Title: Phage predation is a biomarker for disease severity and shapes pathogen genetic diversity in cholera patients

Authors: N. Madi[1,2], E. T. Cato[3], Md. Abu Sayeed[3], A. Creasy-Marrazzo[3], A. Cuénod[1,2], K. Islam[4], Md. I.UI. Khabir[4], Md. T. R. Bhuiyan[4], Y. Begum[4], E. Freeman[3], A. Vustepalli[3], L. Brinkley[3], Kamat[4], L. S. Bailey[5], K. B. Basso[5], F. Qadri[4], A. I. Khan[4*], B. J. Shapiro[1,2,6*], E. J. Nelson[3*]

Affiliations:

[1] Department of Microbiology & Immunology, McGill University, Montréal, Canada
[2] McGill Genome Centre, McGill University, Montréal, Canada
[3] Departments of Pediatrics and Environmental and Global Health, University of Florida, Gainesville, FL, USA
[4] Infectious Diseases Division (IDD) & Nutrition and Clinical Services Division (NCSD), International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh
[5] Department of Chemistry, University of Florida, Gainesville, FL, USA
[6] McGill Centre for Microbiome Research, McGill University, Montréal, Canada

[*] Corresponding authors

Email: ashrafk@icddrb.org; jesse.shapiro@mcgill.ca; eric.nelson@ufl.edu

The PDF file includes:

Index
Materials and methods
Results
Figures: S1 to S19
Tables: S1 to S12
References

Keywords: Cholera, diarrhoea, diarrhea, *Vibrio cholerae*, bacteriophage, phage, phage resistance, antibiotics, antimicrobial resistance, AMR, predator, prey, metagenomics, co-evolution, evolution

Enumerations: Abstract: 194 words. Manuscript 4594 words (less figures). Inserts 5. Refs. 58.

Abstract

A century has passed since the discovery that virulent bacteriophages are associated with improved survival of cholera patients. Despite an increasingly detailed picture of the molecular mechanisms that govern phage-bacterial interactions, we lack an understanding of how these interactions impact disease severity. Here we report a year-long, nation-wide study of diarrheal disease patients in Bangladesh. Among the subset of cholera patients, we quantified *Vibrio cholerae* (prey) and its virulent phages (predators) using metagenomics and quantitative PCR, while accounting for antibiotic exposure using quantitative mass spectrometry. Virulent phage (ICP1) and antibiotics suppressed *V. cholerae* to varying degrees and were inversely associated with severe dehydration; aspects of these effects were dependent on resistance mechanisms. In the absence of anti-phage defenses, predation was 'effective' with a high predator to prey ratio that correlated with increased genetic diversity among the prey. In the presence of anti-phage defenses, predation was 'ineffective' with a low predator to prey ratio that correlated with increased genetic diversity among the predators. Our results link phage predation to disease severity, support phage-bacteria coevolution within patients, and suggest the ratio of phage to pathogen can serve as a biomarker for clinical, diagnostic, and epidemiologic applications.

One Sentence Summary

A nation-wide prospective study of cholera patients in Bangladesh identifies effective phage predation and antibiotics as markers of disease severity and selective pressures for cognate resistance mechanisms.

60 **MAIN TEXT**

61 **Introduction**

62 The secretory diarrheal disease cholera is caused by the Gram negative bacterium *Vibrio*
63 *cholerae* (*Vc*) and can progress to life-threatening hypovolemic shock in less than 12 hours (*1*).
64 Cholera remains a major public health problem due to inadequate sanitation and restricted
65 access to safe drinking water. Global estimates of cholera burden are 1.3-4.0 million cases and
66 21,000 to 143,000 deaths annually (*2*). In 2023, there were over 30 countries with active
67 outbreaks necessitating the WHO to escalate the response to its highest level (*3*). Rehydration
68 is the primary life-saving intervention for cholera patients. With effective rehydration, mortality
69 rates fall from over 20% to less than 1%. Antibiotics reduce stool volume and duration of
70 diarrhea but are considered supportive and reserved for patients with more severe disease to
71 reduce selection for antibiotic resistance (*4-8*). Nevertheless, antibiotic-resistant *Vc* have
72 emerged globally (*6, 9, 10*). Mechanisms of resistance are diverse and reside in the core
73 genome, plasmids of the incompatibility type C (*11, 12*) and on mobile genetic elements,
74 including the ~100 kb SXT integrative conjugative element (ICE) that can harbor resistance to
75 sulfamethoxazole and trimethoprim, ciprofloxacin (qnr_{vc}), trimethoprim (*dfra31*) and streptomycin
76 (*aph(6)*) (*13-16*). Recent work has also shown that the ICE can encode diverse phage
77 resistance mechanisms, with distinct hotspots of gene content variation encoding different
78 resistance genes (*17*).

79 With rising levels of antibiotic resistance, virulent bacteriophages are a promising
80 alternative or complementary therapy to antibiotics. Phages and bacteria coevolve, with each
81 partner selecting for adaptations in the other that generates genetic diversity for both predator
82 and prey (*18, 19*). Given the extensive arsenal of resistance and counter-resistance
83 mechanisms among *V. cholerae* and its phages, coevolution between predator and prey occurs
84 (*17, 19-24*). Yet it remains unclear how these interactions impact disease severity during natural
85 infection, with and without antibiotic exposure. Virulent phages specifically targeting *Vc* include

86 ICP1 (*Myoviridae*), ICP2 (*Podoviridae*), and ICP3, which (*Podoviridae*) (*21, 25, 26*); these
87 phages are most common and considered to be found only in symptomatic or asymptomatic
88 patients during acute infection or the convalescent period. The first clinical trials of phage
89 therapy occurred during the Cholera Phage Inquiry from 1927 to 1936 (*27, 28*). In a *proto*
90 randomized controlled trial, the Inquiry found the odds of mortality were reduced by 58% among
91 those with phage therapy, with an absolute reduction in mortality of 10% (95% CI 0.27-0.64)
92 (*29*) (reanalysis by E. J. Nelson). Despite these early findings, there is a lack of evidence in the
93 modern era that links phage predation with disease severity during natural infection in humans.
94 However, indirect studies do support a link: Environmentally, virulent phages in the aquatic
95 environment were negatively correlated with cholera cases in Bangladesh over time, suggesting
96 a role for phages in influencing epidemic dynamics (*30*). Clinically, a higher percentage of
97 cholera patients shed virulent phage towards the end of an outbreak (*31*). Computationally,
98 models predict that phage predation can dampen cholera outbreaks (*32*). Experimentally,
99 animal studies found phage predation was inversely associated with colonization and severity
100 (*25, 33-35*). The key unanswered question is if and how virulent phages, antibiotics, and
101 bacterial evolution interact to impact infection in a meaningful manner for clinicians, public
102 health officials, and most importantly patients.

103 To address this question, we conducted a national prospective longitudinal study in the
104 cholera endemic setting of Bangladesh. Stool samples were collected at hospital admission
105 from diarrheal disease patients and screened for *Vc*, antibiotics, and prevalent virulent cholera
106 phages ICP1, 2 and 3. We hypothesized that: (i) virulent phages and antibiotics would suppress
107 *Vc* and be inversely associated with disease severity, (ii) this suppression would be lifted for *Vc*
108 encoding anti-phage genes on ICEs, (iii) phages would select for other potential resistance
109 mutations in the absence of ICE-encoded resistance, and (iv) phages would be under selection
110 to escape suppression by ICEs. We provide broad support for these hypotheses, paving the
111 way for mechanistic experimental studies, the development of the phage:pathogen ratio as a

112 biomarker of disease severity, and further dissection of the longer-term consequences of phage
113 predation on pathogen evolution.
114
115 **Results and Discussion**
116 ***Study overview.*** A total of 2574 stool samples were collected from enrolled participants
117 admitted for diarrheal disease at seven hospitals across Bangladesh from March to December
118 2018; collection continued to April 2019 at one site (icddr,b). Three groups of cholera samples
119 were analyzed: (i) *Vc* culture-positive (282/2574; 10.9%), (ii) *Vc* culture-negative but phage
120 (ICP1,2,3) PCR-positive (127/2292; 5.5%; 80 included; 47 excluded for DNA < 1ng/μl), and (iii) a
121 random 10% of *Vc* culture-negative and phage PCR-negative samples that were *Vc* PCR-
122 positive (14.8%; 37/250; 27 included; 10 excluded for DNA < 1ng/μl). Stool metagenomes were
123 successfully sequenced from 88.4% (344/389) of samples from these three groups, 35% of
124 which were from the icddr,b site. Detection rates by PCR were low for ICP2 (0.5%) and ICP3
125 (1.6%) in culture-negative samples and were not targeted for sequencing. Based on
126 metagenomic read mapping to a taxonomic database, detection rates for *Vc,* ICP1, ICP2, and
127 ICP3 were 55%, 18%, 1% and 8%, respectively. These ranked detection rates were supported
128 by an analysis of *Vc* phages identified in metagenomic assemblies. As expected, the prophage
129 encoding the cholera toxin (CTXphi) was identified in most assemblies, with ICP1 being the next
130 most prevalent. While some additional putative phages were detected, none were prevalent
131 enough to merit further analysis (**Fig S1**). For both *Vc* and ICP1, relative abundances in
132 metagenomes correlated with absolute quantification by qPCR (**Fig S2**). Five antibiotics (**Table**
133 **S2**) were prioritized for detection in stool using liquid chromatography-mass spectrometry
134 (LCMS); of these, azithromycin, ciprofloxacin and doxycycline were quantified.
135 ***Metagenomic correlates of disease severity and succession.*** At hospital admission, we
136 expected patients to present at different stages of disease, with an ecological succession of *Vc*
137 followed by the facultative anaerobe *Escherichia coli,* then by a healthy flora of mostly anaerobic

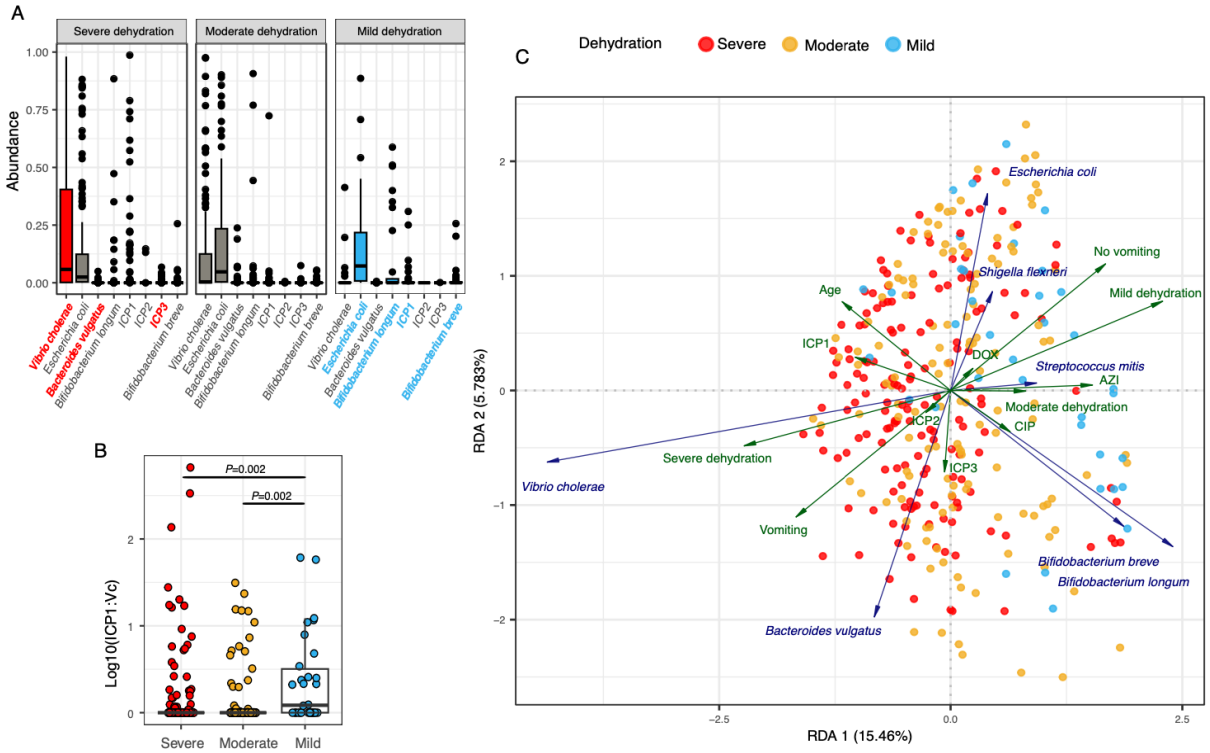
138 bacteria (*36*). We hypothesized that these stages of succession would be associated with
139 different disease severity. As expected, we identified *Vc* as an indicator species of severe
140 dehydration. *Vc* was relatively more abundant in patients with severe dehydration, while two
141 *Bifidobacterium* species, *E. coli*, and ICP1 were indicators of mild dehydration (**Fig. 1A, Table**
142 **S3**). ICP3 was an indicator of severe dehydration despite being less frequently detected in our
143 study (28 samples with >0.1% ICP3 reads, compared to 61 samples with >0.1% ICP1). This
144 suggests that different phages can have contrasting disease associations. Given the robust
145 sample size for ICP1, we focus on ICP1 for subsequent analyses.

146 The distribution of ICP1 relative abundance was variable and less clearly associated
147 with dehydration status than *Vc* (**Fig. S3**). Deeper investigation revealed that it was not simply
148 the presence of phage that mattered. Rather, the ratio of ICP1 to *Vc* was the critical
149 determinant. Higher ratios were inversely associated with dehydration severity (**Fig. 1B**); the
150 same results were obtained using qPCR instead of metagenomics to quantify the ratio (**Fig S4**).
151 This simple ratio is a marker of 'effective' phage predation that could be used in clinical,
152 diagnostic and epidemiological applications.

153 These observations suggest that ICP1 plays a role in modulating disease severity, or
154 could be associated with later stages of disease, when patients are recovering. Consistent with
155 the former possibility, ICP1 was previously associated with early, rather than late stages of
156 disease, peaking on the first day of infection in a time series of cholera patients (*36*). Given that
157 we collected only one sample per patient (at hospital admission), we were limited in
158 disentangling these possibilities. Nevertheless, we recorded self-reported duration of diarrhea,
159 providing an approximate proxy for disease progression. We found that higher relative
160 abundances of ICP1 were associated with mild dehydration at early stages of disease (duration
161 of diarrhea <72h) but not at late stages (**Fig. S3B and D**). We therefore put forth a model in
162 which ICP1 suppresses *Vc* at early disease stages in a way that reduces disease severity by

163 suppressing *Vc*, or by other indirect mechanisms. In either case, ICP1 serves as a meaningful

164 biomarker of disease severity.

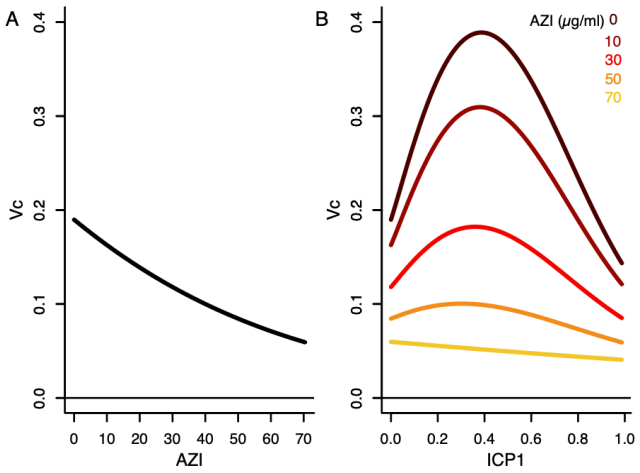


165 **Fig. 1. Dehydration severity is inversely associated with higher ICP1:*Vc* ratios in stool**
166 **metagenomes**. (A) Relative abundances of the most prevalent species in patients with severe, moderate,
167 or mild dehydration; these conventions equate to the World Health Organization (WHO) conventions of
168 severe, some and no dehydration, respectively. Significant indicator species for severe or mild
169 dehydration are shown in red or blue bold text, respectively (P<0.05 in a permutation test with 9999
170 iterations as implemented in the indicator species function in R). See Table S3 for full indicator species
171 results. (B) The ICP1:*V. cholerae* ratio is higher in patients with mild dehydration. P-values are from a
172 Kruskal-Wallis test with Dunn's post-hoc test, adjusted for multiple tests using the Benjamini-Hochberg
173 (BH) method. Only significant P-values (<0.05) are shown. Only 323 samples with *V. cholerae*>0% of
174 metagenomic reads were included, with 165 from severe, 128 from moderate, and 30 from mild cases. A
175 pseudocount of one was added to the ratio before log transformation. For supporting analyses using
176 qPCR data, see Fig. S4. In (A) and (B) the solid horizontal line is the median and the boxed area is the
177 interquartile range. (C) Redundancy analysis (RDA) showing relationships among the seven most
178 dominant bacterial species and explanatory variables: phages, patient metadata, and antibiotic
179 concentration. Angles between variables (arrows) reflect their correlations; arrow length is proportional to
180 effect size. Samples (points) are colored by dehydration severity. All displayed variables are significant
181 (P<0.05, permutation test) except ICP2, ICP3, and doxycycline (Table S4). For the RDA: R^2=0.25 and
182 adjusted R^2=0.184, permutation test P = 0.001. To improve readability, collection date and location are
183 not shown (see Fig. S6 for these details).
184
185 ***Antibiotic exposure is inversely associated with disease severity.*** To visualize the complex

186 relationships between disease severity, bacteria, and phage in the context of antibiotic

187 exposure, we used redundancy analysis (RDA; **Fig. 1C; Table S4**). For simplicity, the seven

188 most dominant bacterial species identified by principal components analysis were included (**Fig.**
189 **S5**). As explanatory variables, we visualized those clinical data with strong effects, chosen by
190 forward selection beginning with phages and antibiotic concentrations (**Fig. S6**). In accordance
191 with the indicator species analysis (**Fig. 1A, Table S3**), higher *Vc* relative abundance was
192 positively correlated with severe dehydration (**Fig. 1C**). ICP1 was moderately associated with
193 *Vc*, consistent with a phage's reliance on its host for replication, but less associated with severe
194 dehydration. The antibiotics azithromycin (AZI) and, to a lesser extent ciprofloxacin (CIP), were
195 negatively correlated with *Vc* and severe dehydration, suggesting their role in suppressing
196 cholera infection and disease. Supporting previous reports that AZI suppresses *Vc* (*37*), AZI
197 was most strongly anticorrelated with *Vc* in our cohort (**Fig. 1C**). We did not identify annotated
198 antibiotic resistance genes associated with AZI exposure (**Fig. S7**) at established thresholds
199 (**Table S5**), suggesting that AZI resistance is relatively rare in our cohort. This does not exclude
200 the possibility that resistance to AZI is controlled by non-canonical resistance genes absent
201 from standard databases. In contrast, CIP exposure was significantly associated with the
202 resistance genes *dfrA* and *aph6* (**Fig. S8),** which are both associated with *Vc* in our
203 metagenomes (**Fig S9**) and have previously been linked with CIP resistance in *Vc (16, 38)*.
204 Taken together, these results suggest CIP exposure selects for resistance genes within
205 patients, possibly explaining why CIP may be less effective at suppressing *Vc* than AZI.



206 **Fig. 2. Interactions between *V. cholerae*, phage ICP1, and azithromycin**. GAM results, fit with *Vc* as a
207 function of antibiotic concentrations and ICP1 relative abundance in 344 metagenomes. (A) *V. cholerae*
208 declines in relative abundance with higher relative abundance of azithromycin (AZI) in μg/ml. (B) The
209 relationship between ICP1 and *V. cholerae* is affected by AZI concentration (μg/ml); the illustrated AZI
210 concentrations show regular intervals between the minimum (0 μg/ml) and maximum (78 μg/ml) observed
211 values. Both effects of AZI (A) and the ICP1-AZI interaction (B) are significant (Chi-square test, $P<0.05$).
212 For details on GAM outputs see Table S6. For equivalent analyses using qPCR data, see Fig. S10.
213

214 ***Antibiotic exposure suppresses predator-prey dynamics.*** We next asked if and how
215 antibiotics interact with phage to suppress *Vc* within patients. To do so, we modeled the
216 relationships between ICP1, *Vc* and antibiotic exposure within each patient. We fit generalized
217 additive models (GAMs) of *Vc* (relative abundance from metagenomics or absolute abundance
218 from qPCR) as a function of ICP1, antibiotic exposure, and their interaction, including
219 dehydration status as a random effect. We fit GAMs with all antibiotics and their interaction with
220 ICP1, as well as separate models with each antibiotic, alone or in combination, and compared
221 them based on their Akaike Information Criterion (AIC; **Tables S6, S7**). The most parsimonious
222 models (with the lowest AIC) using either metagenomics or qPCR data showed a significant
223 negative relationship between *Vc* and AZI (**Fig. 2A, S10**), consistent with the RDA results (**Fig.
224 1C**) and with known patterns of *Vc* suppression by AZI during infection (*37*). The relationship
225 between *Vc* and ICP1 was quadratic in both metagenomics- and qPCR-based models: at low
226 ICP1 abundance, the relationship was positive but became negative at higher ICP1 abundance
227 (**Fig. 2B, S10**). This alternation between positive and negative correlations is consistent with
228 predator-prey dynamics within infected patients (*39*). However, at higher concentrations of AZI,
229 the quadratic relationship flattened, effectively suppressing the phage-bacteria interaction. In the
230 future, these interactions could be interrogated using patient time-series data and laboratory
231 experiments challenging *Vc* with antibiotics and phages.

232 ***Integrative and conjugative elements (ICEs) are associated with phage suppression.***
233 SXTs are approximately 100kb integrative and conjugative elements (ICEs) that are associated
234 with resistance to both antibiotics and phages and were first discovered in *Vc* (*13*). SXTs have

235 conserved ‘core’ genes along with variable ‘hotspots’ encoding different genes; for example,
236 hotspot 5 is a ~17kb region associated with phage resistance. At the time of our sampling,
237 ICEVchInd5 (abbreviated here as *ind5*) and ICEVchInd6 (*ind6*) were the two most prevalent ICE
238 types in Bangladesh (*17*). These ICEs differ in their anti-phage systems: *ind5* encodes a type 1
239 bacteriophage exclusion (BREX) system while *ind6* encodes several other different restriction-
240 modification systems (*17*).

241 We enumerated ICEs in metagenomes by mapping reads against reference sequences
242 for *ind5* (NCBI accession GQ463142.1) and *ind6* (accession MK165649.1). An ICE was defined
243 as present when 90% of its length was covered by at least one metagenomic read (**Fig. S11A**).
244 We found that 59% (144/244) of samples with *Vc*>0.5% or phages >0.1% of metagenomic
245 reads contained *ind5*, 10.6% (26/244) contained *ind6*, and 22.1% (54/244) had no detectable
246 ICE. The lack of ICE detections was not due to the lack of *Vc* in a metagenome because ICE-
247 negative samples did not contain fewer *Vc* reads (**Fig. S11B**).

248 Resistance mechanisms on ICEs have been shown to suppress phage *in vitro (17)*, but
249 their relevance within human infection remains unclear. We found that metagenomes without a
250 detectable ICE (denoted as ICE-) were associated with higher phage to *Vc* ratios (**Fig. 3**). The
251 effect was strongest for ICP1, which had the largest sample size (**Fig. 3A**). This observation is
252 consistent with ICE-encoded mechanisms suppressing phage within patients. Higher ICP1:*Vc*
253 ratios, which occurred more often in ICE- patients, were also associated with mild dehydration
254 (**Fig. 3B**). The phage:*Vc* ratios varied according to the precise phage-ICE combination,
255 suggesting that different ICE variants might confer resistance to different phages (**Fig. 3, Fig.
256 S12**). The same associations were evident using qPCR-based quantification of phage:bacteria
257 ratios (**Fig S13**). Together, these results implicate ICEs in suppressing phages during human
258 infections, complementing and generally confirming the predictions of laboratory experiments
259 (*17*). That said, the suppression is not complete, and further experiments are needed to dissect

the underlying causal relationships.

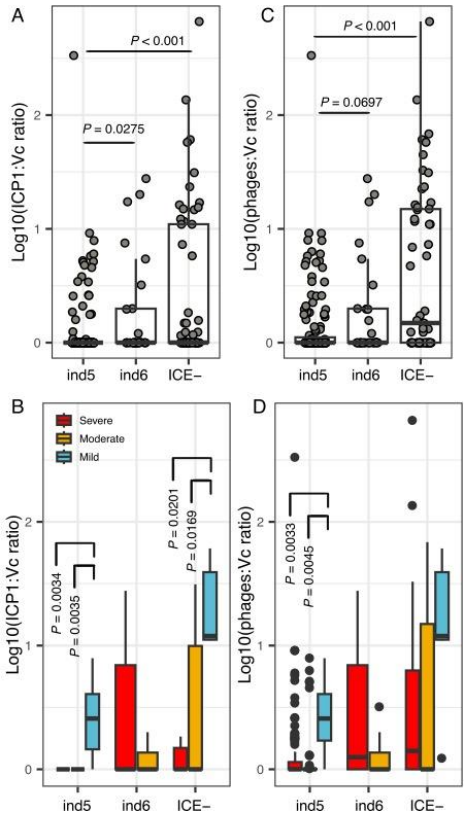


Fig. 3. Integrative conjugative elements (ICEs) are associated with lower ICP1:*V. cholerae* ratios in patient metagenomes. (A) Distribution of ICP1:*Vc* ratios across patients with different ICE profiles. (B) The same data as (A) binned into boxplots according to dehydration status. (C) Distribution of phage:*Vc* ratios, including the sum of all phages (ICP1, ICP2, ICP3). (D) The same data as (C) binned into boxplots according to dehydration status. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test adjusted with the Benjamini-Hochberg (BH) method. Only *P*-values < 0.1 are shown. Only samples with appreciable *Vc* or ICP1 were included (224 samples with *Vc*>0.5% or phages >0.1% of metagenomic reads), of which 54 samples were ICE-, 26 were ind6+ and 144 were ind5+. The Y-axes were log10 transformed after adding one to the ratios. The solid horizontal line is the median and the boxed area is the interquartile range. For supporting analyses using qPCR data, see Fig. S13.

Hypermutation generates V. cholerae genetic diversity. In addition to variation in gene content in ICEs and other mobile elements, we hypothesized that resistance to phages and antibiotics may have been conferred by point mutations (single nucleotide variants; SNVs) that existed before or emerged *de novo* during infection. Although we cannot fully exclude mixed infections by different *Vc* strains as a source of within-patient diversity, we found no evidence for more than one strain co-infecting a patient in our study population (**Fig. S14**). We previously

279 found a low level of *Vc* genetic diversity within individual cholera patients (*40*) – on the order of
280 0-3 detectable SNVs – with the exception of hypermutation events characterized by DNA repair
281 defects and hundreds of SNVs in the *Vc* genome, primarily transversion mutations (*41*).
282 Hypermutation generates deleterious mutations, but may also rapidly confer phage resistance,
283 as shown experimentally with *Pseudomonas fluorescens (18)*. Here, we were able to better
284 quantify the frequency of hypermutators because of sufficient sample size, and test if within-host
285 *Vc* diversity is associated with phage or antibiotic exposure – both of which could potentially
286 select for resistance mutations to each factor.

287 To identify hypermutators in metagenomes, we identified *Vc* populations with defects
288 (nonsynonymous mutations) in any of 65 previously defined DNA repair genes (*42*) or with a
289 relatively high number of SNVs (25 or more) (*41*). We used InStrain (*43*) to quantify *Vc* within-
290 host diversity in 133 samples passing stringent quality filters (Methods) and found that 35% of
291 samples (47/133) had both a high SNV count and nonsynonymous mutations in DNA repair
292 genes – making them likely to contain hypermutators. Higher SNV counts were significantly
293 associated with DNA repair defects (Fisher's exact test, $P<2.2e-16$), consistent with their
294 resulting in higher mutation rates within patients. The number of SNVs was not confounded by
295 *Vc* genome coverage (**Fig. S15A**). Consistent with our previous study (*41*), putative
296 hypermutators had a distinct mutational profile, enriched in transversions (**Fig. S15B,C**). This
297 suggests that hypermutators are common, occurring in over one third of cholera patients. We
298 also observed that samples with high numbers of SNVs but no DNA repair defects had higher
299 levels of phage ICP1 (**Fig. S15D**), suggesting that phage may select for resistance mutations in
300 hypermutators and non-mutator populations alike. For subsequent analysis, we considered all
301 SNVs together, regardless of whether they were generated by hypermutation.

302 ***Phages, not antibiotics, are associated with Vc within-host diversity.*** We asked if *Vc*
303 within-host diversity was dependent on phage or antibiotic exposure – both strong selective

304 pressures. To test this hypothesis, we fit GLMMs with phages and antibiotics as predictors of
305 the number of high-frequency nonsynonymous (NS) SNVs in the *Vc* population within a patient.
306 We focused on higher-frequency SNVs (>10% within a sample) as likely beneficial mutations
307 (unlikely to rise to such high frequency by chance if neutral or deleterious), and on NS SNVs as
308 these are more likely to have fitness effects. We fit several models with different combinations of
309 predictors: from a model with all antibiotics and their interaction with ICP1 to separate models
310 with each antibiotic and its interaction with ICP1. We added *Vc* abundance as a fixed effect to
311 the model to control for any coverage bias in SNV calling (**Table S8**). The most parsimonious
312 model included *Vc* abundance and the interaction between *Vc* and ICP1 as predictors of the
313 number of high-frequency NS SNVs. Adding antibiotics, or their interaction with ICP1, did not
314 improve the model (**Table S8**), suggesting a limited role for antibiotics in selecting for *Vc* point
315 mutations within patients.

316 In the model, *Vc* relative abundance and the interaction between *Vc* and ICP1 both had
317 significant effects (GLMM, Wald test, P=0.00246 and P=0.00494 respectively). The negative
318 relationship between *Vc* and the number of high-frequency NS SNVs (**Fig. 4A**) was not easily
319 explained by sequencing coverage, since the total number of SNVs is not associated with *Vc*
320 relative abundance (**Fig. S15A**). The number of high-frequency NS SNVs rose with increasing
321 ICP1 – but only when *Vc* abundance was relatively high (**Fig. 4A**). As a control, we ran the
322 same GLMM on NS SNVs without a minimum frequency cutoff and found no significant effects,
323 suggesting that the interaction between ICP1 and *Vc* on SNV count is specific to high-frequency
324 mutations, rather than low-frequency mutations that are more likely neutral or deleterious.
325 These data support a scenario in which ICP1 selects for NS SNVs when the *Vc* population is
326 large enough to respond efficiently to selection – for example, at the beginning of an infection or
327 in the absence of antibiotics.

328 If phages select for beneficial mutations, we expect these mutations to increase in
329 frequency at higher phage predation. We lack time-series data from individual patients, but the

330 relative abundance of phage provides a proxy for the combined effects of the strength and
331 duration of phage selection. To test this hypothesis, we fit a GAM with the average within-
332 sample frequency of SNVs as a function of ICP1, antibiotics, and their interactions. We included
333 the fixed effect of ICE presence/absence as another factor that could provide phage or antibiotic
334 resistance, as well as mutation type to differentiate among non-synonymous (NS), synonymous
335 (S), and intergenic (I) mutations. We fit GAMs with all antibiotics and their interaction with ICP1,
336 as well as models with each antibiotic separately (**Table S9**). The most parsimonious model
337 included the interaction between ICP1, ICE and mutation type, but not antibiotics. ICP1 was a
338 strong predictor of higher frequency NS SNVs in the absence of a detectable ICE (**Fig. 4B**).
339 Samples in this analysis were unambiguous in terms of their ICE presence/absence patterns
340 (**Fig. S16**).

341 To confirm this model prediction, we compared the distribution of the average frequency
342 of NS SNVs between ICP1-positive and ICP1-negative samples. Consistent with the model, NS
343 SNV frequency was significantly higher in ICP1-positive samples when the ICE was not
344 detected (Wilcoxon test, P=0.0094) making them the likely targets of selection by ICP1
345 predation (**Fig. 4C**). Qualitatively similar results were found when ICP1 was quantified by qPCR
346 instead of metagenomics (**Fig S17**). Together, these results suggest that, in the absence of
347 detectable ICE-encoded phage resistance, ICP1 selects for nonsynonymous point mutations
348 (**Fig 5A**). We identified several *Vc* genes, including some with membrane or virulence-related
349 functions, that were mutated at higher levels ICP1:*Vc* ratios (**Table S10**); these provide
350 candidate phage resistance mechanisms that can be explored in future experiments. In
351 contrast, the secreted hemolysin gene, *hlyA,* that we previously observed (*41*) to be mutated
352 more often by chance in cholera patients was among the genes most frequently mutated in
353 patients with relatively low ICP1:*Vc* ratios (**Table S11**). This suggests that *hlyA* sequence
354 evolution may be affected directly or indirectly by phage predation.

355 As *Vc* evolves as a function of ICP1, we expect ICP1 evolution to be impacted by *Vc*
356 (**Fig 5A**). Despite the generally low genetic diversity of ICP1 (*21)*, were were able to quantify
357 SNVs in the ICP1 genome in 45 samples. This sample size was too low to fit sophisticated
358 models, but simple correlations allowed us to draw tentative conclusions. First, we ruled out
359 sequencing coverage as source of bias in SNV calling (Spearman correlation between ICP1
360 relative abundance and number of SNVs, $p > 0.1$ for all SNV categories). Next, we observed a
361 negative correlation between the ICP1:*Vc* ratio and the number of NS SNVs in the ICP1
362 genome – a correlation that is only significant when *Vc* encodes an *ind5* ICE (**Fig 5B, S18**). This
363 is consistent with our observation that *ind5* is associated with lower ICP1:*Vc* ratios in our cohort
364 (**Fig. 3**), potentially suppressing ICP1 and applying selection to escape suppression via NS
365 mutations. Some of these NS mutations may be beneficial to the phage, rising to high frequency
366 along with ICP1 – which is indeed what we observe: the mean frequency of NS SNVs in ICP1
367 increases with the ICP1:*Vc* ratio, only in the presence of *ind5* (**Fig 5C, S19**). Several ICP1
368 genes, mostly hypothetical proteins, repeatedly acquired NS mutations in the presence of *ind5*,
369 providing candidate escape mutations to test in future work (**Table S12**).

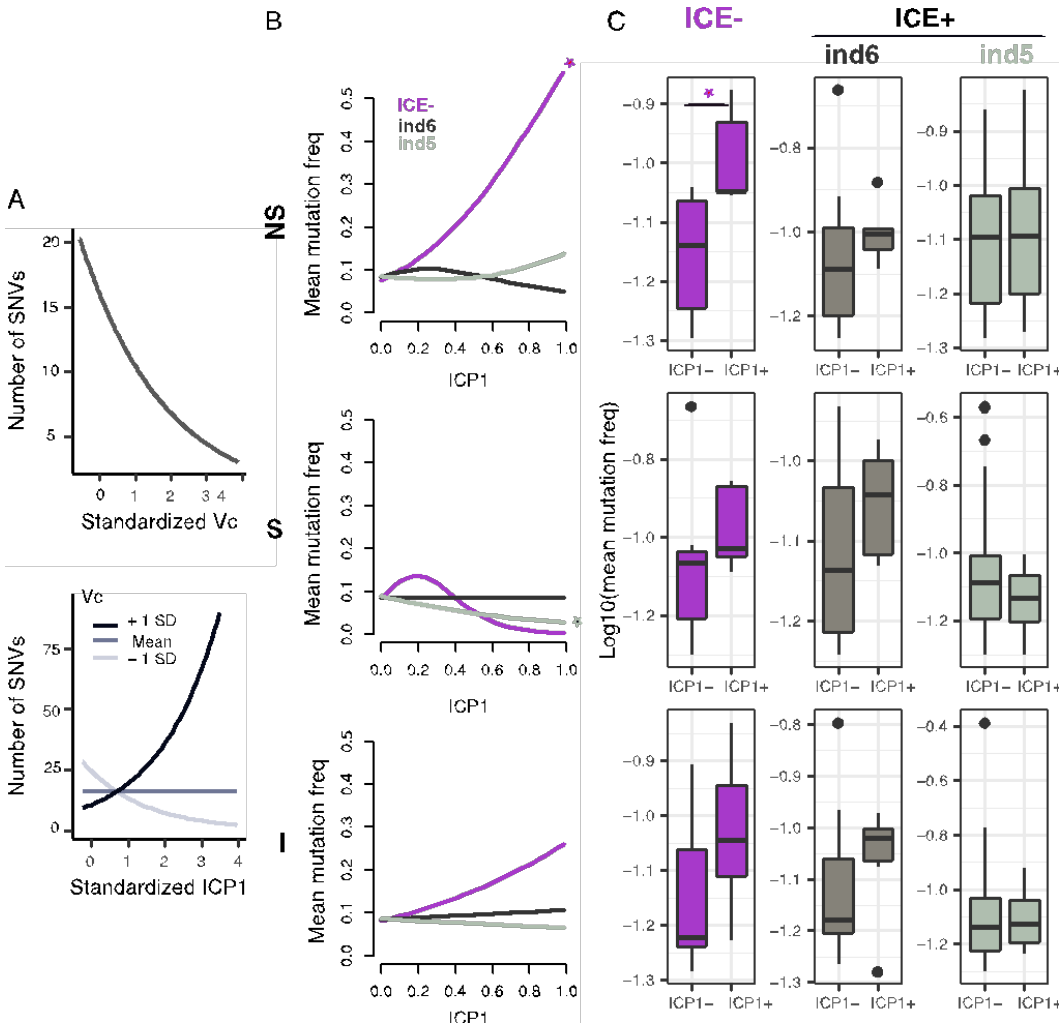
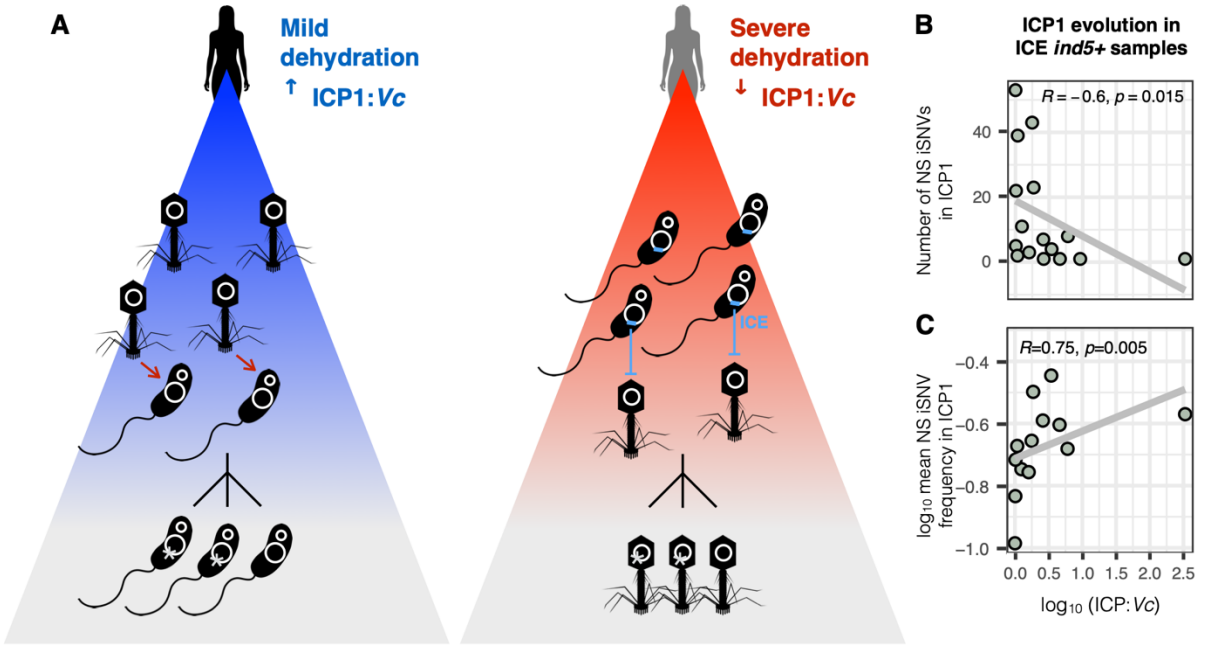


Fig. 4. ICP1 selects for non-synonymous point mutations in the *V. cholerae* genome in the absence of ICE. (A) Results of a GLMM modeling high frequency nonsynonymous SNV counts as a function of *V. cholerae* (*Vc*) and ICP1 standardized relative abundances. In the bottom panel, shades of gray indicate *Vc* relative abundance at the mean or +/- 1 standard deviation (SD) across samples. (B) GAM output with the mean mutation frequency as a function of the interaction between ICP1, ICE and mutation type (non-synonymous; NS, synonymous; S, or intergenic; I). Significant effects are shown with a star (Chi-square test, $P<0.05$). The model was fit using 130 samples that passed the post-InStrain filter (Methods). (C) Boxplots of mutation frequency in the presence or absence of ICP1 and/or ICEs. The single significant comparison is indicated with a star (Wilcoxon test, $P=0.0094$). Boxplots include 130 samples, of which 32 are ICP1+ (ICP1>=0.1% of reads) and 98 are ICP- (ICP1<0.1% of reads). The solid horizontal line is the median and the boxed area is the interquartile range. For supporting analysis using qPCR data, see Fig. S17.

The tripartite interactions between pathogens, phage, and antibiotics have been studied in the laboratory, *in silico* with mathematical models, and to a lesser extent in the field, but how these factors interact during human infection remains an open question. Our objective was to characterize these interactions in the context of cholera. We analyzed more than 300 stool

388 metagenomes from cholera patients enrolled at hospital admission across Bangladesh during
389 an entire outbreak season. We found that high predator (ICP1) to prey (*Vc*) ratios were
390 associated with less severe dehydration. We also demonstrate how *Vc* and ICP1 interact within
391 patients, with ICP1 selecting for potential phage resistance point mutations in the absence of
392 ICE-based anti-phage defenses, and ICE-encoding *Vc* selecting for point mutations in the
393 phage genome (**Figure 5**). This apparent coevolution between predator and prey likely has
394 longer-term consequences for cholera disease severity and transmission. Antibiotics,
395 particularly azithromycin, also played a role in suppressing *Vc* and could mask phage-bacteria
396 interactions. Ciprofloxacin was associated with known antibiotic resistance genes, but we found
397 no evidence that antibiotics select, as ICP1 does, for high-frequency nonsynonymous point
398 mutations. Thus, although resistance mechanisms to certain phages and antibiotics colocalize
399 to the ICE (*17),* they impose distinct selective pressures that could be exploited to improve the
400 efficacy of antibiotics by combining them with phage therapy.
401



402 **Fig. 5. Phage-bacteria coevolution and disease severity in cholera patients. (A)** The illustration
403 summarizes the main conclusions supported by our analyses. Patients with mild dehydration (left) have
404 higher ratios of phage ICP1 to pathogen *Vibrio cholerae*, while the opposite is true in patients with severe
405 dehydration (right). Although the causal relationship between the ICP1:*Vc* ratio and dehydration remains
406 unclear, it provides a potential biomarker for disease severity. At high ICP1:*Vc* ratios, and in the absence
407 of anti-phage genes encoded on the ICE, a higher frequency of nonsynonymous point mutations
408 (asterisks on white circular genomes) are selected for in the *Vc* genome, with potential evolutionary
409 consequences: some mutations may provide phage resistance while others may simply provide greater
410 genetic diversity upon which future natural selection can act. When ICE (primarily the *ind5* variant) is
411 encoded by *Vc*, its anti-phage properties impose strong selection on ICP1, selecting for nonsynonymous
412 point mutations in the ICP1 which may eventually rise to high frequency. For simplicity, the effects of
413 antibiotics are not shown, but antibiotics such as azithromycin can suppress *Vc*, thereby masking
414 interactions with ICP1. **(B)** The number of nonsynonymous (NS) SNVs detected in the ICP1 genome is
415 negatively correlated with the ICP1:*Vc* ratio in the presence of *ind5.* The Spearman correlation coefficient
416 and *p*-value are shown. **(C)** The mean frequency of NS SNVs in the ICP1 genome is positively correlated
417 with the ICP1:*Vc* ratio in the presence of *ind5.* See Figures S18 and S19 for equivalent plots in ICE- and
418 *ind6* samples, and for synonymous and intergenic SNVs.
419

420 Our study has several limitations. First, samples were collected at a single time point
421 (upon hospital admission) which allowed us to establish statistical correlations, but we cannot
422 infer causality in the absence of time-series or interventional clinical studies. Second, our
423 sample size allowed us to study the interaction between *Vc* and ICP1, but the sample size for
424 ICP2 and ICP3 was insufficient for similar statistical analyses. Third, we prioritized common
425 antibiotics for mass spectrometry, and we cannot exclude a role for other unmeasured
426 antibiotics. Fourth, due to logistical limitations, we extracted DNA from a bacterial pellet plus a
427 small amount of supernatant from each sample. This allowed us to capture both intracellular
428 and cell-bound phages, along with free phage particles, but sequencing each fraction separately
429 could yield further insights into distinct phage populations. From a clinical perspective, the
430 measurement of dehydration status was categorical and could be improved in future studies by
431 incorporating digital tools to quantify the degree of dehydration (*44-46*). Finally, our study lacked
432 information about host genetics or immunity, which also influence disease severity (*8, 47*).
433 Future studies combining rich patient metadata, a time-series design, long-read metagenomics,
434 and isolate genome sequencing will complement and expand upon these findings.
435

436 **CONCLUSION**

437 We propose that an index of effective phage predation, quantified as the phage:bacteria
438 ratio, might be used as a tool for physicians to assess disease severity, and potentially
439 prognosticate a disease course. While we show here that this ratio is associated with cholera
440 disease severity, its predictive value should be studied in larger cohorts sampled over the
441 course of infection. Whether this biomarker can be generalized to phages other than ICP1 and
442 diseases other than cholera, and whether the association with disease severity changes as
443 predator and prey coevolve over time, remains unclear. The potential of phage therapy and
444 prophylaxis has long been recognized, and a combination of ICP1, 2, and 3 prevents cholera in
445 animal models (*35*). However, just as hypermutators can drive the evolution of resistance to
446 combinations of antibiotics (*48*), they may also help pathogen populations survive combinations
447 of phages, while increasing their potential to evolve resistance to future antimicrobial
448 treatments. Phages therapy cocktails will therefore need to be updated regularly to remain
449 effective against currently circulating coevolved bacteria, and creative new strategies are
450 needed to minimize the unwanted evolution of resistance.

451

452 **Materials and Methods**

453 **Ethics Statement**. The samples analyzed were collected within two previously published IRB
454 approved clinical studies in Bangladesh: (i) The mHealth Diarrhea Management (mHDM) cluster
455 randomized controlled trial (IEDCR IRB/2017/10; icddr,b ERC/RRC PR-17036; University of
456 Florida IRB 201601762) (*46*). (ii) The National Cholera Surveillance (NCS) study (icddr,b
457 ERC/RRC PR-15127) (*49*); See supplementary materials for further details.

458 **Study Design.** The study design was a prospective longitudinal study of patients presenting
459 with diarrheal disease at five Bangladesh Ministry of Health and Family Welfare district hospitals
460 (both mHDM and NCS sites) and two centralized NCS hospitals (BITID; icddr,b) from March

461 2018 to December 2018. Sites were distributed geographically nation-wide (*50*). See
462 supplementary materials.
463 **Sample collection.** Stool samples were collected at hospital admission. Aliquots for transport
464 and subsequent culture were stabbed into Cary-Blair transport media; aliquots for molecular
465 analysis were preserved in RNAlater (Invitrogen). See supplementary materials.
466 **Microbiological and molecular analysis.** Culture was performed via standard methods (*51*);
467 total nucleic acid (tNA) was extracted from the RNAlater preserved samples using standard
468 methods. Criteria for subset subsequent shotgun metagenomic sequencing were: (i) culture
469 positivity, (ii) Phage (ICP1,2,3) detection by PCR among culture-negative samples, or (iii) *Vc*
470 detection by PCR among a random 10% of culture-negative and phage (ICP1,2,3) negative
471 samples. Sequencing libraries were prepared using the NEB Ultra II shotgun kit and sequenced
472 on illumina NovaSeq 6000 S4, pooling 96 samples per lane, yielding a mean of >30 million
473 paired-end 150bp reads per sample. Among samples identified for metagenomic analysis,
474 qPCR was performed to determine absolute abundances of *Vc*, ICP1, ICP2, and ICP3 using
475 standard methods.
476 **Antibiotic detection by liquid chromatography mass spectrometry (LC-MS/MS).** Those
477 cholera samples identified for metagenomic analysis were analyzed by qualitative and
478 quantitative LC-MS/MS for antibiotics. Based on prior research (*16, 37*), the target list prioritized
479 5 common antibiotics: ciprofloxacin, doxycycline/tetracycline, and azithromycin were analyzed
480 quantitatively, and metronidazole and nalidixic acid were analyzed qualitatively. Standard
481 curves were made for each quantitative target by preparing a dilution series of the three native
482 and isotopic forms of the quantitative targets; clinical samples were spiked with the isotopes of
483 the quantitative targets as internal references. See supplementary materials.
484 **Metagenomic data analysis.** We taxonomically classified short reads using Kraken2 (*52*) and
485 Bracken v.2.5 (*53*). Reads were assembled using MEGAHIT v.1.2.9 (*50*) and binned with DAS

tool (*54*). We inferred probably phage contigs using geNomad v1.7.0 (*55*) and predicted their likely bacterial hosts with iPhoP v1.3.1 (*56*). To characterize intra-patient *Vc* diversity, we used StrainGE (*57*) and InStrain v.1.5.7 (*43*). To identify antibiotic resistance genes in metagenomes, we used deepARG v 1.0.2 (*58*). See supplementary materials for details.

Statistical analyses. Statistics and visualizations were done in R studio version 1.2.5042. See supplementary materials for details.

Data availability. All sequencing data are deposited in the NCBI SRA under BioProject PRJNA976726. See supplementary materials for further information.

Code availability. Computer code for this paper will be available at https://github.com/Naima16/Cholera-phage-antibiotics.

Acknowledgements: We thank the patients for participating in this study as well as the clinical and laboratory teams who collected the samples. We are grateful to S. Flora and colleagues at the Institute of Epidemiology, Disease Control and Research (IEDCR), Ministry of Health and Family Welfare, Government of Bangladesh who collaborated on the original clinical studies in which the samples analyzed herein were collected. We are also grateful to R. Autrey, B. Johnson and K. Berquist for their administrative expertise at the University of Florida. AIK and FQ were the principal investigators in Bangladesh and PIs of the ERC/RRC approvals at the icddr,b. This collective research infrastructure and support was invaluable to the success of this study. We thank members of the Nelson, Khan, Qadri and Shapiro labs, and Frederique Le Roux, for discussions that improved the manuscript.

Funding: This work was supported by the National Institutes of Health grants to EJN [R21TW010182] and KBB [S10 OD021758-01A1] and internal support from the Emerging Pathogens Institute at the University of Florida and the Departments of Pediatrics/ Children's Miracle Network (Florida). BJS and NM were supported by a Canadian Institutes for Health Research Project Grant. AC was supported by a Postdoctoral Mobility Fellowship of the Swiss National Science Foundation [P500PB_214356]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions:

Conceptualization: FW, AIK, BJS, EJN
Methodology: NM, MAS, ETC, MTRB, YB, ACM, MK, AC, EF, AV, LSB, KBS, BJS, EJN
Investigation: NM, MAS, ETC, KI, MIUK, YB, ACM, MK, AC, EF, AV, LSB, BJS, EJN
Visualization: NM, BJS, EJN
Funding acquisition: KBS, FW, AIK, BJS, EJN
Project administration: MTRB, YB, KBS, AIK, BJS, EJN
Supervision: KBS, FQ, AIK, BJS, EJN
Writing – original draft: NM, ETC, MAS, ACM, MK, LSB, KSB, AIK, BJS, EJN
Writing – review & editing: NM, AIK, AC, BJS, EJN

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary materials:

Materials and methods

Figures: S1 to S19

Tables: S1 to S12

References

Data Files S1 to S5

536 **References**

537 1. J. R. Andrews *et al.*, Determinants of severe dehydration from diarrheal disease at
538 hospital presentation: Evidence from 22 years of admissions in Bangladesh. *PLoS Negl*
539 *Trop Dis* **11**, e0005512 (2017).
540 2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, Updated global burden of cholera in
541 endemic countries. *PLoS Negl Trop Dis* **9**, e0003832 (2015).
542 3. Cholera – Global situation (2023). WHO Report. (available at
543 https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON437).
544 4. M. S. Islam *et al.*, Microbiological investigation of diarrhoea epidemics among Rwandan
545 refugees in Zaire. *Trans R Soc Trop Med Hyg* **89**, 506 (1995).
546 5. J. A. Dromigny, O. Rakoto-Alson, D. Rajaonatahina, R. Migliani, J. Ranjalahy, P.
547 Mauclere, Emergence and rapid spread of tetracycline-resistant *Vibrio cholerae* strains,
548 Madagascar. *Emerg Infect Dis* **8**, 336-338 (2002).
549 6. F. X. Weill *et al.*, Genomic history of the seventh pandemic of cholera in Africa. *Science*
550 **358**, 785-789 (2017).
551 7. E. J. Nelson, D. S. Nelson, M. A. Salam, D. A. Sack, Antibiotics for both moderate and
552 severe cholera. *N Engl J Med* **364**, 5-7 (2011).
553 8. E. J. Nelson, J. B. Harris, J. G. Morris, Jr., S. B. Calderwood, A. Camilli, Cholera
554 transmission: the host, pathogen and bacteriophage dynamic. *Nat Rev Microbiol* **7**, 693-
555 702 (2009).
556 9. B. Das, J. Verma, P. Kumar, A. Ghosh, T. Ramamurthy, Antibiotic resistance in *Vibrio*
557 *cholerae*: Understanding the ecology of resistance genes and mechanisms. *Vaccine* **38**
558 **Suppl 1**, A83-A92 (2020).
559 10. F. Lassalle *et al.*, Genomic epidemiology reveals multidrug resistant plasmid spread
560 between *Vibrio cholerae* lineages in Yemen. *Nat Microbiol* **8**, 1787-1798 (2023).

561 11. N. Rivard, R. R. Colwell, V. Burrus, Antibiotic Resistance in *Vibrio cholerae*: Mechanistic
562 Insights from IncC Plasmid-Mediated Dissemination of a Novel Family of Genomic
563 Islands Inserted at trmE. *mSphere* **5**, (2020).
564 12. S. J. Ambrose, C. J. Harmer, R. M. Hall, Compatibility and entry exclusion of IncA and
565 IncC plasmids revisited: IncA and IncC plasmids are compatible. *Plasmid* **96-97**, 7-12
566 (2018).
567 13. M. K. Waldor, H. Tschape, J. J. Mekalanos, A new type of conjugative transposon
568 encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio*
569 *cholerae* O139. *J Bacteriol* **178**, 4157-4165 (1996).
570 14. A. Dalsgaard, A. Forslund, N. V. Tam, D. X. Vinh, P. D. Cam, Cholera in Vietnam:
571 changes in genotypes and emergence of class I integrons containing aminoglycoside
572 resistance gene cassettes in *Vibrio cholerae* O1 strains isolated from 1979 to 1996. *J*
573 *Clin Microbiol* **37**, 734-741 (1999).
574 15. J. W. Beaber, B. Hochhut, M. K. Waldor, Genomic and functional analyses of SXT, an
575 integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J*
576 *Bacteriol* **184**, 4259-4269 (2002).
577 16. A. Creasy-Marrazzo *et al.*, Genome-wide association studies reveal distinct genetic
578 correlates and increased heritability of antimicrobial resistance in *Vibrio cholerae* under
579 anaerobic conditions. *Microbial Genomics* **8**, (2022).
580 17. K. LeGault *et al.*, Temporal shifts in antibiotic resistance elements govern phage-
581 pathogen conflicts. *Science* **373**, 2020.2012.2016.423150 (2021).
582 18. C. Pal, M. D. Macia, A. Oliver, I. Schachar, A. Buckling, Coevolution with viruses drives
583 the evolution of bacterial mutation rates. *Nature* **450**, 1079-1081 (2007).
584 19. M. Yen, A. Camilli, Mechanisms of the evolutionary arms race between *Vibrio cholerae*
585 and Vibriophage clinical isolates. *Int Microbiol* **20**, 116-120 (2017).

586 20. K. D. Seed, S. M. Faruque, J. J. Mekalanos, S. B. Calderwood, F. Qadri, A. Camilli,
587 Phase variable O antigen biosynthetic genes control expression of the major protective
588 antigen and bacteriophage receptor in *Vibrio cholerae* O1. *PLoS Pathog* **8**, e1002917
589 (2012).
590 21. C. M. Boyd, A. Angermeyer, S. G. Hays, Z. K. Barth, K. M. Patel, K. D. Seed,
591 Bacteriophage ICP1: A Persistent Predator of *Vibrio cholerae*. *Annu Rev Virol* **8**, 285-
592 304 (2021).
593 22. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its
594 own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489-
595 491 (2013).
596 23. K. D. Seed *et al.*, Evolutionary consequences of intra-patient phage predation on
597 microbial populations. *Elife* **3**, e03497 (2014).
598 24. R. C. Molina-Quiroz, A. Camilli, C. A. Silva-Valenzuela, Role of Bacteriophages in the
599 Evolution of Pathogenic Vibrios and Lessons for Phage Therapy. *Adv Exp Med Biol*
600 **1404**, 149-173 (2023).
601 25. E. J. Nelson *et al.*, Transmission of *Vibrio cholerae* is antagonized by lytic phage and
602 entry into the aquatic environment. *PLoS Pathog* **4**, e1000187 (2008).
603 26. K. D. Seed *et al.*, Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic
604 bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh.
605 *MBio* **2**, e00334-00310 (2011).
606 27. W. C. Summers, Cholera and plague in India: the bacteriophage inquiry of 1927-1936. *J*
607 *Hist Med Allied Sci* **48**, 275-301 (1993).
608 28. F. D'Herelle, R. Malone, A preliminary report of work carried out by the cholera
609 bacteriophage enquiry. *Indian Medical Gazette*, 614-617 (1927).
610 29. C. L. Pasricha, A. J. H. de Monte, E. G. O'Flynn, Bacteriophage in the treatment of
611 cholera. *Ind. Med. Gaz.* **71**, 61-68 (1936).

612 30. S. M. Faruque *et al.*, Seasonal epidemics of cholera inversely correlate with the
613 prevalence of environmental cholera phages. *Proc Natl Acad Sci U S A* **102**, 1702-1707
614 (2005).
615 31. S. M. Faruque *et al.*, Self-limiting nature of seasonal cholera epidemics: Role of host-
616 mediated amplification of phage. *Proc Natl Acad Sci U S A* **102**, 6119-6124 (2005).
617 32. M. A. Jensen, S. M. Faruque, J. J. Mekalanos, B. R. Levin, Modeling the role of
618 bacteriophage in the control of cholera outbreaks. *Proc Natl Acad Sci U S A* **103**, 4652-
619 4657 (2006).
620 33. M. S. Zahid, S. M. Udden, A. S. Faruque, S. B. Calderwood, J. J. Mekalanos, S. M.
621 Faruque, Effect of phage on the infectivity of *Vibrio cholerae* and emergence of genetic
622 variants. *Infect Immun* **76**, 5266-5273 (2008).
623 34. A. Jaiswal, H. Koley, A. Ghosh, A. Palit, B. Sarkar, Efficacy of cocktail phage therapy in
624 treating *Vibrio cholerae* infection in rabbit model. *Microbes Infect* **15**, 152-156 (2013).
625 35. M. Yen, L. S. Cairns, A. Camilli, A cocktail of three virulent bacteriophages prevents
626 *Vibrio cholerae* infection in animal models. *Nat Commun* **8**, 14187 (2017).
627 36. L. A. David *et al.*, Gut microbial succession follows acute secretory diarrhea in humans.
628 *MBio* **6**, e00381-00315 (2015).
629 37. L. Alexandrova *et al.*, Identification of Widespread Antibiotic Exposure in Patients With
630 Cholera Correlates With Clinically Relevant Microbiota Changes. *J Infect Dis* **220**, 1655-
631 1666 (2019).
632 38. M. M. Monir *et al.*, Genomic attributes of *Vibrio cholerae* O1 responsible for 2022
633 massive cholera outbreak in Bangladesh. *Nat Commun* **14**, 1154 (2023).
634 39. A. Carr, C. Diener, N. S. Baliga, S. M. Gibbons, Use and abuse of correlation analyses
635 in microbial ecology. *ISME J* **13**, 2647-2655 (2019).
636 40. I. Levade *et al.*, *Vibrio cholerae* genomic diversity within and between patients. *Microb*
637 *Genom* **3**, (2017).

638 41. I. Levade *et al.*, A Combination of Metagenomic and Cultivation Approaches Reveals
639 Hypermutator Phenotypes within *Vibrio cholerae*-Infected Patients. *mSystems* **6**,
640 e0088921 (2021).
641 42. A. Jolivet-Gougeon *et al.*, Bacterial hypermutation: clinical implications. *J Med Microbiol*
642 **60**, 563-573 (2011).
643 43. M. R. Olm, A. Crits-Christoph, K. Bouma-Gregson, B. A. Firek, M. J. Morowitz, J. F.
644 Banfield, inStrain profiles population microdiversity from metagenomic data and
645 sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727-736 (2021).
646 44. A. C. Levine *et al.*, A comparison of the NIRUDAK models and WHO algorithm for
647 dehydration assessment in older children and adults with acute diarrhoea: a prospective,
648 observational study. *Lancet Glob Health* **11**, e1725-e1733 (2023).
649 45. A. C. Levine *et al.*, External validation of the DHAKA score and comparison with the
650 current IMCI algorithm for the assessment of dehydration in children with diarrhoea: a
651 prospective cohort study. *Lancet Glob Health* **4**, e744-751 (2016).
652 46. A. I. Khan *et al.*, Electronic decision support and diarrhoeal disease guideline adherence
653 (mHDM): a cluster randomised controlled trial. *Lancet Digit Health* **2**, e250-e258 (2020).
654 47. J. B. Harris *et al.*, Susceptibility to *Vibrio cholerae* Infection in a Cohort of Household
655 Contacts of Patients with Cholera in Bangladesh. *PLoS Negl Trop Dis* **2**, e221 (2008).
656 48. B. Seed, Purification of genomic sequences from bacteriophage libraries by
657 recombination and selection in vivo. *Nucleic Acids Res* **11**, 2427-2445 (1983).
658 49. A. I. Khan, F. Qadri, Epidemiology of cholera in Bangladesh: Findings from Nationwide
659 Hospital-based Surveillance, 2014-2018. *CID*, (2019).
660 50. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: an ultra-fast single-node
661 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
662 *Bioinformatics* **31**, 1674-1676 (2015).

663 51. *Manual of Clinical Microbiology. 8th ed.* P. Murray, E. Baron, J. Jorgensen, M. Pfaller, R.
664 Yolken, Eds., (American Society for Microbiology Press, Washington, D.C., 2003).
665 52. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2.
666 *Genome Biol* **20**, 257 (2019).
667 53. J. Lu, F. P. Breitwieser, P. Thielen, S. L. Salzberg, Bracken: estimating species
668 abundance in metagenomics data. *PeerJ Comput. Sci* **3**, (2017).
669 54. C. M. K. Sieber *et al.*, Recovery of genomes from metagenomes via a dereplication,
670 aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).
671 55. A. P. Camargo *et al.*, Identification of mobile genetic elements with geNomad. *Nat*
672 *Biotechnol*, (2023).
673 56. S. Roux *et al.*, iPHoP: An integrated machine learning framework to maximize host
674 prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**,
675 e3002083 (2023).
676 57. L. R. van Dijk *et al.*, StrainGE: a toolkit to track and characterize low-abundance strains
677 in complex microbial communities. *Genome Biol* **23**, 74 (2022).
678 58. G. Arango-Argoty, E. Garner, A. Pruden, L. S. Heath, P. Vikesland, L. Zhang, DeepARG:
679 a deep learning approach for predicting antibiotic resistance genes from metagenomic
680 data. *Microbiome* **6**, 23 (2018).
681

Supplementary Materials for

Phage predation is a biomarker for disease severity and shapes pathogen genetic diversity in cholera patients

Authors: N. Madi[1,2], E. T. Cato[3], Md. Abu Sayeed[3], A. Creasy-Marrazzo[3], A. Cuénod[1,2], K. Islam[4], Md. I.UI. Khabir[4], Md. T. R. Bhuiyan[4], Y. Begum[4], E. Freeman[3], A. Vustepalli[3], L. Brinkley[3], Kamat[4], L. S. Bailey[5], K. B. Basso[5], F. Qadri[4], A. I. Khan[4*], B. J. Shapiro[1,2,6*], E. J. Nelson[3*]

Affiliations:

[1] Department of Microbiology & Immunology, McGill University, Montréal, Canada
[2] McGill Genome Centre, McGill University, Montréal, Canada
[3] Departments of Pediatrics and Environmental and Global Health, University of Florida, Gainesville, FL, USA
[4] Infectious Diseases Division (IDD) & Nutrition and Clinical Services Division (NCSD), International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh
[5] Department of Chemistry, University of Florida, Gainesville, FL, USA
[6] McGill Centre for Microbiome Research, McGill University, Montréal, Canada

[*] Corresponding authors

Email: ashrafk@icddrb.org; jesse.shapiro@mcgill.ca; eric.nelson@ufl.edu

The PDF file includes:

Index
Materials and methods
Results
Figures: S1 to S19
Tables: S1 to S12
References

Other Supplementary Materials for this manuscript include the following:
Data S1 to S5

Materials and Methods

Ethics Statement. The samples analyzed were collected within the context of two previously published IRB approved clinical studies conducted in Bangladesh: the mHealth Diarrhea Management (mHDM) cluster randomized controlled trial (IEDCR IRB/2017/10; icddr,b ERC/RRC PR-17036; University of Florida IRB 201601762; clinicaltrials.gov NCT03154229) (*1*) and a hospital-based National Cholera Surveillance (NCS) study (icddr,b ERC/RRC PR-15127) (*2*); the consent processes are described within these published studies.

Study Design. The study design was a national prospective longitudinal study of patients presenting with diarrhoeal disease at five Bangladesh Ministry of Health and Family Welfare district hospitals (both mHDM and NCS sites) and two centralized hospitals (NCS sites alone) (BITID, icddr,b) from March 2018 to December 2018. Recruitment was based on the census of diarrheal disease patients that sought care at the hospitals. For the mHDM study, inclusion criteria were patients two-months of age or older, and presentation with acute diarrhea defined as three or more episodes of loose stools in the 24 hours prior to admission and a duration of disease less than seven days. For the NCS study, patients with diarrheal disease of all ages were included. For those under 2 months of age, diarrhea was defined as a change in stool habit from 'usual' (increased frequency with less formed stool). For those 2 months and older, diarrhea was defined as 3 or more loose or liquid stools within 24 hours or 3 loose/liquid stools or fewer causing dehydration in the last 24 hours. There was no exclusion based on enrollment in a second study (e.g., mHDM vs NCS). Patient metadata is in Data File S1.

Sample Collection. For the mHDM study, the intent was to collect four stool samples per study site per day. For the NCS study, the intent was to collect daily stool samples from two participants less than 5 years old and two participants 5 years of age and older; if the target number for an age group was not met, samples were collected from the other age group to achieve a total of 4 samples per day. For the samples collected at the five district hospitals and BITID, an aliquot was placed in transport media (Cary-Blair) and transferred to the icddr,b laboratory for culture, and a 0.5 ml aliquot was added to 1.3 ml of RNAlater to stabilize the sample for subsequent analysis; cold-chain below 4°C was not consistently available at these sites. For samples collected at the icddr,b, samples were stabilized in RNAlater; culture was performed directly. Samples were stored at the centralized icddr,b laboratory at -80°C.

Antibiotic detection by liquid chromatography mass spectrometry (LC-MS/MS).

LC-MS/MS methodology for both qualitative and quantitative approaches. The approach was based on prior studies (*3, 4*). Stool supernatants from the primary collection were obtained by centrifugation without filtration to minimize loss. Proteins were precipitated (1:7 ratio (v/v) of

water:methanol). Supernatants were diluted with methanol and water (1:1 v/v) in 1% formic acid for liquid chromatography, and 5 µl of supernatant was injected for analysis. LC-MS/MS was performed on a 2.1 x 150-mm Hypersil Gold aQ column (particle size, 3 µm) using a high-performance liquid chromatography system (Thermo UltiMate 3000 series) with an LTQ XL ion trap mass spectrometer (Thermo Fisher Scientific). Mobile phases were 1% formic acid in water (A) and 1% formic acid in methanol (B) and held at a constant 5%B for 2 minutes before ramping to 95%B at 15 minutes where it was held for an additional minute before returning to starting conditions for a total run time of 25 min. Blanks were run in between every two samples, as well as before and after quality control samples and standards.

Eluent was ionized using electrospray ionization (ESI) in positive mode at a spray voltage of 5 kV, a nitrogen sheath gas flow rate of 8 L min^{-1}, and capillary temperature of 300°C. Two scan events were programmed to perform an initial scan from *m/z* 100 to 1000, which was followed by targeted collision induced dissociation based on a retention time and mass list. Retention time windows ranged from 1.7 minutes to 5.1 min, depending on the elution range of the standards at high and low concentrations. Masses were targeted for the most abundant adduct or ion associated with each antibiotic (typically the $[M+H]^+$ ion) with a *m/z* 1 window.

LC-MS/MS methodology specific to qualitative analysis. The target list for the qualitative analysis was narrowed from prior work that assayed for antibiotics and non-antibiotics (*3, 4*): acetaminophen, Metronidazole, Zofran, Furazolidone, Nalidixic Acid, Sulfamethoxazole, Trimethoprim, Omeprazole, Ciprofloxacin, Cephalexin, Penicillin V (M+H), Amoxicillin, Penicillin V, Doxycycline or Tetracycline, Ceftriaxone, Erythromycin, and Azithromycin. Based on this prior work, we narrowed our analysis to the common clinical relevant agents of ciprofloxacin, azithromycin, doxycycline/tetracycline (same mass), nalidixic acid and metronidazole. Qualitative data analysis was performed using extracted ion chromatograms and MSMS matching with a control antibiotic MSMS library using Xcalibur 2.2 SP 1.48 (Thermo Fisher Scientific).

LC-MS/MS methodology specific to quantitative analysis. The target list for the quantitative analysis was furthered narrowed to: Ciprofloxacin, Azithromycin and Tetracycline/Doxycycline. A standard curve was made for each quantitative target by preparing a dilution series of a mix of the three native forms of the quantitative targets (Ciprofloxacin, Doxycycline, Azithromycin). The dilutions were: 0.5, 0.25, 0.125, 0.063, 0.05, 0.02, 0.01 µg/ml. All samples, including standards, were spiked with a 0.25 µM mix of isotope labeled control antibiotics (Ciprofloxacin d-8 (Millipore-Sigma), doxycycline d-5 (Isosciences), Azithromycin d-5 (LGC standards).

A calibration curve was generated by plotting the ratio of the area under the curve for analyte peaks in the ion chromatograms to the AUC of the isotope peaks against known

concentrations of the seven standards. The linear line of best fit produced by this plot was generated and used to extrapolate the quantity of drug in mHDM samples. The AUC of chromatogram peaks produced by the target compound was compared to that of the peak produced by the isotope spike in to generate a ratio. This ratio was then converted to a concentration (in units of µg/ml) using the calibration curve equation. Data analysis was performed manually by viewing the extracted ion chromatograms and MSMS from each sample and matching with an in-house antibiotic library using 2.2 SP 1.48 (Thermo Fisher Scientific). The detection of a drug was confirmed when a peak in the chromatogram at the proper retention time window was identified and the most abundant adduct or ion associated with that drug (typically the $[M+H]^+$ ion) with a m/z 1 window. Measured antibiotic concentrations are in Data File S2.

Molecular analyses. Standard methods were used to extract total nucleic acid (tNA) from the samples stabilized in RNAlater (Invitrogen). In brief, the samples were thawed at room temperature and centrifuged for 5 minutes at max speed. All but 50 µl of the supernatant was removed; the supernatant was refrozen for subsequent mass spectrometry. The remaining pellet, with the 50 µl supernatant, was combined with 380 mg of glass beads and 1 ml InhibitEx buffer and processed using the QIAamp Fast DNA Stool Mini Kit. Elutions were performed with 50 µl of Qiagen ATE buffer followed by a second elution with 50 µl ATE, quantification (Nanodrop) and storage at -80°C. For qPCR, we used established *tcpA* primers for *V. cholerae* (*4, 5*) and 16S ribosomal gene (rDNA) primers as a positive control for sample processing and extraction (*6*). Primer pairs were developed for the virulent bacteriophages ICP1 (gp58.2), ICP2 (gp24), and ICP3 (gp19); primer sequences are conserved between Southeast Asia and African isolates. Reactions used Luna qPCR master mix (New England Biolabs) with 25 µl volumes, and the method was a two-step reaction with denaturation at 98°C and annealing/extension at 60°C for all primers except ICP2 GP24 F/R (annealing/extension 63°C). Assays were performed on an Azure Cielo Real-time qPCR System and two Bio-Rad CFX96 Touch Real-Time PCR Detection Systems. Samples were analyzed in 96-well formats with two replicates except for 16SrDNA (one replicate); a Ct threshold less than 28 was used to define positivity (*4, 5, 7*). For quality control, samples that were negative for 16S rDNA were excluded from subsequent analyses. A third replicate was performed when: both replicates were positive but the difference in Ct values was more than 4, or one replicate was positive and one was negative and the difference in Ct values was greater than four. After the third replicate was performed, outliers were removed and the Ct values were averaged and the final determination of positivity was scored (Ct<28). Primers were validated using standard curves of (i) synthetic templates and (ii) biologic agents for *V. cholerae*, ICP1, ICP2, ICP3. Conventional methods were used for PCR.

Metagenomic analyses

Short read classification using kraken2/braken. We classified all short-read data with a Kraken2 database (Wood, Lu, and Langmead 2019) containing bacterial, archaeal, and viral domains, along with the human genome and a collection of known vectors (UniVec_Core) from NCBI, in July 2020. A Bracken database (Lu et al. 2017) was also built with a read length of 150 bp and the default k-mer length of 35. Kraken2 and Bracken version 2.5 were run with default parameters.

Phage identification from metagenomic assemblies. We characterised phages in 224 metagenomes with either >0.5% *Vc* or >0.1% ICP phage relative abundances. We used geNomad (v1.7.0, min-score 0.7, max-fdr 0.1) (*8*) to identify (pro)-phages from 83,928 contigs identified as 'viral'. iPhoP (v1.3.1, min-score 90%) (*8*) was used to predict the bacterial hosts of the identified phage contigs, resulting in 627 contigs classified as '*Vibrio* phages'. We used two different approaches to further characterise the predicted '*Vibrio* phage' sequences: (i) BLASTn to compare these sequences against known *Vibrio* phages in the INPHARED (*9*) database and reference sequences of the TLC (*10*) and PLE satellite phages (*11*) (% identity >= 80%, query coverage >= 90%), and (ii) vContact2 (*12*) to build similarity networks including the putative *Vibrio* phages and genomes from the INPHARED NCBI RefSeq databases. We used clusters in the vContact2 network to harmonise the naming of the INPHARED database genomes (eg. all INPHARED genomes which clustered together with the ICP2 reference genomes, were called 'ICP2').

Quality filtering. Before calling SNVs within the *V. cholerae* population, metagenomes were decontaminated of human and PhiX reads by mapping the reads against the GRCh 38 assembly of the human genome and the PhiX genomes, with bowtie2 version 2.3.5 (*13*). Unmapped reads were used for subsequent analyses. Then, we used iu_filter_quality_minoche from the Illumina Utils package (v2.10) with default parameters (*14*) to trim adapters, remove duplicates and for quality filtering on the short reads.

SNV profiling. Short reads were assembled with MEGAHIT (*15*) version 1.2.9 with the default parameters (k-mers length: k=21 et maxk=99 and min contig length of 1000. The contigs were grouped into bins with concoct (v1.0.0) (*16*) and MetaBAT 2 (v2.12.1) (*17*) with default parameters. Bins were then aggregated together using DAS Tool version 1.1.0 (*18*). Then we used drep (*19*) version 2.0.0 with S_ani 0.98, to de-replicate the bins and identify unique metagenome-assembled genomes (MAGs) with a secondary clustering threshold of 98% and minimum completeness of 75% and maximum contamination of 25%. To prevent mismapping of reads from other species to the *V. cholerae* genome, we competitively mapped reads from all

samples with sufficient *V. cholerae* or phages (*Vc*>0.5% of reads or phages>0.1% of metagenomic reads) against the concatenation of the 677 unique MAGs and ICP1 genome from NCBI (reference NC_015157.1, accessed 2021/12/01), using bowtie2 v 2.3.5. To characterize within-patient *V. cholerae* and ICP1 genetic diversity, we profiled the resulting bam files using inStrain v1.5.7 (*20*) with default parameters (minimum coverage of 5 to call a variant). We also ran gene-level profiling with instrain using genes annotated with prodigal. We used prodigal v2.6.3 (*21*) with the default parameters to predict genes in the *Vc* MAG and ICP1 and annotated them with eggNOG-Mapper v2 with default parameters and eggNOG database v2 downloaded on April 2021 (*22*).

To infer potential mixed infections, we used strainGST from the strainGE toolkit (*23*). To do so, we compiled a reference database of publicly available *V. cholerae* assemblies, including (i) all complete genomes from NCBI RefSeq (n=106, accessed 2023/05/12) and (ii) all assemblies published on NCBI Genbank between 2015/01/01 and 2023/05/12 and originating from samples collected between 2015/01/01 and 2019/12/31 (n=758).

To reduce false positive SNVs, we applied a stringent post-InStrain filter: all positions with coverage <20 were removed. We removed the first and last 100 positions of every scaffold, as well as positions with coverage below 0.3*median and above 3*median (median coverage across all the positions in the sample). We also removed sites that did not pass the coverage filter in more than 2 samples. Finally, we removed 420 sites that were variable in a *V. cholerae* isolate genome sequenced alongside the metagenomes. These sites were considered prone to sequencing error since they varied in an isolate genome that theoretically should contain no variation. After applying all these quality filters, we were left with 130 samples with SNV calls.

Hypermutator definition. Hypermutators were identified as samples with one or more nonsynonymous mutations in DNA repair genes (DNA repair defects) in *V. cholerae*, defined previously (*24*).

Antibiotic resistance gene identification. Antibiotic resistance genes (ARGs) were predicted from short reads using deepARG (*25*) v 1.0.2 with default parameters. Deeparg is a deep learning model that can predict ARGs from short-read metagenomic data; it uses DeepARG-DB, a merged database from 3 databases: Antibiotic Resistance Genes Database [ARDB], Comprehensive Antibiotic Resistance Database [CARD], and UniProt. This database was downloaded in December 2021. We used the relative abundance of ARGs normalized by the 16S rRNA gene content in the sample. We assessed the relative abundances of 634 known ARGs conferring resistance to 34 antibiotic classes. To identify associations between these ARGs and

gut microbes, we ran a multiple factor analysis (MFA) with the 37 most dominant species from a PCA (20 highest coordinates on both axes) and all the ARGs.

SXT ICE identification. We used Bowtie2/2.3.5 with the 'very-sensitive' option to map short reads against the two most prevalent SXT ICEs in Bangladesh at the time of our sampling (*26*): ICEVchind5 (NCBI accession GQ463142.1) and ICEVchind6 (accession MK165649.1) reference sequences downloaded from NCBI. The ICE was considered as present in a metagenome when 90% of its length was covered by at least one read. Using this criterion, 144 samples (59%) were ind5+, 26 (10.6%) ind6+, and 54 (22.1%) ICE-.

Statistical analyses. All statistics and visualizations were done in R studio version 1.2.5042.To reduce dimensions of the species composition table, we ran a principal component analysis (PCA) on the Hellinger transformed abundance data (5719 species*344 samples) (decostand function and rda function in the vegan R library) and selected the 20 most dominant species on each axis (37 in total).

To understand relationships between taxa in the microbiome, phages, antibiotics and patient metadata, we used a Redundancy analysis (RDA) on bacterial species abundances (rda function from vegan R package). Only the 7 most dominant species identified with the PCA were included: *v*, *Escherichia coli*, *Bacteroides vulgatus*, *Bifidobacterium longum*, *Shigella flexneri*, *Bifidobacterium breve*, and *Streptococcus mitis*. As explanatory variables, we used three antibiotics: azithromycin (AZI), ciprofloxacin (CIP), Doxycycline (DOX) as well as the three phages (ICP1, ICP2, ICP3), along with the most contributing patient metadata, selected with forward selection method (ordiR2step function from the vegan R library). These metadata are: the area where the sample was collected, date of sampling, age in years, vomiting state (yes or no) and dehydration status (severe, moderate or mild). We began the forward selection with a model with phages and antibiotic concentrations (µg/ml). The RDA was run on log-chord transformed abundances, and a permutation test with 999 iterations was used to assess the statistical significance of both the model and of each explanatory variable (anova.cca function from vegan package in R). The explained variation R^2 and adjusted R^2 were estimated with the RsquareAdj function in the vegan R package.

To identify species associated with each degree of dehydration, we used the indicator species analysis (*27*) as implemented in the multipatt R function from the indicspecies R library with (no group combination, duleg parameter set to TRUE and 9999 permutations) on the log-chord transformed species table (334 samples, 37 dominant species from the PCA) using the decostand and rda functions from the R vegan library). Reported P-values correspond to a permutation test with 999 iterations.

To study associations between ARGs and species, we ran a Multiple factor analysis (MFA) (*28*) using the mfa function from the FactoMineR R package, with the 37 most dominant species and all 634 resistance genes identified by deepARG. The visualization was done with the fviz_mfa_var function from the factoextra R library.

To model the relationships between phage, antibiotics and *V. cholerae* within each patient, we fit a generalized additive model (GAM) with *V. cholerae* relative abundance as a function of ICP1, antibiotics, and the interaction between them. We added the degree of dehydration as a random effect because it improved the model (smallest Akaike information criteria (AIC) compared to the other models). We fit several GAMs with different combinations of predictors: from a model with all antibiotics and their interaction with ICP1 to separate models with each antibiotic and its interaction with ICP1, and compared them based on their AIC (AICtab function from the bblme R package). The most parsimonious model was retained (the one with the smallest AIC). The GAMs were fitted using a beta error distribution with log-link function because *Vc* relative abundance is a continuous value between 0 and 1. We evaluated the selected model fit by inspecting residual distributions and fitted-observed value plots using the gam.check function from the mgcv R package. All P-values reported for the GAMs correspond to the Chi-square test from gam.summary function from mgcv R package. We also reported the adjusted R^2 (from the same function) as an evaluation of the goodness of fit (**Table S6**). To support the results from the models fitted on metagenomic relative abundances, we repeated the same methods with the qPCR-based absolute abundances of *Vc* and ICP1. Since the response (*Vc* absolute abundance from qPCR) is a continuous variable, we used additive models (Gaussian GAMs) after log-transforming the response (Figure S10, Table S7).

We tested whether phages or antibiotics select for potentially adaptive mutations in *Vc* by fitting generalized linear mixed models (GLMMs) (glmmtmb function from glmmTMB R package) with phages and antibiotics as predictors of the number of high-frequency nonsynonymous SNVs in the *V. cholerae* genome. We added *Vc* abundance as a fixed effect to the model to control for any coverage effects. We focused on higher-frequency SNVs (>10% within a sample) as more likely to be beneficial, and nonsynonymous SNVs as those more likely to have fitness effects. We fit several models with different combinations of predictors: from a model with all antibiotics and their interaction with ICP1 to separate models with each antibiotic and its interaction with ICP1, and compared them based on their AIC (AICtab function from bblme R package). The most parsimonious model was retained (the one with the smallest AIC). Because the response is count data, we fit three different count GLMMs: Poisson, negative binomial1 and negative binomial2, implemented in the glmmTMB R package, and then selected the model with the lowest AIC as

described (*29*). The selected model was fitted with the nbinom2 error distribution. All continuous predictors were standardized to zero mean and unit variance before analyses to improve the model's convergence. We evaluated the selected model fits by inspecting the residuals using the DHARMa library in R (simulateResiduals and plot functions). All the *P*-values reported for the GLMMs correspond to the Wald *P*-values reported by the glmm.summary function from the R package glmmTMB. As an evaluation of the goodness of fit, we compared the GLMMs and the corresponding null models (the same model but with no fixed effects other than the intercept) with likelihood-ratio tests (anova function from the stats R package), and reported the corresponding P-values (**Table S8**). We also reported the marginal R^2 as a measure of the variance explained by fixed effects, estimated with the r2 function from the performance R package. We used GLMMs because they are currently more flexible than GAMs in the range of count models that they can fit (https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html). Furthermore, they can deal with overdispersion with two versions of negative binomial distributions: negative binomial 1 and negative binomial 2, respectively with linear and quadratic parametrization (*30*).

To further test the hypothesis that phages or antibiotics select for adaptive mutations in *V. cholerae*, we examined the relationship of the frequency of nonsynonymous SNVs in *V. cholerae* with phages and antibiotics. We fit a generalized additive model (GAM) (gam function in mgcv R library) with the average frequency of nonsynonymous SNVs as a function of ICP1, antibiotics, and their interactions, and *Vc* and ICP1 relative abundances from metagenomic data. We also included the fixed effect of ICE presence/absence as another factor that could provide phage or antibiotic resistance as well as mutation type to differentiate among non-synonymous (NS), synonymous (S), and intergenic (I) mutations. We fit GAMs with all antibiotics and their interaction with ICP1, as well as several simpler models with each antibiotic separately, and compared them based on their AIC. The GAMs were fitted using a beta error distribution with a log-link function because the average minor allele frequency of nonsynonymous SNVs is a continuous value between 0 and 0.5. We evaluated the selected model fit by inspecting residual distributions and fitted-observed value plots using the gam.check function from the mgcv R package. All the P-values reported for the GAMs correspond to the Chi-square test from the gam.summary function of the mgcv R package. We also reported the adjusted R2 (from the same function) as an evaluation of the goodness of fit (**Table S9**).

SUPPLEMENTARY FIGURES

FIGURE S1

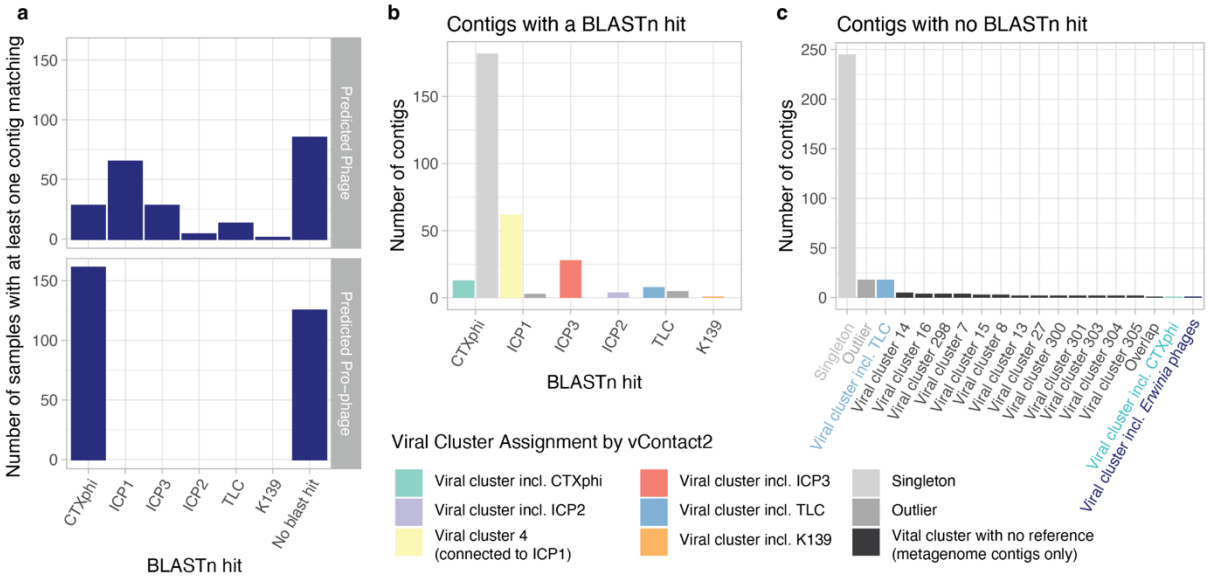


Fig S1: Inferred *Vibrio* phages in 224 metagenome assemblies. (a) At least one predicted *Vibrio* phage contig matched a CTXphi, ICP1, ICP3, TLC, ICP2 and K139 genome using BLASTn in 189, 63, 28, 13, 4 and 1 samples, respectively. **(b)** Phages ICP3, ICP2 and K139 are unambiguously assigned using vContact2. Contigs matching the CTXphi reference genome using BLASTn clustered either with the CTXphi reference viral cluster, or formed singletons (observed in only one sample). Similarly, most contigs matching an ICP1 reference genome using BLASTn were assigned by vContact2 to Viral Cluster 4, which did not include any reference phage genomes, but numerous connections to the reference ICP1 genome. **(c)** For 321 inferred *Vibrio* phage contigs from 210 samples, there was no BLASTn hit to a known *Vibrio* phage. Of these, 245 formed singletons, 18 were labeled as 'outliers' and one was placed between two viral sequence clusters. The remaining 57 contigs with no BLASTn hit clustered into 16 viral clusters, 14 of which exclusively comprised '*Vibrio* phages' identified in this analysis (mean cluster size = 2.8 contigs, range 2-5). Eighteen contigs with no BLASTn hit clustered with the TLC satellite phage (which flanks the CTXphi prophage in the *V. cholerae* genome), and one each with CTXphi and in a cluster predominantly including *Erwinia* phages.

FIGURE S2

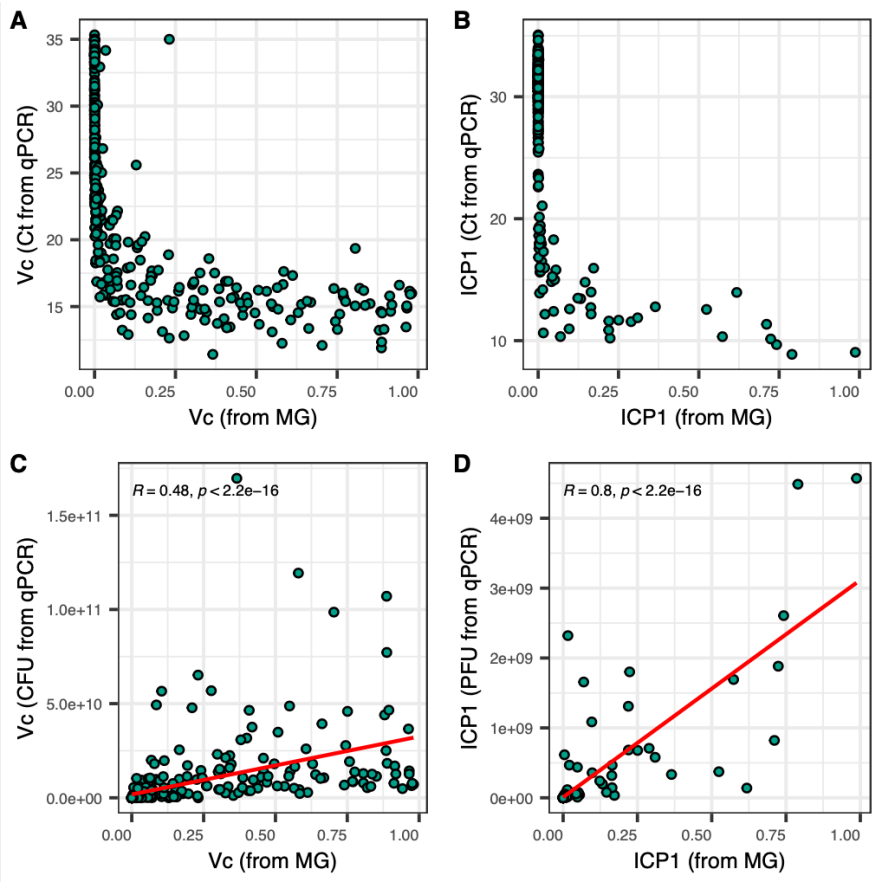


Fig S2. Correlations between qPCR- and metagenomics-based quantification of *Vc* and ICP1. Raw qPCR cycle threshold (Ct) values plotted against relative abundances in metagenomes for (A) *Vc* and (B) ICP1. Inferred number of (C) *Vc* colony forming units (CFU) per ml of stool and (D) ICP1 plaque forming units (PFU) per ml, both at a Ct threshold of 28. Red lines show Pearson correlations, with coefficients and *p*-values noted in the figure panels.

FIGURE S3

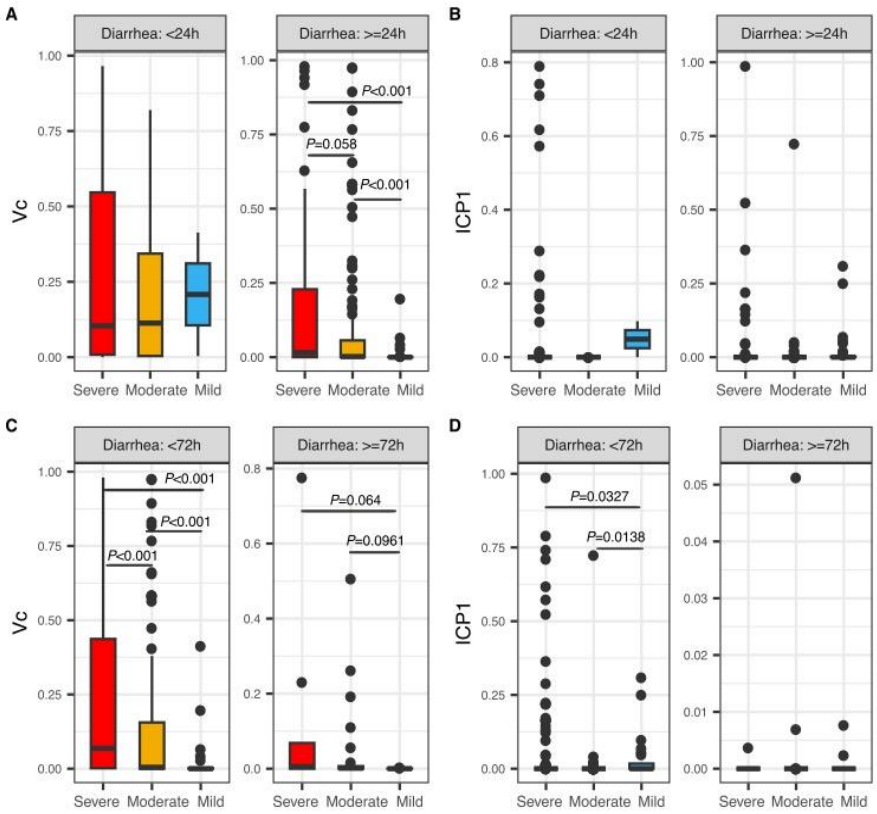


Fig S3. ICP1:*V. cholerae* ratios among patients with different dehydration status binned by self-reported duration of diarrhea. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test, adjusted for multiple tests using the Benjamini-Hochberg (BH) method. Only significant ($P<0.05$) and marginally significant *P*-values (<0.1) are shown. 323 samples with *V. cholerae*>0% of metagenomic reads were included, with 165 from severe, 128 from moderate, and 30 from mild cases. The solid horizontal line is the median and the boxed area is the interquartile range.

FIGURE S4

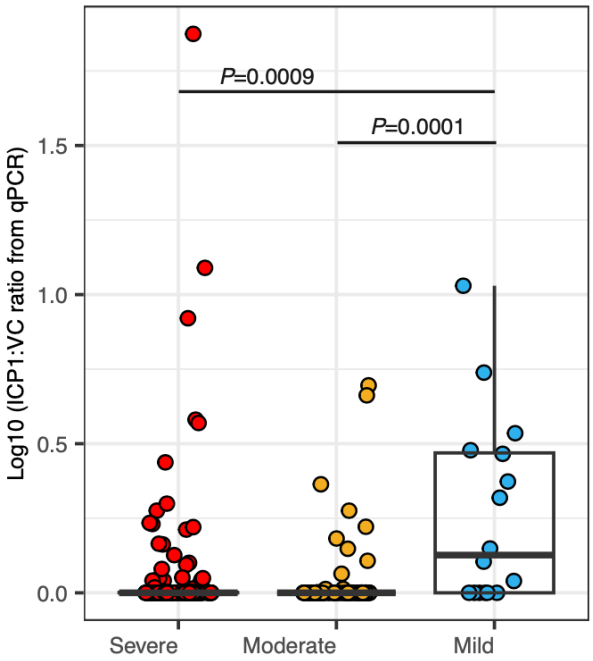


Fig S4. The ICP1:*V. cholerae* ratio quantified with qPCR is higher in patients with mild dehydration. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test, adjusted for multiple tests using the Benjamini-Hochberg (BH) method. Only significant *P*-values (<0.05) are shown. Only 248 samples with *V. cholerae*>0% from qPCR data were included to exclude zeroes from the denominator of the ratio, with 138 from severe, 94 from moderate, and 16 from mild cases. A pseudocount of one was added to the ratio before log transformation. This figure is a version of Figure 1B using qPCR data.

FIGURE S5

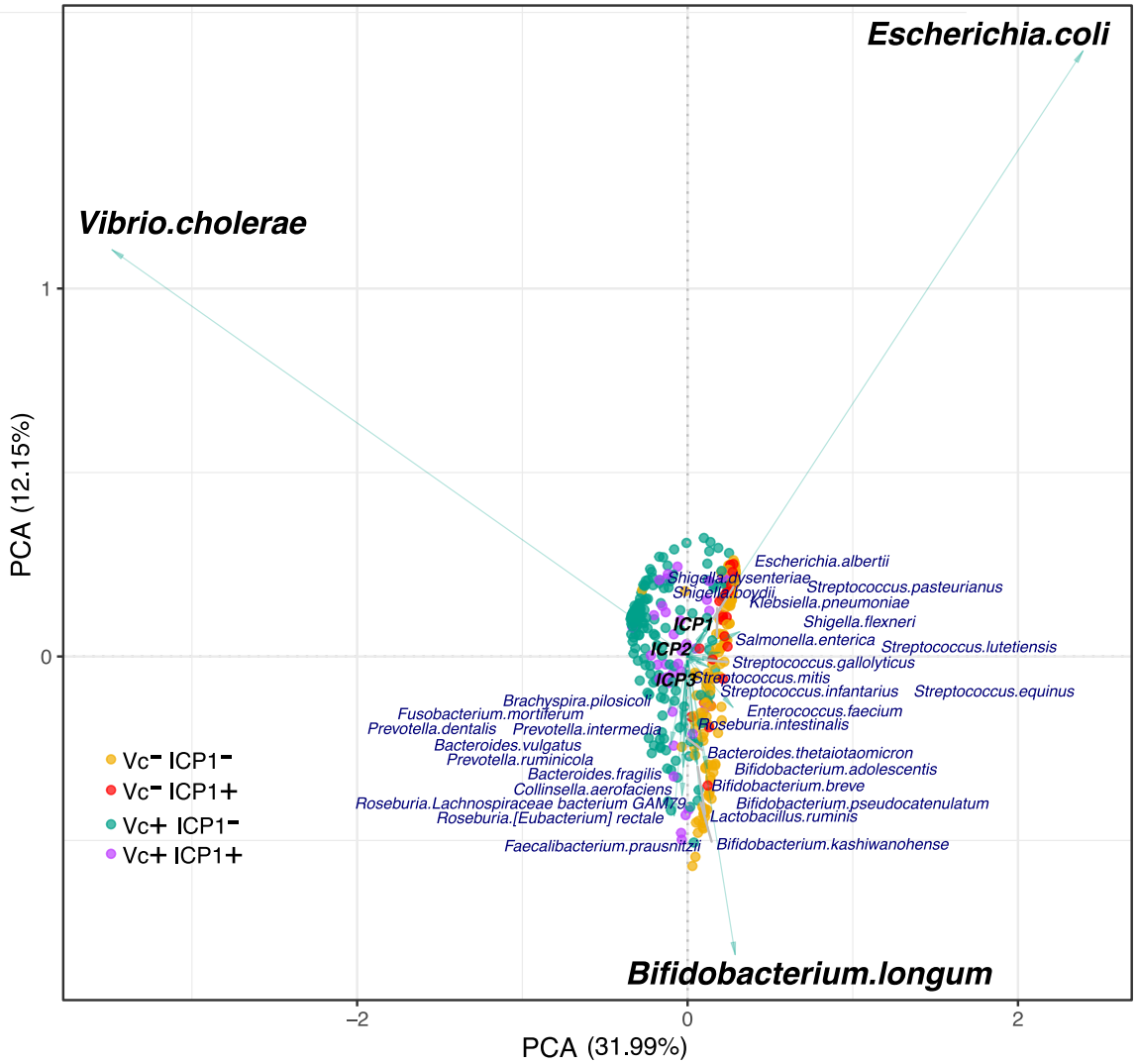


Fig. S5. Principal component analysis (PCA) of the taxonomic composition of cholera patient stool samples. This analysis was run to reduce species composition dimensions from 5719 annotated with kraken2/braken to the 37 most dominant (with the 20 most dominant on the two first axes). Circles indicate samples; arrows indicate species. Samples are colored by their *V. cholerae* and ICP1 relative abundances (percentage of metagenomic reads). Yellow: patients with *V. cholerae* < 0.5% and ICP1 < 0.1%, red: *V. cholerae* < 0.5% and ICP1 >= 0.1%, green: *V. cholerae* >= 0.5% and ICP1 < 0.1%, purple: *V. cholerae* >= 0.5% and ICP1 >= 0.1%.

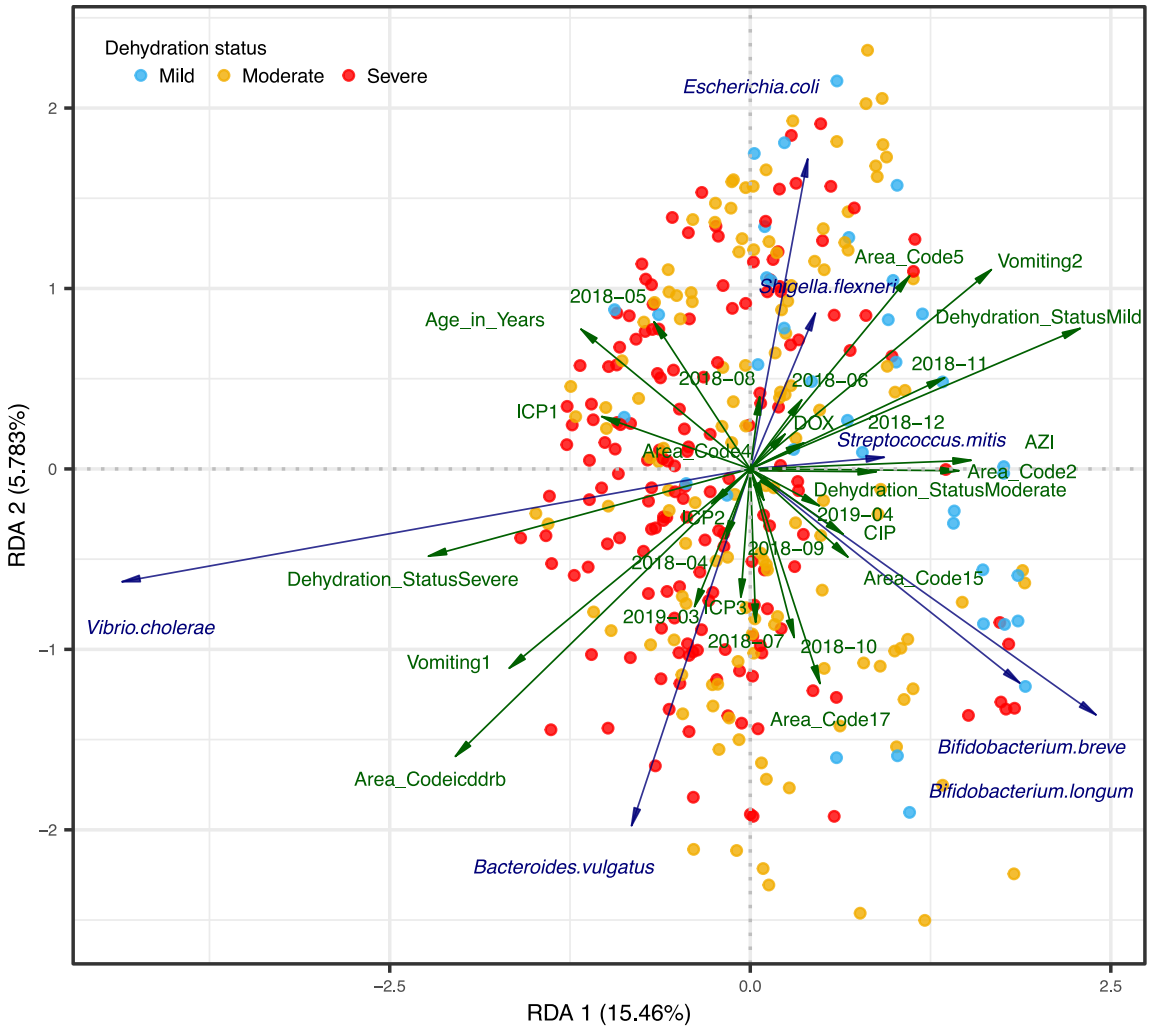


Fig S6. The two first axes from the RDA on the 7 most dominant species identified with the PCA (Figure S4), phages (ICP1, ICP2, ICP3), patient metadata, and antibiotic concentrations. Three quantified antibiotics are represented: azithromycin (AZI), ciprofloxacin (CIP), Doxycycline (DOX). Patient metadata are: the area where the sample was collected, date of sampling, age in years, vomiting state (yes or no) and dehydration status (severe, moderate or mild). All variables are shown here; the collection date and area code were omitted in Fig. 1C for clarity.

FIGURE S7

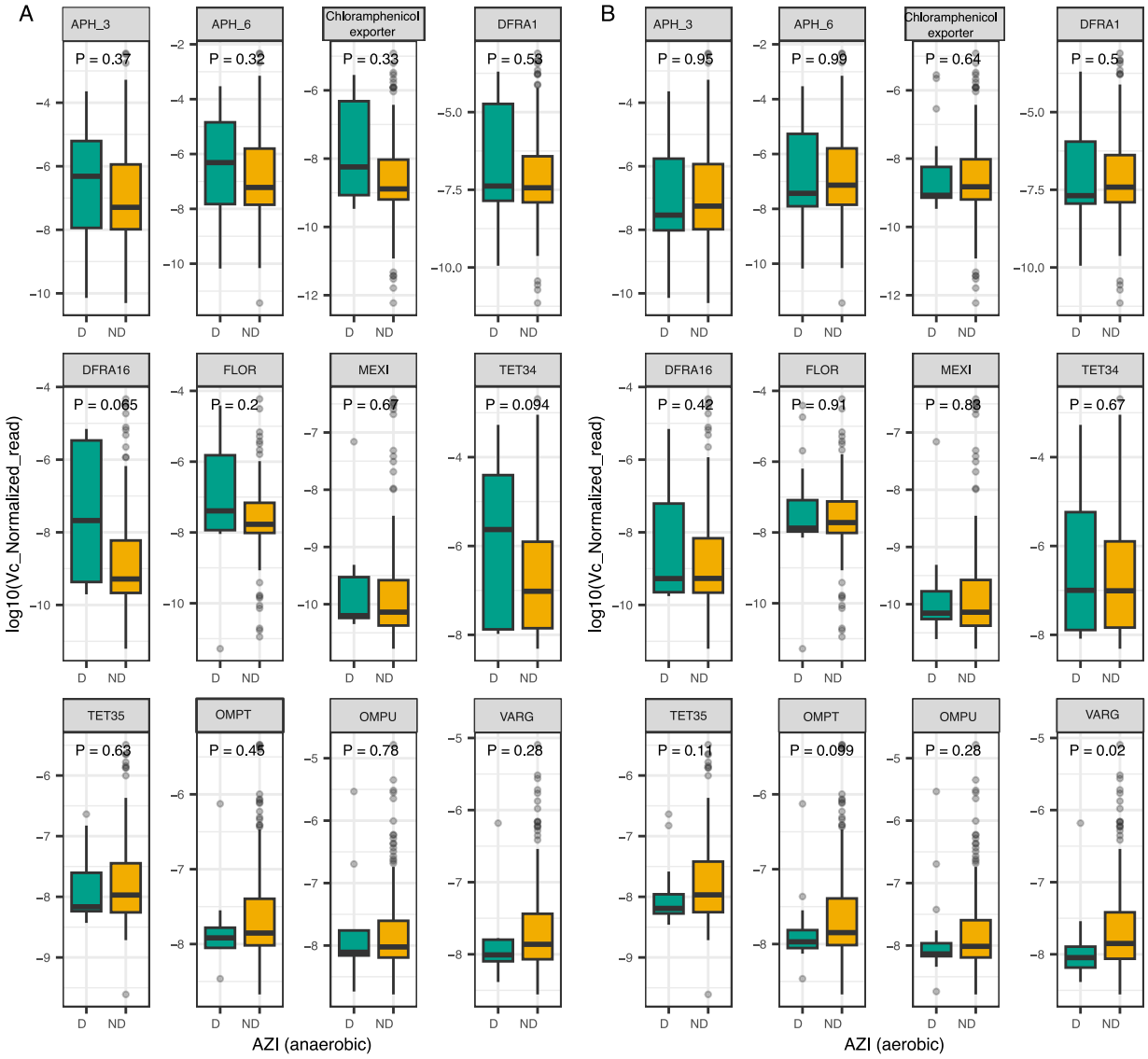


Fig. S7. Distribution of the relative abundance of *V. cholerae*-associated ARGs in patients with different exposures to azithromycin (AZI). AZI>=MIC (detected; D) and AZI<MIC (not detected; ND) under anaerobic conditions (A) or aerobic conditions (B). AZI MIC=1 μg/ml under aerobic conditions and 8 μg/ml under anaerobic conditions. The Y-axis is the relative abundance of ARGs in metagenomes normalized by 16S rRNA gene reads and by *V. cholerae* reads. Only V. cholerae-positive samples are plotted (V. cholerae>0 reads). P-values are from a Wilcoxon test. No comparisons (D vs. ND) are significant (P<0.05) after BH correction for 24 tests.

FIGURE S8

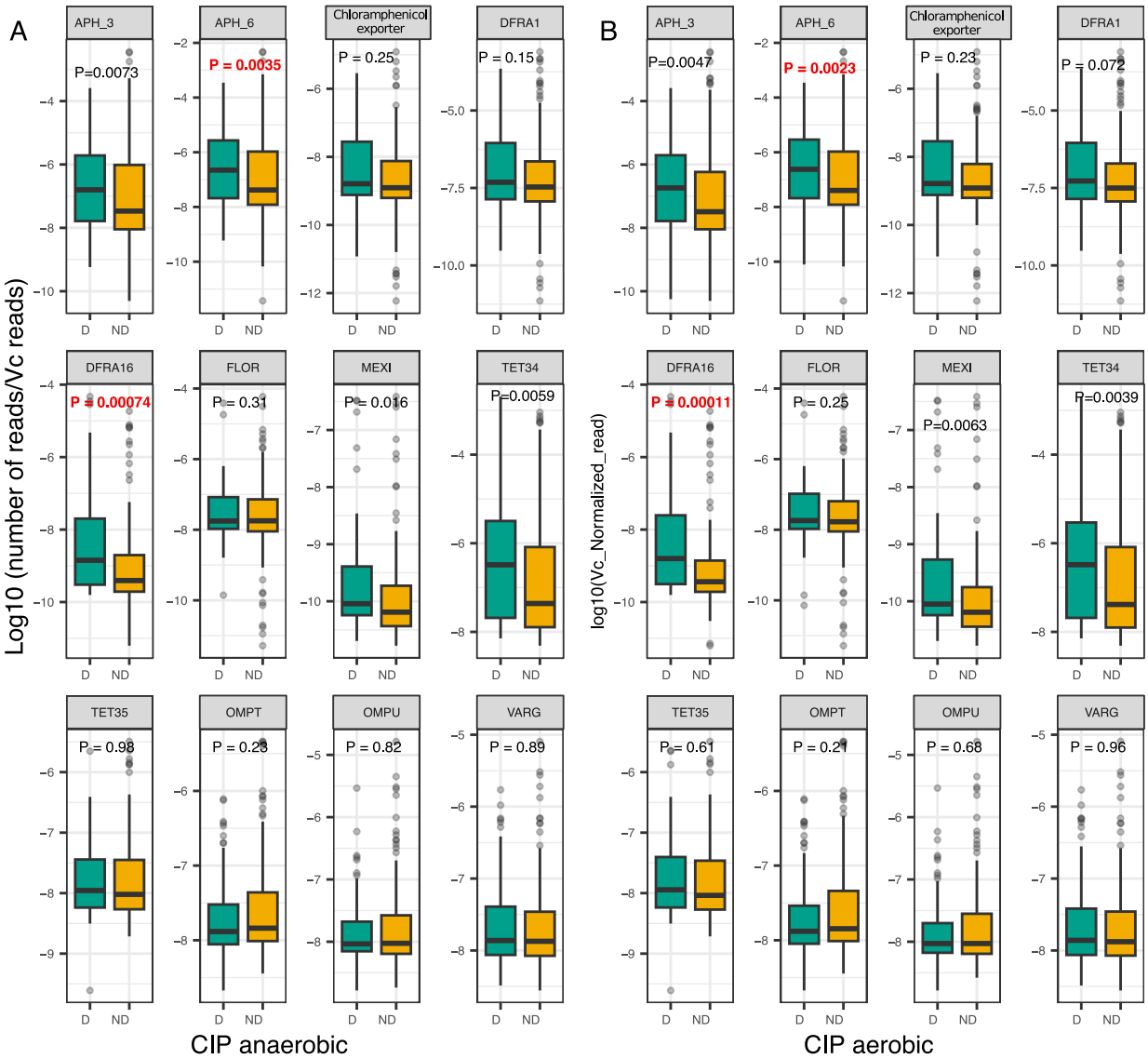


Fig. S8. Distribution of the relative abundance of *V. cholerae*-associated ARGs in patients with different exposures to ciprofloxacin (CI). CIP>=MIC (detected; D) and CIP<MIC (not detected; ND) under anaerobic conditions (A) or aerobic conditions (B). CIP MIC=0.016 µg/ml under aerobic conditions and 0.063 µg/ml under anaerobic conditions. The Y-axis is the relative abundance of ARGs in metagenomes normalized by 16S rRNA gene reads and by *V. cholerae* reads. Only *V. cholerae*-positive samples are plotted (*V. cholerae*>0 reads). *P*-values are from a Wilcoxon test. In red: BH corrected *P*<0.05 after correction for 24 tests.

FIGURE S9

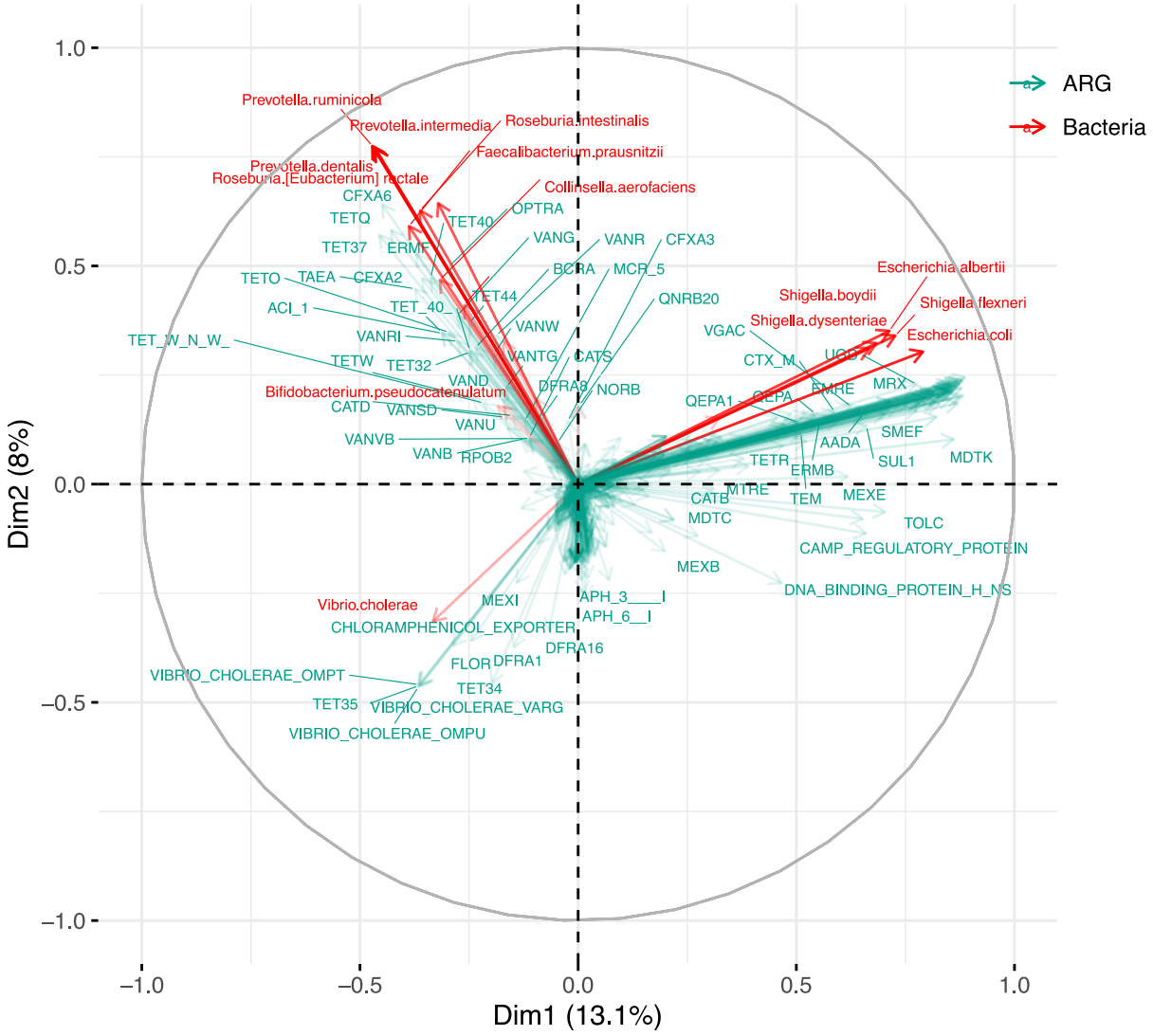


Fig S9. MFA showing correlations between bacterial species and antimicrobial resistance genes (ARGs) in patient gut microbiomes. The 37 most dominant species from the PCA (Fig S5) and all the resistance genes annotated with deepARG were included in the ordination. 12 ARGs clustered with *V. cholerae*: OMPU, OMPT, VARG, TET34, TET35, DFRA1, DFRA16, FLOR, Chloramphenicol exporter, Aph6, Aph3, MEXI. Only the greatest correlations on each axis are shown for clarity.

FIGURE S10

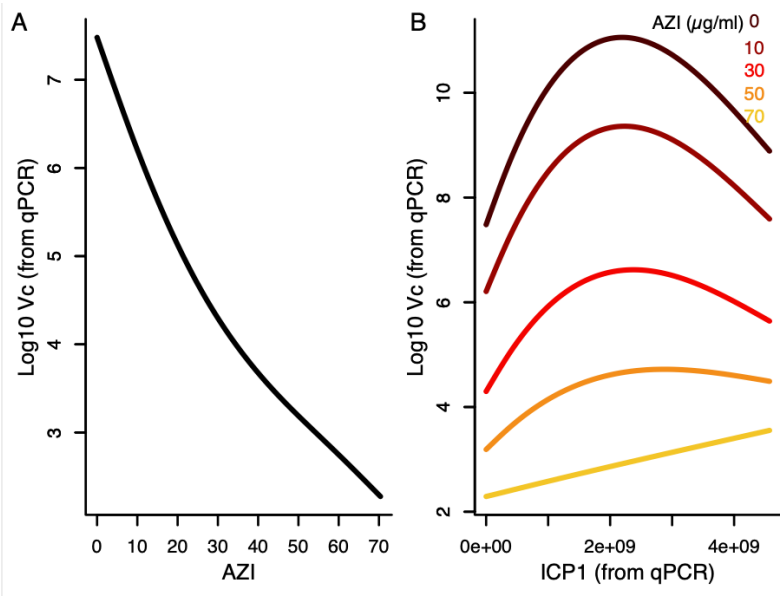


Fig S10. Interactions between *V. cholerae*, phage ICP1, and azithromycin tested with qPCR data. (A) *V. cholerae* declines in absolute abundance with higher relative abundance of azithromycin (AZI) in µg/ml. (B) The relationship between ICP1 and *V. cholerae* abundances estimated with qPCR is affected by AZI concentration (µg/ml); the illustrated AZI concentrations show regular intervals between the minimum (0 µg/ml) and maximum (78 µg/ml) observed values. GAM results fit to data from 329 samples with qPCR data. The response variable (*Vc* absolute abundance from qPCR) was log-transformed after adding a pseudocount of 1 and a Gaussian distribution was used. The predictors are ICP1 abundance from qPCR and antibiotics concentration (µg/ml). As with metagenomics data, we fit several GAMs with different combinations of predictors: from a model with all antibiotics and their interaction with ICP1 to separate models with each antibiotic and its interaction with ICP1 and compared them based on their Akaike information criteria (AIC). The most parsimonious model (shown) was retained (the one with the smallest AIC). Model details are provided in Table S7.

FIGURE S11

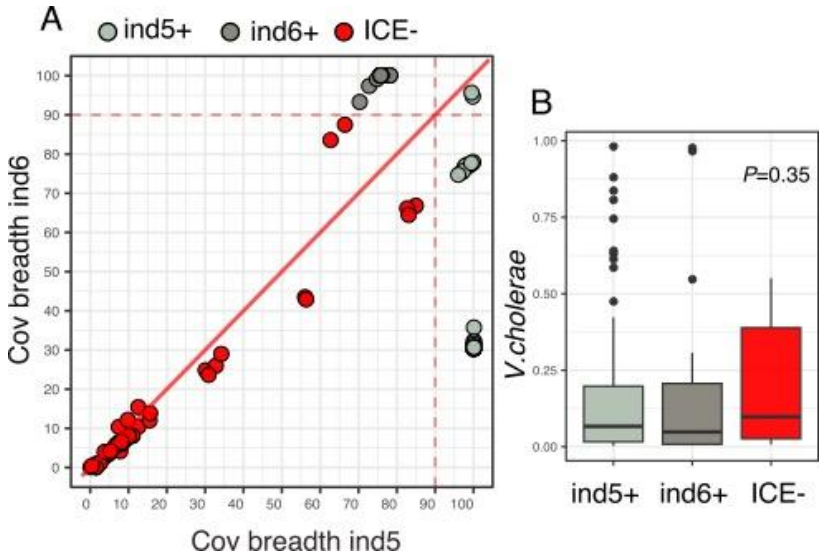


Fig S11. A) Breadth of metagenomic read coverage of the ICEs ind5 and ind6. Samples were identified as ind5-positive or ind6-positive based on read mapping against the reference ICE sequences. The ICE was considered as present when 90% of the reference ICE length was covered by at least one read. Only 2 samples had 100% ind5 and >90% ind6, which were considered ind5-positive. B) Boxplot showing that ICE-negative samples are not associated with lower relative abundance of *V. cholerae*, suggesting that the lack of detecting ICEs is not due to low *Vc* abundance. The *P*-value is from a Kruskal-Wallis test.

FIGURE S12

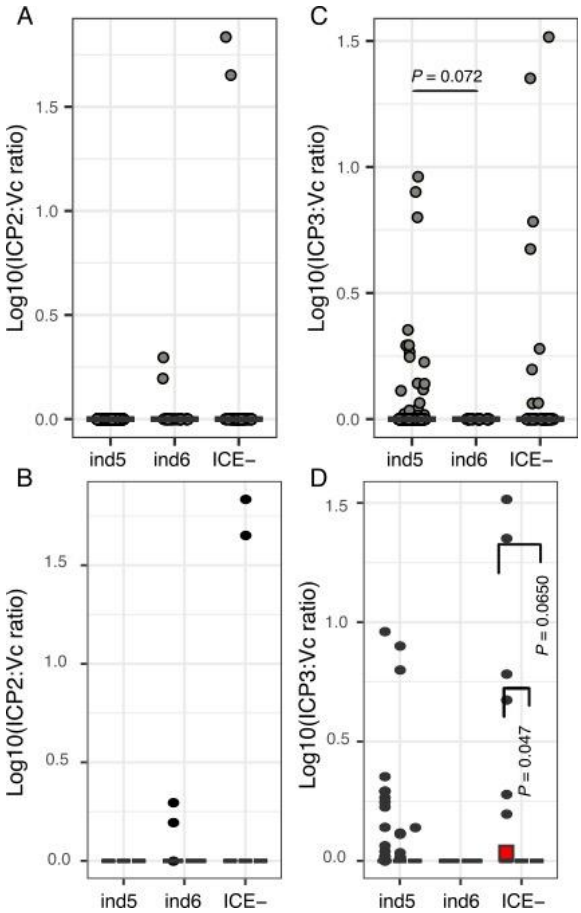


Fig S12. (A) Distribution of ICP2:*Vc* ratios across patients with different ICE profiles. (B) The same data as (A) binned into boxplots according to dehydration status. (C) Distribution of ICP3:*Vc* ratios across patients with different ICE profiles, (D) The same data as (C) binned into boxplots according to dehydration status. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test adjusted with the Benjamini-Hochberg (BH) method. Only *P*-values < 0.1 are shown. Only samples with sufficient *Vc* or ICP were included. For clarity, the Y-axes were log10 transformed after adding one to the ratios. The solid horizontal line is the median and the boxed area is the interquartile range.

FIGURE S13

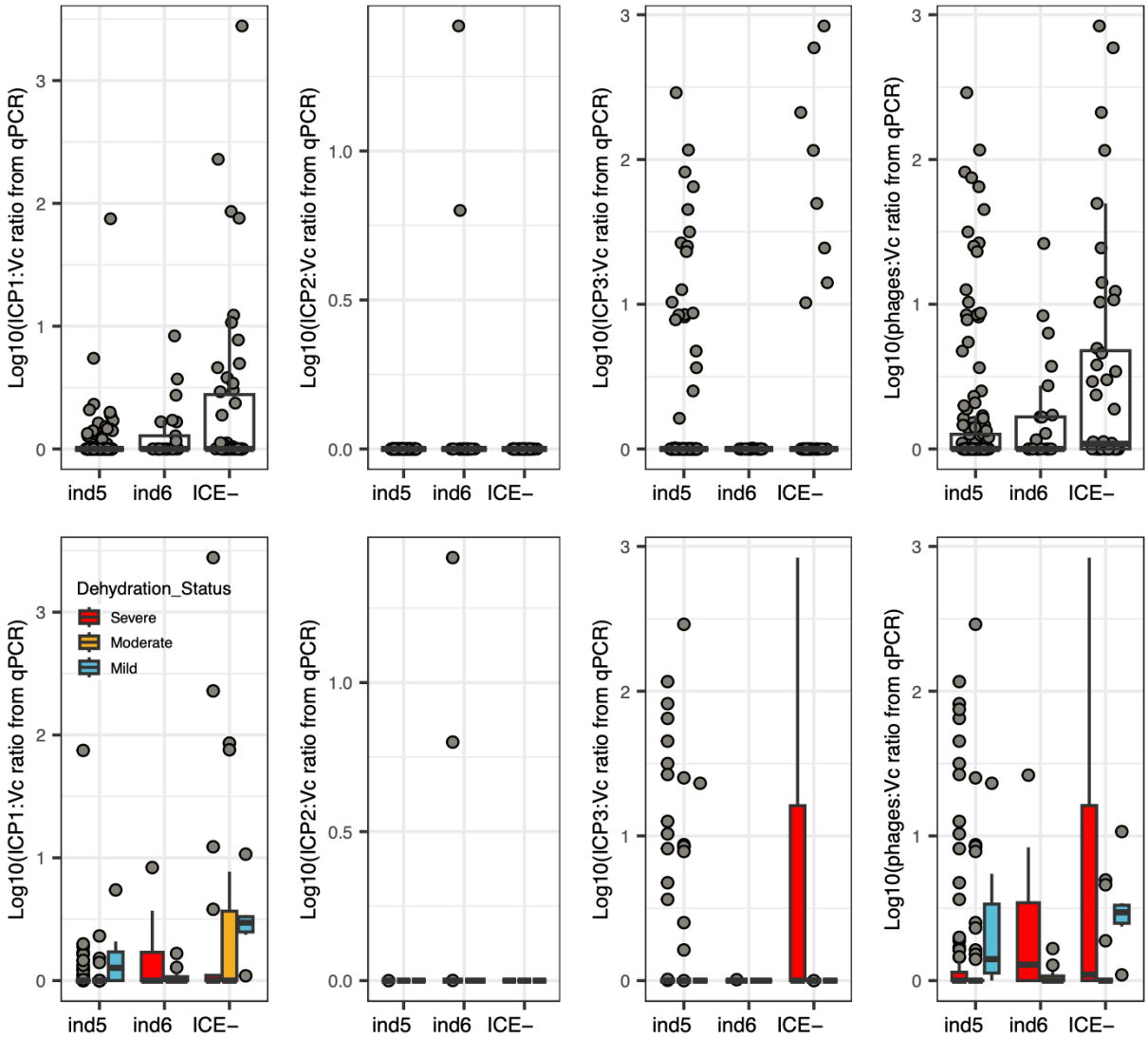


Fig S13. Integrative conjugative elements (ICEs) are associated with lower ICP1:*V. cholerae* ratios using qPCR data. (A) Distribution of ICP1:*Vc* ratios across patients with different ICE profiles. (B) The same data as (A) binned into boxplots according to dehydration status. (C) Distribution of ICP2:*Vc* ratios. (D) The same data as (C) binned into boxplots according to dehydration status. (E) Distribution of ICP3:*Vc* ratios, (F) The same data as (E) binned into boxplots according to dehydration status. (G) Distribution of phage:*Vc* ratios, including the sum of all phages (ICP1, ICP2, ICP3). (H) The same data as (G) binned into boxplots according to dehydration status. Only samples with appreciable *Vc* or ICP1 were included (215 samples with *Vc*>0 or phages >0 from qPCR data), of which 50 samples were ICE-, 25 were ind6+ and 140 were ind5+. The Y-axes were log10 transformed after adding one to the ratios to improve readability. The solid horizontal line is the median and the boxed area is the interquartile range. These results support the findings based on metagenomic data shown in Figure 3 and S12 using qPCR data.

FIGURE S14

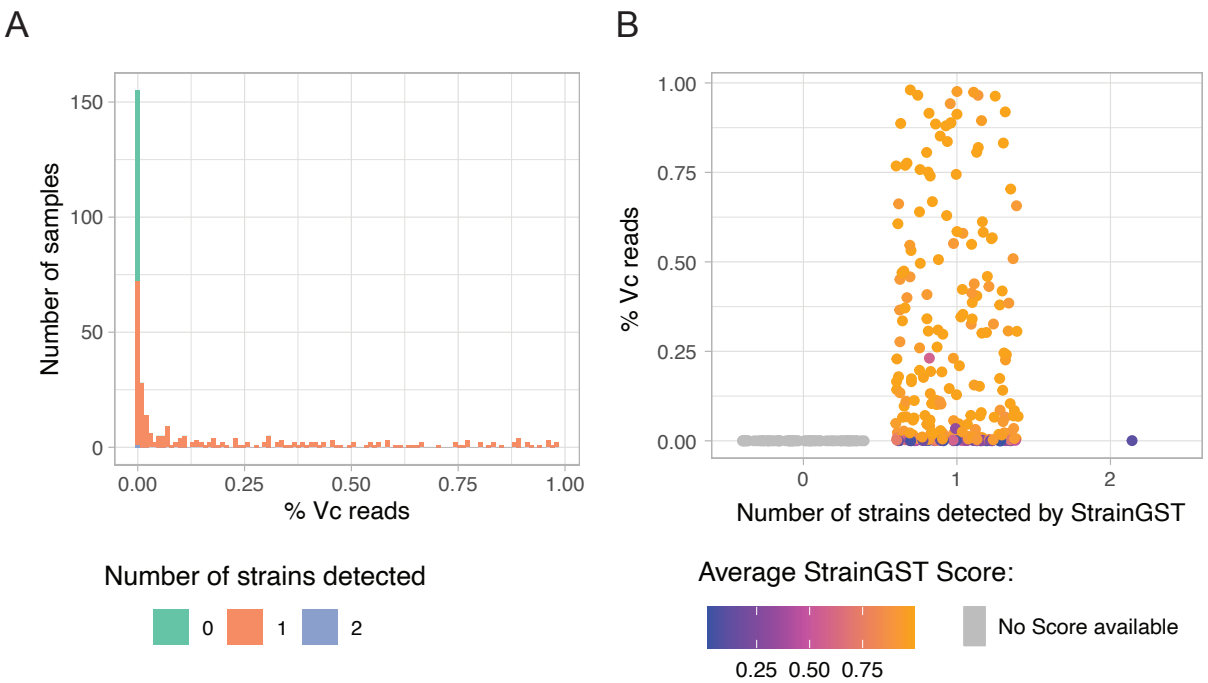


Fig S14. Mixed infections by more than on *V. cholerae* strain is unlikely in our patients. (A) The distribution of the number of distinct strains detected across samples. In 260/344 samples, strainGST identified only one reference strain. In 83/344 samples, no reference strain could be identified with confidence, likely due to low coverage of *V. cholerae* (these samples all had <1% *Vc* reads). (B) The relative abundance of *Vc* (% reads) in samples with zero, one, or two strains identified. In one sample, strainGST identified two reference strains, but with a low confidence score and at low *Vc* abundance.

FIGURE S15

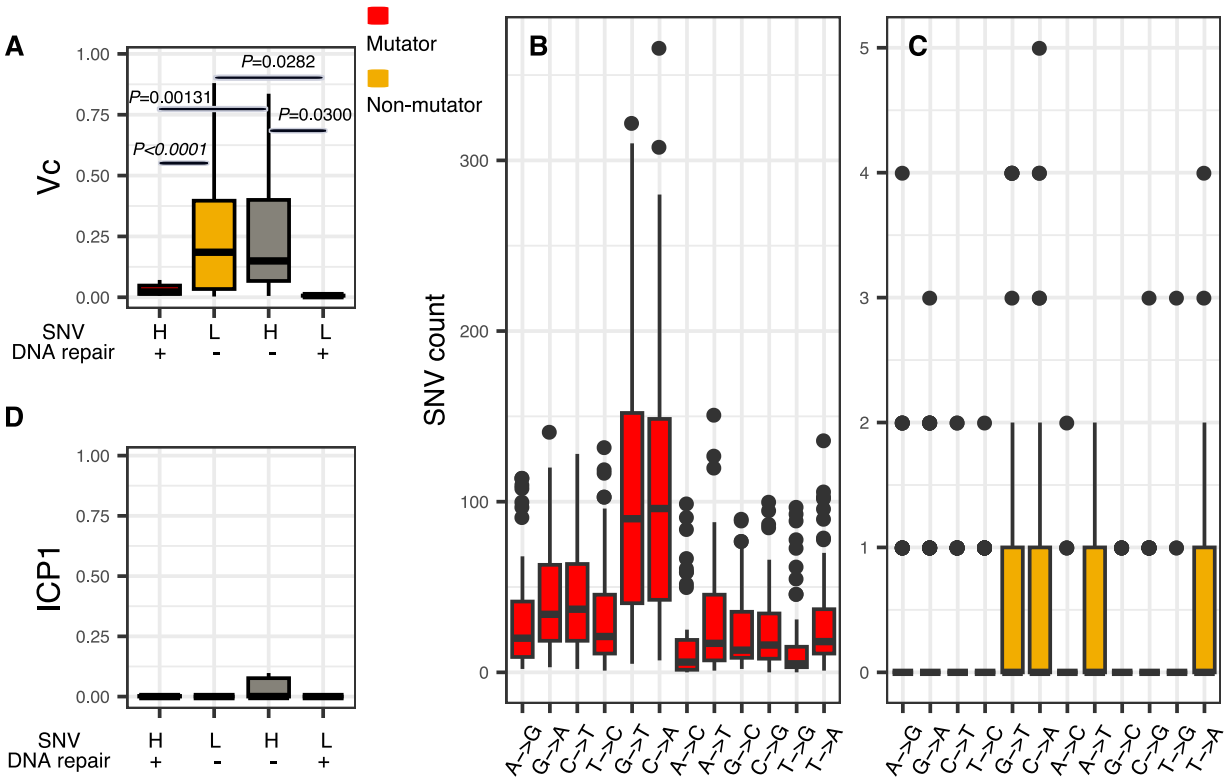


Fig S15. A) Low *V. cholerae* relative abundance is associated with DNA repair mutations independently of the number of SNVs. Mutators are defined as having a high (H) number of SNVs (25 or more) in the *V. cholerae* genome, along with one or more nonsynonymous mutations in a DNA repair gene (+) resulting in a predicted defect in DNA repair. Non-mutators have neither a high number of SNVs nor a DNA repair defect. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test adjusted with the Benjamini-Hochberg (BH) method. Only *P*-values < 0.1 are shown. (B and C) Transversion mutations, particularly G->T and C->A, are more common in mutators compared to non-mutators. D) Relative abundance of phage ICP1 in samples with different *V. cholerae* mutation profiles. Kruskall-wallis test (*P*= 0.07197). n=133, with 47 belonging to mutators, 70 to non-mutators, 14 to "high SNVs/no DNA repair defect" and 2 to "low SNVs/with DNA repair defect" groups.

FIGURE S16

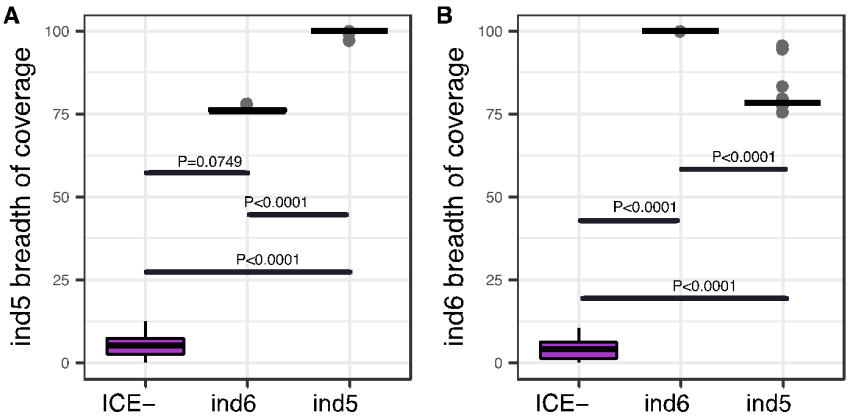


Fig S16. ICE- are readily distinguishable from ind5+ and ind6+ samples. Breadth of ICE coverage across patients with different ICE profiles. (A) Breadth of ind5 coverage. (B) Breadth of ind6 coverage. Note that the few samples with ambiguous breadth of ICE coverage (in the 40-80% range) were not included in the InStrain analysis, and are not included here. *P*-values are from a Kruskal-Wallis test with Dunn's post-hoc test adjusted with the Benjamini-Hochberg (BH) method. n=131 with 24 belonging to ICE-, 18 to ind6+, and 89 to ind5+ groups.

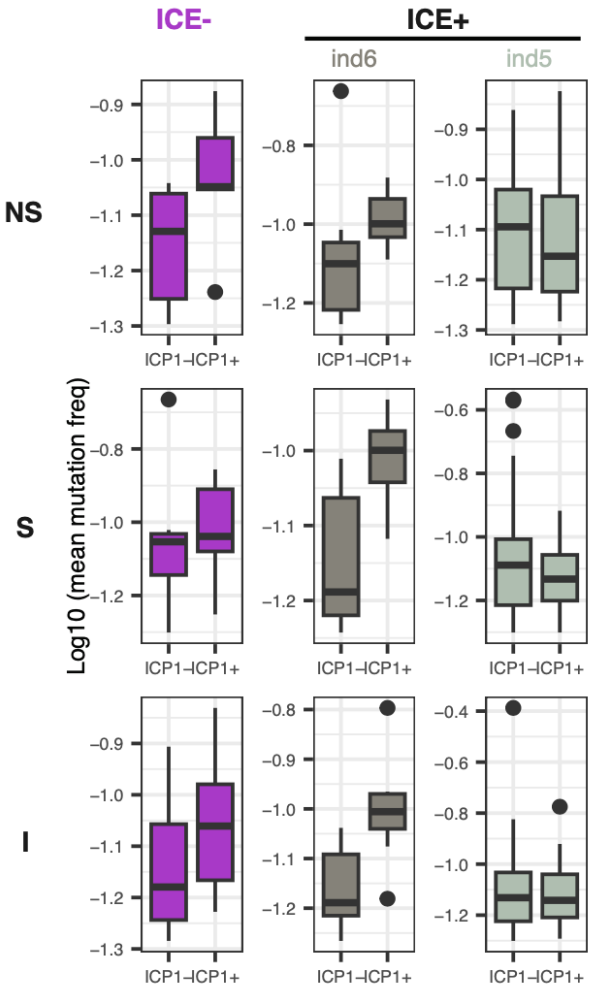


Fig S17. Boxplots of mutation frequency in the presence or absence of ICP1 and/or ICEs. Boxplots include 128 samples, of which 41 are ICP1+ (ICP1 detected by qPCR at a Ct cutoff of 28) and 87 are ICP- (Ct below 28). The Y-axes were log10 transformed after adding a pseudocount of 1 to improve readability. The solid horizontal line is the median and the boxed area is the interquartile range. This figure is a version of Fig 4C with qPCR data.

FIGURE S18

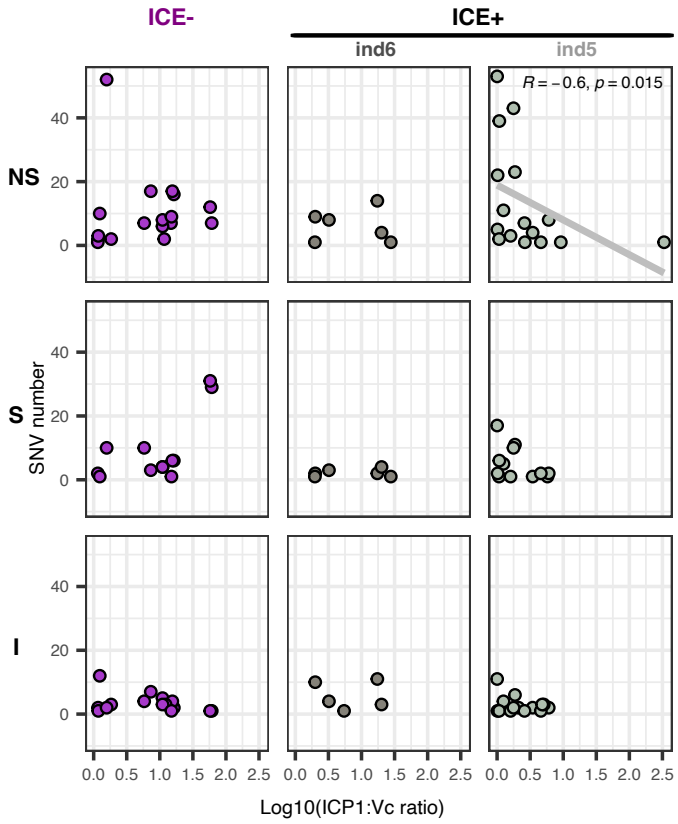


Fig S18. Number of SNVs in ICP1 as a function of the ICP1:*Vc* ratio and the ICE type encoded by *Vc*. Single nucleotide variant (SNV) count per sample in the ICP1 genome is plotted as a function of the ICP1:*Vc* ratio for different ICE profiles. All 45 samples for which mutations in ICP1 were called with inStrain were included in this analysis, of which 18 ICE-, 20 ind5 and 7 ind6. Non-synonymous (NS), synonymous (S), and intergenic (I) SNVs are plotted in separate rows. The only significant correlation ($p < 0.05$; top right) is shown with the trendline in gray and Spearman correlation coefficient and *p*-value.

FIGURE S19

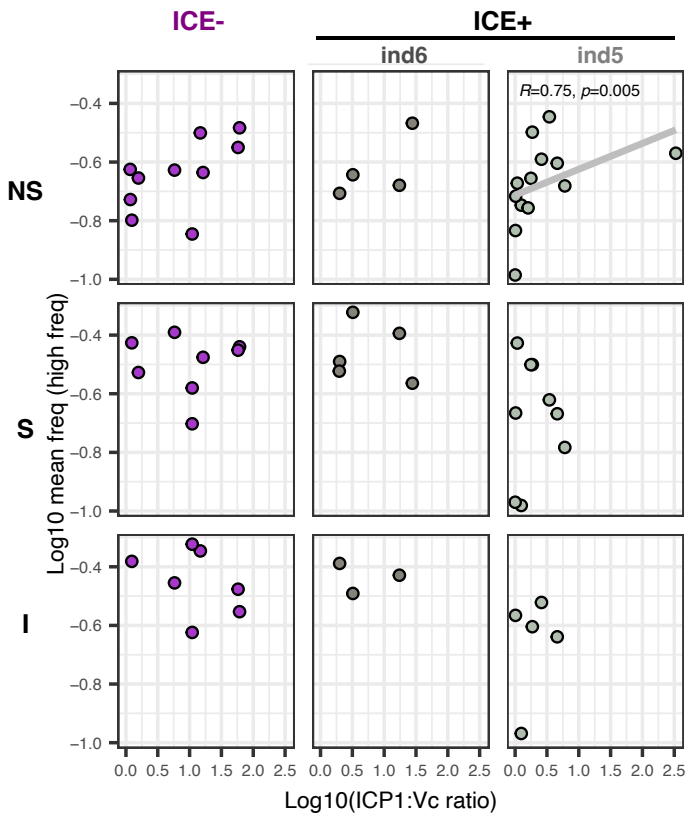


Fig S19. SNVs in ICP1 as a function of the ICP1:Vc ratio and the ICE type encoded by *Vc*. The mean single nucleotide variant (SNV) frequency per sample in the ICP1 genome is plotted as a function of the ICP1:*Vc* ratio for different ICE profiles. This analysis includes 29 samples for which high frequency mutations in ICP1 were called by inStrain, of which 11 ICE-, 13 *ind5* and 5 *ind6*. Non-synonymous (NS), synonymous (S), and intergenic (I) SNVs are plotted in separate rows. The only significant correlation ($p < 0.05$; top right) is shown with the trendline in gray and Spearman correlation coefficient and *p*-value. Only SNV frequencies >0.1 are included.

SUPPLEMENTARY TABLES

Table S1. Primers used for PCR and qPCR.

Target	Primer Name	Sequence 5'-3'	Reference
ompW *(Vibrio cholerae)*	ompW_F ompW_R	CACCAAGAAGGTGACTTTAT GAACTTATAACCACCCGCG	(*31*)
tcpA *(Vibrio cholerae)*	*tcpA set2*_F	ACACGATAAGAAAACCGGTCA	(*4*)
	*tcpA set2*_R	GCCTTGGTCATATTCTGCGA	
ICP1	gp58.2_F	CAAAGGCAGCAGGTAGGACA	This study
	gp58.2_R	CCCTTCAAGCCGTAGTTGGT	
ICP2	gp24_F	AGAAGTCGCAAACGGGGTAC	This study
	gp24_R	AACGTGGTTCTCGTGAGTGG	
ICP3	gp19_F	AGACCAACGCCGACTGTTAG	This Study
	gp19_R	CGATACCACGGAAAGCCTGT	
16S rDNA1	Maeda_1048_1067_F	GTGSTGCAYGGYTGTCGTCA	(*6*)
	Maeda_1175_1194_R A	ACGTCRTCCMCACCTTCCTC	

Table S2. LC MS/MS targets and parameters.

Antibiotic	RT (min)	Extraction Ions
Azithromycin	11.4 – 13.1	750,591
Ciprofloxacin	9.5 – 11.2	332
Doxycycline	8.9 – 14.0	445,428
Nalidixic Acid	12.3 – 14.4	233
Azithromycin-d5	11.4 – 13.1	754,595
Ciprofloxacin-d8	9.5 – 11.2	339
Doxycycline-d5	8.9 – 14.0	450,433
Metronidazole	3.5-5.5	172

Table S3. Indicator species analysis. For each group, we report the indicator value (stat) ranging between 0-1, with 1 being the perfect indicator species meaning that the species occurs exclusively in one group. P-values are from a permutation test with 9999 iterations as implemented in the indicator species function in R (multipatt from indicspecies library). The analysis was run on all the data (344 samples) and the 37 most dominant species from the PCA in Figure S5, and 24 species were selected as indicators of the patient's dehydration status.

Group	stat	*p*
Group 1: mild dehydration		
1. Bifidobacterium longum	0.372	**1e-05**
2. Bifidobacterium breve	0.304	**6e-05**
3. Escherichia coli	0.266	**0.00057**
4. Enterococcus aecium	0.214	0.00437
5. Bifidobacterium kashiwanohense	0.2	0.00818
6. Salmonella enterica	0.183	0.01262
7. ICP1	0.167	**0.02641**
Group 2: moderate dehydration		
1. Streptococcus pasteurianus	0.23	0.00272
2. Roseburia intestinalis	0.21	0.00596
3. Streptococcus gallolyticus	0.16	0.03152
Group 3: severe dehydration		
1. Vibrio cholerae	0.333	**6e-05**
2. Fusobacterium mortiferum	0.31	0.00011
3. Faecalibacterium prausnitzii	0.291	0.00020
4. Prevotella ruminicola	0.284	0.00028
5. Prevotella dentalis	0.281	0.00030
6. Brachyspira pilosicoli	0.249	0.00084
7. Collinsella aerofaciens	0.246	0.00149
8. Bacteroides thetaiotaomicron	0.234	0.00221
9. Bacteroides vulgatus	0.226	**0.00309**
10. Prevotella intermedia	0.209	0.00560
11. Bacteroides fragilis	0.201	0.00823
12. Roseburia[Eubacterium] rectale	0.194	0.01063
13. ICP3	0.189	**0.01050**
14. Roseburia lachnospiraceae bacterium GAM79	0.168	0.02409

Table S4. Variables used in the Redundancy analysis (RDA) showing relationships among the seven most dominant bacterial species and explanatory variables: phages, patient metadata, and antibiotic concentration. First column: explanatory variables (CIP:, Ciprofloxacin, DOX:Doxycycline, and AZI: Azithromycin), second column: p-values from a permutation test with 999 iterations.

Variable	*p*[a]
CIP	0.073
DOX	0.701
AZI	0.001***
ICP1	0.012*
ICP2	0.336
ICP3	0.104
Area_Code	0.001***
Dehydration status	0.001***
Collection date	0.003**
Vomiting	0.002**
Age_in_Years	0.006**

[a] Permutation test (anova function from the vegan R package).

Table S5. Antibiotic resistance thresholds. Minimal inhibitory concentrations (MICs) were established under aerobic or anaerobic conditions in (Creasy-Marrazzo et al. 2022). Concentrations are in units of µg/ml.

	Aerobic		Anaerobic
Ciprofloxacin	0.016		0.063
Azithromycin	1		8
Doxycycline	0.13		0.13

Table S6. Generalized additive models included in the model selection using metagenomic data. Generalized additive models included in the model selection. GAMs were fit with *Vc* relative abundance as a function of ICP1, antibiotics and their interactions. The selection was based on ∆AIC. We report ∆AIC, R syntax for the formula, the predictors (fixed effects) and the corresponding P-values (Chi-square test), as well as the *P*-value corresponding to dehydration random effect (RE) and the adjusted r-squared of the model. nt: not tested.

model	∆AIC	Formula	ICP1	AZI	ICP1*AZI	Dehyd (RE)	R2
Gam1 (selected)	0	Vc ~ s(ICP1) + s(AZI) + te(ICP1, AZI) + s(Dehydration_Status, bs = "re")	0.614	0.002	0.026	3.66e-07	0.031
Gam2	1.3	Vc ~ s(ICP1) + s(AZI) + s(Dehydration_Status, bs = "re")	0.077	0.004	Not tested	2.76e-07	0.035
Gam3	3.9	Vc ~ s(ICP1) + s(AZI) + s(CIP) + s(DOX) + te(ICP1, AZI) + te(ICP1, CIP) + te(ICP1, DOX) +s(Dehydration_Status, bs = "re")	0.606	0.002	0.025	4.08e-07	0.027
Gam4	5.2	Vc ~ s(ICP1) + s(AZI) + s(CIP) + s(DOX) + s(Dehydration_Status, bs = "re")	0.078	0.003	nt	3.06e-07	0.032
Gam5	11.7	Vc ~ s(ICP1) + s(DOX) + te(ICP1, DOX) + s(Dehydration_Status, bs = "re")	0.059	nt	nt	3.65e-08	0.035
Gam6	11.9	Vc ~ s(ICP1) + s(CIP) + te(ICP1, CIP) + s(Dehydration_Status, bs = "re")	0.664	nt	4.35e-08	4.35e-08	0.022
Gam7	12.4	Vc ~ s(ICP1) + s(AZI) + te(ICP1, AZI, by = Dehydration_Status)					
Gam1(B)[a]	29.3	Vc ~ s(ICP1) + s(AZI) + te(ICP1, AZI)	0.914	0.003	0.015	nt	-0.017

[a] Without dehydration RE

Table S7. Generalized additive models included in the model selection using qPCR data. Generalized additive models included in the model selection. GAMs were fit with *Vc* as a function of ICP1 (absolute abundances from qPCR with CT=28), antibiotics and their interactions, the same methods as for the models fit with metagenomics data were used. The selection was based on ∆AIC. We report ∆AIC, R syntax for the formula, the predictors (fixed effects) and the corresponding P-values (Chi-square test), as well as the *P*-value corresponding to dehydration random effect (RE) and the adjusted r-squared of the model. nt: not tested.

model	∆AIC	formula	ICP1	AZI	ICP1*AZI	Dehyd(RE)	R2
Gam3	0	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + s(CIP) + s(DOX) + te(qPCR_ICP1_PFUml_CT28, AZI) +te(qPCR_ICP1_PFUml_CT28, CIP) + te(qPCR_ICP1_PFUml_CT28, DOX) + s(Dehydration_Status,bs = "re")	0.622	0.001	0.017	< 2e-16	0.172
Gam4	0.4	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + s(CIP) + s(DOX) + s(Dehydration_Status, bs = "re")	0.023	0.001	nt	< 2e-16	0.172
Gam1	0.9	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + te(qPCR_ICP1_PFUml_CT28, AZI) + s(Dehydration_Status, bs = "re")	0.636	0.001	0.015	< 2e-16	0.163
Gam2	1.3	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + s(Dehydration_Status, bs = "re")	0.018	0.001	nt	< 2e-16	0.163
Gam7	5.3	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + te(qPCR_ICP1_PFUml_CT28, AZI, by = Dehydration_Status)	0.265	0.004	<0.003 *	nt	0.169
Gam5	11.8	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(DOX) + te(qPCR_ICP1_PFUml_CT28, DOX) + s(Dehydration_Status, bs = "re")	0.019	nt	nt	< 2e-16	0.136
Gam6	11.9	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(CIP) + te(qPCR_ICP1_PFUml_CT28, CIP) + s(Dehydration_Status, bs = "re")	0.786	nt	nt	< 2e-16	0.134
Gam8 (Gam1 without RE)	30.6	log_qpcr_vc ~ s(qPCR_ICP1_PFUml_CT28) + s(AZI) + te(qPCR_ICP1_PFUml_CT28, AZI)	0.393	0.0001	0.031	nt	0.079

*For all dehydration status levels

Table S8. Generalized linear mixed models. GLMMs with *Vc* SNV count as a function of phage relative abundance from metagenomics and antibiotics and their interaction. We report ∆AIC, the R syntax for the formula, the predictors and the corresponding P-values (Wald test). As an evaluation of the goodness of fit, we report the adjusted R-squared (R2) and *P*-value from the comparison of the models and null models (exactly the same model but with no fixed terms) (likelihood-ratio test, anova function from stats package in R). Antibiotics terms are not shown (not significant, *P*>0.05, GLMM, wald test). nt : not tested.

Model	**dAIC**	**Formula**	**Vc**	**ICP1**	**ICP1*Vc**	**R2**	**P/null model**
Mod1 (Selected)	0.0	SNV_nb ~ Vc + Vc:ICP1	0.002	nt	0.004	0.377	7e-04
Mod2	0.6	SNV_nb ~ ICP1 + Vc + Vc:ICP1	0.007	0.266	0.012	0.408	1.19e-03
Mod3	1.6	SNV_nb ~ ICP1 + Vc + AZI + Vc:ICP1	0.008	0.267	0.015	0.409	0.002
Mod4	3.2	SNV_nb ~ ICP1 + Vc + CIP + AZI + Vc:ICP1	0.007	0.273	0.015	0.412	0.004
Mod5	5.0	SNV_nb ~ Vc	0.002	nt	nt	0.228	0.006
Mod6	5.2	SNV_nb ~ ICP1 + Vc + CIP + DOX + AZI + Vc:ICP1	0.00745	0.27359	0.01554	0.412	0.008
Mod6	5.9	SNV_nb ~ ICP1 + Vc + CIP + DOX + AZI + ICP1:CIP + ICP1:AZI + Vc:ICP1	0.0428	0.1634	0.0385	0.458	0.008
Mod7	7.0	SNV_nb ~ ICP1 + Vc + CIP + AZI + DOX + Vc:ICP1:CIP + Vc:ICP1:AZI + Vc:ICP1:DOX	0.015	0.732	nt	0.442	0.012
Mod8	7.1	SNV_nb ~ ICP1 + Vc + CIP + DOX + AZI + ICP1:AZI + Vc:ICP1	0.047	0.275	0.059	0.422	0.015
Mod9	12.1	SNV_nb ~ ICP1	nt	0.506	nt	0.018	0.527

Table S9. Generalized additive models. GAMs used to select the most parsimonious model. The response is the average frequency of NS mutations in *V. cholerae* and the predictors are ICP1 (relative abundance from metagenomics), antibiotics, ICE presence/absence and mutation type (NS: non-synonymous, S:synonymous and I:intergenic) as well as their interactions. We defined a variable (ICE.by.mut) as the combination between ICE and mutation type with 9 levels (mutation type*ICE) to represent the interaction between mutation type and ICE factors. Model selection was based on ∆AIC. We report ∆AIC, the R syntax for the formula, the predictors and the corresponding P-values (Chi-square test) and the adjusted r-squared of the model. nt: not tested.

	∆AIC	**formula**	***P* (ICP1*ICE*mut)**			***P* (antbx *mut)**	**R2**
			Ind5*NS	Ind6*NS	ICE-*NS		
Mod1	0.0	mean_freq ~ s(ICP1, by = ICE.by.mut)	0.1336	0.642	0.0469	nt	0.021
Mod2	5.8	mean_freq ~ s(ICP1, by = ICE.by.mut) + s(AZI, by = mutation_type) + s(CIP, by = mutation_type) + DOX * mutation_type	0.13242	0.58891	0.04677	>0.05	0.060
Mod3	9.0	mean_freq ~ s(ICP1, by = ICE.by.mut) + s(AZI, by = mutation_type) + s(CIP, by = mutation_type) + DOX * mutation_type + te(ICP1, AZI, by = mutation_type) + te(ICP1, CIP, by = mutation_type) + DOX:ICP1:mutation_type	<2e-16	<2e-16	<2e-16	<2e-16 (AZI)	0.067
Mod4	10.9	mean_freq ~ s(ICP1, by = ICE.by.mut) + s(AZI, by = mutation_type) + s(CIP, by = mutation_type)	0.1223	0.6346	0.0496	>0.05	0.002

Table S10. Most frequently mutated *Vc* genes in samples with high ICP1:*Vc* ratios.

Mutation count denotes the total number of SNVs and patient count denotes the number of patients in which a mutation was observed. Genes were identified with prodigal v2.6.3 and annotated with eggNOG-Mapper v2 as well as manually with blastx against NCBI database within the Basic Local Alignment Search Tool (BLAST) online tool. Genes that contain mutations in samples with ICP1>*Vc* exclusively are shown in bold. Only the top 15 genes are shown, See Data File S3 for complete list.

COG (eggnog)	PFAM (eggnog) annotation	NCBI annotation	Mutation count	Patient count
Q	HemolysinCabind_Peptidase_M10_C	retention module-containing protein WP_001191814.1	5	2
M	**Peptidase_S13**	**D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase (genbank : AAF93798.1)**	**4**	**1**
V	ACR_tran	multidrug resistance protein, putative AAF93795.1	4	1
O	PPC_Peptidase_M9_Peptidase_M9_N	TPA: collagenase HAS4622795.1	4	1
P	BPD_transp_1	ABC transporter permease subunit WP_000252168.1	4	1
K	**MerR_1**	**MerR family transcriptional regulator WP_000226962.1**	**3**	**1**
C	CCG_FADoxidase_C_FAD_binding_4_Fer4_7_Fer4_8	FAD-binding and (Fe-S)-binding domain-containing protein WP_000188699.1	3	1
C	**Gp_dh_C_Gp_dh_N**	**glyceraldehyde 3-phosphate dehydrogenase GenBank: AAF96741.1**	**3**	**1**
S	DUF3302	DUF3302 domain-containing protein WP_000478180.1	2	1
IQ	PPbinding	Chain A, 3-oxoacyl-[acyl-carrier-protein] synthase 2 PDB: 4JRH_A	2	1
C	**Fer4_12_Radical_SAM**	**methyl-accepting chemotaxis protein WP_000383592.1**	**2**	**2**
Q	FtsX_MacB_PCD	ABC transporter permease WP_000645916.1	2	1
C	**CCG_Fer4_8**	**anaerobic glycerol-3-phosphate dehydrogenase subunit GlpC WP_001014995.1**	**2**	**1**
GM	CoA_binding_3_Polysacc_synt_2	nucleoside-diphosphate sugar epimerase/dehydratase WP_000494952.1	2	1
F	PpxGppA	guanosine-5'-triphosphate,3'-diphosphate diphosphatase WP_000076046.1	2	1

Table S11. Most frequently mutated *Vc* genes in samples with low ICP1:*V.c* ratios. Mutation count denotes the total number of SNVs and patient count denotes the number of patients in which a mutation was observed. Genes were identified with prodigal v2.6.3 and annotated with eggNOG-Mapper v2 as well as manually with blastx against NCBI database within the Basic Local Alignment Search Tool (BLAST) online tool. Genes that contain mutations in samples with ICP1<*Vc* exclusively are shown in bold. Only the top 15 genes are shown, see Data File S4 for complete list.

COG category (eggnog)	PFAM (eggnog) annotation	NCBI annotation	Mutation count	Patient count
O	PPC_Peptidase_M9_Peptidase_M9_N	TPA: collagenase HAS4622795.1	56	15
MQ	**ACD_ADPrib_exo_Tox_Anthrax_toxA_Hydrolase_4_MLD_Peptidase_C80_RtxA**	**TPA: MARTX multifunctional-autoprocessing repeats-in-toxin holotoxin RtxA HAS4620517.1**	**14**	**6**
PT	**GGDEF_Hemerythrin**	**GGDEF domain-containing protein WP_001190450.1**	**12**	**4**
T	**EAL_GGDEF**	**EAL domain-containing protein WP_000160066.1**	**11**	**2**
S	**Glyco_hydro_129**	**hypothetical protein VC_A0254 AAF96165.1**	**11**	**2**
S	**Betaprism_lec_Hemolysin_N_Leukocidin**	**Cytolysin vcc (ncbi) WP_001125271.1 (hlyA gene) [Levade et al 2021]**	**10**	**3**
C	FixO	cytochrome-c oxidase, cbb3-type subunit II WP_000097777.1	9	1
L	**Fapy_DNA_glyco_H2TH_zfFPG_IleRS**	**bifunctional DNA-formamidopyrimidine glycosylase/DNA-(apurinic or apyrimidinic site) lyase WP_001114647.1**	**9**	**3**
I	AcylCoA_dh_1_AcylCoA_dh_M_AcylCoA_dh_N_DUF1974	acyl-CoA dehydrogenase FadE WP_000404358.1	8	4
L	Phage_int_SAM_4	hypothetical protein WP_000222725.1	8	4
M	ACD_PG_binding_1_Pesticin_Phage_GPD	Type VI secretion system tip protein VgrG WP_000212125.1	7	1
K	LacI_Peripla_BP_3	DNA-binding transcriptional regulator CytR WP_000224452.1	7	1
T	EAL	EAL domain-containing protein WP_000547542.1	7	2
E	SDH_alpha_SDH_beta	sodium:solute symporter WP_001882525.1	7	4
S	Inovirus_Gp2	inovirus Gp2 family protein WP_000149629.1	7	4

Table S12. Most frequently mutated ICP1 genes in samples with ICE *ind5* detected. Mutation count denotes the total number SNVs (NS SNVs with frequency>0.1) and patient count denotes the number of patients in which a mutation was observed. Only samples with the ICE *ind5* are included (13 samples out of the 27 samples for which high frequency NS mutations were called in ICP1 with inStrain). Genes were annotated manually with blastx against NCBI database within the Basic Local Alignment Search Tool (BLAST) online tool. Most of the genes had no hit with eggNOG-Mapper. Only the top 15 genes are shown; see Data File S5 for the complete list (41 genes).

NCBI annotation	Mutation count	Patient count
hypothetical protein ViPhICP1_gp184 YP_004251125.1	33	3
hypothetical protein ViPhICP1_gp147 YP_004251088.1	5	2
anaerobic ribonucleotide reductase small subunit YP_004251055.1	4	3
hypothetical protein ViPhICP1_gp150 YP_004251091.1	4	2
acyl carrier protein YP_004251038.1	4	3
hypothetical protein ViPhICP1_gp226 YP_004251167.1	3	3
nicotinamide phosphoribosyl transferase YP_004251171.1	3	3
hypothetical protein ViPhICP1_gp139 YP_004251080.1	2	1
hypothetical protein ViPhICP1_gp166 YP_004251107.1	2	2
hypothetical protein ViPhICP1_gp018 YP_004250959.1	2	2
hypothetical protein ViPhICP1_gp175 YP_004251116.1	2	1
hypothetical protein ViPhICP1_gp199 YP_004251140.1	2	2
ribonucleoside diphosphate reductase, beta chain YP_004251147.1	2	2
hypothetical protein ViPhICP1_gp068 YP_004251009.1	2	2
hypothetical protein ViPhICP1_gp070 YP_004251011.1	2	1

REFERENCES

1. A. I. Khan *et al.*, Electronic decision support and diarrhoeal disease guideline adherence (mHDM): a cluster randomised controlled trial. *Lancet Digit Health* **2**, e250-e258 (2020).
2. A. I. Khan, F. Qadri, Epidemiology of cholera in Bangladesh: Findings from Nationwide Hospital-based Surveillance, 2014-2018. *CID*, (2019).
3. A. Creasy-Marrazzo *et al.*, Genome-wide association studies reveal distinct genetic correlates and increased heritability of antimicrobial resistance in *Vibrio cholerae* under anaerobic conditions. *Microbial Genomics* **8**, (2022).
4. L. Alexandrova *et al.*, Identification of Widespread Antibiotic Exposure in Patients With Cholera Correlates With Clinically Relevant Microbiota Changes. *J Infect Dis* **220**, 1655-1666 (2019).
5. J. A. Grembi, K. Mayer-Blackwell, S. P. Luby, A. M. Spormann, High-Throughput Multiparallel Enteropathogen Detection via Nano-Liter qPCR. *Front Cell Infect Microbiol* **10**, 351 (2020).
6. H. Maeda *et al.*, Quantitative real-time PCR using TaqMan and SYBR Green for Actinobacillus actinomycetemcomitans, Porphyromonas gingivalis, Prevotella intermedia, tetQ gene and total bacteria. *FEMS Immunol Med Microbiol* **39**, 81-86 (2003).
7. K. E. Flaherty *et al.*, High-throughput low-cost nl-qPCR for enteropathogen detection: A proof-of-concept among hospitalized patients in Bangladesh. *PLoS One* **16**, e0257708 (2021).
8. S. Roux *et al.*, iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**, e3002083 (2023).

9. R. Cook *et al.*, INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *Phage (New Rochelle)* **2**, 214-223 (2021).

10. F. Hassan, M. Kamruzzaman, J. J. Mekalanos, S. M. Faruque, Satellite phage TLCphi enables toxigenic conversion by CTX phage through dif site alteration. *Nature* **467**, 982-985 (2010).

11. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489-491 (2013).

12. H. Bin Jang *et al.*, Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* **37**, 632-639 (2019).

13. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

14. A. E. Minoche, J. C. Dohm, H. Himmelbauer, Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**, R112 (2011).

15. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).

16. J. Alneberg *et al.*, Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**, 1144-1146 (2014).

17. D. D. Kang *et al.*, MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

18. C. M. K. Sieber *et al.*, Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).

19. M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868 (2017).

20. M. R. Olm, A. Crits-Christoph, K. Bouma-Gregson, B. A. Firek, M. J. Morowitz, J. F. Banfield, inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727-736 (2021).

21. D. Hyatt, G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

22. C. P. Cantalapiedra, A. Hernandez-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825-5829 (2021).

23. L. R. van Dijk *et al.*, StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. *Genome Biol* **23**, 74 (2022).

24. A. Jolivet-Gougeon *et al.*, Bacterial hypermutation: clinical implications. *J Med Microbiol* **60**, 563-573 (2011).

25. G. Arango-Argoty, E. Garner, A. Pruden, L. S. Heath, P. Vikesland, L. Zhang, DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).

26. K. LeGault *et al.*, Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* **373**, 2020.2012.2016.423150 (2021).

27. M. Dufrene, Pireere Legendre, Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach. *Ecological Monographs* **67**, 345-366 (1997).

28. P. J., Analyse Factorielle Multiple Appliquée Aux Variables Qualitatives et Aux Données Mixtes. *Revue de Statistique Appliquee* **4**, 5-37 (2002).

29. M. E. Brooks *et al.*, Modeling zero-inflated count data with glmmTMB. *bioRxiv*, 132753 (2017).

30. J. W. Hardin, J. M. Hilbe, *Generalized Linear Models and Extensions*. (Stata Press), vol. Vol. 4th ed.

31. B. Nandi, R. K. Nandy, S. Mukhopadhyay, G. B. Nair, T. Shimada, A. C. Ghose, Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. *J Clin Microbiol* **38**, 4145-4151 (2000).

DATA FILES

File S1. Patient metadata.

File S2. Antibiotic concentrations in stool samples.

File S3. Full list of *Vc* genes containing mutations in samples where ICP1 > *Vc* (% of reads).

File S4. Full list of *Vc* genes containing mutations in samples where ICP1 < *Vc* (% of reads).

File S5. Full list of ICP1 genes containing mutations in samples containing ICE *ind5*.