize that DcdV is activated upon phage infection by disruption of DifV inhibition, and we
454 are currently performing experiments to test this hypothesis (graphical abstract). Our systemic
455 search for DcdV homologs containing at least a single PLK and DCD domain revealed hundreds
456 of examples in a variety of bacteria beyond the Proteobacteria phylum including Bacteroidetes
457 and Actinobacteria and a few homologs in archaea and eukaryota.

458 Phage defense mechanisms are often found clustered together in mobile genetic
459 elements called defense islands [63, 64] and we speculate that the co-occurrence of DcdV and
460 DncV (along with the rest of the CBASS system) in bacterial genomes is a result of their shared
461 anti-phage activity. Our results indicate that DcdV reduces the available dC pool, and we
462 hypothesize that this activity delays phage genome replication potentially decreasing phage
463 burst size. Although the *S. flexneri* host population expressing DcdV eventually collapses, we

464 speculate that the delay in phage replication could provide an opportunity to prompt other phage
465 defense systems, such as CBASS or a restriction modifications system to further target invading
466 phages [65, 66].

467 Our study reveals that bacteria, like eukaryotes, also use CDA enzymes to protect
468 against biological invasion although through different mechanisms. The eukaryotic APOBEC
469 proteins deaminate ssRNA, leading to increased mutation and decreased genome stability of
470 RNA viruses, whereas the substrates of DcdV are free deoxynucleotides. Further studies are
471 required to determine if these two biological defense systems evolved from a common CDA
472 ancestor.

473

474 **ACKNOWLEDGEMENTS**

475 We thank Shannon Manning (STEC Center, Michigan State University), Jessica Jones (U.S
476 FDA), and Allison Brown (U.S. CDC), for providing us with *E. coli* ETEC, *P. mirabilis*, and *V.*
477 *parahaemolyticus* strains, respectively. We thank Wei Leung Ng (Tuft University) for providing
478 us *V. cholerae* ICP1 and ICP3 phages. We thank Kefei Yu and Dohun Pyeon for valuable
479 suggestions and Dan Jones and Lijun Chen from the MSU RTSF mass spectrometry facility
480 core for their technical support. This work was supported by National Institutes of Health (NIH)
481 grants GM109259, GM110444, GM139537, AI143098, and AI158433 and National Science
482 Foundation (NSF) grant DBI-0939454 to C.M.W, the NIH grant GM110185 and the NSF
483 CAREER Award 1750125 to K.N.P., and the NSF Graduate Research Fellowship Grant No.
484 1842399 to C.A.E. Any opinions, findings, and conclusions or recommendations expressed in
485 this material are those of the author(s) and do not necessarily reflect the views of the National
486 Science Foundation.

487

488

489

490 **MATERIALS AND METHODS**

491 The strains, plasmids, and primers used in this study are listed in Supplementary Table
492 S1, S2, and S3, respectively. Unless otherwise stated, cultures were grown in Luria-Bertani (LB)
493 at 35°C and supplemented with the following as needed: ampicillin (100 µg/mL), kanamycin
494 (100 µg/mL), tetracycline (10 µg/mL), and isopropyl-β-D-thiogalactoside (IPTG) (100 µg/mL). *E.*
495 *coli* BW29427, a diaminopimelic acid (DAP) auxotroph, was additionally supplemented with 300
496 µg/mL DAP. The *V. cholerae* El Tor biotype strain C6706str2 was utilized in this study and
497 mutant strains were generated using the pKAS32 suicide vector [67] using three fragments: 500
498 bp of sequence upstream of the gene of interest, 500 bp of sequence downstream of the gene
499 of interest and cloned into the KpnI and SacI restriction sites of pKAS32 using by Gibson
500 Assembly (NEB). P_{tac} inducible expression vectors were constructed by Gibson Assembly with
501 inserts amplified by PCR and pEVS143 [68] or pMMB67EH [69] each linearized by EcoRI and
502 BamHI, as well as pET28b digested with Ncol and Xhol. pEVS141 [70] is used as an empty
503 vector control for experiments using pEVS143 derived constructs. Site-directed mutagenesis
504 was performed using the SPRINP method [71]. Plasmids were introduced into *V. cholerae*
505 through biparental conjugation using an *E. coli* BW29427 donor. Transformation of *E. coli* for
506 ectopic expression experiments was performed using electroporation with DH10b for expression
507 of pEVS143 and pMMB67EH derived plasmids and BL21(DE3) for pET28b based constructs.
508

509 **Correology Bioinformatics Analysis**

510 Our Correology software package is built on Kim and Price's approach [32] to calculate
511 genetic co-occurrence. The source code, documentation, and a Docker container for this
512 Python3 package are available at <https://github.com/clinte14/correology>. While VSP-1 is used to
513 simplify the description of the method detailed below, both VSP-1 and 2 were independently
514 analyzed in the same fashion. To establish maximum related subnetworks (MRS) for the
515 genomic region of the VSP-1 island, a BLASTP amino acid sequence was performed to search

516 for each VSP-1 gene against the NCBI non-redundant protein database with an E-value cutoff
517 of 10^{-4} . The BLAST results were limited to bacterial genomes, and all taxa belonging to the
518 genus *Vibrio* were removed to avoid bias from closely related vertical inheritance. The BLAST
519 results were used to generate a presence or absence matrix of VSP-1 homologues with all
520 species along one axis and VSP-1 genes along the other axis. Next, a pairwise Pearson
521 correlation value was calculated between all VSP-1 genes i and j using binary data from the
522 above-mentioned presence/absence matrix:

523

$$r_{ij} = \frac{C_{ij}N - E_iE_j}{\sqrt{E_iE_j(N - E_i)(N - E_j)}}$$

524 where N is the total number of unique species returned from the BLAST search and C_{ij} the
525 number of species with co-occurrence of genes i and j . While a Pearson correlation is warranted
526 for a normally distributed binary data set, it does not account for indirect correlation. For
527 example, if genes i and j individually associate with a third gene, a Pearson correlation will
528 incorrectly calculate a correlation between i and j . To help correct for indirect correlation we
529 calculate a partial correlation w_{ij} from the Pearson r_{ij} :

530

$$w_{ij} = \frac{P_{ij}}{\sqrt{P_{ii}P_{jj}}}$$

531 where the (i, j) element of the inverse matrix of Pearson r_{ij} is P_{ij} [32].

532 The partial correlation correction w_{ij} has the advantage of generating a normalized output
533 that ranges between -1 to 1. For example, a w_{ij} of -1 reveals genes i and j never occur in the
534 same species, while a value of 1 demonstrates genes i and j always co-occur in the same
535 species. A w_{ij} of 0 is the amount of co-occurrence expected between unrelated genes i and j
536 drawn from a normal distribution. Using the above-mentioned approach, a partial correlation
537 value w_{ij} was calculated for all genes i to j in VSP-1 and VSP-2 (Supplemental Files 1 and 2).
538 The single highest w_{ij} value for each VSP-1 gene was represented as an edge (i.e., line) in our

539 visualization (Fig. 1B, S1A, and S1B). Any set of genes that contains no further edges were
540 assigned to a unique MRS that suggests functional association of the gene products within a
541 unique gene network.

542

543 **Genomic Identification, Structural, and Sequence Analyses of DcdV & DifV Homologs**

544 DcdV from *V. cholerae* El Tor N16961 ([WP_001901328.1](#)) was identified as locus tag
545 *vc0175*. DcdV and homologs profiles are performed using translated BLAST tblastn and run
546 against the Nucleotide collection (nr/nt) database of National Center for Biotechnology
547 Information (NCBI), using >40% similarities cutoff. For previously annotated domains, the Pfam
548 feature in KEGG [72, 73] were utilized as a guide to determine DcdV homologs. Out of all the
549 DcdV homologs, DcdV homologs from *Vibrio parahaemolyticus* O1: Kuk str. FDA_R31
550 ([WP_020839904.1](#)), *Proteus mirabilis* AR_379 ([WP_108717204.1](#)), and *E. coli* O78:H11
551 H10407 (ETEC) ([WP_096882215.1](#)) were analyzed in this study. Genomic contextual
552 information from prokaryotic gene neighborhoods was retrieved from NCBI genome graphics
553 feature to uncover *difV*-like gene, encoded as a hypothetical ORF 5' of the *dcdV* locus. If
554 unannotated, the ORFfinder feature from NCBI was used to determine the location and size of
555 the putative *difV* locus. To predict the structure of DcdV from *V. cholerae*, the amino acid
556 sequence was submitted to Phyre2 [39] and structural visualization was performed using PyMol
557 (<https://pymol.org>). The amino acid and nucleotide alignments were analyzed using ClustalW
558 Omega from EMBL-EBI web services [74] and LocARNA [75], respectively.

559

560 **Identification and Characterization of Protein Homologs**

561 **Homology searches:** To ensure the identification of a comprehensive set of homologs
562 (close and remote), we started with six representative DcdV proteins across proteobacteria from
563 *V. cholerae*, *V. parahaemolyticus*, *P. mirabilis*, and *E. coli* described above along with *E.*
564 *cloacae* ([WP_129996984.1](#)), and *A. veronii* ([WP_043825948.1](#)) and performed homolog

565 searches using DELTABLAST [76] against all sequenced genomes across the tree of life in the
566 NCBI RefSeq database [77–79]. Homology searches were conducted for each protein and the
567 search results were aggregated; the numbers of homologs per species and of genomes carrying
568 each of the query proteins were recorded. These proteins were clustered into orthologous
569 families using the similarity-based clustering program BLASTCLUST [76].

570 **Characterizing homologous proteins:** Phyre2, InterProScan, HHpred, SignalP,
571 TMHMM, Phobius, Pfam, and custom profile databases [39–41, 80–85] were used to identify
572 signal peptides, transmembrane (TM) regions, known domains, and secondary structures of
573 proteins in every genome. Custom scripts were written to consolidate the results [86–91], and
574 the domain architectures and protein function predictions were visualized using the MolEvolVR
575 web-app (<http://iravilab.org/molevolvr/>).

576 **Phylogenetic analysis (MSA and Tree):** Thousands of homologs from all six starting
577 points for DcdV proteins were consolidated and representatives were chosen from distinct
578 Lineages and Genera, containing both the N- and C-terminal DcdV domains (PLK and DCD
579 domains). Multiple sequence alignment (MSA) of the identified homologs was performed using
580 Kalign [89] and MUSCLE [92, 93] (msa R package [94]). The phylogenetic trees were
581 constructed using FastTree [95] FigTree [96] and the R package, ape [97].

582

583 **Growth Curve Assays**

584 Overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG
585 in a 96-well microplate (Costar®). Growth was monitored by measuring OD₆₀₀ every 15 minutes
586 for 15 hour (h) using a BioTek plate reader with continuous, linear shaking.

587

588 **Fluorescence Microscopy and Analysis**

589 Cells were imaged as previously described [34]. Briefly, overnight cultures were diluted
590 1:1000 into LB supplemented with antibiotics and IPTG. Cultures were grown and induced for 7-

591 8 h, at which point cells were diluted to an OD₆₀₀ of 0.5 in 1X PBS, then membrane stain FM4-
592 64 dye (ThermoFisher Scientific) was added to a final concentration of 20 µg/mL. 1% agarose
593 pads in deionized water were cut into squares of approximately 20 x 20 mm and placed on
594 microscope slides. 2 µl of diluted cultures were spotted onto a glass coverslip and then gently
595 placed onto the agarose pad. FM4-64 signal was visualized using a Leica DM5000b
596 epifluorescence microscope with a 100X-brightfield objective under RFP fluorescence channel.
597 Images were captured using a Spot Pursuit CCD camera and an X-cite 120 Illumination system.
598 Each slide was imaged with at least 20 fields of view for each biological replicate. Cell lengths
599 were processed using the Fiji plugin MicrobeJ [98, 99], and data were visualized and analyzed
600 using R [90] by quantifying the length of the curvilinear (medial) axis of detected cells.

601

602 **Construction and screening of mutant gene libraries**

603 DifV-insensitive DcdV constructs were generated by error-prone PCR (epPCR) using
604 pDcdV (pCMW204) as the template. Three different concentrations of MnCl₂ (12.5 mM, 1.25
605 mM, and 125 µM) were used in triplicate using Taq polymerase (Invitrogen) and reactions
606 containing the same MnCl₂ concentration were pooled. The PCR products were purified, using
607 The Wizard® SV Gel and PCR Clean-Up Kit (Promega), and ligated to pEVS143 via Gibson
608 Assembly. The assembled reactions were electroporated to *E. coli* DH10b and plasmid libraries
609 were collected from ~ 30,000 representative colonies for each MnCl₂ concentration. Plasmid
610 libraries were harvested using the Wizard® Plus SV Minipreps DNA purification Kit (Promega).
611 Plasmid libraries were subsequently electroporated to *E. coli* BW29427 which were again plated
612 and pooled to contain ~ 30,000 representative colonies. The *E. coli* BW29427 random mutant
613 pDcdV libraries were conjugated with Δ dcdV *V. cholerae* on LB agar plates for 8 h, harvested,
614 diluted, and spread on LB agar plates containing 1 mM IPTG and antibiotics, and grown
615 overnight. ~ 5,000 colonies were screened in each library and all colonies exhibiting a wrinkled
616 and small colony morphology, indicative of cell filamentation, were isolated and filamentation

617 was confirmed by fluorescence microscopy. Mutant pDcdV plasmids recovered from cells
618 exhibiting cell filamentation were sequenced by Sanger sequencing. Mutations were
619 reintroduced individually into the WT pDcdV construct using SPRINP mutagenesis [71] and
620 reevaluated using fluorescence microscopy to confirm the DcdV variant's ability to remain
621 constitutively active in $\Delta dcdV$ *V. cholerae*.

622

623 **RNA Isolation, qRT-PCR, and Co-transcription Analysis**

624 RNA isolation and qRT-PCR analysis were carried out as previously described [100,
625 101]. Briefly, triplicate overnight cultures were subcultured 1:1000 in 10 mL LB and grown to
626 three different OD₆₀₀: 0.2 (Early Exponential), 1.0 (Late Exponential), and 2.5 (Stationary). 1 mL
627 of each replicate was pelleted, and RNA was extracted using TRIzol[®] reagent following the
628 manufacturer's directions (Thermo Fischer Scientific). RNA quality and quantity were
629 determined using a NanoDrop spectrophotometer (Thermo Fischer Scientific). 5 μ g of purified
630 RNA was treated with DNase (TurboTM DNase, Thermo Fischer Scientific). cDNA synthesis was
631 carried out using SuperScriptTM III Reverse Transcriptase (Thermo Fischer Scientific). cDNA
632 was diluted 1:64 into molecular biology grade water and amplification was quantified using 2x
633 SYBR Green (Applied BiosystemsTM). For measuring gene expressions or determining *ori/ter*
634 ratios, 25 μ L reactions consisted of 5 μ L each of 0.625 μ M primers 1 and 2, 12.5 μ L of 2X SYBR
635 master mix, and 2.5 μ L of template (0.78 ng/ μ L cDNA for gene expression and 0.25 ng/ μ L
636 genomic DNA for *ori/ter*). qRT-PCR reactions were performed in technical duplicates for
637 biological triplicate samples and included no reverse transcriptase reaction controls ("no RT") to
638 monitor for contaminating genomic DNA in purified RNA samples. qRT-PCR reaction thermo
639 profile was 95°C for 20 seconds (s) then 40 cycles of 95°C for 2 s and 60°C for 30 s in the
640 QuantStudio 3 Real-Time PCR system (Applied BiosystemsTM). The *gyrA* gene was used as an
641 endogenous control to calculate relative quantification (ΔC_t).

642 To determine the co-transcription of *difV* and *dcdV*, PCR amplification was performed in
643 25 μ L volumes using Q5 polymerase (NEB), 0.5 μ M each of the forward and reverse primers as
644 indicated, 0.2 mM dNTPs, and 3.5 μ L of cDNA or no RT control templates (0.78 ng/ μ L) from
645 RNA purified from WT and Δ ig²²² *V. cholerae* grown to late exponential-phase in biological
646 triplicate. The thermal profile was 98°C for 30 s, 30 cycles of 98°C for 10 s, 55 °C for 30 s, 72 °C
647 for 10 sec and one cycle of 72 °C for 2 min. PCR products were loaded on a 1% agarose gel
648 and stained with EZ-Vision® (VWR). Images were taken using the GelDoc system (Bio-Rad).

649

650 **In-vitro Nucleic Acid Deamination Assay**

651 **Cell Lysate Preparation:** Overnight cultures were subcultured 1:333 and grown to an
652 OD₆₀₀ of ~0.5 - 1.0. Cultures were induced with 1 mM IPTG, supplemented with 100 μ M ZnSO₄,
653 and grown for an additional 3 hr. Cell pellets from 100 mL of induced cultures were harvested in
654 two successive 15 min centrifugation steps at 4,000 x g and 4°C. Supernatants were decanted
655 and pellets were snap frozen in an ethanol and dry ice bath and stored at -80° C. Pellets were
656 thawed on ice and suspended in 2 mL of lysis buffer A (50 mM NaPO₄, pH 7.3, 300 mM NaCl, 2
657 mM β -mercaptoethanol, 20% glycerol and Roche cOmplete protease inhibitor (1 tablet per 10
658 mL)). 1 mL of cell suspension was transferred to a microcentrifuge tube and sonicated on ice
659 using a Branson 450 Digital Sonifier (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off).
660 Crude lysates were centrifuged at 15,000 x g for 10 min at 4°C and clarified lysates were
661 transferred to fresh microcentrifuge tubes on ice. Clarified lysates were normalized for total
662 protein to 1.9 mg/mL using Bradford reagents and a BSA standard. 26.5 μ L reactions composed
663 of lysis buffer A, nucleic acid substrates, and 3.5 μ L of normalized clarified lysates were
664 assembled in PCR strip tubes, mixed by gentle pipetting, and incubated at room temperature
665 (~23°C) for 60 minutes. NH₄Cl solutions at the indicated concentration were dissolved in lysis
666 buffer A and substituted for nucleic acid substrates as positive ammonium controls.

667 **Ammonium Detection:** The evolution of NH₄⁺ by deamination of the nucleic acid
668 substrates was observed using a phenol-hypochlorite reaction to produce indophenol in a clear
669 96-well microtiter plate and modified from Dong et al. 2015 [102]. The work of Ngo et al. [103]
670 was considered when designing the lysis buffer so as not to interfere with the phenol-
671 hypochlorite reaction. 50 µL of Reagent A (composition below) was added to each well followed
672 by 20 µL of the completed in vitro deamination reaction described above. The phenol-
673 hypochlorite reaction was initiated by the addition and gentle mixing of 50 µL Reagent B
674 (composition below) to the wells. The reaction was incubated at 35°C for 30 min and the ABS₆₃₀
675 was measured using a plate reader.

676 *Reagent A* = 1:1 (v/v), 6% (w/v) sodium hydroxide (Sigma) in water: 1.5% (v/v) sodium
677 hypochlorite solution (Sigma, reagent grade) in water.

678 *Reagent B* = 1:1:0.04 (v/v/v), water: 0.5% (w/v) sodium nitroprusside (Sigma) in water:
679 phenol solution (Sigma, P4557)

680

681 **Western Blot**

682 Strains containing DcdV- and variant- C-terminal 6x-histidine fusions were grown,
683 induced, and harvested as described previously above (See In-vitro Nucleic Acid Deamination
684 Assay: Cell Lysate Prep), except for the His-tag fusion (pGBS98) which are induced for only 2 h
685 with 100 µM IPTG and not subjected to sonication. The cell pellets were resuspended in 2 mL of
686 chilled 1X PBS and subsequently normalized to OD of 1.0. 1 mL aliquots were collected by
687 centrifugation at 15k x g for 1 min. Cell pellets were subsequently resuspended in 90 µL of lysis
688 buffer A and 30 µL of 4x Laemmli buffer, denatured for 10 minutes at 65°C, and centrifuged at
689 15k x g for 10 minutes. 5 µL of samples were loaded into a precast 4-20% sodium dodecyl
690 sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (Mini-PROTEAN TGX Precast
691 Protein Gels, Bio-Rad) alongside size standards (Precision Protein Plus, Bio-Rad). Gels were
692 run at room temperature for 90 min at 100 V in 1x Tris/glycine/SDS running buffer. Proteins

693 were transferred to nitrocellulose membranes (Optitran). The membranes were blocked using
694 5% skim milk and incubated with 1:5000 THE™ His Tag Antibody, mAb, Mouse (GenScript)
695 followed by 1:4000 Goat Anti-Mouse IgG Antibody (H&L) [HRP], pAb (GenScript), treated with
696 Pierce™ ECL Western Blotting Substrate, and imaged using an Amersham™ Imager 600.

697

698 **UPLC-MS/MS Quantification of In Vitro and In Vivo Deoxynucleotides**

699 Deoxynucleotide concentrations were determined as previously described [104] with
700 minor modifications. For measuring in vivo intracellular deoxynucleotide concentrations,
701 overnight cultures were subcultured 1:1000 and grown to OD₆₀₀ of ~1.0. Plasmid expression
702 was induced by the addition of 1 mM IPTG for 1 h, and 1 mL of cultures were collected by
703 centrifugation at 15,000 x g for 1 min. Cell pellets were resuspended in 200 µL of chilled
704 extraction buffer [acetonitrile, methanol, ultra-pure water, formic acid (2:2:1:0.02, v/v/v/v)]. To
705 normalize in vivo nucleotide samples, an additional cell pellet was collected from 1 mL of culture
706 by centrifugation at 15,000 x g for 1 min, resuspended in 200 µL lysis buffer B (20 mM Tris·HCl,
707 1% SDS, pH 6.8), and denatured for 10 minutes at 60°C. Denatured lysates were centrifuged at
708 15,000 x g for 1 min to pellet cellular debris, and the supernatant was used to quantify the total
709 protein concentration in the sample using the DC protein assay (Bio-Rad) a BSA standard curve
710 [34]. The concentrations of deoxynucleotides detected by UPLC-MS/MS were then normalized
711 to total protein in each sample.

712 For the quantification of deoxynucleotides in vitro *E. coli* BL21(DE3) clarified lysates
713 were prepared as described for the deamination experiment above and normalized to 20 mg/mL
714 of total protein and 200 µL of normalized clarified lysates were assembled in PCR strip tubes.
715 To measure abundance of dUMP and dUTP prior to the addition of 1 µM dCTP, 20 µL of
716 normalized clarified lysates were added to 200 µL of chilled extraction buffer. 20 µL of 10 µM
717 dCTP was then added to the remaining clarified lysates and 20 µL lysates aliquots were

718 removed 1, 5, 10, and 30 minutes after the addition of dCTP and mixed in 200 μ L chilled
719 extraction buffer.

720 All samples resuspended in extraction buffer, in vivo and in vitro, were immediately
721 incubated at -20°C for 30 minutes after collection and centrifuged at 15,000 x g for 1 min. The
722 supernatant was transferred to a new tube, dried overnight in a speed vacuum, and finally
723 resuspended in 100 μ L ultra-pure water. Experimental samples and deoxynucleotides standards
724 [1.9, 3.9, 7.8, 15.6, 31.3, 62.5, and 125 nM of dATP (Invitrogen), dGTP (Invitrogen), dTTP,
725 (Invitrogen), dCTP (Invitrogen), dCMP (Sigma), dUTP (Sigma), and dUMP (Sigma)] were
726 analyzed by UPLC-MS/MS using an Acquity Ultra Performance LC system (Waters) coupled
727 with a Xevo TQ-S mass spectrometer (Waters) with an ESI source in negative ion mode. The
728 MS parameters were as follows: capillary voltage, 1.0 kV; source temperature, 150°C;
729 desolvation temperature, 400°C; cone gas, 120 L/hr. Five microliter of each sample was
730 separated in reverse phase using Acquity UPLC Premier BEH C18, 2.1 x 100 mm, 1.7 μ m
731 particle size, VanGuard FIT at a flow rate of 0.3 mL/min with the following gradient of solvent A
732 (8mM DMHA (N,N-dimethylhexylamine) + 2.8 mM acetic acid in water, pH~9) to solvent B
733 (methanol): $t = 0$ min; A-100%:B-0%, $t = 10$ min; A-60%:B-40%, $t = 10.5$; A-100%:B-0%, $t = 15$
734 min; A-100%:B-0% (end of gradient). The conditions of the MRM transitions were as follows
735 [cone voltage (V), collision energy (eV)]: dATP, 490 > 159 (34, 34); dCTP, 466 > 159 (34, 34);
736 dGTP, 506 > 159 (15, 46); dTTP, 481 > 159 (25, 34); dUTP, 467 > 159 (25, 34); dCMP, 306 >
737 97 (43, 22); dUMP, 306 > 111 (22, 22).

738

739 **Phage Infection and Plaque Assays**

740 *V. cholerae* phages ICP1 and ICP3 were provided by Wai-Leung Ng at Tuft University
741 School of Medicine. ICP1 was propagated on *V. cholerae* E7946, while ICP3 were propagated
742 on *V. cholerae* C6706str2 in LB, and their titer was determined using the small drop plaque
743 assay method, as previously described [10]. Briefly, 1 ml of overnight cultures were mixed with 9

744 ml of MMB agar (LB + 0.1 mM MnCl₂ + 5 mM MgCl₂ + 5 mM CaCl₂ + 0.5% agar), tenfold serial
745 dilutions of phages in MMB were dropped on top of them, and incubated overnight at 35°C. The
746 viral titer is expressed as plaque forming units per mL (pfu/mL). 4 mL of *V. cholerae* overnight
747 cultures were diluted 1:1000 in MMB medium. 145 µL of the diluted cultures, in three sets of
748 biological replicates, were transferred and incubated at 35°C in a 96-well microplate (Costar®).
749 Once the OD₆₀₀ reached ~0.1, 5 µL of phages with a final MOI of 0.1 were added to each
750 biological replicate. Cultures were infected at room temperature (~23°C) for 12 h in a
751 SpectraMax M5 Plate Reader with continuous shaking and OD₆₀₀ measurements taken every
752 2.5 min.

753 *Shigella flexneri* strain PE577 [54] cells transformed with the pVector (pMMB67eh) and
754 each of the associated pDcdV plasmids were grown in LB medium and incubated with aeration
755 at 37° C overnight. The following day, 20 µL of each of the overnight cultures were used to
756 inoculate fresh medium in a 96-well microtiter plate with a final volume of 200 µL/well.
757 Depending on the experimental condition, wells were supplemented with and without IPTG (100
758 µM final concentration) and/or phage Sf6 [55] at an MOI of 0.1 phage per cell. Initial cell
759 densities of the overnight cultures were experimentally determined by plating and found to be
760 within a factor of two of one another. For all experiments, three biological replicates were tested.
761 Additionally, the plates were set up with each unique condition having three technical replicates.
762 Plate reader assays were conducted using a Molecular Devices FilterMax F5 plate reader, as
763 previously described [105]. Briefly, the plates were incubated at 37°C for 6 h. Every five
764 minutes, the plate was mixed and aerated by orbital shaking before an absorbance (595 nm)
765 reading was taken. After the kinetic assay was complete an aliquot from each of the replicates
766 was removed and used to determine the endpoint titer via plaque assay.

767

768

769

770 **Statistical Analysis**

771 As specified in the figure legends, all of the statistical analyses for the violin plots were
772 performed with R statistical computing software [90], while other data were analyzed in
773 GraphPad Prism Software. Statistically significances denote as the following: a single asterisk
774 (*) indicates $p < 0.05$; double asterisks (**) indicate $p < 0.01$; triple asterisks (***) indicate $p <$
775 0.001 ; and quadruple asterisks (****) indicate $p < 0.0001$. Means \pm SEM and specific n values
776 are reported in each figure legend.

777

778 **REFERENCE**

- 779 1. Thompson FL, Iida T, Swings J (2004) Biodiversity of *Vibrios*. *Microbiol Mol Biol Rev*
780 68:403–431 . <https://doi.org/10.1128/MMBR.68.3.403-431.2004>
- 781 2. Son MS, Taylor RK (2011) Genetic Screens and Biochemical Assays to Characterize
782 *Vibrio cholerae* O1 Biotypes: Classical and El Tor. *Curr Protoc Microbiol* 22:6A.2.1-
783 6A.2.17 . <https://doi.org/10.1002/9780471729259.mc06a02s22>
- 784 3. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ (2002)
785 Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera
786 endemic and pandemic disease. *Proc Natl Acad Sci* 99:1556–1561 .
787 <https://doi.org/10.1073/pnas.042667999>
- 788 4. Hu D, Liu B, Feng L, Ding P, Guo X, Wang M, Cao B, Reeves PR, Wang L (2016) Origins
789 of the current seventh cholera pandemic. *Proc Natl Acad Sci* 113:E7730–E7739 .
790 <https://doi.org/10.1073/pnas.1608732113>
- 791 5. O’Shea YA, Finn S, Reen FJ, Morrissey JP, O’Gara F, Boyd EF (2004) The *Vibrio*
792 seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor
793 and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V.*
794 *vulnificus*. *Microbiology* 150:4053–4063 . <https://doi.org/10.1099/mic.0.27172-0>

795 6. Murphy RA, Boyd EF (2008) Three Pathogenicity Islands of *Vibrio cholerae* Can Excise
796 from the Chromosome and Form Circular Intermediates. *J Bacteriol* 190:636–647 .
797 <https://doi.org/10.1128/JB.00562-07>

798 7. Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, Lanata CF, Hall E, Miranda H,
799 Huapaya B, Vargas G. C, Luna MA, Sack DA, Yamasaki S, Nair GB (2009) Peruvian
800 *Vibrio cholerae* O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic
801 island-II that differentiates them from the prototype seventh pandemic El Tor strains. *J*
802 *Med Microbiol* 58:342–354 . <https://doi.org/10.1099/jmm.0.005397-0>

803 8. Beyhan S, Tischler AD, Camilli A, Yildiz FH (2006) Differences in Gene Expression
804 between the Classical and El Tor Biotypes of *Vibrio cholerae* O1. *Infect Immun* 74:3633–
805 3642 . <https://doi.org/10.1128/IAI.01750-05>

806 9. Davies BW, Bogard RW, Young TS, Mekalanos JJ (2012) Coordinated Regulation of
807 Accessory Genetic Elements Produces Cyclic Di-Nucleotides for *V. cholerae* Virulence.
808 *Cell* 149:358–370 . <https://doi.org/10.1016/j.cell.2012.01.053>

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

812 11. Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger A-K, Khataokar A,
813 O'Hara BJ, Behrmann L V., Neiditch MB, Benning C, Waters CM, Ng W-L (2018) Direct
814 activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc Natl*
815 *Acad Sci* 115:E6048–E6055 . <https://doi.org/10.1073/pnas.1801233115>

816 12. O'Donovan GA, Neuhard J (1970) Pyrimidine metabolism in microorganisms. *Bacteriol*
817 *Rev* 34:278–343 . <https://doi.org/10.1128/MMBR.34.3.278-343.1970>

818 13. Weiss B (2007) The Deoxycytidine Pathway for Thymidylate Synthesis in *Escherichia*
819 *coli*. *J Bacteriol* 189:7922–7926 . <https://doi.org/10.1128/JB.00461-07>

820 14. Weiss B, Wang L (1994) De novo synthesis of thymidylate via deoxycytidine in dcd

821 (dCTP deaminase) mutants of *Escherichia coli*. *J Bacteriol* 176:2194–2199 .

822 <https://doi.org/10.1128/JB.176.8.2194-2199.1994>

823 15. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L (2018) Diversification of

824 AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in

825 immunity. *Proc Natl Acad Sci* 115:E3201–E3210 .

826 <https://doi.org/10.1073/pnas.1720897115>

827 16. Chiu Y-L, Greene WC (2008) The APOBEC3 Cytidine Deaminases: An Innate Defensive

828 Network Opposing Exogenous Retroviruses and Endogenous Retroelements. *Annu Rev*

829 *Immunol* 26:317–353 . <https://doi.org/10.1146/annurev.immunol.26.021607.090350>

830 17. Lada AG, Frahm Krick C, Kozmin SG, Mayorov VI, Karpova TS, Rogozin IB, Pavlov YI

831 (2011) Mutator effects and mutation signatures of editing deaminases produced in

832 bacteria and yeast. *Biochem* 76:131–146 . <https://doi.org/10.1134/S0006297911010135>

833 18. Stavrou S, Ross SR (2015) APOBEC3 Proteins in Viral Immunity. *J Immunol* 195:4565–

834 70 . <https://doi.org/10.4049/jimmunol.1501504>

835 19. Sharma S, Patnaik SK, Thomas Taggart R, Kannisto ED, Enriquez SM, Gollnick P,

836 Baysal BE (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes

837 and macrophages. *Nat Commun* 6:6881 . <https://doi.org/10.1038/ncomms7881>

838 20. Kreuzer KN, Brister JR (2010) Initiation of bacteriophage T4 DNA replication and

839 replication fork dynamics: a review in the Virology Journal series on bacteriophage T4

840 and its relatives. *Virol J* 7:358 . <https://doi.org/10.1186/1743-422X-7-358>

841 21. You L, Suthers PF, Yin J (2002) Effects of *Escherichia coli* Physiology on Growth of

842 Phage T7 In Vivo and In Silico. *J Bacteriol* 184:1888–1894 .

843 <https://doi.org/10.1128/JB.184.7.1888-1894.2002>

844 22. Endy D, Kong D, Yin J (1997) Intracellular kinetics of a growing virus: A genetically

845 structured simulation for bacteriophage T7. *Biotechnol Bioeng* 55:375–389 .

846 [https://doi.org/10.1002/\(SICI\)1097-0290\(19970720\)55:2<375::AID-BIT15>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0290(19970720)55:2<375::AID-BIT15>3.0.CO;2-G)

847 23. Zhang Y, Maley F, Maley GF, Duncan G, Dunigan DD, Van Etten JL (2007)
848 Chloroviruses Encode a Bifunctional dCMP-dCTP Deaminase That Produces Two Key
849 Intermediates in dTTP Formation. *J Virol* 81:7662–7671 .
850 <https://doi.org/10.1128/jvi.00186-07>

851 24. Graziani S, Xia Y, Gurnon JR, Van Etten JL, Leduc D, Skouloubris S, Myllykallio H, Liebl
852 U (2004) Functional Analysis of FAD-dependent Thymidylate Synthase ThyX from
853 *Paramecium bursaria* Chlorella Virus-1. *J Biol Chem* 279:54340–54347 .
854 <https://doi.org/10.1074/jbc.M409121200>

855 25. Zhang Y, Moriyama H, Homma K, Van Etten JL (2005) Chlorella Virus-Encoded
856 Deoxyuridine Triphosphatases Exhibit Different Temperature Optima. *J Virol* 79:9945–
857 9953 . <https://doi.org/10.1128/JVI.79.15.9945-9953.2005>

858 26. Chiu CS, Ruettinger T, Flanagan JB, Greenberg GR (1977) Role of deoxycytidylate
859 deaminase in deoxyribonucleotide synthesis in bacteriophage T4 DNA replication. *J Biol
860 Chem* 252:8603–8608 . [https://doi.org/10.1016/S0021-9258\(19\)75263-6](https://doi.org/10.1016/S0021-9258(19)75263-6)

861 27. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W (2003) Bacteriophage T4
862 Genome. *Microbiol Mol Biol Rev* 67:86–156 . <https://doi.org/10.1128/MMBR.67.1.86-156.2003>

863 28. Guzmán EC, Martín CM (2015) Thymineless death, at the origin. *Front Microbiol* 6:1–7 .
864 <https://doi.org/10.3389/fmicb.2015.00499>

865 29. Itsko M, Schaaper RM (2014) dGTP Starvation in *Escherichia coli* Provides New Insights
866 into the Thymineless-Death Phenomenon. *PLoS Genet* 10:e1004310 .
867 <https://doi.org/10.1371/journal.pgen.1004310>

868 30. Kuong KJ, Kuzminov A (2010) Stalled replication fork repair and misrepair during
869 thymineless death in *Escherichia coli*. *Genes to Cells* 15:619–634 .
870 <https://doi.org/10.1111/j.1365-2443.2010.01405.x>

871 31. Ahmad SI, Kirk SH, Eisenstark A (1998) Thymine Metabolism and Thymineless Death in

873 Prokaryotes and Eukaryotes. *Annu Rev Microbiol* 52:591–625 .

874 <https://doi.org/10.1146/annurev.micro.52.1.591>

875 32. Kim PJ, Price ND (2011) Genetic co-occurrence network across sequenced microbes.

876 *PLoS Comput Biol* 7: . <https://doi.org/10.1371/journal.pcbi.1002340>

877 33. Bartlett TM, Bratton BP, Duvshani A, Miguel A, Sheng Y, Martin NR, Nguyen JP, Persat

878 A, Desmarais SM, VanNieuwenhze MS, Huang KC, Zhu J, Shaevitz JW, Gitai Z (2017) A

879 Periplasmic Polymer Curves *Vibrio cholerae* and Promotes Pathogenesis. *Cell* 168:172-

880 185.e15 . <https://doi.org/10.1016/j.cell.2016.12.019>

881 34. Fernandez NL, Hsueh BY, Nhu NTQ, Franklin JL, Dufour YS, Waters CM (2020) *Vibrio*

882 *cholerae* adapts to sessile and motile lifestyles by cyclic di-GMP regulation of cell shape.

883 *Proc Natl Acad Sci* 117:29046–29054 . <https://doi.org/10.1073/pnas.2010199117>

884 35. Raabe CA, Hoe CH, Randau G, Brosius J, Tang TH, Rozhdestvensky TS (2011) The

885 rocks and shallows of deep RNA sequencing: Examples in the *Vibrio cholerae* RNome.

886 *RNA* 17:1357–1366 . <https://doi.org/10.1261/rna.2682311>

887 36. Papenfort K, Förstner KU, Cong J-P, Sharma CM, Bassler BL (2015) Differential RNA-

888 seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation.

889 *Proc Natl Acad Sci* 112:E766–E775 . <https://doi.org/10.1073/pnas.1500203112>

890 37. Krin E, Pierlé SA, Sismeiro O, Jagla B, Dillies M-AA, Varet H, Irazoki O, Campoy S, Rouy

891 Z, Cruveiller S, Médigue C, Coppée J-YY, Mazel D (2018) Expansion of the SOS regulon

892 of *Vibrio cholerae* through extensive transcriptome analysis and experimental validation.

893 *BMC Genomics* 19:1–18 . <https://doi.org/10.1186/s12864-018-4716-8>

894 38. Short FL, Akusobi C, Broadhurst WR, Salmond GPC (2018) The bacterial Type III toxin-

895 antitoxin system, ToxIN, is a dynamic protein-RNA complex with stability-dependent

896 antiviral abortive infection activity. *Sci Rep* 8:1013 . [35](https://doi.org/10.1038/s41598-017-</p><p>897 18696-x</p><p>898 39. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web</p></div><div data-bbox=)

899 portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–58 .

900 <https://doi.org/10.1038/nprot.2015.053>

901 40. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,

902 Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong

903 S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function

904 classification. *Bioinformatics* 30:1236–1240 .

905 <https://doi.org/10.1093/bioinformatics/btu031>

906 41. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi

907 Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I,

908 Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo

909 CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L,

910 Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G,

911 Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell

912 AL (2017) InterPro in 2017—beyond protein family and domain annotations. *Nucleic*

913 *Acids Res* 45:D190–D199 . <https://doi.org/10.1093/nar/gkw1107>

914 42. Leipe DD, Koonin E V., Aravind L (2003) Evolution and classification of P-loop kinases

915 and related proteins. *J Mol Biol* 333:781–815 . <https://doi.org/10.1016/j.jmb.2003.08.040>

916 43. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the

917 alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring

918 enzymes and a common nucleotide binding fold. *EMBO J* 1:945–51 .

919 <https://doi.org/10.1002/j.1460-2075.1982.tb01276.x>

920 44. Khan SR, Kuzminov A (2019) Thymineless Death in *Escherichia coli* Is Unaffected by

921 Chromosomal Replication Complexity. *J Bacteriol* 201: .

922 <https://doi.org/10.1128/JB.00797-18>

923 45. Hou H-F, Liang Y-H, Li L-F, Su X-D, Dong Y-H (2008) Crystal structures of *Streptococcus*

924 *mutans* 2'-deoxycytidylate deaminase and its complex with substrate analog and

925 allosteric regulator dCTP x Mg2+. *J Mol Biol* 377:220–31 .

926 <https://doi.org/10.1016/j.jmb.2007.12.064>

927 46. Marx A, Alian A (2015) The first crystal structure of a dTTP-bound deoxycytidylate

928 deaminase validates and details the allosteric-inhibitor binding site. *J Biol Chem*

929 290:682–90 . <https://doi.org/10.1074/jbc.M114.617720>

930 47. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ,

931 Kashin D, Niemann P, Jain M, Schwede F, Mekalanos JJ, Shao S, Lee ASY, Kranzusch

932 PJ (2020) CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5' and 2'-5'-

933 Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell*

934 182:38-49.e17 . <https://doi.org/10.1016/j.cell.2020.05.019>

935 48. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the

936 AID/APOBEC Family of Polynucleotide (Deoxy)cytidine Deaminases. *Mol Biol Evol*

937 22:367–377 . <https://doi.org/10.1093/molbev/msi026>

938 49. Seed KD, Bodi KL, Kropinski AM, Ackermann H-W, Calderwood SB, Qadri F, Camilli A

939 (2011) Evidence of a Dominant Lineage of *Vibrio cholerae*-Specific Lytic Bacteriophages

940 Shed by Cholera Patients over a 10-Year Period in Dhaka, Bangladesh. *MBio* 2:e00334-

941 10 . <https://doi.org/10.1128/mBio.00334-10>

942 50. Yen M, Cairns LS, Camilli A (2017) A cocktail of three virulent bacteriophages prevents

943 *Vibrio cholerae* infection in animal models. *Nat Commun* 8:14187 .

944 <https://doi.org/10.1038/ncomms14187>

945 51. O'Hara BJ, Barth ZK, McKitterick AC, Seed KD (2017) A highly specific phage defense

946 system is a conserved feature of the *Vibrio cholerae* mobilome. *PLoS Genet*

947 13:e1006838 . <https://doi.org/10.1371/journal.pgen.1006838>

948 52. McKitterick AC, Seed KD (2018) Anti-phage islands force their target phage to directly

949 mediate island excision and spread. *Nat Commun* 9:2348 .

950 <https://doi.org/10.1038/s41467-018-04786-5>

951 53. Hays SG, Seed KD (2020) Dominant *Vibrio cholerae* phage exhibits lysis inhibition
952 sensitive to disruption by a defensive phage satellite. *Elife* 9: .
953 <https://doi.org/10.7554/eLife.53200>

954 54. Morona R, Mavris M, Fallarino A, Manning PA (1994) Characterization of the rfc region of
955 *Shigella flexneri*. *J Bacteriol* 176:733–747 . <https://doi.org/10.1128/JB.176.3.733-747.1994>

956 55. Casjens SR (2004) The Chromosome of *Shigella flexneri* Bacteriophage Sf6: Complete
957 Nucleotide Sequence, Genetic Mosaicism, and DNA Packaging^{*1}. *J Mol Biol* 339:379–
958 394 . [https://doi.org/10.1016/S0022-2836\(04\)00377-8](https://doi.org/10.1016/S0022-2836(04)00377-8)

959 56. Jemielita M, Wingreen NS, Bassler BL Quorum sensing controls *Vibrio cholerae*
960 multicellular aggregate formation. *Elife*. <https://doi.org/10.7554/eLife.42057>

961 57. Johansson E, Fanø M, Bynck JH, Neuhard J, Larsen S, Sigurskjold BW, Christensen U,
962 Willemoes M (2005) Structures of dCTP Deaminase from *Escherichia coli* with Bound
963 Substrate and Product. *J Biol Chem* 280:3051–3059 .
964 <https://doi.org/10.1074/jbc.M409534200>

965 58. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, Rosenberg MM, Tal N,
966 Melamed S, Amitai G, Sorek R (2020) Prokaryotic viperins produce diverse antiviral
967 molecules. *Nature*. <https://doi.org/10.1038/s41586-020-2762-2>

968 59. Ayinde D, Casartelli N, Schwartz O (2012) Restricting HIV the SAMHD1 way: Through
969 nucleotide starvation. *Nat Rev Microbiol* 10:675–680 .
970 <https://doi.org/10.1038/nrmicro2862>

971 60. Shepard C, Xu J, Holler J, Kim DH, Mansky LM, Schinazi RF, Kim B (2019) Effect of
972 induced dNTP pool imbalance on HIV-1 reverse transcription in macrophages.
973 *Retrovirology* 16:29 . <https://doi.org/10.1186/s12977-019-0491-0>

974 61. Oo A, Kim DH, Schinazi RF, Kim B (2020) Viral protein X reduces the incorporation of
975 mutagenic noncanonical rNTPs during lentivirus reverse transcription in macrophages. *J*

976

977 Biol Chem 295:657–666 . <https://doi.org/10.1074/jbc.RA119.011466>

978 62. Harms A, Brodersen DE, Mitarai N, Gerdes K (2018) Toxins, Targets, and Triggers: An
979 Overview of Toxin-Antitoxin Biology. Mol Cell 70:768–784 .
980 <https://doi.org/10.1016/j.molcel.2018.01.003>

981 63. Makarova KS, Wolf YI, Snir S, Koonin E V. (2011) Defense Islands in Bacterial and
982 Archaeal Genomes and Prediction of Novel Defense Systems. J Bacteriol 193:6039–
983 6056 . <https://doi.org/10.1128/JB.05535-11>

984 64. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R (2018)
985 Systematic discovery of antiphage defense systems in the microbial pangenome. Science
986 (80-) 359:eaar4120 . <https://doi.org/10.1126/science.aar4120>

987 65. Bernheim A, Sorek R (2020) The pan-immune system of bacteria: antiviral defence as a
988 community resource. Nat Rev Microbiol 18:113–119 . <https://doi.org/10.1038/s41579-019-0278-2>

990 66. Westra ER, Swarts DC, Staals RHJ, Jore MM, Brouns SJ, van der Oost J (2012) The
991 CRISPRs, They Are A-Changin’: How Prokaryotes Generate Adaptive Immunity. Annu
992 Rev Genet 46:311–339 . <https://doi.org/10.1146/annurev-genet-110711-155447>

993 67. Skorupski K, Taylor RK (1996) Positive selection vectors for allelic exchange. Gene
994 169:47–52 . [https://doi.org/10.1016/0378-1119\(95\)00793-8](https://doi.org/10.1016/0378-1119(95)00793-8)

995 68. Bose JL, Rosenberg CS, Stabb E V. (2008) Effects of luxCDABEG induction in *Vibrio*
996 *fischeri*: Enhancement of symbiotic colonization and conditional attenuation of growth in
997 culture. Arch Microbiol 190:169–183 . <https://doi.org/10.1007/s00203-008-0387-1>

998 69. Fürste JP, Pansegrouw W, Frank R, Blöcker H, Scholz P, Bagdasarian M, Lanka E (1986)
999 Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP
1000 expression vector. Gene 48:119–131 . [https://doi.org/10.1016/0378-1119\(86\)90358-6](https://doi.org/10.1016/0378-1119(86)90358-6)

1001 70. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb E V (2006) and pES213-Derived Tools
1002 for Analyzing Symbiotic. Society 72:802–810 . <https://doi.org/10.1128/AEM.72.1.802>

1003 71. Edelheit O, Hanukoglu A, Hanukoglu I (2009) Simple and efficient site-directed
1004 mutagenesis using two single-primer reactions in parallel to generate mutants for protein
1005 structure-function studies. *BMC Biotechnol* 9:61 . <https://doi.org/10.1186/1472-6750-9-61>

1006 72. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic
1007 Acids Res* 28:27–30 . <https://doi.org/10.1093/nar/28.1.27>

1008 73. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M (2019) New approach for
1009 understanding genome variations in KEGG. *Nucleic Acids Res* 47:D590–D595 .
1010 <https://doi.org/10.1093/nar/gky962>

1011 74. Madeira F, Madhusoodanan N, Lee J, Tivey ARN, Lopez R (2019) Using EMBL-EBI
1012 Services via Web Interface and Programmatically via Web Services. *Curr Protoc
1013 Bioinforma* 66:e74 . <https://doi.org/10.1002/cpbi.74>

1014 75. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate
1015 boundary prediction and improved detection of structural RNAs. *RNA* 18:900–14 .
1016 <https://doi.org/10.1261/rna.029041.111>

1017 76. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL
1018 (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421 .
1019 <https://doi.org/10.1186/1471-2105-10-421>

1020 77. Tatusova T, Ciufo S, Federhen S, Fedorov B, McVeigh R, O'Neill K, Tolstoy I, Zaslavsky
1021 L (2015) Update on RefSeq microbial genomes resources. *Nucleic Acids Res* 43:D599–
1022 D605 . <https://doi.org/10.1093/nar/gku1062>

1023 78. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse
1024 B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V,
1025 Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D,
1026 Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey
1027 KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C,
1028 Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C,

1029 Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD,

1030 Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status,

1031 taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733-45 .

1032 <https://doi.org/10.1093/nar/gkv1189>

1033 79. Federhen S (2012) The NCBI Taxonomy database. *Nucleic Acids Res* 40:D136–D143 .

1034 <https://doi.org/10.1093/nar/gkr1178>

1035 80. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein

1036 homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248 .

1037 <https://doi.org/10.1093/nar/gki408>

1038 81. Nielsen H (2017) Predicting Secretory Proteins with SignalP. *Methods Mol Biol* 1611:59–

1039 73 . https://doi.org/10.1007/978-1-4939-7015-5_6

1040 82. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane

1041 protein topology with a hidden markov model: application to complete genomes¹¹Edited

1042 by F. Cohen. *J Mol Biol* 305:567–580 . <https://doi.org/10.1006/jmbi.2000.4315>

1043 83. Käll L, Krogh A, Sonnhammer ELL (2004) A Combined Transmembrane Topology and

1044 Signal Peptide Prediction Method. *J Mol Biol* 338:1027–1036 .

1045 <https://doi.org/10.1016/j.jmb.2004.03.016>

1046 84. Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane

1047 topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res* 35: .

1048 <https://doi.org/10.1093/nar/gkm256>

1049 85. Mistry J, Finn R (2007) Pfam. *Methods Mol Biol* 396:43–58 . https://doi.org/10.1007/978-1-59745-515-2_4

1050 86. Wilkins D (2020) gggenes. available: <https://github.com/wilcox/gggenes>

1051 87. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G,

1052 Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J,

1053 Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H

1055 (2019) Welcome to the Tidyverse. *J Open Source Softw* 4:1686 .

1056 <https://doi.org/10.21105/joss.01686>

1057 88. Müller K (2017) here: A Simpler Way to Find Your Files. available: <https://cran.r-project.org/web/packages/here/index.html>

1059 89. Lassmann T (2019) Kalign 3: multiple sequence alignment of large datasets.

1060 *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz795>

1061 90. R Core Team (2019) R: A language and environment for statistical computing. In: R

1062 Found. Stat. Comput. available: <https://www.r-project.org/>

1063 91. RStudio Team (2020) RStudio: Integrated Development Environment for R. available:

1064 <https://www.rstudio.com/>

1065 92. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high

1066 throughput. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkh340>

1067 93. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time

1068 and space complexity. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-5-113>

1069 94. Bonatesta E, Horejs-Kainrath C, Bodenhofer U (2020) Multiple Sequence Alignment.

1070 available: <https://www.bioinf.jku.at/software/msa/>

1071 95. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood

1072 trees for large alignments. *PLoS One* 5:e9490 .

1073 <https://doi.org/10.1371/journal.pone.0009490>

1074 96. Rambaut A (2009) FigTree. availabe: <https://tree.bio.ed.ac.uk/software/figtree/>

1075 97. Paradis E, Blomberg S, Bolker B, Brown J, Claramunt S, Claude J, Cuong HS, Desper R,

1076 Didier G, Durand B, Dutheil J, Ewing R, Gascuel O, Guillerme T, Heibl C, Ives A, Jone B,

1077 Krah F, Lawson D, Lefort V, Legendre P, Lemon J, Louvel G, Marcon E, McCloskey R,

1078 Nylander J, Opgen-Rhein R, Popescu A-A, Royer-Carenzi M, Schliep K, Strimmer K, de

1079 Vienne D (2020) ape: Analyses of Phylogenetics and Evolution. available: <https://ape-package.ird.fr/>

1081 98. Ducret A, Quardokus EM, Brun Y V. (2016) MicrobeJ, a tool for high throughput bacterial
1082 cell detection and quantitative analysis. *Nat Microbiol* 1:16077 .
1083 <https://doi.org/10.1038/nmicrobiol.2016.77>

1084 99. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,
1085 Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K,
1086 Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image
1087 analysis. *Nat Methods* 9:676–682 . <https://doi.org/10.1038/nmeth.2019>

1088 100. Fernandez NL, Waters CM (2019) Cyclic di-GMP increases catalase production and
1089 hydrogen peroxide tolerance in *Vibrio cholerae*. *Appl Environ Microbiol* 85: .
1090 <https://doi.org/10.1128/AEM.01043-19>

1091 101. Dryselius R, Izutsu K, Honda T, Iida T (2008) Differential replication dynamics for large
1092 and small *Vibrio* chromosomes affect gene dosage, expression and location. *BMC*
1093 *Genomics* 9:559 . <https://doi.org/10.1186/1471-2164-9-559>

1094 102. Dong H, Liu Y, Zu X, Li N, Li F, Zhang D (2015) An enzymatic assay for high-throughput
1095 screening of cytidine-producing microbial strains. *PLoS One* 10:e0121612 .
1096 <https://doi.org/10.1371/journal.pone.0121612>

1097 103. Ngo TT, Phan APH, Yam CF, Lenhoff HM (1982) Interference in determination of
1098 ammonia with the hypochlorite-alkaline phenol method of Berthelot. *Anal Chem* 54:46–
1099 49 . <https://doi.org/10.1021/ac00238a015>

1100 104. Massie JP, Reynolds EL, Koestler BJ, Cong J-P, Agostoni M, Waters CM (2012)
1101 Quantification of high-specificity cyclic diguanylate signaling. *Proc Natl Acad Sci*
1102 109:12746–12751 . <https://doi.org/10.1073/pnas.1115663109>

1103 105. Dover JA, Burmeister AR, Molineux IJ, Parent KN (2016) Evolved populations of *Shigella*
1104 *flexneri* phage Sf6 acquire large deletions, altered genomic architecture, and faster life
1105 cycles. *Genome Biol Evol* 8:2827–2840 . <https://doi.org/10.1093/gbe/evw177>

1106 106. Evans DJ, Evans DG (1973) Three Characteristics Associated with Enterotoxigenic

1107 *Escherichia coli* Isolated from Man. Infect Immun 8:322–328 .

1108 <https://doi.org/10.1128/IAI.8.3.322-328.1973>

1109 107. Helene Thelin K, Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive

1110 hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139

1111 strains. Infect Immun 64:2853–2856 . <https://doi.org/10.1128/iai.64.7.2853-2856.1996>

1112 108. Miller VL, DiRita VJ, Mekalanos JJ (1989) Identification of *toxS*, a regulatory gene whose

1113 product enhances *ToxR*-mediated activation of the cholera toxin promoter. J Bacteriol

1114 171:1288–1293 . <https://doi.org/10.1128/jb.171.3.1288-1293.1989>

1115 109. Lüdeke CHM, Kong N, Weimer BC, Fischer M, Jones JL (2015) Complete Genome

1116 Sequences of a Clinical Isolate and an Environmental Isolate of *Vibrio parahaemolyticus*.

1117 Genome Announc 3:e00216-15 . <https://doi.org/10.1128/genomeA.00216-15>

1118 110. Lutgring JD, Machado M-J, Benahmed FH, Conville P, Shawar RM, Patel J, Brown AC

1119 (2017) FDA-CDC Antimicrobial Resistance Isolate Bank: a Publicly Available Resource

1120 To Support Research, Development, and Regulatory Requirements. J Clin Microbiol 56: .

1121 <https://doi.org/10.1128/JCM.01415-17>

1122 111. Parent KN, Erb ML, Cardone G, Nguyen K, Gilcrease EB, Porcek NB, Pogliano J, Baker

1123 TS, Casjens SR (2014) OmpA and OmpC are critical host factors for bacteriophage Sf6

1124 entry in *Shigella*. Mol Microbiol 92:47–60 . <https://doi.org/10.1111/mmi.12536>

1125 112. Vanbleu E, Marchal K, Vanderleyden J (2004) Genetic and physical map of the pLAFR1

1126 vector. DNA Seq - J DNA Seq Mapp 15:225–227 .

1127 <https://doi.org/10.1080/10425170410001723949>

1128

1129 **FIGURE LEGENDS-located with the figures**

Fig. 1

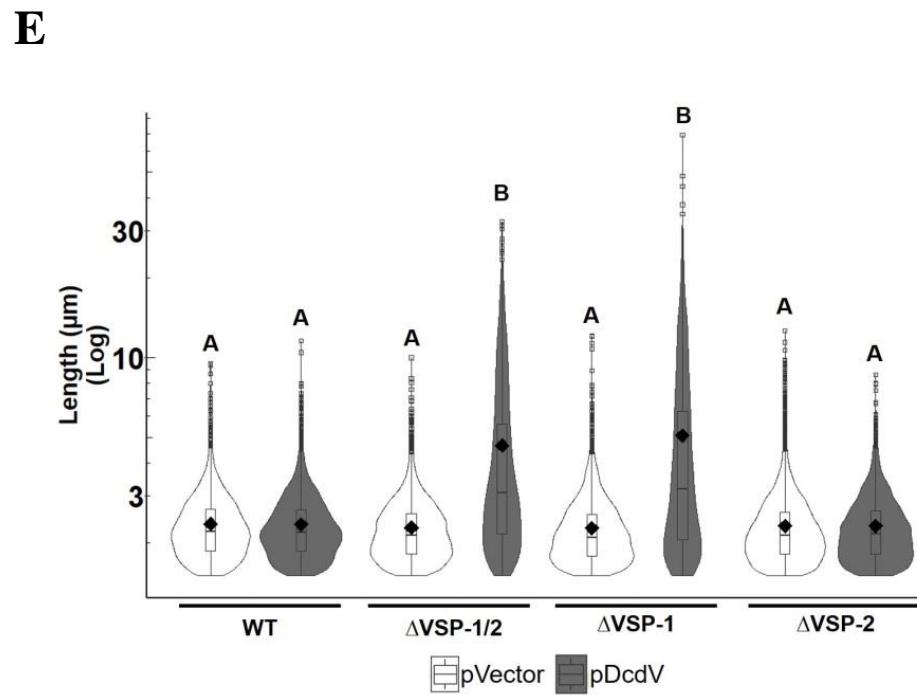
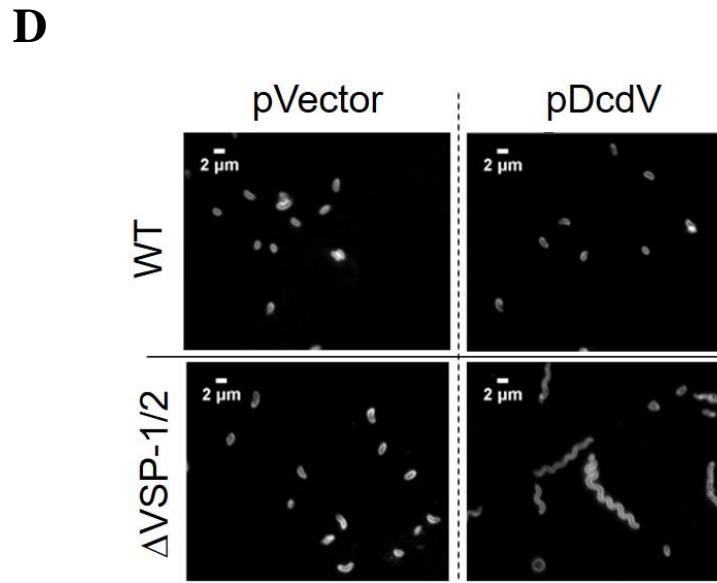
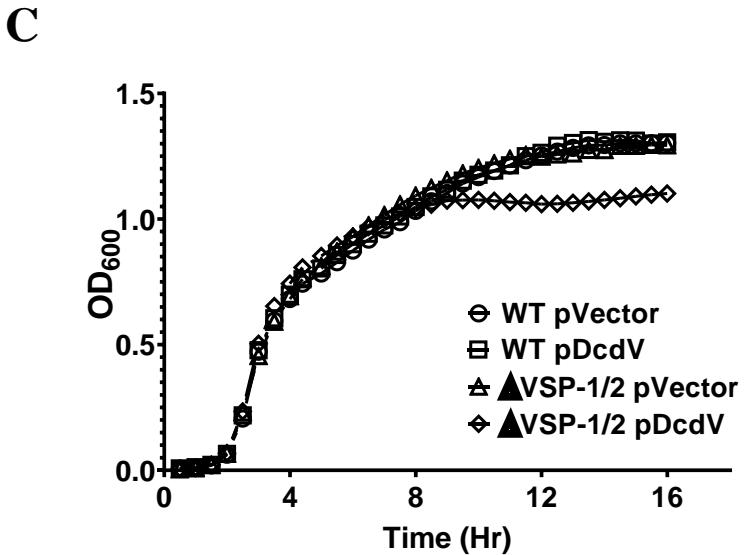
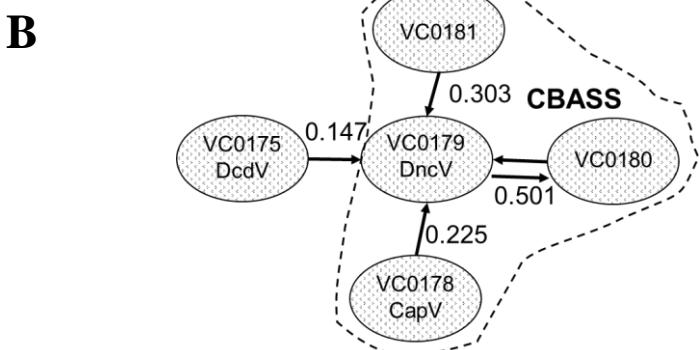
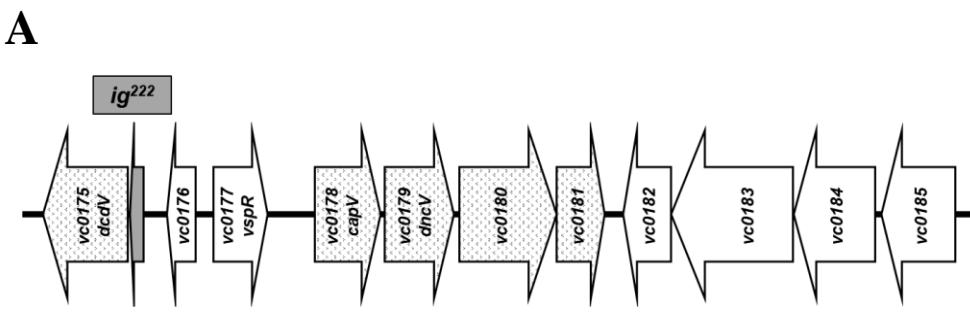
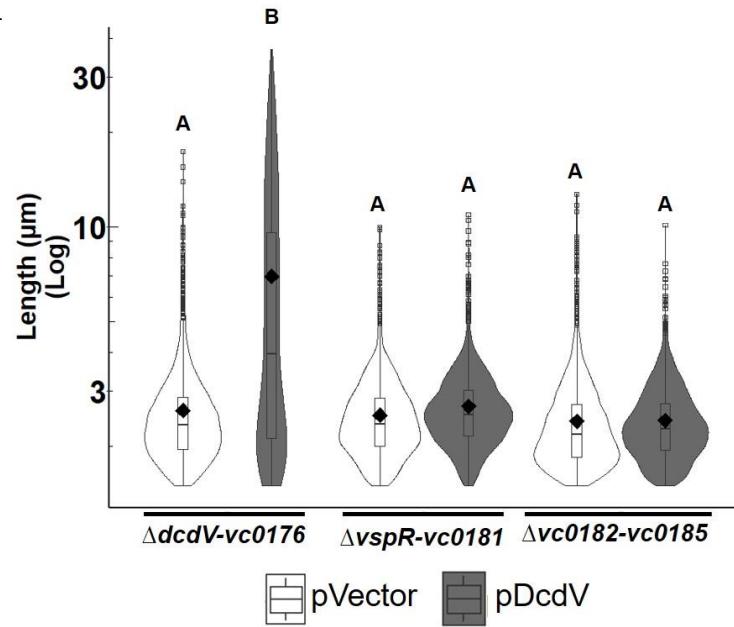


Fig. 1: DcdV promotes filamentation in *V. cholerae* in the absence of VSP-1

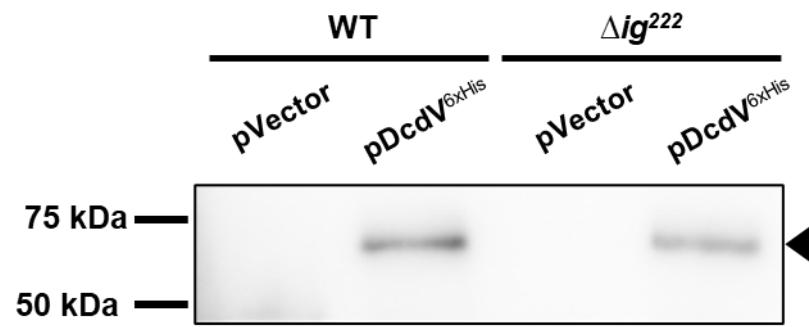
(A) Cartoon schematic of VSP-1 and **(B)** the Correology gene network prediction for *dncV* where arrows show the highest partial correlation W_{ij} each individual VSP-1 gene has to another. **(C)** Growth of WT *V. cholerae* and Δ VSP-1/2 strains with the vector or pDcdV. Data represent the mean \pm SEM, $n=3$. **(D)** Representative images of WT and Δ VSP-1/2 strains with the vector or pDcdV. **(E)** Violin plots of cell length distributions of WT, Δ VSP-1/2, Δ VSP-1, and Δ VSP-2 strains with the vector or pDcdV: summary statistic for this and all following violin plots are mean (diamonds), median (horizontal black line), interquartile range (box), and data below and above the interquartile range (vertical lines). Different letters indicate significant differences ($n=3$) at $p < 0.05$, according to Tukey's post-hoc test.

Fig. 2

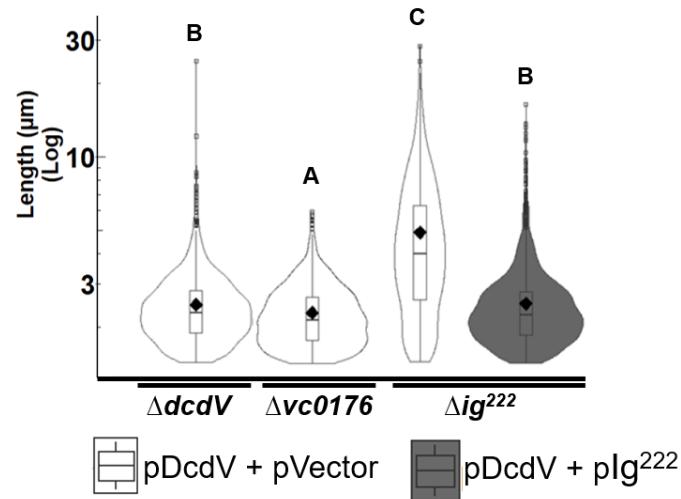
A



C



B



D

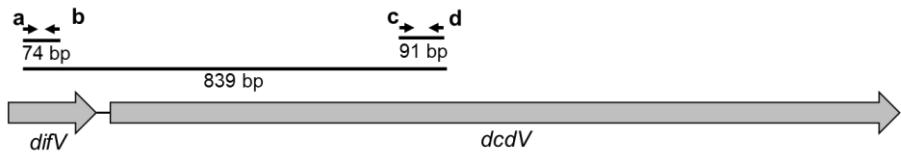
<u><i>difV</i> constructs size (position in <i>ig</i>²²²)</u>	<u>Diagram of <i>difV</i> constructs (5' → 3')</u>	<u>Inhibits DcdV?</u>
222 nt (<i>ig</i> ²²²)	RBS 1 111 222	YES
222 nt ^{STOP}	*	YES
174 nt (49-222)	YES
174 nt ^{RBS}	YES
174 nt ^{STOP} *	YES
174 nt ^{InteriorSTOP} *	YES
138 nt (49-186)	NO
156 nt (49-204)	NO
166 nt (49-214)	NO
170 nt (49-218)	NO
157 nt (66-222)	NO
137 nt (86-222)	NO
100 nt (123-222)	NO
npvR_3991 (102-206)	NO

Fig. 2: DifV is a sRNA that post-translationally regulates DcdV

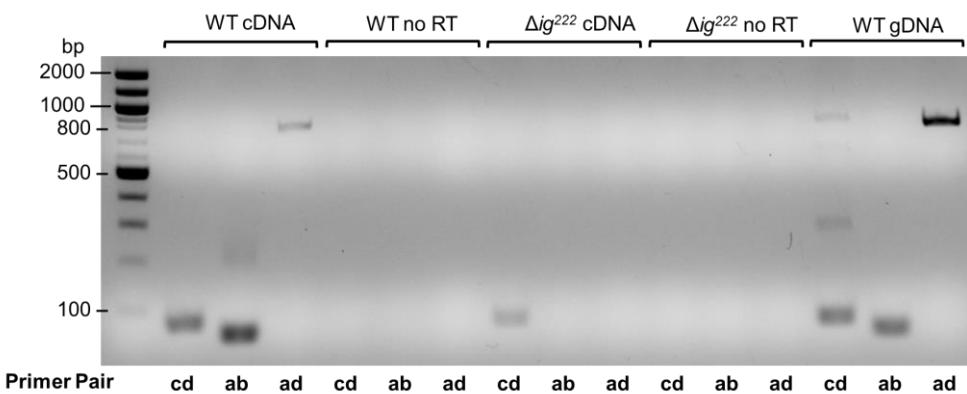
(A) Distribution of cell lengths measured from three biological replicates of gene deletions within VSP-1 or **(B)** individual gene deletions as indicated containing vector or pDcdV grown in the presence of 100 μ M IPTG for 8 h. Different letters indicate significant differences ($n=3$) at $p < 0.05$, according to Tukey's post-hoc test. **(C)** Representative anti-6x His antibody Western blot of whole cell lysates from *V. cholerae* WT and Δig^{222} cultures maintaining vector or pDcdV^{6xHis}. Analysis was performed in triplicate biological samples. Black triangle corresponds to DcdV^{6xHis} (60.6 kDa). **(D)** Table of various *difV* constructs expressed in Δig^{222} under a P_{tac} -inducible promoter with a non-native ribosomal binding site (RBS, denoted by dotted line). DcdV induced filamentation in the presence of these *difV* constructs was assessed using fluorescence microscopy in biological triplicate cultures. “**” indicates a stop codon introduced in place of a putative start codon.

Fig. 3

A



B



C

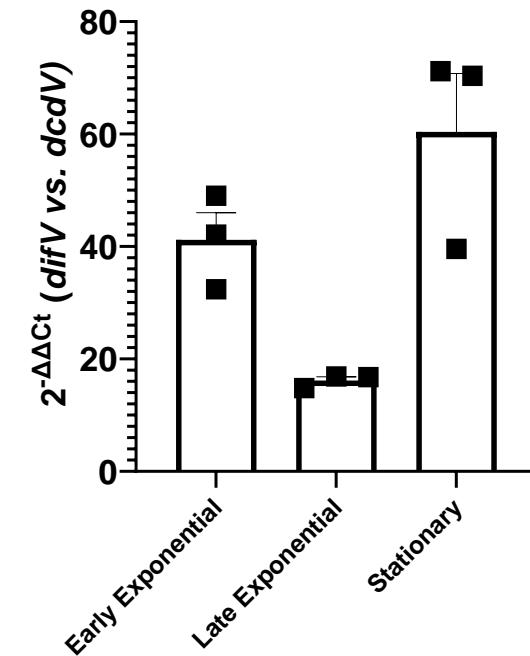
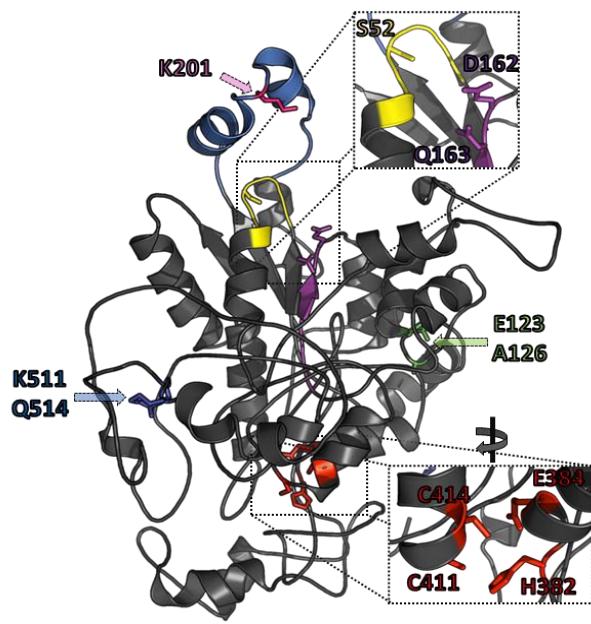


Fig. 3: *difV* and *dcdV* are in an operon and *difV* expression exceeds *dcdV*

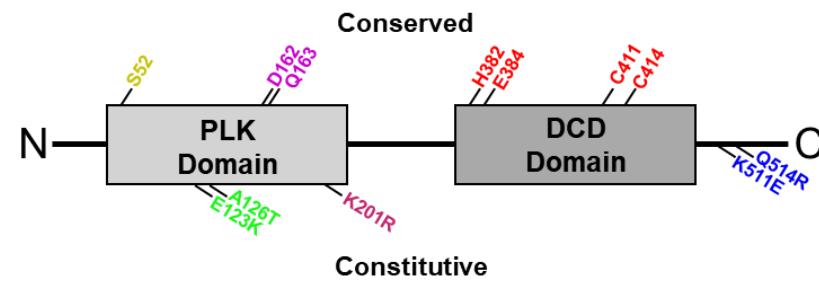
(A) Genomic diagram of *difV* and *dcdV* and the primers (a, b, c, and d) used for generating diagnostic PCR products. **(B)** PCR products amplified from nucleic acid templates (above) using the indicated primer pairs (below) resolved in a 1% agarose gel. All reactions were performed in duplicate using biologically independent samples with similar results. No RT = non-reverse transcribed RNA control. gDNA = genomic DNA control **(C)** qRT-PCR analysis of relative difference between *difV* transcript and *dcdV* transcript levels at different growth phases in WT *V. cholerae* normalized to an endogenous *gyrA* control. Data are graphed as mean \pm SEM, $n=3$.

Fig. 4

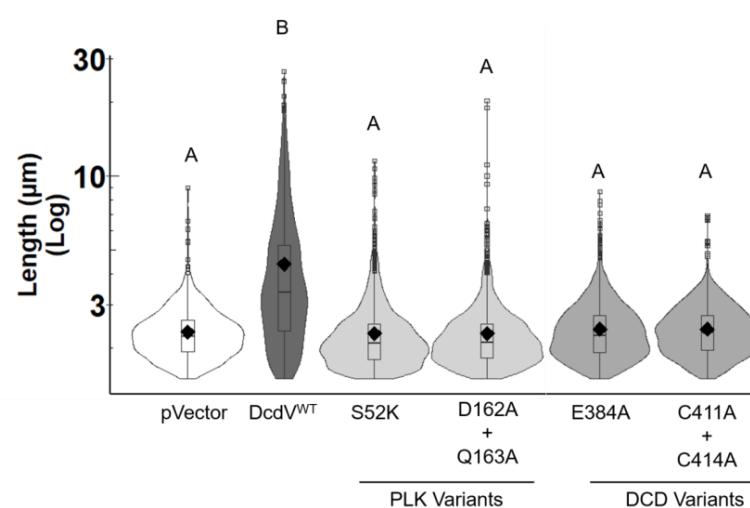
A



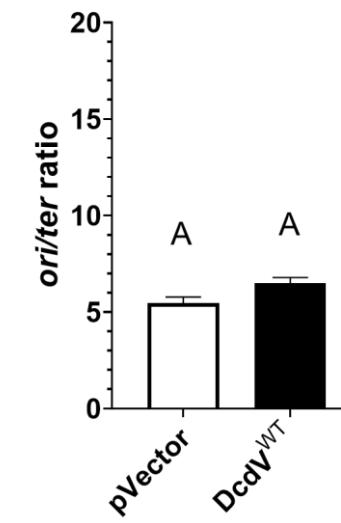
B



C

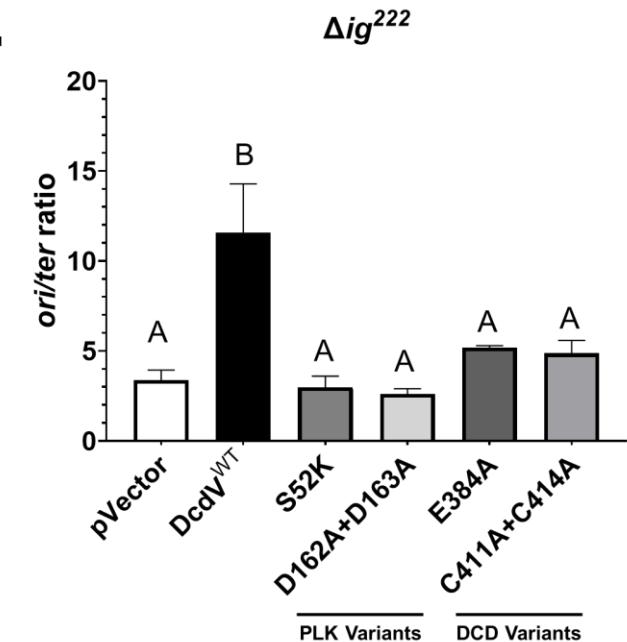


E



WT

F



Δig²²²

D

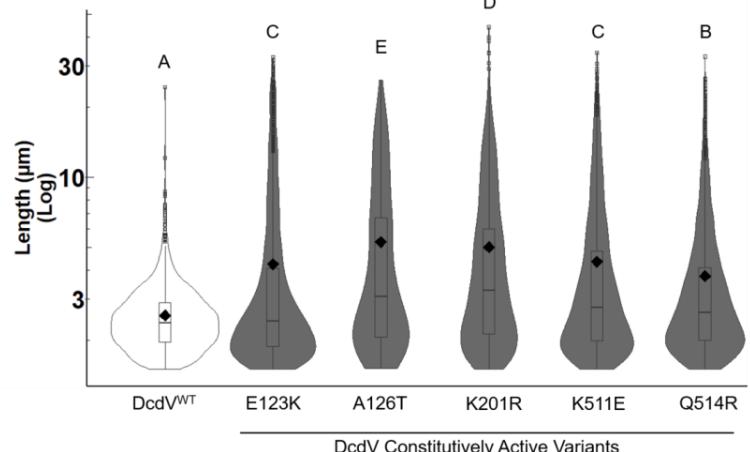


Fig. 4: Both the PLK and DCD domains are required for DcdV induced filamentation

(A) Phyre2 predicted structure of DcdV from *V. cholerae* El Tor. The inset shows the conserved residues of PLK (top) and DCD (bottom) domains. **(B)** Domain organization and conserved residues at each domain of DcdV. Top labeled residues indicate conserved features of both domains, and the bottom labeled residues indicate variants that render DcdV constitutively active. **(C)** Distribution of cell lengths measured from three biological replicates of WT *E. coli* as indicated. **(D)** Distribution of cell lengths measured from three biological replicates of the Δ dcdV *V. cholerae* mutant expressing the indicated DcdV variants. *ori/ter* ratios of Chromosome 1 in **(E)** WT and **(F)** Δ ig²²² *V. cholerae* strains expressing the indicated DcdV construct for 8 h and quantified using qRT-PCR. Each bar represents the mean \pm SEM, $n=3$. Different letters indicate significant differences ($n=3$) at $p < 0.05$, according to Tukey's post-hoc test.

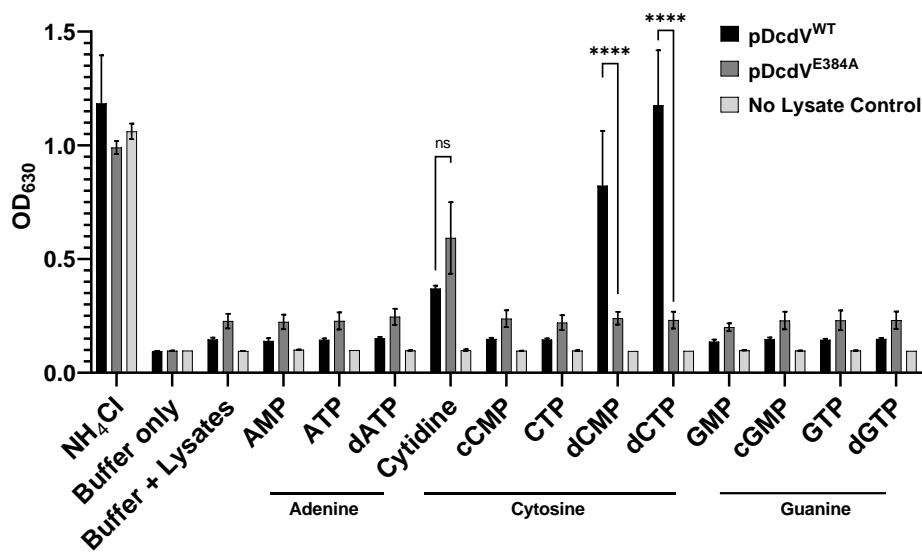
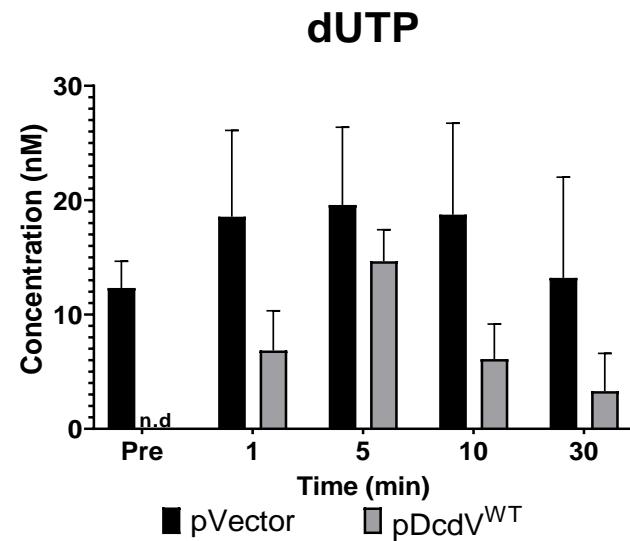
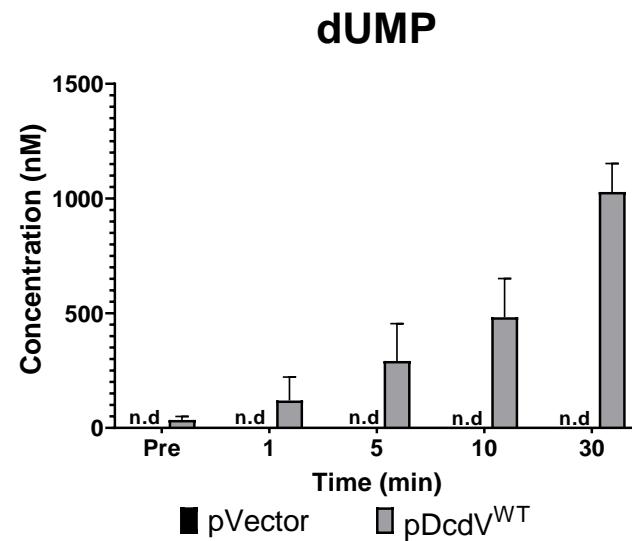
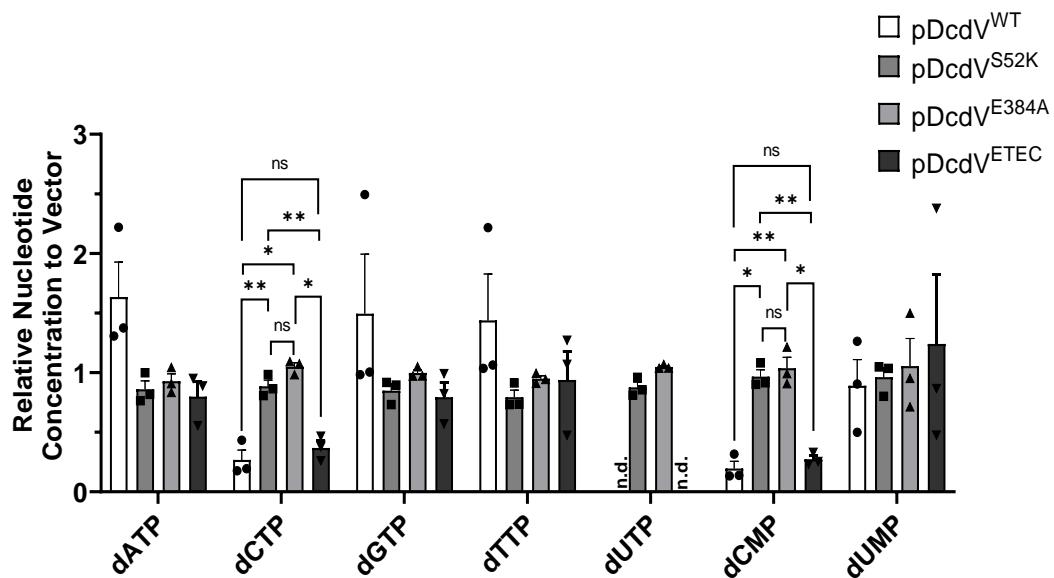
Fig. 5**A****B****C****D**

Fig. 5: DcdV alters cellular nucleotide metabolism.

(A) Lysates collected from *E. coli* expressing DcdV or DcdV^{E384A} and a “no lysate” buffer control incubated with 12 nucleotide substrates (1.9 mM NH₄Cl as a positive control, 37.7 mM cytidine, and 7.5 mM for all other substrates). Data represent the mean \pm SEM, *n*=3. Quantification of dUTP (**B**) and dUMP (**C**) using UPLC-MS/MS, in the indicated cell lysates before (Pre) and after addition of 1 mM dCTP. Each lysate was normalized to 20 mg/mL total protein. Each bar represents mean \pm SEM, *n*=3. (**D**) Quantification of the indicated dNTPs in vivo using UPLC-MS/MS in strains expressing the four DcdV variants, as indicated, normalized to dNTP concentrations measured in a vector control. Data are graphed as mean \pm SEM, *n*= 3, Two-way ANOVA with Tukey’s multiple-comparison test, normalized to pVector, n.d. indicates “none detected”, and ns indicates “not significant”.

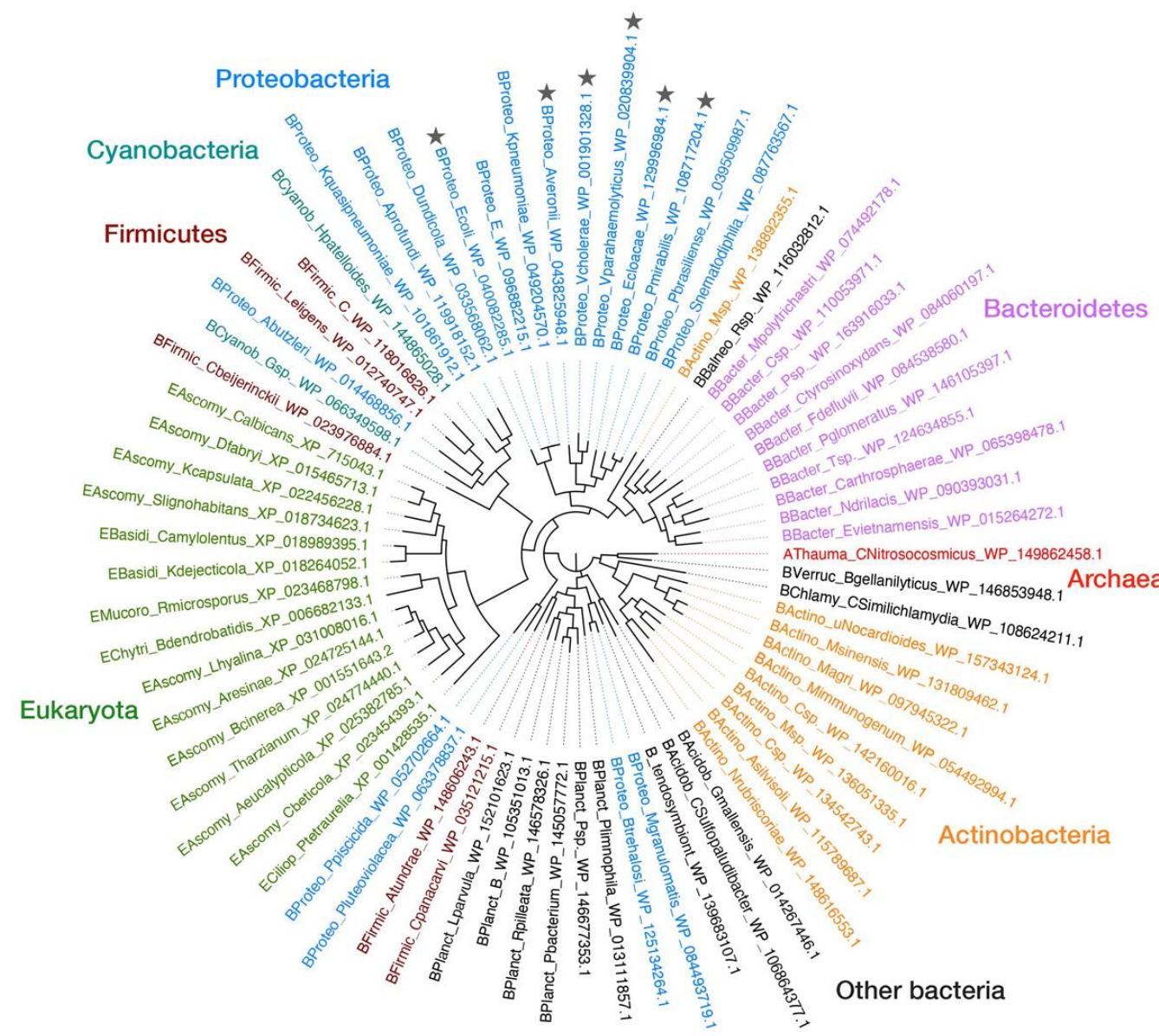
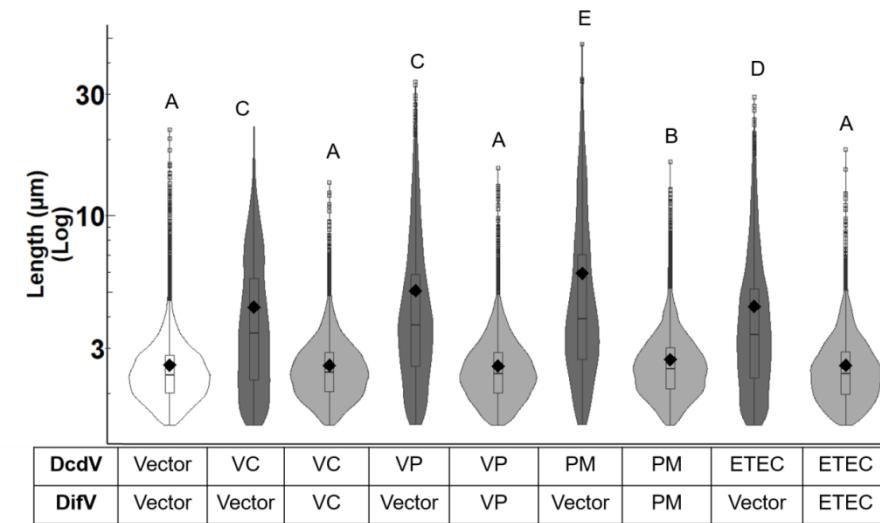
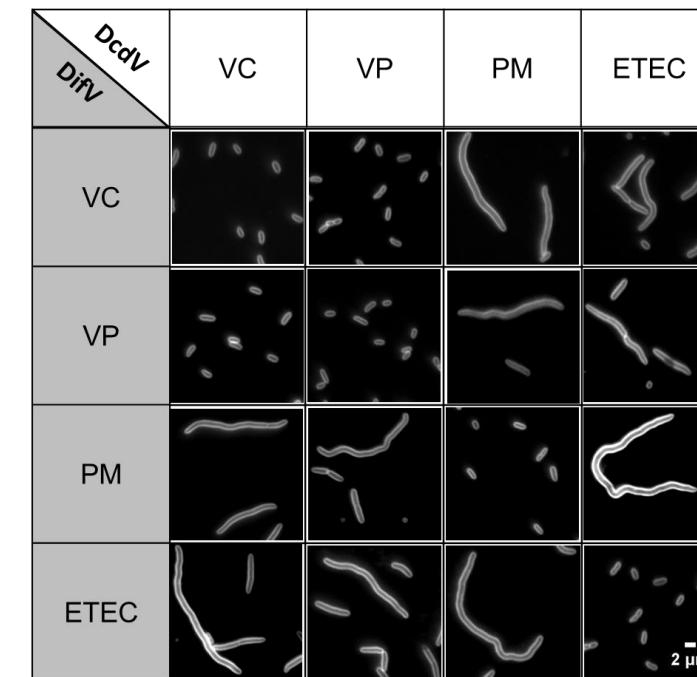
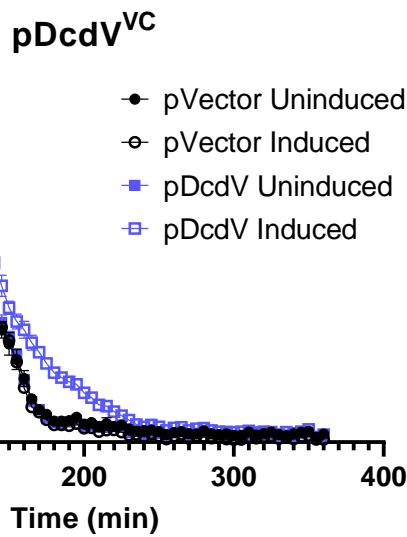
Fig. 6**A****B****C**

Fig. 6: *dcdV* and *difV* are widely conserved

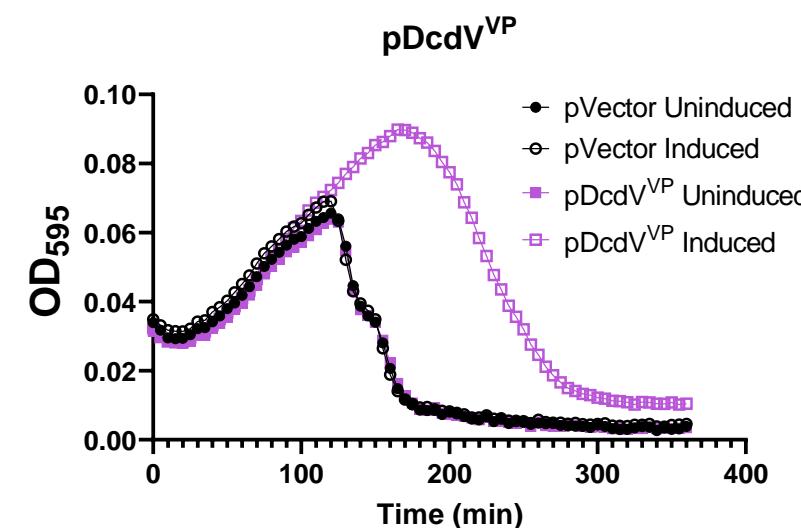
(A) Phylogenetic tree of DcdV homologs containing PLK and DCD domains from representative phyla across the tree of life. Stars indicate query proteins of interest in this study. **(B)** Distribution of cell lengths expressing the indicated DcdV homologs and their cognate DifV or vector control in *E. coli* (n=3). Different letters indicate significant differences (n=3) at $p < 0.05$, according to Dunnett's post-hoc test against the control (pVector^{DcdV} + pVector^{DifV}) strain. **(C)** Representative images of *E. coli* expressing pDcdV/homologs and pDifV/homologs combinations. Scale represents 2 μ m.

Fig. 7

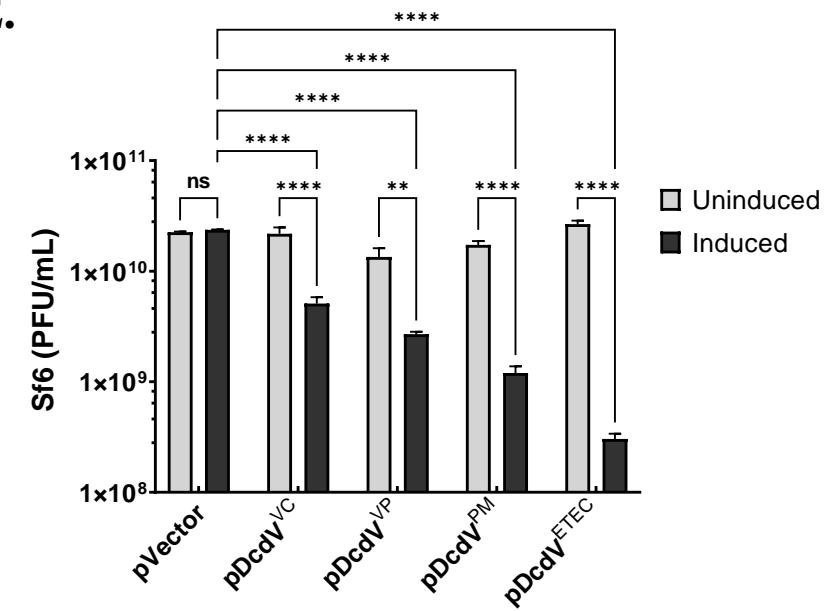
A.



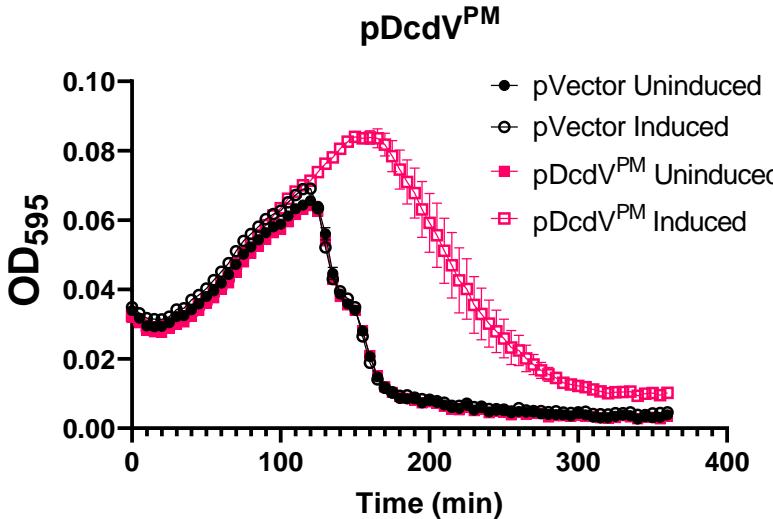
B.



E.



C.



D.

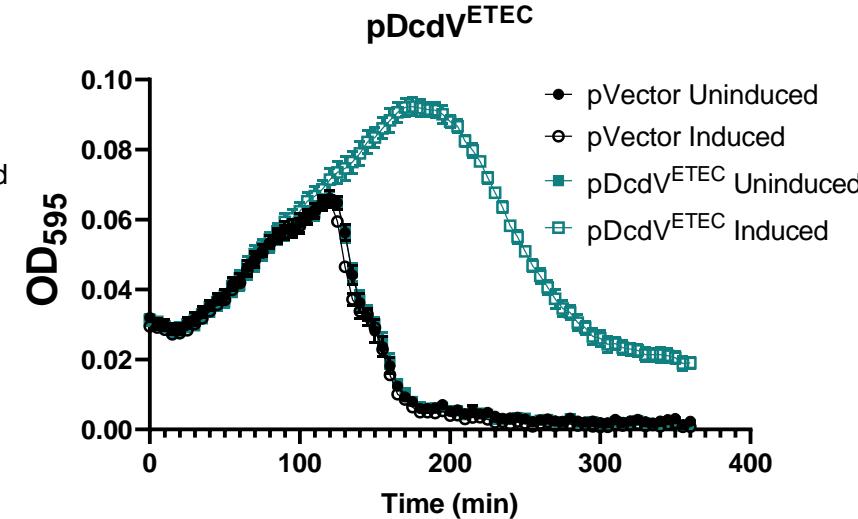
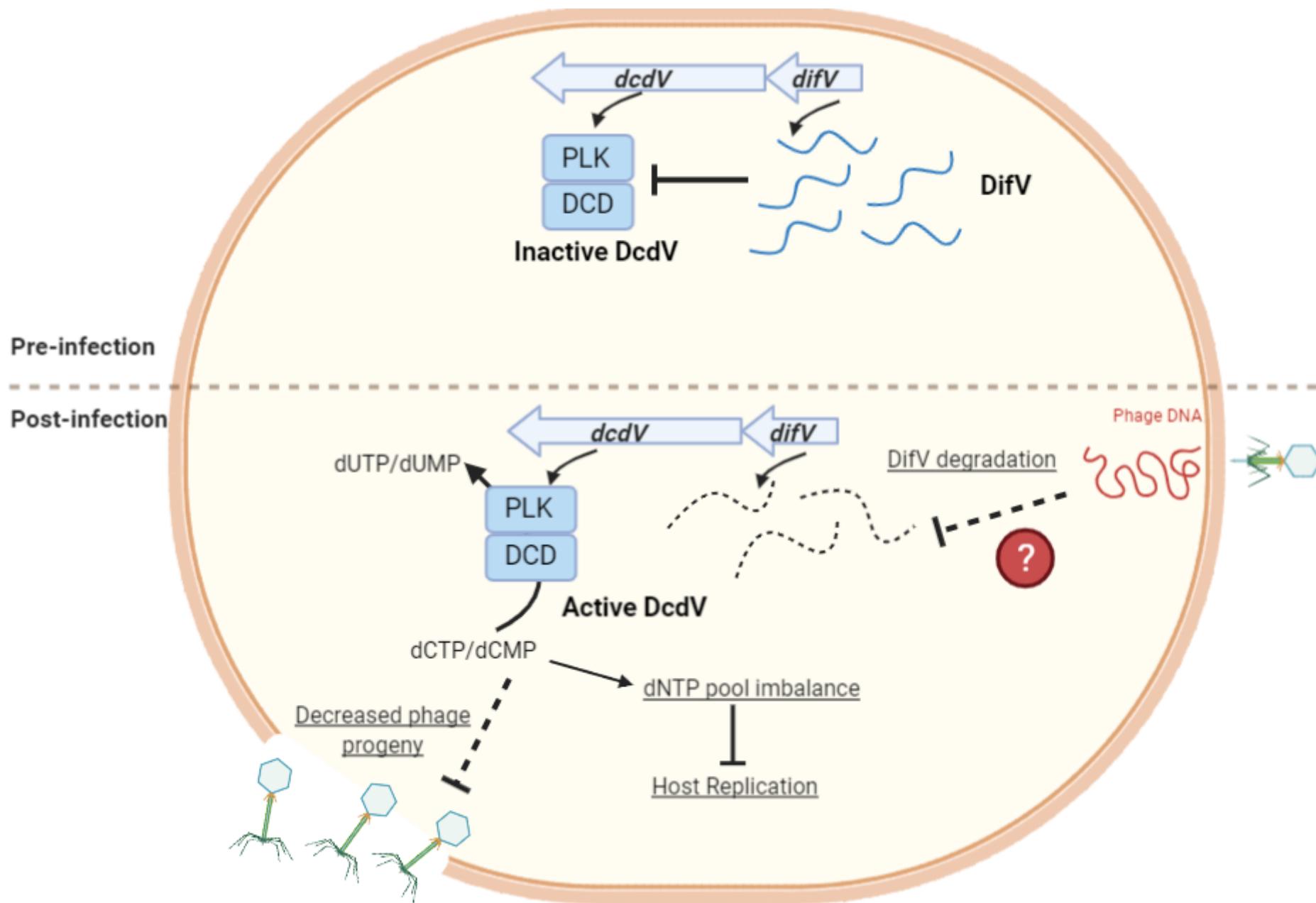


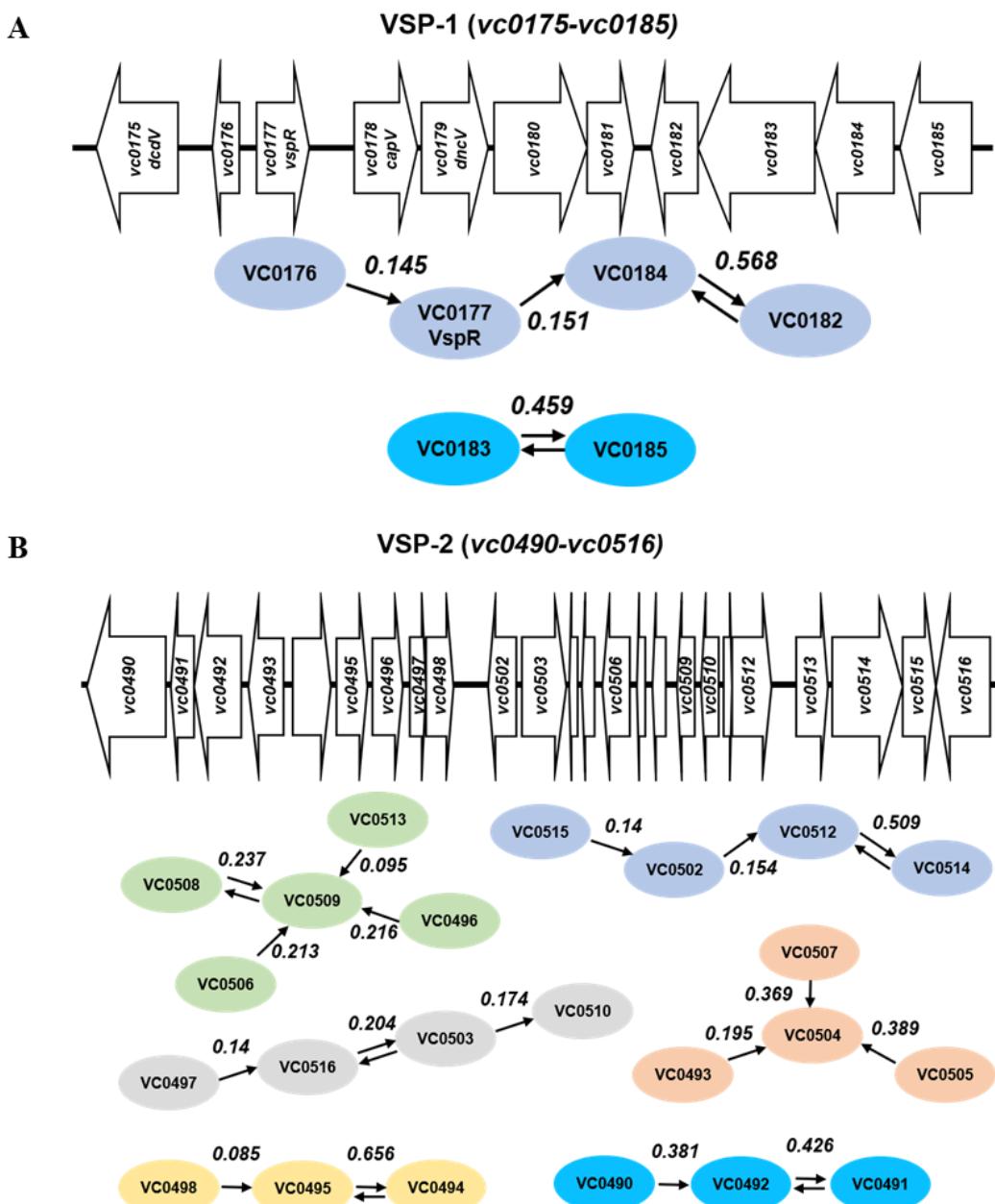
Fig. 7: DcdV mediates phage defense

(A-D) Growth curves for *S. flexneri* containing vector or pDcdV/homologs infected Sf6 at time 0 at an MOI of 0.1 in the presence or absence of 100 μ M IPTG. Each graph represents three biological replicates each with three technical replicates. **(B)** Plaque-forming units (PFU) per mL of phage Sf6 measured at the conclusion of the *S. flexneri* growth curve experiment above. Results are represented as mean \pm SEM, $n= 3$, Two-way ANOVA with Tukey's multiple-comparison test.

Graphical Abstract



1 **SUPPLEMENTAL MATERIAL**

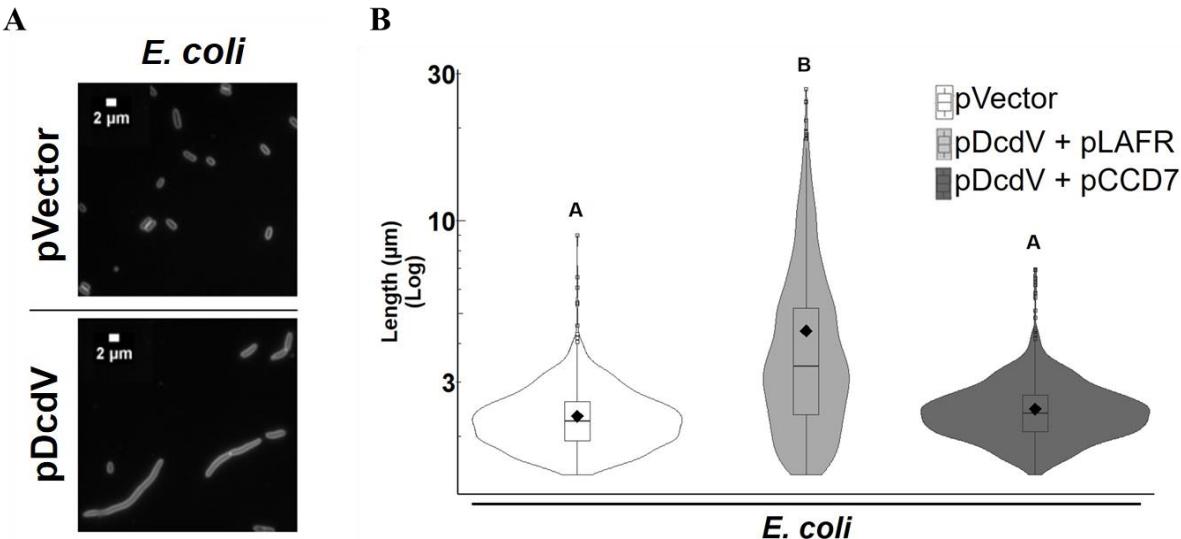


2
3

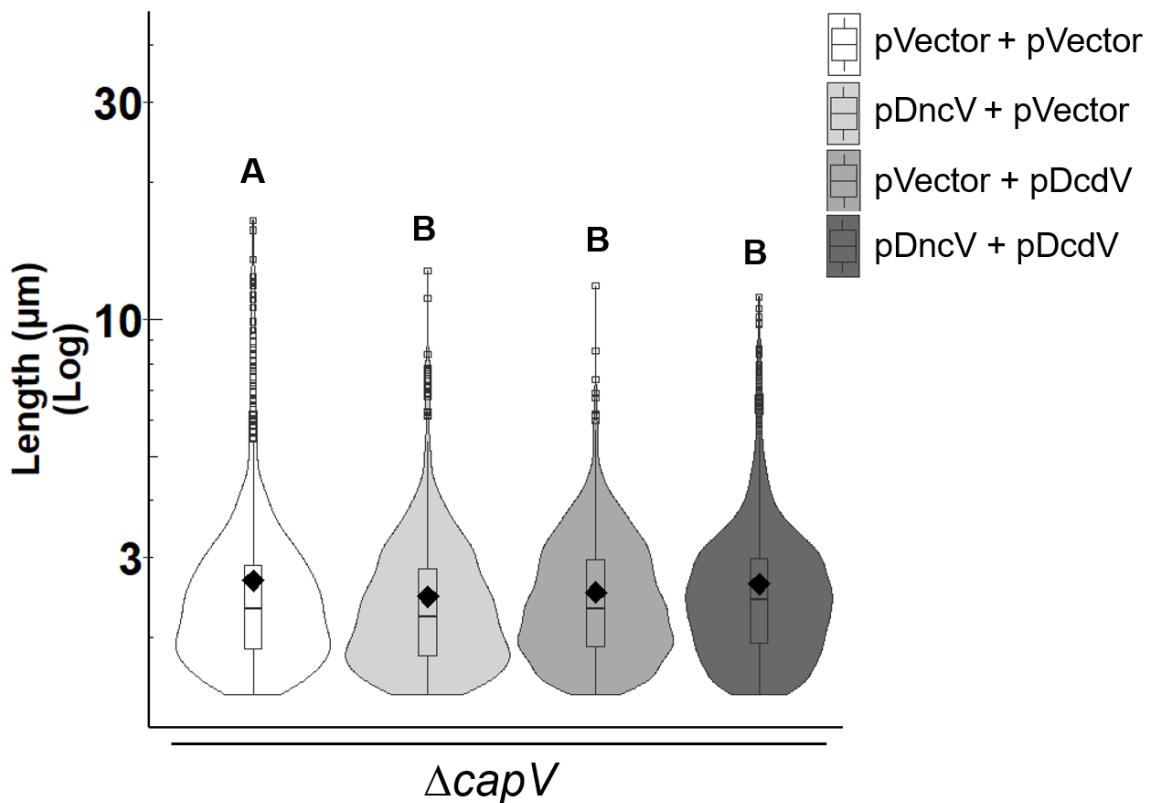
4 **Supplemental Figure 1. VSP-1 and VSP-2 schematic and predicted maximum
5 related subnetworks (MRS).**

6 (A) Cartoon schematic and gene network predictions, other than DcdV and CBASS (see Figs. 1B
7 and 1), of VSP-1 from El Tor *V. cholerae* N16961 (not to scale). (B) Cartoon schematic and gene
8 network predictions of VSP-2 from El Tor *V. cholerae* N16961 (not to scale). Arrows indicate the
9 highest partial correlation W_{ij} of each individual VSP gene to another (represented by ovals). Two
10 arrows pointing in opposing directions indicates the two genes each have their highest correlation
11 to each other.

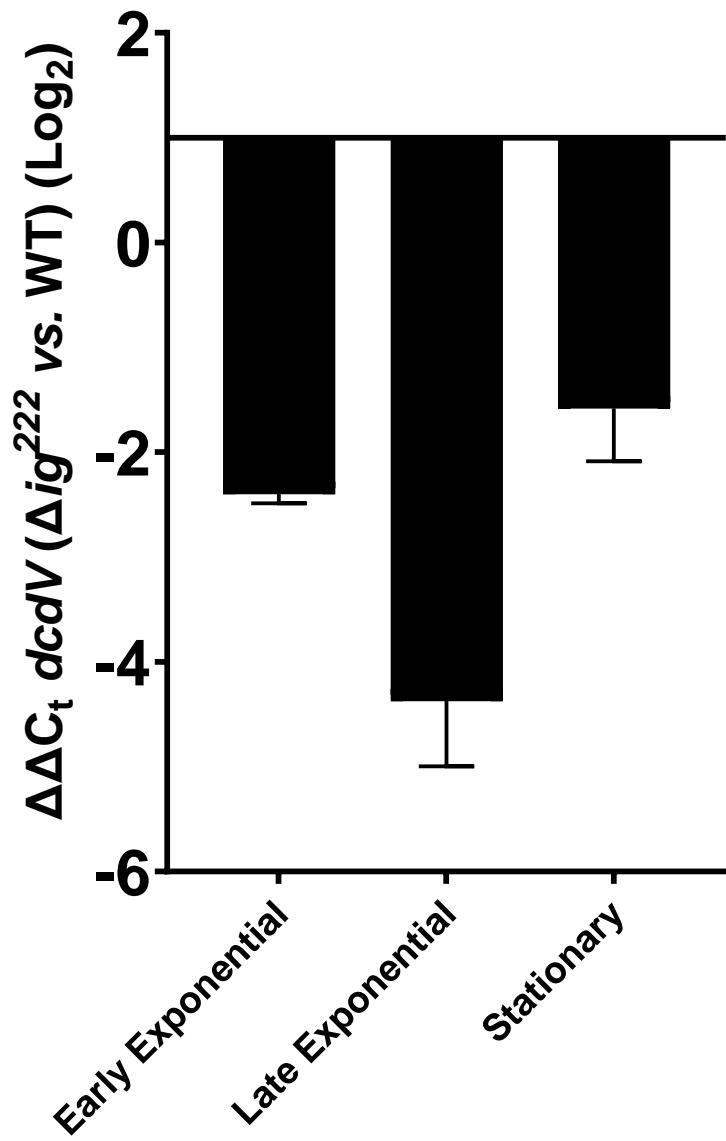
12



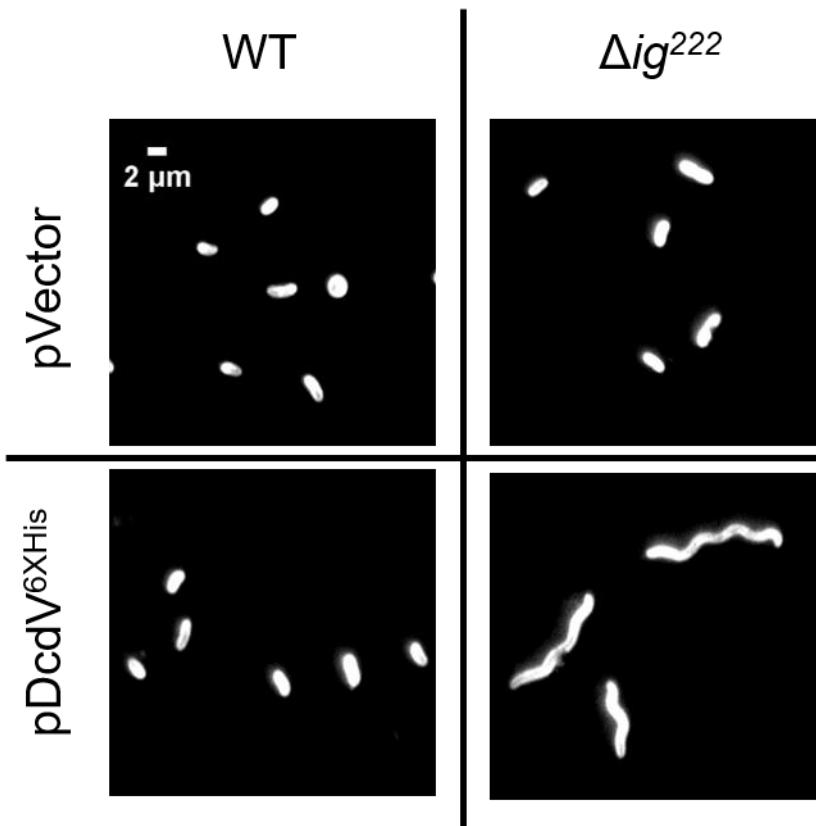
13
14 **Supplemental Figure 2. Ectopic expression of *dcdV* leads to cell filamentation in *E. coli***
15 **that is alleviated by provision of a single copy cosmid containing VSP-1.**
16 (A) Representative images of *E. coli* cultures maintaining an empty vector plasmid (pVector) or
17 P_{tac} -inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 μ M IPTG for 8 h. Cells were
18 stained with FM4-64 prior to imaging. Scale represents 2 μ m. (B) Distribution of cell lengths
19 measured from three biological replicates of *E. coli* cultures carrying an empty vector (Vector) or
20 P_{tac} -inducible *dcdV* plasmid (pDcdV) in addition to either an empty vector single copy cosmid
21 control (pLAFR) or pLAFR containing VSP-1 (pCCD7) grown in the presence of 100 μ M IPTG for
22 8 h. Distributions represent ~1000 to 2000 cells measured per strain. Different letters indicate
23 significant differences at $p < 0.05$, according to Tukey's post-hoc test.



24
25 **Supplemental Figure 3. Ectopic expression of DncV and DcdV does not lead to**
26 **filamentation in the ΔcapV mutant of *V. cholerae*.**
27 Distribution of cell lengths measured from three biological replicates of ΔcapV mutant cultures
28 maintaining the indicated plasmids grown in the presence of 100 μM IPTG for 8 h. Distributions
29 represent ~1200-1700 cells measured per strain. Different letters indicate significant differences
30 at $p < 0.05$, according to Tukey's post-hoc test.



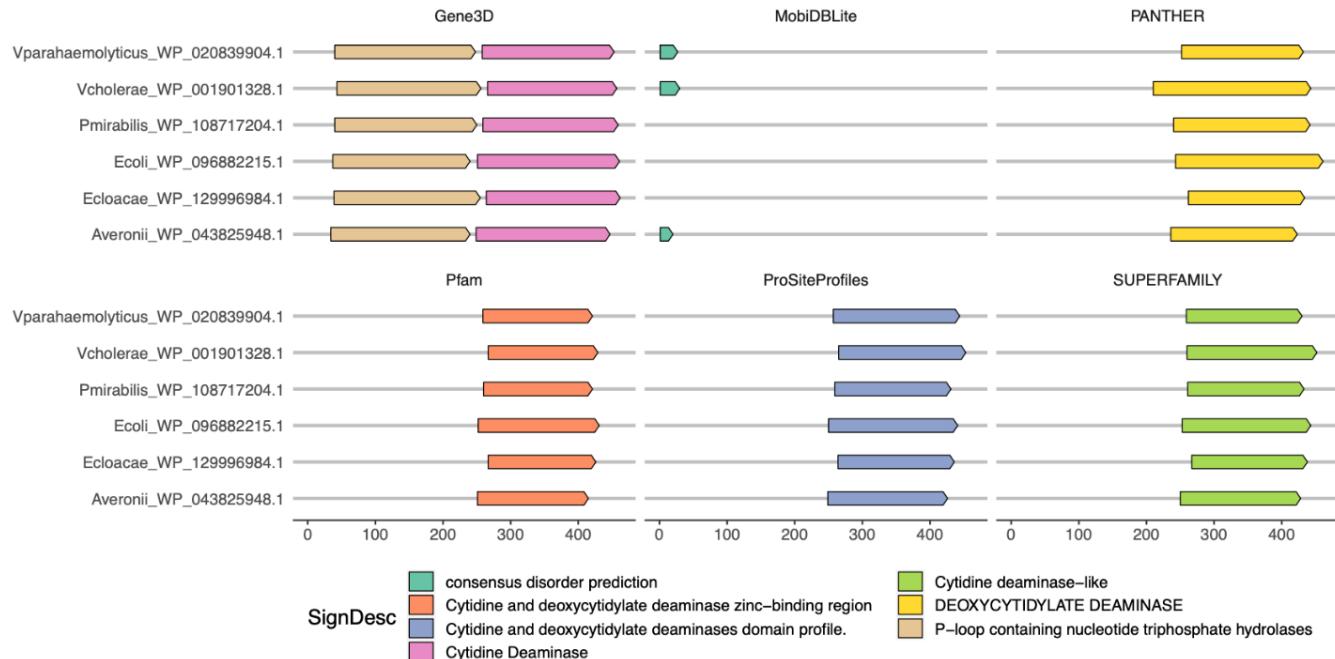
31
32 **Supplemental Figure 4. Δig^{222} has decreased $dcdV$ expression relative to WT *V. cholerae*.**
33 Relative difference in $dcdV$ expression between Δig^{222} and WT *V. cholerae* at three different
34 growth phases using qRT-PCR and an endogenous *gyrA* control. Data represent the mean \pm
35 SEM of three biological replicates.



36
37
38
39
40
41
42
43

Supplemental Figure 5. DcdV C-terminal 6x Histidine fusion maintains the same activity as the WT DcdV enzyme.

Representative images of *V. cholerae* WT and Δig^{222} cultures maintaining an empty vector plasmid (pVector) or P_{tac} -inducible *dcdV-6xHis* plasmid (pDcdV 6xHis) grown in the presence of 100 μM IPTG for 2 h. Cells were stained with FM4-64 prior to imaging and performed in biological triplicate.

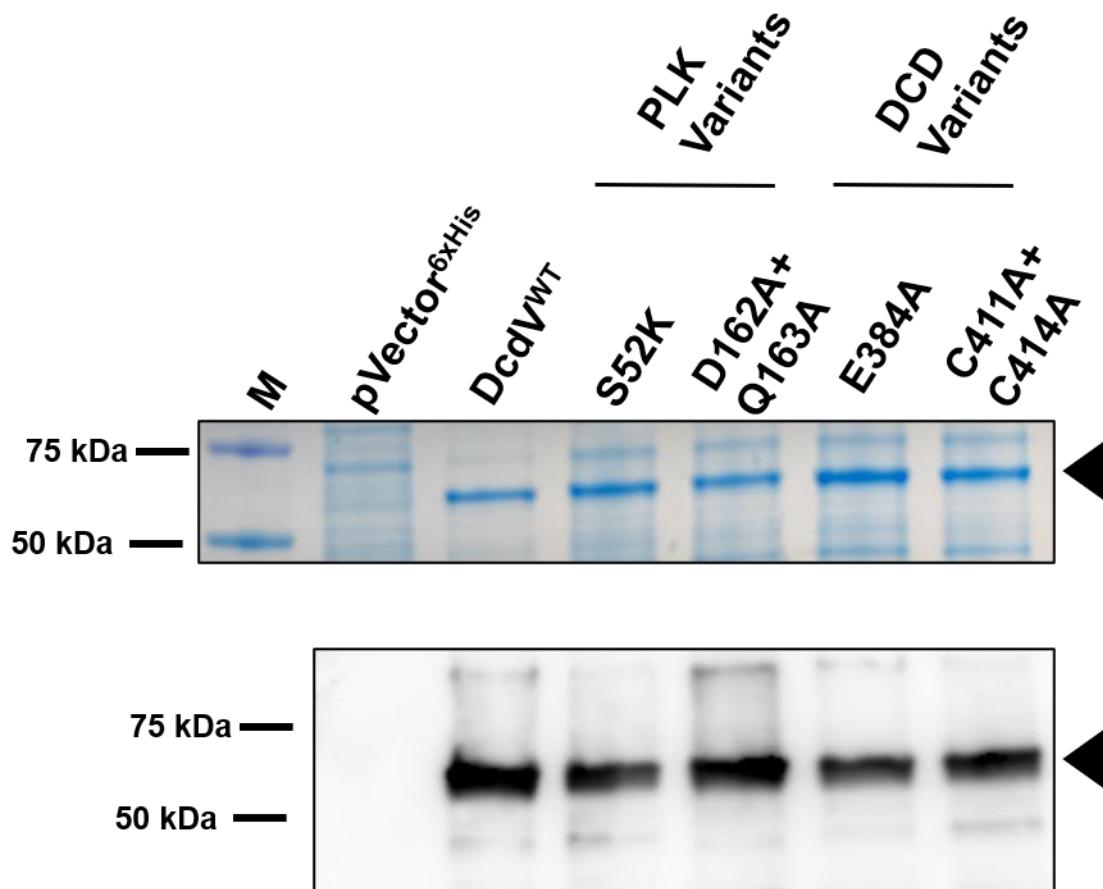


44

45

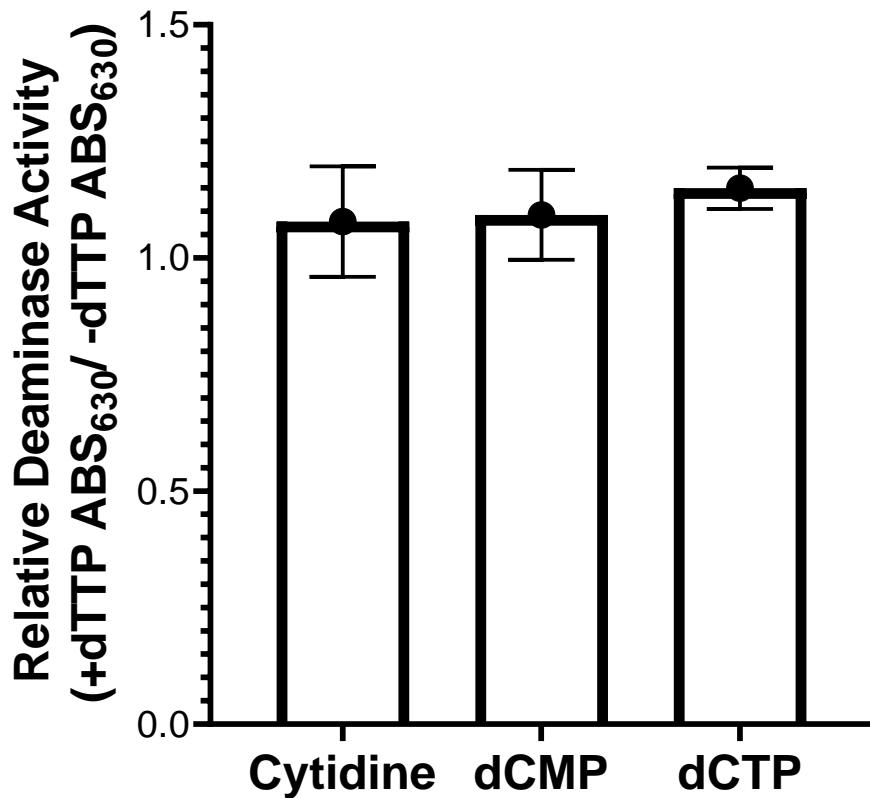
46 **Supplemental Figure 6. Domain architectures of the six DcdV query proteins.**

47 Domain architecture and secondary structure predictions for the six proteobacterial starting points
48 of interest (query proteins) using InterProScan [[40]; see Methods]. Results from six main
49 analyses are shown here for the query proteins: Gene3D (including CATH structure database),
50 Pfam, ProSiteProfiles, PANTHER, and SUPERFAMILY protein domain profile databases, and
51 MobiDBLite for disorder prediction. No transmembrane regions (using TMHMM) or
52 membrane/extracellular localization were predicted for any of the proteins (using Phobius); hence
53 not shown.



54
55
56 **Supplemental Figure 7. Cellular abundance of C-terminal 6x histidine tagged DcdV**
57 **variant fusions analyzed by Coomassie stain and Western blot.**

58 Representative Coomassie stained gel (top) and anti-6x His antibody Western blot (bottom) of
59 whole cell lysates from *E. coli* BL21(DE3) cells maintaining an empty vector (pVector^{6xHis}),
60 inducible C-terminal 6x histidine tagged *dcdV* (WT) or *dcdV* variants (S52K, D162A + Q163A,
61 E384A, and C411A + C414A) grown in the presence of 1 mM IPTG for 3 h. Sample inputs were
62 normalized by culture OD₆₀₀ and resolved by SDS-PAGE. Three biological replicates of each
63 strain were analyzed with similar results. Black triangles correspond to the predicted molecular
64 weight of the DcdV tagged fusions (60.6 kDa). M = molecular weight marker.



65
66
67 **Supplemental Figure 8. Addition of exogenous dTTP does not inhibit DcdV deaminase**
68 **activity in *E. coli* lysates.**
69 Lysates collected from *E. coli* expressing WT DcdV incubated with or without exogenous 7.5 mM
70 dTTP and either 75 mM cytidine, 7.5 mM dCMP, or 7.5 mM dCTP. The evolution of NH₄⁺ resulting
71 from substrate deamination was detected by measuring the solution ABS₆₃₀ after a Berthelot's
72 reaction in microtiter plates. The relative deaminase activity was calculated by dividing the ABS₆₃₀
73 of the +dTTP reaction by the no dTTP control reaction for each lysate. Data represent the mean
74 ± SEM of three biological replicate lysates.
75

76

Nucleotides	Absolute Intracellular dNTP Concentration in pmol/mg of Total Protein				
	pVector	pDcdV ^{WT}	pDcdV ^{S52K}	pDcdV ^{E384A}	pDcdV ^{ETEC}
dATP	75.6 ± 5.3	121.9 ± 18.1	64.6 ± 3.6	70.1 ± 5.7	59.1 ± 5.8
dCTP	6.6 ± 0.7	1.6 ± 0.3	5.8 ± 0.3	6.9 ± 0.7	2.4 ± 0.2
dGTP	42.7 ± 4.5	59.9 ± 13.9	35.7 ± 1.8	42.4 ± 3.7	32.8 ± 1.9
dTTP	11.4 ± 0.7	16.5 ± 4.7	9.0 ± 0.3	10.80 ± 0.4	10.4 ± 2.2
dUTP	1.1 ± 0.1	n.d.	0.9 ± 0.1	1.1 ± 0.10	n.d.
dCMP	7158.3 ± 2485.5	1185.7 ± 227.9	6859.4 ± 2226.4	7027.8 ± 1950.4	1945.1 ± 636.2
dUMP	9.3 ± 2.90	7.1 ± 0.3	9.1 ± 3.3	8.7 ± 1.1	9.9 ± 3.3

77

78

79

80

81

82

83 **Supplemental Figure 9. Absolute intracellular concentration of deoxynucleotides.**

84 Quantification of the indicated dNTPs in vivo, using UPLC-MS/MS, in strains expressing the
85 empty vector and four DcdV variants, as indicated. Data represents mean ± SEM, n=3.

86

87

88

89

90 **Supplemental Figure 10. ClustalW multiple sequence alignment of DcdV homologs**
91 **explored in this study.**

92 Amino acid alignment of DcdV and three homologs using webservice EMBL-EBI [74]. “*” indicates
 93 100% identity, “:” indicates >75%, and “.” Indicates >50% similarity. Open triangles above the
 94 alignments indicate conserved residues of PLK and DCD domains. Closed triangles indicate
 95 amino acids where single amino acid substitutions were found to render *V. cholerae* DcdV
 96 insensitive to DifV inhibition (Figs. 6A, B, and D).

97

98

V.cholerae	---MITSIH-----EYRNASNSDKNRLSHNSKSDTSGNVVLFSQLDAKKE--	KARANLLR	50
V.parahaemolyticus	---MVTNLN-----EYRDASTKSSVSYMS-TRKSASSNVVFTKECPKKA--	KARANLLR	49
P.mirabilis	---MMNVQHTVANDEFCSDEYHESIDCVHRNNRNSSPNVVSIELIREKKIYNDAIKRAAA		57
E.coli_ETEC	MSNQLTVIDGYAEREATNDNAE-----NLNISYICGNVVDFVNFQEKE	TINSFDKILK	55
	.. . * *** : :* .: .		

V.cholerae	AAQKIH 57
V.parahaemolyticus	AAQKIRW 56
P.mirabilis	VADTLDW 64
E.coli_ETEC	KEIILD- 61
	:

99

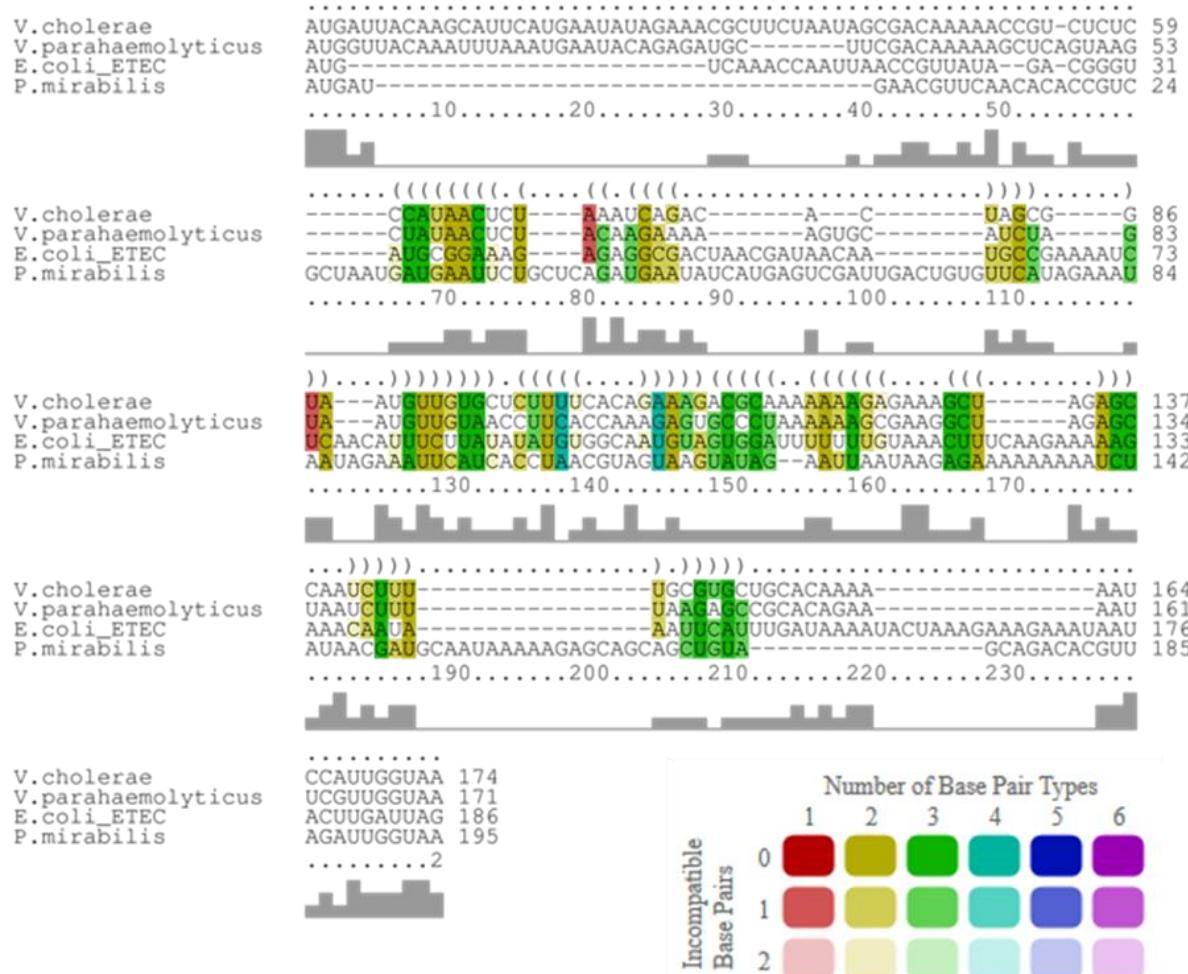
100

**101 Supplemental Figure 11. DifV (174 nt) and the three ORFs encoded upstream of dcdV
102 homologs do not exhibit amino acid similarity.**

103 Amino acid alignment of the *V. cholerae* Ig²²² translated ORF and three ORFs 5' of the *dcdV*
104 homologs using EMBL-EBI ClustalW [74]. “*” indicates 100% identity, “:” indicates >75%, and “.”
105 Indicates >50% similarity.

106

107



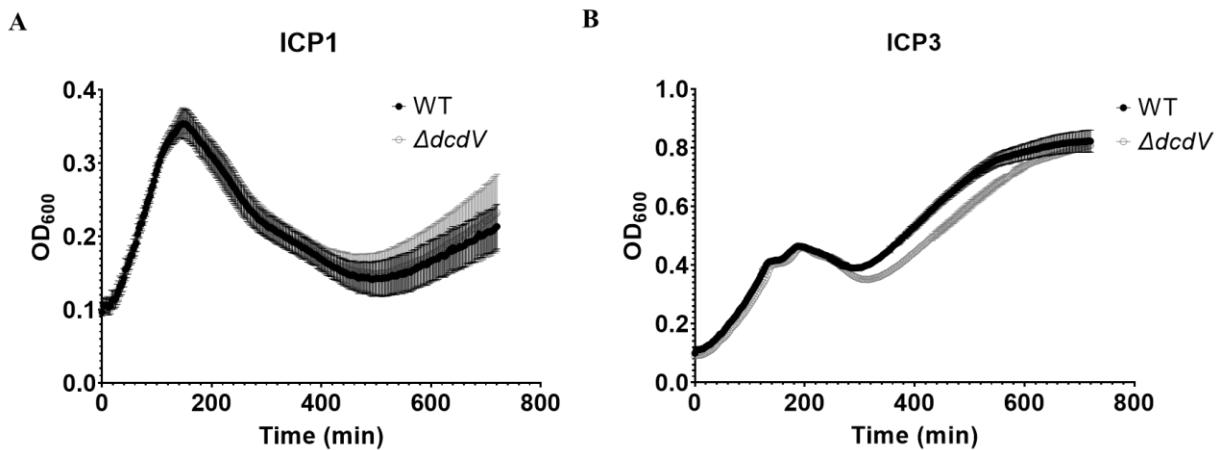
108

109 **Supplemental Figure 12.** DifV (174 nt) and the three ORFs encoded upstream of *dcdV*
110 homologs do not have exhibit similarity.

111 Nucleotide alignment of the *V. cholerae* DifV and the ORFs 5' of *dcdV* homologs using LocARNA
 112 [75]. Consensus identities are correlated with the height of the bars below the corresponding
 113 nucleotide (bottom). The average secondary structure is indicated in dot-bracket notation (top).
 114 Compatible base pairs are colored according to the number of different types C-G (1), G-C (2),
 115 A-U (3), U-A (4), G-U (5) or U-G (6) of compatible base pairs in the corresponding columns. The
 116 saturation decreases with the number of incompatible base pairs.

117

118



119

120

121

122

123

124

125

Supplemental Figure 13. *V. cholerae* lacking *dcdV* do not exhibit enhanced susceptibility to predation by *V. cholerae* lytic phage ICP1 and ICP3.

Growth curves for *V. cholerae* WT and $\Delta dcdV$ infected by lytic phage ICP1 (A) and ICP3 (B). Bacteria were infected at time 0 at an MOI of 0.1 in microtiter plates. Data represent the mean \pm SEM, $n=3$.

126

127 **Supplementary Table 1.** Strains and phages used in this study.

Strains	Name in this Study	Relevant Characteristics	Source or reference
<i>E. coli</i>			
DH10b		<i>F-mcrA Δ(mrr-hsdRMS-mcrBC)</i> <i>Φ80lacZΔM15 ΔlacX74 recA1</i> <i>endA1 araD139Δ(ara, leu)7697</i> <i>galU galK λrpsL nupG</i>	ThermoFisher Scientific
BW29427		<i>RP4-2(TetSkan1360::FRT), thrB1004, lacZ58(M15), ΔdapA1341::[erm</i> <i>pir^r], <i>rpsL</i>(strR), <i>thi</i>-, <i>hsdS</i>-, <i>pro</i>-</i>	Lab Stock
BL21(DE3)		<i>F- ompT hsdSB(rB -mB +) gal dcm</i> (DE3)	Lab Stock
078:H11 H10407 (ETEC)	ETEC	Wild type	[106]
<i>V. cholerae</i>			
C6706str2	WT	Wild type O1 El Tor; Sm ^R	[107]
E7946		Clinical isolate obtained in Bahrain in 1978; Sm ^R	[108]
CR01	ΔVSP-1	O1 El Tor ΔVSP-1	This study
CR02	ΔVSP-2	O1 El Tor ΔVSP-2	This study
CR03	ΔVSP-1/2	O1 El Tor ΔVSP-1/2	This study
BYH206	Δ <i>ig</i> ²²²	O1 El Tor Δ <i>ig</i> ²²² between vc0175-vc0176 position in N16961 chromosome I [177,230-177,008]	This study
BYH207	Δvc0176	O1 El Tor Δvc0176	This study
BYH255	Δvc0175-176	O1 El Tor Δvc0175-176	This study
BYH256	Δvc0177-181	O1 El Tor Δvc0177-181	This study
BYH257	Δvc0182-185	O1 El Tor Δvc0182-185	This study
GS05	Δvc0175	O1 El Tor Δvc0175 (dcdV)	This study
WLN5105	ΔcapV	O1 El Tor ΔcapV	[11]
<i>V. parahaemolyticus</i>			
O1:Kuk str. FDA_R31	VP	Wild type	[109]
<i>P. mirabilis</i>			
AR379	PM	Wild type	[110]
<i>Shigella flexneri</i>			
PE577	Sf	Wild type	[54]
Phages			
ICP1	ICP1	Wild type	[49]
ICP3	ICP3	Wild type	[49]
Sf6	Sf6	Wild type	[111]

128

129

130

131

132

133 **Supplementary Table 2.** Plasmids Descriptions

Plasmids	Name in this Manuscript	Relevant characteristics	Source or Reference
pEVS141	pVector/pVector ^{DifV}	pEVS143 without pTac; Km ^r	[70]
pEVS143		Broad-host range pTac overexpression vector; Km ^r	[68]
pMMB67EH	pVector ^{DcdV}	Broad-host range pTac overexpression vector; Amp ^r	[69]
pKAS32		Suicide vector for mutant construction, Amp ^r	[67]
pET28b	pVector ^{6xHis}	T7 promoter; Km ^r	Novagen
pLAFR	pLAFR	pLAFR; Tet ^r	[112]
pCCD7	pCCD7	pLAFR::VSP-1; Tet ^r	[11]
pBRP353	pDncV	pMMB67EH::dncV; Amp ^r	[11]
pCMW204	pDcdV	pEVS143::dcdV; Km ^r	This study
pGBS87	pDcdV/pDcdV ^{VC}	pMMB67EH::dcdV; Amp ^r	This study
pGBS65	pDcdV ^{6xHis}	pET28b::dcdV-6xHis C-term; Km ^r (*only* in <i>E. coli</i> BL21(DE3))	This study
pGBS98	pDcdV ^{6xHis}	pEVS143::dcdV-6xHis C-term; Km ^r (*only* in <i>V. cholerae</i>)	This study
pGBS71	pDcdV ^{E384A}	pEVS143::dcdV-E384A; Km ^r	This study
pGBS82	pDcdV ^{E384A}	pET28b::dcdV-E384A-6xHis C-term; Km ^r (*only* for in vitro and western blot)	This study
pGBS81	pDcdV ^{C411A+C414A}	pEVS143::dcdV-C411A+C414A; Km ^r	This study
pGBS75	pDcdV ^{C411A+C414A}	pET28b::dcdV-C411A+C414A-6xHis C-term; Km ^r	This study
pGBS103	pDcdV ^{S52K}	pEVS143::dcdV-S52K; Km ^r	This study
pGBS114	pDcdV ^{S52K}	pET28b::dcdV-S52K-6xHis C-term; Km ^r	This study
pGBS106	pDcdV ^{D162A+Q163A}	pEVS143::dcdV-D162A+Q163A; Km ^r	This study
pGBS116	pDcdV ^{D162A+Q163A}	pET28b::dcdV-D162A+Q163A-6xHis C-term; Km ^r	This study
pGBS80	plg ²²²	pEVS143::lg ²²² , (position in N16961 chromosome I [177,230-177,008]); Km ^r	This study
pGBS108	plg ^{222-STOP}	pEVS143::ig ²²² -1C>T, 2T>A; Km ^r	This study
pGBS110	pDifV	pEVS143::difV (position in N16961 chromosome I [177,181-177,008]); Km ^r	This study
pAW01	pDifV ^{RBS-less}	pEVS143::difV without RBS; Km ^r	This study
pGBS111	pDifV ^{STOP}	pEVS143::difV-1A>T, 2T>A, 3G>A; Km ^r	This study
pGBS118	pDifV ^{InteriorSTOP}	pEVS143::difV-17A>T, 18T>A, 19G>A; Km ^r	This study

pBYH49	pDifV ⁴⁹⁻¹⁸⁶	pEVS143:: <i>difV</i> (49-186 NT); Km ^r	This study
pBYH52	pDifV ⁴⁹⁻²⁰⁴	pEVS143:: <i>difV</i> (49-204 NT); Km ^r	This study
pBYH53	pDifV ⁴⁹⁻²¹⁴	pEVS143:: <i>difV</i> (49-214 NT); Km ^r	This study
pBYH54	pDifV ⁴⁹⁻²¹⁸	pEVS143:: <i>difV</i> (49-218 NT); Km ^r	This study
pBYH55	pDifV ⁶⁶⁻²²²	pEVS143:: <i>difV</i> (66-222 NT); Km ^r	This study
pBYH56	pDifV ⁸⁶⁻²²²	pEVS143:: <i>difV</i> (86-222 NT); Km ^r	This study
pBYH57	pDifV ¹²³⁻²²²	pEVS143:: <i>difV</i> (123-222 NT); Km ^r	This study
pBYH50	pNpcR_3991	pEVS143:: <i>npcR</i> _3991; Km ^r	This study
pGBS120	pDcdV ^{E123K}	pEVS143:: <i>dcdV</i> -E123K; Km ^r	This study
pGBS131	pDcdV ^{A126T}	pEVS143:: <i>dcdV</i> -A126T; Km ^r	This study
pGBS128	pDcdV ^{K201R}	pEVS143:: <i>dcdV</i> -K201R; Km ^r	This study
pGBS129	pDcdV ^{K511R}	pEVS143:: <i>dcdV</i> -K511R; Km ^r	This study
pGBS130	pDcdV ^{Q514R}	pEVS143:: <i>dcdV</i> -Q514R; Km ^r	This study
pGBS124	pDcdV ^{ETEC}	pEVS143:: <i>dcdV</i> from <i>Escherichia coli</i> O78:H11 H10407 (ETEC); Km ^r (*only* for mass spec experiment)	This study
pGBS125	pDifV ^{ETEC}	pEVS143:: <i>difV</i> from <i>Escherichia coli</i> O78:H11 H10407 (ETEC); Km ^r	This study
pGBS126	pDcdV ^{ETEC}	pMMB67EH:: <i>dcdV</i> from <i>Escherichia coli</i> O78:H11 H10407 (ETEC); Amp ^r	This study
pAW07	pDifV ^{VP}	pEVS143:: <i>difV</i> from <i>V. parahaemolyticus</i> O1:Kuk str. FDA_R31; Km ^r	This study
pAW06	pDcdV ^{VP}	pMMB67EH:: <i>dcdV</i> from <i>V. parahaemolyticus</i> O1:Kuk str. FDA_R31; Amp ^r	This study
pAW02	pDifV ^{PM}	pEVS143:: <i>difV</i> from <i>P. mirabilis</i> AR379; Km ^r	This study
pAW04	pDcdV ^{PM}	pMMB67EH:: <i>dcdV</i> from <i>P. mirabilis</i> AR379	This study
pCRR01		Deletion construct for ΔVSP-1, Amp ^r	This study
pCRR02		Deletion construct for ΔVSP-2, Amp ^r	This study
pBYH36		Deletion construct for Δ <i>ig</i> ²²² , Amp ^r	This study
pBYH37		Deletion construct for Δ <i>vc0176</i> , Amp ^r	This study
pBYH40		Deletion construct for Δ <i>dcdV-vc0176</i> , Amp ^r	This study
pBYH41		Deletion construct for Δ <i>vc0177-vc0181</i> , Amp ^r	This study
pBYH42		Deletion construct for Δ <i>vc0182-VC0185</i> , Amp ^r	This study
pGBS88		Deletion construct for Δ <i>dcdV</i> , Amp ^r	This study

135 **Supplementary Table 3.** Oligonucleotides used in this study.

Name	Primer use	Sequence	Reference
Vector Construction			
CMW3009	<i>dcdV</i> F ¹ EcoRI + RBS ³ (pEVS143-DcdV)	GGAAACAGCCTCGACAGGCCTAGGAG GAAGCTAATTGTTACAATGAATAAGT CCTCCG	This study
CMW3010	<i>dcdV</i> R ² BamHI (pEVS143-DcdV)	CATAAAGCTTGCTCAATCAATCACCAG ATCCTAGTCTGGATGCTCTTC	This study
CMW3025	<i>dcdV</i> F EcoRI + RBS (pMMB67EH-DcdV)	ATTCACACAGGAAACAGAGGGAGCTAA GGAAGCTAATTGTTACAATGAATAAG TCCTC	This study
CMW3026	<i>dcdV</i> R BamHI (pMMB67EH-DcdV)	CCTGCAGGTCGACTCTAGAGCTAGTCT TGGATGCTCTC	This study
CMW3066	<i>dcdV</i> +6His R BamHI (pEVS143-DcdV-His ₆)	GCTTGCTCAATCAATCACCAGGTTAGTGG TGGTGGTGGTGGTGGCTCGATGTCTTGG	This study
CMW3079	<i>Ig</i> ²²² F EcoRI + RBS (pEVS143- <i>Ig</i> ²²²)	CAGCCTCGACAGGCCTAGGAGGGAGCT AAGGAAGCTAAACTGTTGCAAATCAT ACTTTAG	This study
CMW3080	<i>Ig</i> ²²² R BamHI (pEVS143- <i>Ig</i> ²²² , pEVS143-DifV & pEVS143-DifV 3' end truncations and interior stop codon)	GCTTGCTCAATCAATCACCAGGTTACCAAT GGATTTTTGTG	This study
CMW3081	<i>Ig</i> ^{222-STOP} F EcoRI + RBS (pEVS143- <i>Ig</i> ^{222- STOP})	CAGCCTCGACAGGCCTAGGAGGGAGCT AAGGAAGCTAAATAGTTGCAAATCAT AGTTAG	This study
CMW3093	<i>dcdV</i> F Ncol (pET28b-DcdV-His ₆)	AACTTAAGAAGGAGATATACATGTTA CAATGAATAAGTCCTCCGC	This study
CMW3094	<i>dcdV</i> R Xhol (pET28b-Dcdv-His ₆)	CTCAGTGGTGGTGGTGGTGGTGGCTCG ATGTCCTGGATGCTCTCTTCACTCG ATGG	This study
CMW3102	<i>difV</i> F EcoRI + RBS (pEVS143-DifV & pEVS143-DifV 5' end truncations)	CTCGACAGGCCTAGGAGGGAGCTAAGG AAGCTAAAATGATTACAAGCATTGATGA ATATAG	This study
CMW3103	<i>difV</i> F EcoRI + RBS (pEVS143-DifV ^{STOP})	CTCGACAGGCCTAGGAGGGAGCTAAGG AAGCTAAAATGATTACAAGCATTGATGA ATATAG	This study
CMW3128	<i>difV</i> ⁴⁹⁻¹⁸⁶ F EcoRI + RBS (pEVS143-DifV ^{49- 186})	ACAGCCTCGACAGGCCTAGGAGGGAGC TAAGGAAGCTAAAATGATTACAAG	This study
CMW2129	<i>difV</i> ⁴⁹⁻¹⁸⁶ R BamHI (pEVS143-DifV ⁴⁹⁻¹⁸⁶)	GCTTGCTCAATCAATCACCAGGCTCTA GCTTTCTCTTTTGCCTCTTC	This study
CMW3130	<i>npvR</i> _3991 F EcoRI + RBS (pEVS143- <i>npvR</i> _3991)	ACAGCCTCGACAGGCCTAGGAGGGAGC TAAGGAAGCTAAAATCTCCCATGACTC	This study

CMW3131	<i>npcR_3991</i> R BamHI (pEVS143- <i>npcR_3991</i>)	GCTTGCTCAATCAATCACCGTGTGCAG CACGCAAAAGATTGGCTCTAGCT	This study
CMW3162	<i>dcdV^{ETEC}</i> F EcoRI + RBS (pEVS143- <i>DcdV^{ETEC}</i>)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGGCTATAGCTTT GAAAAAG	This study
CMW3163	<i>dcdV^{ETEC}</i> R BamHI (pEVS143- <i>DcdV^{ETEC}</i>)	GCTTGCTCAATCAATCACCGTTAAATCA AGTCATCTTGTGTTG	This study
CMW3164	<i>dcdV^{ETEC}</i> F EcoRI + RBS (pMMB67EH- <i>DcdV^{ETEC}</i>)	AATTCACACAGGAAACAGAGGGAGCTA AGGAAGCTAAAATGGCTATAGCTTGA AAAAGG	This study
CMW3165	<i>dcdV^{ETEC}</i> F BamHI (pMMB67EH- <i>DcdV^{ETEC}</i>)	CCTGCAGGTCGACTCTAGAGTTAAATC AAGTCATCTTGTGTTGG	This study
CMW3166	<i>difV^{ETEC}</i> F EcoRI + RBS (pEVS143- <i>DifV^{ETEC}</i>)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGTCAAACCAATTAA ACCG	This study
CMW3167	<i>dcdV^{ETEC}</i> F BamHI (pEVS143- <i>DifV^{ETEC}</i>)	GCTTGCTCAATCAATCACCGCTAATCA AGTATTATTCTTTCTTAGTATTTATC	This study
CMW3180	<i>difV^{WP}</i> F EcoRI + RBS (pEVS143- <i>DifV^{WP}</i>)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGGTTACAAATTAA AATG	This study
CMW3181	<i>difV^{WP}</i> R BamHI (pEVS143- <i>DifV^{WP}</i>)	GCTTGCTCAATCAATCACCGTTACCAA CGAATTTCTGTGCGGCTCTAAAAG	This study
CMW3184	<i>dcdV^{WP}</i> F EcoRI + RBS (pMMB67EH- <i>DcdV^{WP}</i>)	CAATTTCACACAGGAAACAGAGGGAGCT AAGGAAGCTAAAATGGAAAATCCTCTA	This study
CMW3185	<i>dcdV^{WP}</i> R BamHI (pMMB67EH- <i>DcdV^{WP}</i>)	CCTGCAGGTCGACTCTAGAGTTATTCA ATAGTGGCTTCACTTGTGCTTGTGA ATG	This study
CMW3189	<i>difV</i> F EcoRI (pEVS143- <i>DifV</i>)	ACAGCCTCGACAGGCCTAGGATGATTA CAAGCATTATGAATATAGAAACGCTTC	This study
CMW3192	<i>difV^{PM}</i> F EcoRI + RBS (pEVS143- <i>DifV^{PM}</i>)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATGAACGTTCAAC	This study
CMW3193	<i>difV^{PM}</i> R BamHI (pEVS143- <i>DifV^{PM}</i>)	GCTTGCTCAATCAATCACCGTTACCAAT CTAACGTGCTGCTACAGCTGC	This study
CMW3196	<i>dcdV^{WP}</i> F EcoRI + RBS (pMMB67EH- <i>DcdV^{PM}</i>)	CAATTTCACACAGGAAACAGAGGGAGCT AAGGAAGCTAAAATGGGTAAATCC	This study
CMW3197	<i>dcdV^{WP}</i> R BamHI (pMMB67EH- <i>DcdV^{PM}</i>)	CCTGCAGGTCGACTCTAGAGTTAACTT CTCTCTCACCTAACGAAGATTAC	This study
CMW3200	<i>difV⁴⁹⁻²⁰⁴</i> R BamHI (pEVS143- <i>DifV⁴⁹⁻²⁰⁴</i>)	GCTTGCTCAATCAATCACCGTGCAGCA CGCAAAAGATTG	This study
CMW3201	<i>difV⁴⁹⁻²¹⁴</i> R BamHI (pEVS143- <i>DifV⁴⁹⁻²¹⁴</i>)	GCTTGCTCAATCAATCACCGGGATTGTT TGTGCAGCAC	This study
CMW3202	<i>difV⁴⁹⁻²¹⁸</i> R BamHI (pEVS143- <i>DifV⁴⁹⁻²¹⁸</i>)	GCTTGCTCAATCAATCACCGCAATGGA TTTTTGTGCAAGCACGCAAAAGA	This study
CMW3203	<i>difV⁶⁶⁻²²²</i> F EcoRI + RBS (pEVS143- <i>DifV⁶⁶⁻²²²</i>)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAGAATATAGAAACG	This study

CMW3204	<i>difV</i> ⁸⁶⁻²²² F EcoRI + RBS (pEVS143-DifV ⁸⁶⁻²²²)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAATAGCGACAAAAAC	This study
CMW3205	<i>difV</i> ¹²³⁻²²² F EcoRI + RBS (pEVS143-DifV ¹²³⁻²²²)	ACAGCCTCGACAGGCCTAGGAGGAGC TAAGGAAGCTAAAAGACACTAGCG	This study
Site-directed Mutagenesis			
CMW3011	<i>dcdV</i> (E384A) F (pEVS143-DcdV ^{E384A} & pET28b-DcdV ^{E384A})	CAAGAGCGGTTCATGCTGCAATGGATT CTCTTATAGC	This study
CMW3012	<i>dcdV</i> (E384A) R (pEVS143-DcdV ^{E384A} & pET28b-DcdV ^{E384A})	GCTATAAGAGAATCCATTGCAGCATGA ACCGCTCTTG	This study
CMW3013	<i>dcdV</i> (C411A + C414A) F (pEVS143-DcdV ^{C411A+C414A})	TATATGTTACGACATATCCGGCTCACAA CGCTGCGCGACACATCGTTGCTG	This study
CMW3014	<i>dcdV</i> (C411A + C414A) R (pEVS143-DcdV ^{C411A+C414A})	CAGCAACGATGTGTCGCGCAGCGTTGT GAGCCGGATATGTCGTAACATATA	This study
CMW3021	<i>dcdV</i> (K55A) F (pEVS143-DcdV ^{K55A})	GCTATTGGCTCTGGTAGCGGCATTA AAAGAGAGTTAGTTAGTTCTCTTGAGA CATAT	This study
CMW3022	<i>dcdV</i> (K55A) R (pEVS143-DcdV ^{K55A})	ATATGTCTCAAGAGAACTAACTAAACTC TCTTTAATGCCGCTACACCAGAGCCA ATAGC	This study
CMW3104	<i>dcdV</i> (D162A + Q163A) F (pEVS143-DcdV ^{D162A+Q163A})	CGCATACATCATCGCGCGTAAAGCA CCCTGATGAAATCAAATTCC	This study
CMW3105	<i>dcdV</i> (D162A + Q163A) R (pEVS143-DcdV ^{Q162A+Q163A})	GGAATTGATTCATCAGGGTGCTTAA CGCCCGCGATGATGTATGCG	This study
CMW3110	<i>dcdV</i> (S52K) F (pEVS143-DcdV ^{S52K})	CCTCTGTGGGGCTATTGGCAAAGGTGT AAAGGCATTAAGAGAG	This study
CMW3111	<i>dcdV</i> (S52K) R (pEVS143-DcdV ^{S52K})	CTCTCTTTAATGCCTTACACCTTGCA AATAGCCCCACAGAGG	This study
CMW3112	<i>dcdV</i> (S52P) F (pEVS143-DcdV ^{S52P})	CCTCTGTGGGGCTATTGGCCCGGGTG TAAAGGCATTAAGAGAG	This study
CMW3113	<i>dcdV</i> (S52P) R (pEVS143-DcdV ^{S52P})	CTCTCTTTAATGCCTTACACCCGGG CCAATAGCCCCACAGAGG	This study
CMW3114	<i>dcdV</i> (S52W) F (pEVS143-DcdV ^{S52W})	CCTCTGTGGGGCTATTGGCTGGGTG TAAAGGCATTAAGAGAG	This study
CMW3115	<i>dcdV</i> (S52K) R (pEVS143-DcdV ^{S52W})	CTCTCTTTAATGCCTTACACCCAGC CAATAGCCCCACAGAGG	This study
CMW3118	<i>difV</i> (interior alternative frame stop) F (pEVS143-DifV17A>T, 18T>A, 19G>A)	AAGGAAGCTAAAATGATTACAAGCATT CTAAAATATAGAAACGCTTCTAATAGCG	This study
CMW3119	<i>difV</i> (interior alternative frame stop)	CGCTATTAGAAGCGTTCTATATTTAG AATGCTTGTAAATCATTAGCTTCCTT	This study

	R (pEVS143-DifV17A>T, 18T>A, 19G>A)		
CMW3124	<i>dcdV</i> (E123K) F (pEVS143-DcdV ^{E123K})	GCAGCCTGTGCTATCAAAGAAATTGCGCTGG	This study
CMW3125	<i>dcdV</i> (E123K) R (pEVS143-DcdV ^{E123K})	CCAGCGCAATTCTTTGATAGCACAGGCTGC	This study
CMW3172	<i>dcdV</i> (A126T) F (pEVS143-DcdV ^{A126T})	GCTATCGAAGAAATTACGCTGGAAAGAACATTAATCTGTC	This study
CMW3173	<i>dcdV</i> (A126T) R (pEVS143-DcdV ^{A126T})	GACAGATTAATGTTCTTCCAGCGTAATTCTTCGATAGC	This study
Gene Deletion			
CMW2794	ΔVSP-2 up ⁴ F; CR02 & CR03	GTGGAATTCCCGGGAGAGCTCGGCTT GTTCACTATCGTAATAATGC	This study
CMW2795	ΔVSP-2 up R; CR02 & CR03	GGAGGGGCCACCACTGGGAGGGCACCAGATTC	This study
CMW2796	ΔVSP-2 down ⁵ F; CR02 & CR03	GCCCTCCCAGTGGTGGCCCCCTCCCAGGT	This study
CMW2797	ΔVSP-2 down R; CR02 & CR03	AGCTATAGTTCTAGAGGTACGGCATT AAGGTGGTGGAAACCG	This study
CMW2814	ΔVSP-1 up F; CR01 & CR03	GTGGAATTCCCGGGAGAGCTGGCTTTA CTGTTATTCGC	This study
CMW2815	ΔVSP-1 up R; CR01 & CR03	TACCATGTAGTAGCGGTATCGAGATTCC	This study
CMW2816	ΔVSP-1 down F; CR01 & CR03	GATACCGCTACTACATGGAACGAACTCTTC	This study
CMW2817	ΔVSP-1 down R; CR01 & CR03	AGCTATAGTTCTAGAGGTACCGCTAAGTTTGTGGATGC	This study
CMW2970	Δvc0176 up F; BYH207	ATAACAATTGTGGAATTCCCGGGAGAGCTGGGAATCGAACATTGAGAG	This study
CMW2971	Δvc0176 up R; BYH207	ATATAGTGTCTCTATTATGGCTCATAAATCTTGAAG	This study
CMW2972	Δvc0176 down F; BYH207	GATTATGAGGCCATAAATAGAGACACTATATTAGTGTAAATTAAAC	This study
CMW2973	Δvc0176 down R; BYH207	TGCGCATGCTAGCTATAGTTCTAGAGGTTACTATGAAACTTATTCTATACTCTCAG	This study
CMW3035	Δvc0176-vc0175 up F; BYH255	ATAACAATTGTGGAATTCCCGGGAGAGCTGGGAATCGAACATTGAGAG	This study
CMW3036	Δvc0176-vc0175 up R; BYH255	TTTCCAGACTAAAGTTATGGCTCATAAATCTTGAAG	This study
CMW3037	Δvc0176-vc0175 down F; BYH255	GATTATGAGGCCATAACTTAGCTCTGGAAATTCACTTTTC	This study
CMW3038	Δvc0176-vc0175 down R; BYH255	TGCGCATGCTAGCTATAGTTCTAGAGGTACACATGGAGCATGATCAGG	This study
CMW3039	Δvc0177-vc0181 up F; BYH256	ATAACAATTGTGGAATTCCCGGGAGAGCTGTTGTATGTTGGGTG	This study
CMW3040	Δvc0177-vc0181 up R; BYH256	AATGAATTAGTATACGTATTCTAATACCACTAAAAACTAAG	This study

CMW3041	$\Delta vc0177$ -vc0181 down F; BYH256	TGGTATTAGAAATACGTATACTAATTCA TTCACTGTACTTC	This study
CMW3042	$\Delta vc0177$ -vc0181 down R; BYH256	TGCGCATGCTAGCTATAGTTCTAGAGG TACAAAGTTCTCCACAAATTTCAG	This study
CMW3043	$\Delta vc0182$ -vc0185 up F; BYH257	ATAACAATTGTGGAATTCCCGGGAGA GCTGCTGACTCCGGTGGCCGT	This study
CMW3044	$\Delta vc0182$ -vc0185 up R; BYH257	CTTAGGTATACTAATTGTATTGTGATATA CATAGAGGCTAGTATGGTTCCAGAGT TTAC	This study
CMW3045	$\Delta vc0182$ -vc0185 down F; BYH257	TGTATATCAAATACAATTAGTATACCTA AGATTGATTTTC	This study
CMW3046	$\Delta vc0182$ -vc0185 down R; BYH257	TGCGCATGCTAGCTATAGTTCTAGAGG TACTTCTCAGGATGTAATATTGTG	This study
CMW3067	$\Delta vc0175$ up F; GS05	GTGGAATTCCCGGGAGAGCTACTATAT TTAGTGTAAATTACAAAAAAC	This study
CMW3068	$\Delta vc0175$ up R; GS05	CAGACTAAAGCCTGAAATTATGAAACTT ATTTCTATAC	This study
CMW3069	$\Delta vc0175$ down F; GS05	TAATTCAGGCTTAGTCTGGAAAATTC ACTTTTC	This study
CMW3070	$\Delta vc0175$ down R; GS05	AGCTATAGTTCTAGAGGTACACATGGA GCATGATCAGG	This study
CMW3071	ΔIg^{222} up F; BYH206	ATAACAATTGTGGAATTCCCGGGAGA GCTTCTCAAAGAACGACGTAAAAAAG	This study
CMW3072	ΔIg^{222} up R; BYH206	CAAGAATTAAACGTGGTAAAGTGCAC ATTCTAC	This study
CMW3073	ΔIg^{222} down F; BYH206	AATGTGCGCACTTACCAACGTTAATTCT TGATTAGC	This study
CMW3074	ΔIg^{222} down R; BYH206	TGCGCATGCTAGCTATAGTTCTAGAGG TACTCATTCTCTGAGGTTTC	This study

qRT-PCR

CMW2926	<i>gyrA</i> F	TGGCCAGCCAGAGATCAAG	This study
CMW2927	<i>gyrA</i> R	ACCCGCAGCGGTACGA	This study
CMW3206	<i>dcdV</i> F	TCGACCAGTTAAAGCACCCCT	This study
CMW3207	<i>dcdV</i> R	CCTTCTGTACGGATCAAGCCA	This study
CMW3208	<i>difV</i> F	GTGAATGGATATTCGGTGGGA	This study
CMW3209	<i>difV</i> R	TTGTCGCTATTAGAACGCGTT	This study
CMW3288	<i>ori</i> F	CAGGTGAACCAGCAAAATCGA	[101]
CMW3289	<i>ori</i> R	TGGTATTGAAGCTCAATGCGG	[101]
CMW3290	<i>ter</i> F	TTCAAGCTGAGGCGGGATTG	[101]
CMW3291	<i>ter</i> R	GCTCATTGGCTTCTGTGCTT	[101]

1¹F = Forward

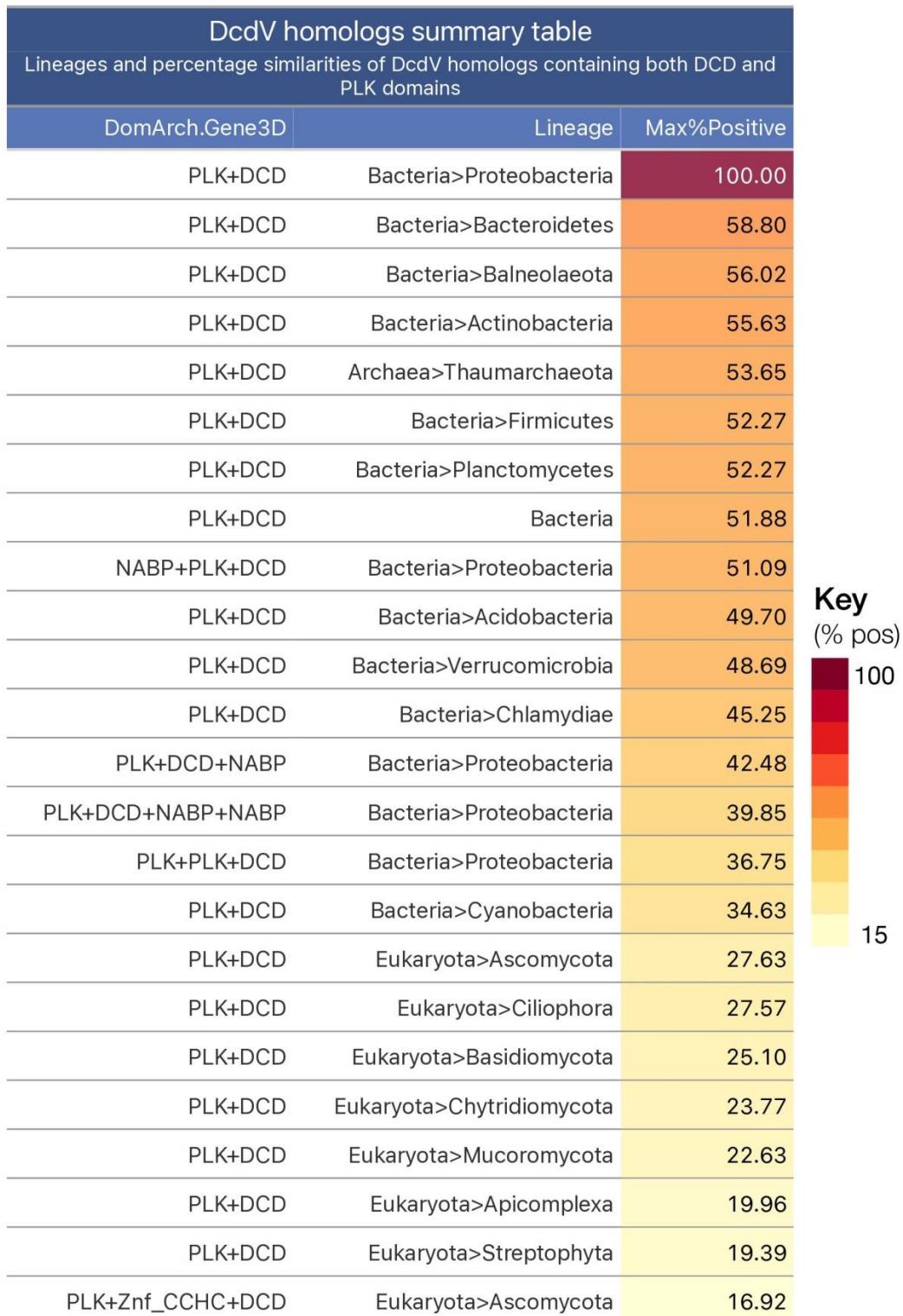
2²R= Reverse

3³RBS= Ribosomal Binding Site

4⁴Up= Amplifies Upstream Fragment

5⁵Down= Amplifies Downstream Fragment

142 **Supplementary Table 4.** This table sorts the indicated lineages by the DcdV homolog in that
 143 group with the maximum amino acid similarity to *V. cholerae* DcdV.



Abbreviations. PLK, P-loop containing nucleotide triphosphate hydrolases; DCD, Cytidine Deaminase domain 2; NABP, Nucleic acid-binding proteins; Znf_CCHC, Zinc finger CCHC-type

144

Partial Correlation Value w_{ij} of VSP-1 Genes i to j (Supplemental File 1)

	VC0175	VC0176	VC0177	VC0178	VC0179	VC0180	VC0181	VC0182	VC0183	VC0184	VC0185
VC0175	1	-0.301	-0.034	-0.459	0.147	-0.085	-0.072	0.089	-0.068	0.02	-0.099
VC0176	-0.301	1	0.145	-0.394	0.116	-0.048	-0.036	0.077	-0.055	0.028	-0.031
VC0177	-0.034	0.145	1	-0.043	0.026	-0.038	0.03	0.095	0.069	0.151	0.021
VC0178	-0.459	-0.394	-0.043	1	0.225	-0.098	-0.057	0.086	-0.11	0.027	-0.146
VC0179	0.147	0.116	0.026	0.225	1	0.501	0.303	-0.035	0.059	-0.002	0.062
VC0180	-0.085	-0.048	-0.038	-0.098	0.501	1	0.293	0.025	-0.024	0.001	-0.008
VC0181	-0.072	-0.036	0.03	-0.057	0.303	0.293	1	0.042	-0.005	0.007	-0.035
VC0182	0.089	0.077	0.095	0.086	-0.035	0.025	0.042	1	0.088	0.568	0.088
VC0183	-0.068	-0.055	0.069	-0.11	0.059	-0.024	-0.005	0.088	1	0.166	0.459
VC0184	0.02	0.028	0.151	0.027	-0.002	0.001	0.007	0.568	0.166	1	0.141
VC0185	-0.099	-0.031	0.021	-0.146	0.062	-0.008	-0.035	0.088	0.459	0.141	1

$w_{ij} = -1$ genes i and j never occur in the same species

$w_{ij} = 0$ expected co-occurrence between unrelated genes i and j drawn from a normal distribution

$w_{ij} > 0.045$ suggests shared biological function (Kim and Peterson 2011)

$w_{ij} = 1$ genes i and j always occur in the same species

Partial Correlation Value w_{ij} of VSP-2 Genes i to j (Supplemental File 2)

	VC0490	VC0491	VC0492	VC0493	VC0494	VC0495	VC0496	VC0497	VC0498	VC0502	VC0503	VC0504
VC0490	1	0.035	0.381	0.021	-0.067	0.071	0.032	-0.016	-0.124	-0.05	-0.033	0.011
VC0491	0.035	1	0.426	-0.021	-0.001	-0.008	0.014	0.008	0.014	-0.019	0.004	0
VC0492	0.381	0.426	1	0.027	0.006	0.003	0.013	0.004	0.004	0.01	-0.013	-0.021
VC0493	0.021	-0.021	0.027	1	0.008	-0.007	0.062	0.006	0.02	0.003	0.018	0.195
VC0494	-0.067	-0.001	0.006	0.008	1	0.656	0.084	0.008	-0.11	-0.022	0.03	-0.028
VC0495	0.071	-0.008	0.003	-0.007	0.656	1	0.104	0.118	0.085	-0.007	-0.009	0.02
VC0496	0.032	0.014	0.013	0.062	0.084	0.104	1	0.009	0.031	-0.039	-0.008	0.102
VC0497	-0.016	0.008	0.004	0.006	0.008	0.118	0.009	1	-0.058	-0.043	0.028	-0.004
VC0498	-0.124	0.014	0.004	0.02	-0.11	0.085	0.031	-0.058	1	-0.014	-0.028	0.028
VC0502	-0.05	-0.019	0.01	0.003	-0.022	-0.007	-0.039	-0.043	-0.014	1	0.088	0.013
VC0503	-0.033	0.004	-0.013	0.018	0.03	-0.009	-0.008	0.028	-0.028	0.088	1	-0.018
VC0504	0.011	0	-0.021	0.195	-0.028	0.02	0.102	-0.004	0.028	0.013	-0.018	1
VC0505	-0.011	0.003	0.013	0.097	-0.02	0.017	0.108	0.018	-0.022	-0.004	-0.007	0.389
VC0506	-0.006	0.001	0.002	-0.004	-0.026	0.024	0.045	0.018	-0.005	0.01	0.064	0.09
VC0507	-0.018	0.009	0.035	0.179	-0.003	0.009	0.03	-0.011	0.005	-0.01	-0.003	0.369
VC0508	0.053	0.007	-0.002	-0.029	0.028	0.111	0.073	0.06	0.017	0.123	0.098	0.003
VC0509	0.012	-0.001	-0.023	0.037	0.057	-0.02	0.216	0.023	0.058	0.095	-0.022	0.161
VC0510	0.015	0.011	0.001	-0.012	0.049	0.067	-0.005	0.085	0.064	0.128	0.174	-0.004
VC0512	-0.015	-0.006	0	0.004	-0.006	0.028	0.041	0.037	0.012	0.154	0.1	-0.002
VC0513	0.001	0.01	0.004	0.03	-0.013	-0.007	0.049	-0.008	-0.01	0.09	0.045	-0.01
VC0514	0.015	-0.007	-0.009	0.001	-0.03	0.026	-0.034	0.005	-0.052	0.012	0.1	0.028
VC0515	-0.106	0.006	0.008	0.007	-0.079	0.029	0.035	-0.065	-0.145	0.14	-0.011	0.03
VC0516	-0.018	0	0.006	0.005	0.061	0.066	0.02	0.14	-0.06	-0.046	0.204	0.021

$w_{ij} = -1$ genes i and j never occur in the same species

$w_{ij} = 0$ expected co-occurrence between unrelated genes i and j drawn from a normal distribution

$w_{ij} > 0.045$ suggests shared biological function (Kim and Peterson 2011)

$w_{ij} = 1$ genes i and j always occur in the same species

Partial Correlation Value w_{ij} of VSP-2 Genes i to j (Supplemental File 2) (con't)

	VC0505	VC0506	VC0507	VC0508	VC0509	VC0510	VC0512	VC0513	VC0514	VC0515	VC0516
VC0490	-0.011	-0.006	-0.018	0.053	0.012	0.015	-0.015	0.001	0.015	-0.106	-0.018
VC0491	0.003	0.001	0.009	0.007	-0.001	0.011	-0.006	0.01	-0.007	0.006	0
VC0492	0.013	0.002	0.035	-0.002	-0.023	0.001	0	0.004	-0.009	0.008	0.006
VC0493	0.097	-0.004	0.179	-0.029	0.037	-0.012	0.004	0.03	0.001	0.007	0.005
VC0494	-0.02	-0.026	-0.003	0.028	0.057	0.049	-0.006	-0.013	-0.03	-0.079	0.061
VC0495	0.017	0.024	0.009	0.111	-0.02	0.067	0.028	-0.007	0.026	0.029	0.066
VC0496	0.108	0.045	0.03	0.073	0.216	-0.005	0.041	0.049	-0.034	0.035	0.02
VC0497	0.018	0.018	-0.011	0.06	0.023	0.085	0.037	-0.008	0.005	-0.065	0.14
VC0498	-0.022	-0.005	0.005	0.017	0.058	0.064	0.012	-0.01	-0.052	-0.145	-0.06
VC0502	-0.004	0.01	-0.01	0.123	0.095	0.128	0.154	0.09	0.012	0.14	-0.046
VC0503	-0.007	0.064	-0.003	0.098	-0.022	0.174	0.1	0.045	0.1	-0.011	0.204
VC0504	0.389	0.09	0.369	0.003	0.161	-0.004	-0.002	-0.01	0.028	0.03	0.021
VC0505	1	0.026	0.162	0.005	0.081	-0.003	0.025	0.031	-0.022	-0.025	-0.036
VC0506	0.026	1	0.024	0.008	0.213	0.008	-0.048	0.035	-0.008	-0.009	0.044
VC0507	0.162	0.024	1	0.016	-0.023	0.003	-0.025	0.018	0.032	0.004	-0.008
VC0508	0.005	0.008	0.016	1	0.237	0.047	0.048	0.044	-0.069	0.061	0.111
VC0509	0.081	0.213	-0.023	0.237	1	0.002	-0.086	0.095	0.113	0.042	0.018
VC0510	-0.003	0.008	0.003	0.047	0.002	1	0.153	0.005	0.026	0.081	0.014
VC0512	0.025	-0.048	-0.025	0.048	-0.086	0.153	1	-0.006	0.509	0.112	0.032
VC0513	0.031	0.035	0.018	0.044	0.095	0.005	-0.006	1	0.044	0.034	-0.004
VC0514	-0.022	-0.008	0.032	-0.069	0.113	0.026	0.509	0.044	1	-0.105	0.056
VC0515	-0.025	-0.009	0.004	0.061	0.042	0.081	0.112	0.034	-0.105	1	-0.059
VC0516	-0.036	0.044	-0.008	0.111	0.018	0.081	0.014	0.032	-0.004	-0.004	1

$w_{ij} = -1$ genes i and j never occur in the same species

$w_{ij} = 0$ expected co-occurrence between unrelated genes i and j drawn from a normal distribution

$w_{ij} > 0.045$ suggests shared biological function (Kim and Peterson 2011)

$w_{ij} = 1$ genes i and j always occur in the same species