TITLE: A combination of metagenomic and cultivation approaches reveals hypermutator phenotypes within *Vibrio cholerae* infected patients.

AUTHORS: Inès Levade[1], Ashraful I. Khan[2], Fahima Chowdhury[2], Stephen B. Calderwood[3,4,5], Edward T. Ryan[3,4], Jason B. Harris[3,6], Regina C. LaRocque[3,4,], Taufiqur R. Bhuiyan[2], Firdausi Qadri[2], Ana A. Weil[7], B. Jesse Shapiro[1,8,9*]

[1]Department of Biological Sciences, University of Montreal, Montreal, Quebec, Canada.

[2]Center for Vaccine Sciences, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

[3]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA

[4]Department of Medicine, Harvard Medical School, Boston, MA USA

[5]Department of Microbiology, Harvard Medical School, Boston, MA USA

[6]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

[7]Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA, USA

[8]Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada

[9]McGill Genome Centre, Montreal, QC, Canada

CORRESPONDING AUTHOR: jesse.shapiro@mcgill.ca

ABSTRACT

Vibrio cholerae can cause a range of symptoms in infected patients, ranging from severe diarrhea to asymptomatic infection. Previous studies using whole genome sequencing (WGS) of multiple bacterial isolates per patient have shown that *Vibrio cholerae* can evolve a modest amount of genetic diversity during symptomatic infection. Little is known about *V. cholerae* genetic diversity within asymptomatic infected patients. To achieve increased resolution in the detection of *Vibrio cholerae* diversity within individual infections, we applied culture-based population genomics and metagenomics to a cohort of symptomatic and asymptomatic cholera patients. While the metagenomic approach allowed us to detect more mutations in symptomatic patients compared to the culture-based approach, WGS of isolates was still necessary to detect *V. cholerae* diversity in asymptomatic carriers, likely due to their low *Vibrio cholerae* load. We found that symptomatic and asymptomatic patients contain similar levels of within-patient diversity, and discovered *V. cholerae* hypermutators in some patients. While hypermutators appeared to generate mostly selectively neutral mutations, non-mutators showed signs of convergent mutation across multiple patients, suggesting *V. cholerae* adaptation within hosts. Our results highlight the power of metagenomics combined with isolate sequencing to characterize within-patient diversity in acute *V. cholerae* infection and asymptomatic infection, while providing evidence for hypermutator phenotypes within cholera patients.

IMPORTANCE

Pathogen evolution within patients can impact phenotypes such as drug resistance and virulence, potentially affecting clinical outcomes. *V. cholerae* infection can result in life-threatening diarrheal disease, or asymptomatic infection. Here we describe whole-genome sequencing of *V. cholerae* isolates and culture-free metagenomic sequencing from stool of symptomatic cholera patients and asymptomatic carriers. Despite the acuteness of cholera infections, we found evidence for adaptive mutations in the *V. cholerae* genome that occur independently and repeatedly within multiple symptomatic patients. We also identified *V. cholerae* hypermutator phenotypes within 6 out of 14 patients, which appear to generate mainly neutral or deleterious mutations. Our work sets the stage for future studies of the role of hypermutators and within-patient evolution in explaining the variation from asymptomatic carriage to symptomatic cholera.

KEYWORDS: *Vibrio cholerae,* cholera, metagenomics, within-patient evolution, hypermutation, asymptomatic carriage

55 **INTRODUCTION**

56 Infection with *Vibrio cholerae,* the etiological agent of cholera, causes a clinical spectrum
57 of symptoms that range from asymptomatic colonization of the intestine to severe watery diarrhea
58 that can lead to death. Although absent from most resource-rich countries, this severe diarrheal
59 disease still plagues many developing nations. According to the WHO, there are an estimated 1.3
60 to 4.0 million cases of cholera each year, with 21,000 to 143,000 deaths worldwide (Ali et al.
61 2015). Cholera predominantly occurs in endemic areas, but can also cause explosive outbreaks as
62 seen in Haiti in 2010 or in Yemen, where over 2.2 million cases are suspected since 2016 (Weil,
63 Ivers, et Harris 2011; Camacho et al. 2018). Although cholera vaccines have reduced disease in
64 some areas, the increasing number of people lacking access to sanitation and safe drinking water,
65 the emergence of a pandemic lineage of *V. cholerae* with increased virulence (Satchell et al.
66 2016), and environmental persistence of this waterborne pathogen underscore the need to
67 understand and interrupt transmission of this disease.

68 Cholera epidemiology and evolutionary dynamics have been studied by high-throughput
69 sequencing technologies and new modeling approaches, at a global and local scale (Weil et Ryan
70 2018; Domman et al. 2018). Yet, many questions remain regarding asymptomatic carriers of *V.*
71 *cholerae*, including their role and importance in the transmission chain during an epidemic (King
72 et al. 2008; Phelps, Simonsen, et Jensen 2019). Numerous observational studies have identified
73 host factors that could impact the severity of symptoms, including lack of pre-existing immunity,
74 blood group O status, age, polymorphisms in genes of the innate immune system, or variation in
75 the gut microbiome (Harris et al. 2005; 2008; Weil, Khan, et al. 2009; Midani et al. 2018; Levade
76 et al. 2020).

77 Recent studies have shown that despite the acute nature of cholera infection, which
78 typically lasts only a few days, genetic diversity can appear and be detected in a *V. cholerae*

79 population infecting individual patients (Seed et al. 2014; Levade et al. 2017). In a previous
80 study, we sampled multiple *V. cholerae* isolates from each of eight patients (five from
81 Bangladesh and three from Haiti) and sequenced 122 bacterial genomes in total. Using stringent
82 controls to guard against sequencing errors, we detected a few (0-3 per patient) within-patient
83 intra-host single nucleotide variants (iSNVs), and a greater number of gene content variants (gene
84 gain/loss events within patients) (Levade et al. 2017). This variation may affect adaptation to the
85 host environment, either by resistance to phage predation (Seed et al. 2014) or by impacting
86 biofilm formation (Levade et al. 2017), but it is not known how within-patient diversity affects
87 disease severity.

88 Several pathogens are known to evolve within human hosts (Didelot et al. 2016), and
89 hypermutation has been observed in some cases (Jolivet-Gougeon et al. 2011; Lieberman et al.
90 2013; Marvig et al. 2013). Hypermutation is a phenotype whereby a strain loses the function of
91 its mismatch repair machinery and thus becomes a hypermutator. While these hypermutators may
92 quickly acquire adaptive mutations, they also bear a burden of deleterious mutations (Giraud et
93 al. 2001). For the population to survive the burden of deleterious mutations, hypermutators may
94 revert to a non-mutator state, or may recombine their adaptive alleles into the genomes of non-
95 mutators in the population (Denamur et al. 2000; Jolivet-Gougeon et al. 2011). The hypermutator
96 phenotype has been observed in vibrios in the aquatic environment (Chu et al. 2017), and induced
97 in *V. cholerae* in an experimental setting (Wang et al. 2018), but not clearly documented within
98 infected patients. There is some evidence for hypermutation in *V. cholerae* clinical strains
99 isolated between 1961 and 1965 (Didelot et al. 2015); however, the authors recognized that these
100 hypermutators could also have emerged during long-term culture (Eisenstark 2010). It therefore
101 remains unclear if hypermutators readily emerge within cholera patients, nor their consequences
102 for disease outcomes and transmission.

103 When within-patient pathogen populations are studied with culture-based methods, their
104 diversity may be underestimated because the culture process can select isolates more suited to
105 growth in culture, and due to undersampling of rare variants. In this study, we used a combination
106 of culture-free metagenomics and WGS of cultured isolates to characterize the within-patient
107 diversity of *V. cholerae* in individuals with different clinical syndromes ranging from
108 symptomatic to asymptomatic infection. We found that previous culture-based analyses likely
109 underestimated the true variation within infected hosts. However, asymptomatic persons yielded
110 too few metagenomic reads to assess within-patient variation, and this could only be accessed
111 using cultured *V. cholerae* isolates. Using this culture-based approach to compare symptomatic to
112 asymptomatic contacts from three households, we found similar levels of within-patient diversity
113 regardless of disease severity. Using both approaches, we also describe the presence of
114 hypermutator *V. cholerae* within both symptomatic and asymptomatic infected patients. These
115 hypermutators are characterized by a high mutation rate, and accumulation of an excess of likely
116 neutral or deleterious mutations in the genome. Finally, we provide evidence of adaptive
117 mutations occurring during non-mutator *V. cholerae* infections.
118

119 **RESULTS**

120

121 **Taxonomic analyses of metagenomics sequences from *Vibrio cholerae* infected index cases**

122 **and household contacts.**

123 To evaluate the level of within-patient diversity of *Vibrio cholerae* populations infecting

124 symptomatic and asymptomatic patients in a cohort in Dhaka, Bangladesh, we used both culture-

125 based whole genome sequencing and culture-free shotgun metagenomic approaches (Fig. 1). We

126 performed metagenomic sequencing of 22 samples from 21 index cases and 11 samples from 10

127 household contacts infected with *Vibrio cholerae,* of which two remained asymptomatic during

128 the follow-up period (Table S1). After removal of reads mapping to the human genome, we used

129 Kraken2 and MIDAS to taxonomically classify the remaining reads and identify samples with

130 enough *Vibrio cholerae* reads to reconstruct genomes. Among symptomatic patients (index cases

131 and household contacts), 15 samples from 14 patients contained enough reads to reconstruct the

132 *Vibrio cholerae* genome with a mean coverage > 5X. Neither of the two asymptomatic patients

133 had enough *Vibrio cholerae* reads in their metagenomic sequences to reconstruct genomes by

134 mapping or *de novo* assembly (mean coverage <0.05X). We also detected reads from two *Vibrio*

135 phages (ICP1 and ICP3) in some of these samples (Table S1).

136

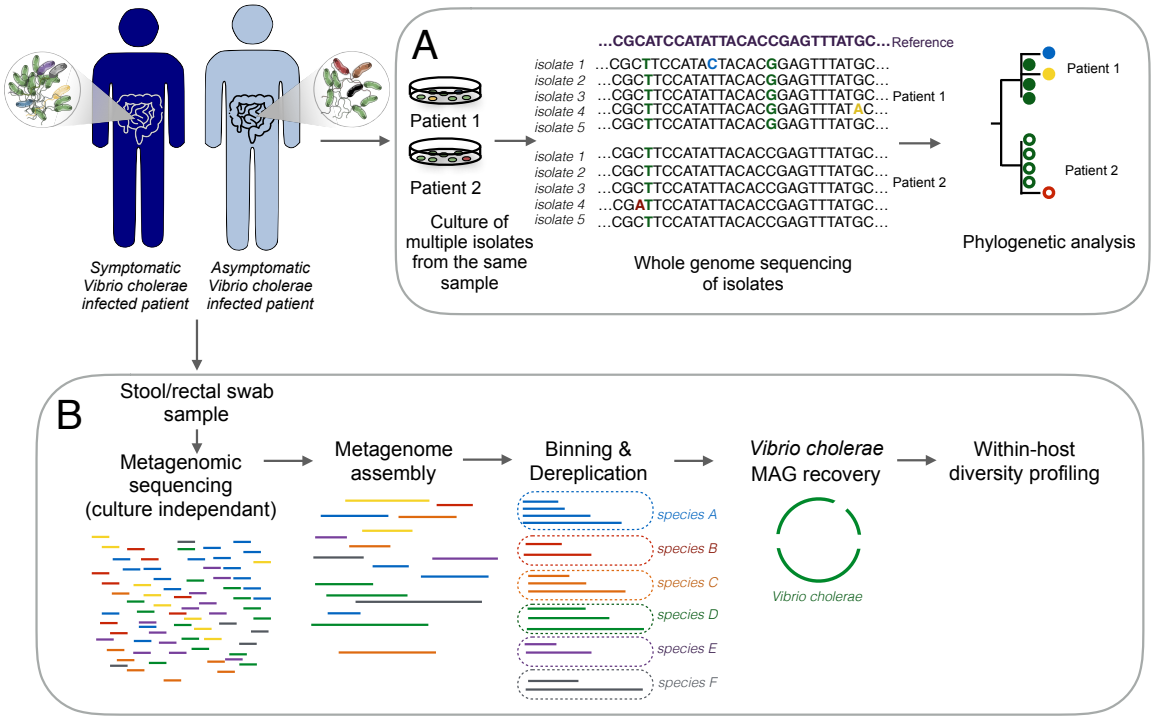


Figure 1. Summary of the culture-dependent and the culture-free metagenomics workflows for the characterization of the *Vibrio cholerae* within-patient diversity. Stool or rectal swab samples were collected from symptomatic and asymptomatic *Vibrio cholerae* infected individuals and processed using two different approaches: (A) Culture, DNA extraction and whole genome sequencing of multiple isolates per patient; (B) Genome-resolved metagenomics involves DNA extraction directly from a microbiome sample followed by DNA sequencing, assembly, genome binning and dereplication to generate metagenome-assembled genomes (MAGs), and within-host diversity profiling by mapping reads back to the MAGs.

Recovery of high quality *Vibrio cholerae* MAGs from metagenomic samples

To reconstruct *Vibrio cholerae* metagenomic assembled genomes (MAGs) from the 11 samples with coverage > 10X, we *de novo* assembled each sample individually except for patient

151 E, for whom we co-assembled two samples from two consecutive sampling days. High quality
152 MAGs identified as *Vibrio cholerae* were obtained from each assembly, with no redundancy, and
153 completeness ranging from 91 to 100% (Table S2). We dereplicated the set of bins and removed
154 all but the highest quality genome from each redundant set, identifying the bin from patient J as
155 the best quality MAG overall.
156

157 ***Vibrio cholerae* within patient nucleotide diversity estimated from metagenomic data**

158 All metagenomes with *Vibrio cholerae* mean coverage >5X were mapped against the
159 dereplicated genome set, and we assessed within-patient genetic diversity using inStrain (Olm et
160 al. 2020). We identified both single nucleotide polymorphisms (SNPs) that varied between
161 patients (Table S3), and intra patient single nucleotide variants (iSNVs) that varied within
162 patients (Table S4). We found a total of 39 SNPs between patients, and a range of two to 207
163 iSNVs within each sample (Table 1, Fig. 2). Given the wide variation in coverage across samples,
164 we checked for any bias toward detecting iSNVs in high-coverage samples. We observed no
165 correlation between the number of detected iSNVs and coverage values ($\rho = -0.12$, $P > 0.05$,
166 Pearson correlation), suggesting no coverage bias, and that diversity levels are comparable across
167 samples.
168

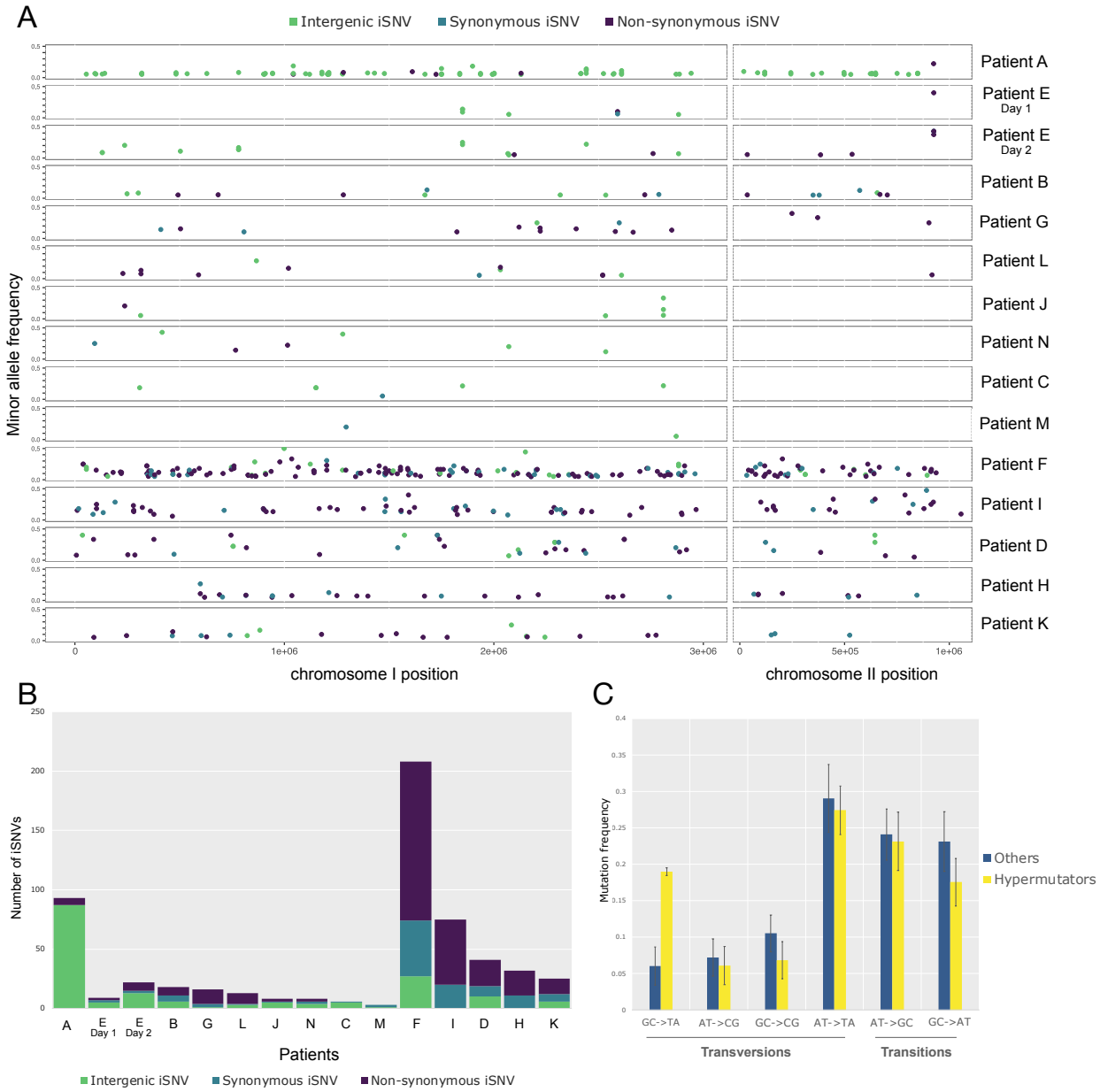


Figure 2. Within-patient *Vibrio cholerae* diversity quantified from metagenomic data. (A) Minor allele frequency and distribution of intergenic, synonymous and non-synonymous iSNVs across the two *Vibrio cholerae* chromosomes for 14 patients with over 5X coverage of the *V. cholerae* genome. (B) Numbers of intergenic, synonymous and non-synonymous iSNVs for each patient. (C) Spectrum of within-patient mutation in hypermutators compared to non-mutators. Error bars represent standard error of

175 the mean across the group of hypermutators or non-mutators. Only samples with 6 or more iSNVs were
176 included to reduce noise from low counts.
177
178 Several mechanisms could account for the origins of the observed iSNVs, including *de novo*
179 mutation within a patient, co-infection by divergent *V. cholerae* strains, or homologous
180 recombination. Most iSNVs had low-frequency minor alleles (Fig S1), consistent with recent
181 mutations occurring within individual patients, rather than co-infection by a roughly equal
182 mixture of distantly related strains. No iSNVs were observed at the exact same nucleotide
183 position in different patients, suggesting that iSNVs rarely spread by homologous recombination,
184 and are never precisely recurrent in our dataset. In patient E, sampled on two consecutive days,
185 we detected eight iSNVs on the first day, of which four were again detected on the second day,
186 along with 13 additional iSNVs. This suggests that iSNV allele frequencies can fluctuate
187 significantly over time. Moreover, iSNVs were distributed across the genome (Fig. 2A), rather
188 than clustered in hotspots, which would be expected if iSNVs arose from recombination events
189 (Croucher et al. 2011). Although we cannot strictly exclude co-infection or recombination events
190 as sources of diversity, most of the observed iSNVs are consistent with *de novo* mutation within
191 patients.
192
193 **Evidence for *V. cholerae* hypermutators within patients**
194 In 5 of the 6 patients with a high number of iSNVs (>25), we identified non-synonymous
195 (NS) mutations in genes involved in DNA mismatch repair pathways, including the DNA
196 polymerase II in patient D, or proteins of the methyl-directed mismatch repair (MMR) system in
197 patient F, I and K (Table 1). These NS variants could explain why these samples seem to have a
198 higher level of within-host diversity, and suggest the presence of *Vibrio cholerae* with a

199 hypermutator phenotype (Jolivet-Gougeon et al. 2011). In the patient harboring the highest
200 number of variants (Patient F, 207 iSNVs), we detected two NS mutations in two different genes
201 coding for proteins involved in DNA repair: the DNA mismatch repair endonuclease MutL
202 (Jolivet-Gougeon et al. 2011), and the nuclease SbcCD subunit C (Didelot et al. 2015)(Lovett
203 2011)(Darmon et al. 2007). The patient with the second highest number of iSNVs, patient A,
204 contained a high number of intergenic variants (87 out of 96 iSNVs, Fig. 2B), but no apparent NS
205 mutations in genes involved in DNA repair. This large number of intergenic iSNVs are unlikely
206 due to read mapping errors, since the same iSNV calls were obtained when using the MAG from
207 patient A as a reference genome. In patient I, where we also detected a high number of iSNVs, a
208 NS mutation in the gene coding for the MutT/nudix protein, involved in the repair of oxidative
209 DNA damage (Lu et al. 2001), could also cause a strong hypermutation phenotype. Patient D, H
210 and K presented fewer iSNVs but also showed NS mutations in genes involved in DNA damage
211 repair (Foster et al. 1995) (Lee, Sung, et Verdine 2019). However, some of these genes have been
212 shown to play less critical roles in bacterial DNA repair than MutSLH (Jolivet-Gougeon et al.
213 2011)(Kunkel et Erie 2005), which could lead to a weaker hypermutator phenotype.
214 Previous studies have noted mutational biases in hypermutators, such as an increase of
215 transition over transversion mutations in a *Burkholderia dolosa* mutator with a defective MutL
216 (Lieberman et al. 2013), or an excess of G:C→T:A transversions in a *Bacillus anthracis*
217 hypermutator (Zeibell et al. 2007), and in members of the gut microbiome (Zhao et al. 2019).
218 When we compared the spectrum of mutations observed in suspected hypermutators to non-
219 mutator samples, we found a significance difference (Chi-square test, $P<0.01$) due to an apparent
220 excess of G:C→T:A transversions in hypermutators (Fig. 2C).
221 Current theory suggests that hypermutators may be adaptive under novel or stressful
222 environmental conditions because they more rapidly explore the mutational space and are the first

223 to acquire adaptive mutations. However, hypermutation comes at the cost of the accumulation of
224 deleterious mutations. To test the hypothesis that hypermutation leads to fitness costs due to these
225 deleterious mutations, we used iRep (Brown et al. 2016) to estimate *V. cholerae* replication rates
226 in each sample, and test whether replication rate was negatively associated with the number of
227 iSNVs. iRep infers replication rates from MAGs and metagenomic reads (Brown et al. 2016). For
228 instance, an iRep value of 2 would indicate that most of the population is replicating one copy of
229 its chromosome. In our data (Table 1), iRep values varied from 1.23 (patient E at day 2) to 5.43
230 (patient D), and we did not find any association between the replication rate of *Vibrio cholerae*
231 and the number of iSNVs detected within each subject (Fig. S2B, Pearson correlation, $\rho = 0.15$, P
232 > 0.05). This suggests that deleterious mutations in hypermutators could be counterbalanced by
233 adaptive mutations that maintain growth. Alternatively, higher iRep values could be associated
234 with larger *V. cholerae* population sizes, which could support greater genetic diversity and yield a
235 positive correlation between iRep values and the number of iSNVs. These hypotheses merit
236 testing in larger patient cohorts.
237

238 **Convergent evolution suggests adaptation of non-mutator *V. cholerae* within patients**

239 While none of the patients shared iSNVs at the exact same nucleotide position, some
240 contained parallel (or convergent) mutations in the same gene (Table 2). To determine whether
241 genes that acquired multiple mutations could be under positive selection within the host, we
242 performed permutation tests for hypermutator and non-mutator samples separately (Methods).
243 Among the hypermutator samples, we identified five genes with NS mutations in two or more
244 patients (Table 2), which was not an unexpectedly high level of convergence given the large
245 number of mutations in hypermutators (permutation test, P=0.97). That the P-value approaches 1

246 suggests either that the hypermutators are actually selected against mutating the same genes in
247 different patients, or – more likely – that the permutation test is conservative. For the samples
248 with no evidence of hypermutator phenotypes, we identified two genes with NS mutations in two
249 patients. The first gene, *hlyA*, encodes a hemolysin that causes cytolysis by forming heptameric
250 pores in human cell membranes (Olson et Gouaux 2005), while the second gene encodes a
251 putative ABC transporter ferric-binding protein (Table 2). Observing convergent mutations in
252 two different genes is unexpected (permutation test, P=0.039) in a test that is likely to be
253 conservative. We also note that the three iSNVs in *hlyA* have relatively high minor allele
254 frequencies (0.22-0.43) compared to other convergent NS mutations (median minor allele
255 frequency of 0.11; Table 2) and to NS mutations overall (median of 0.12; Table S4). Together,
256 these analyses suggest that *V. cholerae* hypermutators produce NS mutations that are
257 predominantly deleterious or neutral, while evidence for within-patient positive selection on
258 certain genes in non-mutators merits further investigation.

259 To further explore differential selection at the protein level within and between patients,
260 we applied the McDonald-Kreitman test (McDonald et Kreitman 1991) to the 9 patients with no
261 evidence for hypermutation, and to the five patients harboring potential hypermutators. Based on
262 whole-genome sequences of *V. cholerae* isolates, we previously found an excess of NS mutations
263 fixed between patients in Bangladesh, based on a small sample of five patients (Levade et al.
264 2017). Here, based on metagenomes from a larger number of patients, we found the opposite
265 pattern of a slight excess of NS mutations segregating as iSNVs within patients, consistent with
266 slightly deleterious mutations occurring within patients and purged over evolutionary time.
267 However, the difference between NS:S ratios within and between patients was not statistically
268 significant (Fisher's exact test, P>0.05; Table S5); thus, the evidence for differential selective
269 pressures within versus between cholera patients remains inconclusive.

270 Many NS mutations occurred in genes involved in transmembrane transport,
271 pathogenesis, response to antibiotics, secretion systems, chemotaxis, and metabolic processes
272 (Figure S3). Both hypermutator samples (Fig. S3B) and non-mutators (Fig. S3C) have a high
273 NS:S ratio in genes of unknown function, while hypermutators have many NS mutations in
274 transmembrane proteins, which are absent in non-mutators. However, non-mutator samples have
275 more NS mutations in genes involved in pathogenesis and secretion systems. Most of the NS
276 mutations involved in pathogenesis were found in the gene *hlyA* (a target of convergent
277 evolution, mentioned above).
278
279 **Whole genome sequencing of *Vibrio cholerae* isolates confirms hypermutator phenotypes**
280 **and suggests similar diversity levels in symptomatic and asymptomatic patients**
281 In addition to metagenomic analyses, we performed whole genome sequencing of
282 multiple *Vibrio cholerae* clinical isolates from index cases and asymptomatic contacts (Fig. 1A)
283 from three households (56, 57, and 58, Table S1). As noted above, asymptomatic infected
284 contacts did not yield sufficient metagenomic reads to assemble the *V. cholerae* genome or call
285 iSNVs, but their stool cultures yielded colonies for whole-genome sequencing. The first
286 asymptomatic contact, 58.01, tested positive for *Vibrio cholerae* on day 4 after the presentation of
287 the index case to the hospital, and *Vibrio cholerae* was cultured from the stool on days 4, 6, 7 and
288 8. We sequenced five isolates respectively from day 4 and 6 samples, and four isolates from each
289 of the subsequent days. For households 56 and 57, five isolates were sequenced from each
290 sample, at day 1 for the index cases and day 2 for the asymptomatic carriers (Table S6).
291 The index case from household 58 (called 58.00 or patient N) was also included in the
292 metagenomic analysis described above, allowing a comparison between culture-dependent and -
293 independent assessments of within-patient diversity. We did not detect any iSNVs in patient

294 58.00, as the five isolates sequenced were isogenic. In contrast, the metagenomic analysis of
295 patient N revealed seven iSNVs (Table 1), suggesting a higher sensitivity for the detection of rare
296 variants.

297 In contrast to metagenomes consisting of many unlinked reads, whole-genome sequencing
298 allows the reconstruction of a phylogeny describing the evolution of *V. cholerae* within and
299 between patients (Fig. 3). As described previously (Domman et al. 2018), isolates from members
300 of the same household tended to cluster together. In index case 57.00, four of the isolates were
301 isogenic, and one isolate was identical to the five isolates sequenced from the asymptomatic
302 contact from the same household, patient 57.01 (Table 3, Fig. 3). This shared genotype between
303 the two individuals was unexpected, and could suggest a potential transmission event from the
304 asymptomatic contact to the index case, followed by a mutational event and the spreading of the
305 new variant in the index case. The only mutation found in four of the five isolates from the index
306 case was a non-synonymous mutation in a gene coding for a cyclic-di-GMP-modulating response
307 regulator, which could have an impact on the regulation of biofilm formation in the host (Tischler
308 et Camilli 2004). However, the hypothesis that this was a transmission event is only supported by
309 one mutation, and therefore remains uncertain. Among the other index cases, we found no iSNVs
310 in patient 58.00 and two iSNVs in patient 56.00. One isolate from this patient had a synonymous
311 mutation in a hypothetical protein, and another isolate had a non-synonymous mutation in a
312 UDP-N-acetylglucosamine 4,6-dehydratase gene (Table 3). We detected iSNVs in the other
313 asymptomatic contacts, with one synonymous and one intergenic mutation in contact 58.02, and
314 one non-synonymous mutation in one isolate from contact 56.01 (Table 3, Fig. 3).
315

Figure 3. Phylogeny and pan-genome of 48 *Vibrio cholerae* isolates from index cases and their asymptomatic contacts. The phylogeny was inferred using Maximum Parsimony. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Filled circles represent isolates from index cases and empty circles represent isolates from their asymptomatic contacts. The heatmap of gene presence-absence is based on 102 genes in the flexible genome. Colored blocks in the heatmap indicate gene presence; white indicates gene absence. Each row corresponds to an isolate from the phylogenetic tree and each column represents an orthologous gene family. Each unique color represents a different individual.

326 Notably, we also found evidence for a hypermutator in contact 58.01. One isolate sampled
327 from this contact had the highest number of mutations seen in any branch in the phylogeny (five
328 NS mutations) which could be explained by a NS mutation in the gene encoding MutS, another
329 key component of the methyl-directed mismatch repair (MMR) system (Table 3, Fig. 3). The
330 mutation in this gene could explain the accumulation of a surprising number of mutations in this
331 isolate, which is likely a hypermutator. This contact presented no variants in the isolates sampled
332 at day 4 and 6, but we found this hypermutator isolate on day 7. However, this genotype was not
333 found at day eight, either due to the lower resolution in the detection of variants with the WGS of
334 cultured isolates, or the disappearance of this mutant from the population.
335
336 **Pan-genome analyses**
337 Whole-genome isolate sequencing also provides the opportunity to study variation in
338 gene content (the pangenome) within and between patients. We identified a total of 3523 core
339 genes common to all *V. cholerae* genomes, and 102 flexible genes present in some but not all
340 genomes (Figure 3; Table S7). We also found an additional 214 genes present uniquely in isolate
341 56.00C4, assembled into one single contig identified as the genome of the lytic Vibrio phage
342 ICP1, which was assembled alongside the *Vibrio cholerae* genome. This phage contig contained
343 the ICP1 CRISPR/Cas system, which consists of two CRISPR loci (designated CR1 and CR2)
344 and six *cas* genes, as previously described (Seed et al. 2011; 2013). These genes were excluded
345 from subsequent *V. cholerae* pan genome analyses.
346 Among the 102 flexible genes, some varied in presence/absence within a patient, ranging
347 from twelve to 53 genes gained or lost per patient (Table S7; Fig. 3). The majority of these
348 flexible genes (78%) were annotated as hypothetical, and several were transposases or prophage
349 genes. One gene known to be critical for streptogramin antibiotic resistance, a streptogramin A

350 acetyltransferase, was absent in five isolates from contact 58.01 (Alcala et al. 2020). A large
351 deletion of 24 genes was detected in patient 58.00, in an 18kb phage-inducible chromosomal
352 island (PICI) previously shown to prevent phage reproduction, and which is targeted by the ICP1
353 CRISPR/Cas system (Seed et al. 2013). These PICI-like elements are induced during phage
354 infection, and interfere with phage reproduction via multiple mechanisms (Ram et al. 2012;
355 O'Hara et al. 2017). The deletion of this PICI element in the *V. cholerae* genome may be a
356 consequence of an ongoing evolutionary arms race between *V. cholerae* and its phages.
357

358 **DISCUSSION**

359 Although within-patient *Vibrio cholerae* genetic diversity has been reported previously
360 (Seed et al. 2012; 2014; Kendall et al. 2010; Levade et al. 2017), our results confirmed that
361 within-patient diversity is a common feature observed in both symptomatic patients with cholera
362 but also in asymptomatically infected individuals. In this study, we used a combination of
363 metagenomic and WGS sequencing technologies to characterize this within-patient diversity,
364 revealing evidence for hypermutator phenotypes in both symptomatic and asymptomatic
365 infections.

366 We showed that metagenomics has a higher sensitivity to detect rare genetic variants in
367 the within-patient *Vibrio cholerae* population. In our previous study, we detected between zero
368 and three iSNVs in cultured isolates from patients with acute infection (Levade et al. 2017). In
369 contrast, metagenomic analyses allowed us to detect two iSNVs in the patient with the lowest
370 level of diversity, but up to 207 iSNVs in another individual (Table 1). In the only patient for
371 which we were able to characterize *Vibrio cholerae* intra-host diversity both from the
372 metagenome and from cultured isolates, we did not identify any iSNVs in the isolates, but
373 detected 7 iSNVs from the metagenomic analyses, even with a coverage <10X. These results

374 highlight one of the potential limitations of the colony sequencing-based approach for the study
375 of within-host diversity: the difficulty of recovering rare members of the population (Brenzinger
376 et al. 2019).

377 Despite better sensitivity to detect rare variants, metagenomics has limitations. Within-
378 sample diversity profiles cannot be established for low-abundance microbes with insufficient
379 sequence coverage (< 5X) and depth, and this level of coverage is difficult to obtain in diverse
380 microbial communities. In this study, only 48% of the samples from patients with acute
381 symptoms, known to harbor a high fraction of vibrios in their stool (10^{10}-10^{12} vibrios per liter of
382 stool), contained enough reads to reconstruct *Vibrio cholerae* MAGs and to quantify within-
383 patient diversity. Asymptomatic patients typically shed even less *V. cholerae* in their stool
384 (Nelson et al. 2009), making it even more challenging to assemble their genomes using
385 metagenomics without depletion of host DNA or targeted sequence capture techniques
386 (Bachmann et al. 2018; Vezzulli et al. 2017).

387 Hypermutation has been defined as an excess of mutations due to deficiency in DNA
388 mismatch repair, and hypermutator strains have been described in diverse pathogenic infections
389 and *in vivo* experiments, including *Pseudomonas aeruginosa, Haemophilus influenzae* and
390 *Streptococcus pneumoniae* in cystic fibrosis patients, or *E. coli* in diverse habitats (Jolivet-
391 Gougeon et al. 2011; Oliver et Mena 2010; Labat et al. 2005). In *Vibrio cholerae,* a previous
392 study of 260 clinical isolate genomes identified 17 isolates with an unusually high number of
393 SNPs uniformly distributed along the genome (Didelot et al. 2015). Most of these genomes
394 contained mutations in one or more of four genes (*mutS, mutH, mutL* and *uvrD*) that play key
395 roles in DNA mismatch repair (Didelot et al. 2015). These authors cautiously suggested that this
396 apparent high frequency of hypermutators could be associated with the rapid spread of the
397 seventh cholera pandemic, particularly because hypermutators may be a sign of population

398 bottlenecks and recent selective pressure. However, they also hypothesized that these high
399 mutation rates could be artefactual because the *V. cholerae* isolates had been maintained in stab
400 cultures for many years. It thus remains unclear if a hypermutator phenotype was derived within
401 patients or during culture (Didelot et al. 2015; Eisenstark 2010). Using our metagenomic
402 approach, we showed that hypermutators can indeed emerge during infection. Using culture-
403 based whole genome sequencing, using only a brief overnight culture, we confirmed that
404 hypermutators occur in asymptomatic patients as well. Future work will be required to determine
405 any impacts of hypermutation on cholera disease severity or transmission.

406 Hypermutator phenotypes are believed to be advantageous for the colonization of new
407 environments or hosts, allowing the hypermutator bacteria to generate adaptive mutations more
408 quickly, which leads to the more efficient exploitation of resources or increased resistance to
409 environmental stressful conditions, such as antibiotics (Jolivet-Gougeon et al. 2011; Oliver et
410 Mena 2010; Labat et al. 2005; Giraud et al. 2001). However, this high mutation rate can have a
411 negative impact on fitness in the long term, with most of the mutations being neutral or
412 deleterious (Funchain et al. 2000; Giraud et al. 2001; Chu et al. 2017). A mouse model study
413 showed that hypermutation can be an adaptive strategy for *V. cholerae* to resist host-produced
414 reactive oxygen induced stress, and lead to a colonization advantage by increased catalase
415 production and increased biofilm formation (Wang et al. 2018). In our study of convergent
416 evolution, we found no evidence for adaptive mutations in the hypermutators. This could be
417 because the signal from a small number of adaptive mutations are obscured by overwhelming
418 noise from a large number of neutral or deleterious mutations. Further work is therefore needed
419 to determine if *V. cholerae* mutators produce adaptive mutations during human infection.

420 In contrast, we did find evidence for an excess of convergent mutations occurring
421 independently in the same genes in different patients, suggesting parallel adaptation in non-

422 mutator *V. cholerae* infections. Specifically, two patients contained mutations in the same
423 hemolysin gene, *hlyA*, which codes for a toxin that has both vacuolating and cytocidal activities
424 against a number of cell lines, including human intestinal cells (Tsou et Zhu 2010), and is known
425 to be an important virulence factor in *Vibrio cholerae* El Tor O1 and a major target of immune
426 responses during acute infection (Olivier et al. 2007; Weil, Arifuzzaman, et al. 2009). Previous
427 studies of within-patient *V. cholerae* evolution did not identify mutations in *hlyA*, and instead
428 identified different mutations possibly under selection for biofilm formation (Levade et al. 2017)
429 or phage resistance phenotypes (Seed et al. 2014). This lack of concordance could be explained
430 by relatively modest sample sizes of cholera patients in these studies but could also suggest that
431 selective pressures may be idiosyncratic and person-specific across *Vibrio cholerae* infections.
432 In conclusion, our results illustrate the potential and limitations of metagenomics as a
433 culture-independent approach for the characterization of within-host pathogen diversity, and that
434 this diversity is likely to be underestimated by traditional culture-based techniques. We also
435 provide evidence that hypermutators emerge within human *V. cholerae* infection, and their
436 evolutionary dynamics and relevance to disease progression merits further study.
437
438

439 ## MATERIALS AND METHODS

440 **Sample collection, clinical outcomes and metagenomic sequencing**

441 To study within-host diversity of *V. cholerae* during infection, we used stool and rectal
442 swab samples collected from cholera patients admitted to the icddr,b (International Center for
443 Diarrheal Disease Research, Bangladesh) Dhaka Hospital, and from their household contacts, as
444 previously described (Midani et al. 2018). Index cases were defined as patients presenting to the
445 hospital with severe acute diarrhea and a stool culture positive for *V. cholerae*. Individuals who
446 shared the same cooking pot with an index patient for three or more days are considered
447 household contacts and were enrolled within 6 hours of the presentation of the index patient to
448 the hospital. Rectal swabs were collected each day during a ten-day follow up period after
449 presentation of the index case. Household contacts underwent daily clinical assessment of
450 symptoms and collection of blood for serological testing. Contacts were determined to be
451 infected if any rectal swab culture was positive for *V. cholerae* or if the contact developed
452 diarrhea and a 4-fold increase in vibriocidal titer during the follow-up period (Harris et al. 2008;
453 Weil, Khan, et al. 2009). If they developed watery diarrhea during the follow up period, contacts
454 with positive rectal swabs were categorized as symptomatic and those without diarrhea were
455 considered asymptomatic. We excluded patients with age below two and above 60 years old, or
456 with major comorbid conditions (Harris et al. 2008; Weil et al. 2009).

457

458 Fecal samples and rectal swabs from the day of infection and follow up timepoints were
459 collected and immediately placed on ice after collection and stored at -80°C until DNA
460 extraction. DNA extraction was performed with the PowerSoil DNA extraction kits (Qiagen)
461 after pre-heating to 65°C for 10 min and to 95°C for 10 min. Sequencing libraries were
462 constructed for 33 samples from 31 patients, for which we obtained enough DNA. We used the

463 NEBNext Ultra II DNA library prep kit and sequenced the libraries on the Illumina HiSeq 2500
464 (paired-end 125 bp) and the Illumina NovaSeq 6000 S4 (paired-end 150 bp) platforms at the
465 Genome Québec sequencing platform (McGill University).
466
467 **Metagenomic analyses**
468 ***Sequence preprocessing and assembly***
469 Sequencing fastq files were quality checked with FastQC
470 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We removed human and technical
471 contaminant DNA by aligning reads to the PhiX genome and the human genome (hg19) with
472 Bowtie2 (Langmead et Salzberg 2012) , and used the iu-filter-quality-minoche script of the
473 illumina-utils program with default parameters to filter the reads (Eren et al. 2013).
474
475 ***Taxonomic assignment***
476 Processed paired-end metagenomics sequences were classified using two taxonomic
477 profilers: Kraken2 v.2.0.8_beta (a k-mer matching algorithm) (Wood, Lu, et Langmead 2019)
478 and MIDAS v.1.3.0 (a read mapping algorithm) (Nayfach et al. 2016). Kraken 2 examines the k-
479 mers within a query sequence and uses the information within those k-mers to query a database,
480 then maps k-mers to the lowest common ancestor (LCA) of all genomes known to contain a
481 given k-mer. Kraken2 was run against a reference database containing all RefSeq viral, bacterial
482 and archaeal genomes (built in May 2019), with default parameters. MIDAS uses a panel of 15
483 single-copy marker genes present in all of ~31,000 bacterial species included in its database to
484 perform taxonomic classification, and maps metagenomic reads to this database to estimate the
485 read depth and relative abundance of 5,952 bacterial species. We identified metagenomic samples

486 containing *V. cholerae* and vibriophage reads, and computed the mean coverage (number of reads
487 per base-pair) of the *V. cholerae* pangenome in the MIDAS database (Table 1).
488
489 ***Assembly and binning of Vibrio cholerae genomes***
490 To recover good quality metagenome-assembled genomes (MAGs) of *V. cholerae,* we
491 selected metagenomic samples with coverage >10X against the *V. cholerae* pangenome in the
492 MIDAS database, and used MEGAHIT v.1.2.9 (Li et al. 2016) to perform *de novo* assembly. For
493 9 of the 11 selected samples, we independently assembled the genome of each sample, and co-
494 assembled the two remaining samples, which belong to the same patient (a symptomatic infected
495 contact on days 9 and 10). Contigs of <1.5 kb were discarded.
496 We extracted MAGs by binning of our metagenomic assemblies. Because no single
497 binning approach is superior in every case, with performance of the algorithms varying across
498 samples, we used different binning tools to recover MAGs. The quality of a metagenomic bin is
499 evaluated by its completeness (the level of coverage of a population genome), and the
500 contamination level (the amount of sequence that does not belong to this population from another
501 genome). These metrics can be estimated by counting the frequency of single-copy marker genes
502 within each bin (Parks et al. 2015). We inferred bins using CONCOCT v.1.1.0 (Alneberg et al.
503 2014), MaxBin 2 v.2.2.7 (Wu, Simmons, et Singer 2016) and MetaBAT 2 v.2.12.1 (Kang et al.
504 2019), with default parameters. We then used DAS_Tool v.1.1.1 on the results of these three
505 methods, to select a single set of non-redundant, high-quality bins per sample (Sieber et al. 2018).
506 DAS_Tool is a bin consolidation tool, which predicts single-copy genes in all the provided bin
507 sets, aggregates bins from the different binning predictions, and extracts a more complete
508 consensus bin from each aggregate such that the resulting bin has the most single-copy genes
509 while having a reasonably low number of duplicate genes (Sieber et al. 2018). We then used

510 Anvi'o v.6.1 (Eren et al. 2015) to manually refine the bins with contamination higher than 10%
511 and Centrifuge v.1.0.4_beta (Kim et al. 2016) to determine the taxonomy of all bins in each
512 sample, in order to identified *V. cholerae* MAGs.

513 Bins with completeness > 60% and contamination <10% were first selected, and those
514 assigned to *V. cholerae* were further filtered (completeness > 90% and contamination <1% for
515 the *V. cholerae* bins). We dereplicated the entire set of bins with dRep v.2.2.3 using a minimum
516 completeness of 60%, the ANImf algorithm, 99% secondary clustering threshold, maximum
517 contamination of 10%, and 25% minimum coverage overlap, and obtained 79 MAGs displaying
518 the best quality and representing individual metagenomic species (MGS).

519

520 ***Detection of Vibrio cholerae genetic diversity within and between metagenomic samples***

521 We created a bowtie2 index of the 79 representative genomes from the dereplicated set,
522 including a single high-quality *Vibrio cholerae* MAG, and mapped reads from each sample to this
523 set. By including many diverse microbial genomes in the bowtie2 index, we aimed to avoid the
524 mismapping of reads from other species to the *V. cholerae* genome, and to reduce potential false
525 positive intra-host single nucleotide variant (iSNV) calls. We mapped the metagenomics reads of
526 each sample with a *V. cholerae* coverage value >5X (obtained with MIDAS) against the set of 79
527 MAGs, using Bowtie2 (Langmead et Salzberg 2012) with the --very-sensitive parameters. We
528 also used Prodigal (Hyatt et al. 2010) on the concatenated MAGs, in order to predict open
529 reading frames using default metagenomic settings.

530 We then used inStrain on the 15 selected samples
531 (https://instrain.readthedocs.io/en/latest/index.html). This program aims to identify and compare
532 the genetic heterogeneity of microbial populations within and between metagenomic samples
533 (Olm et al. 2020). "InStrain profile" was run on the mapping results, with the minimum percent

534 identity of read pairs to consensus set to 99%, minimum coverage to call a variant of 5X, and
535 minimum allele frequency to confirm a SNV equal to 0.05. All non-paired reads were filtered
536 out, as well as reads with an identity value below 0.99. Coverage and breadth of coverage
537 (percentage of reference base pairs covered by at least one read) were computed for each
538 genome. InStrain identified both bialllelic and multiallelic SNV frequencies at positions where
539 phred30 quality filtered reads differ from the reference genome and at positions where multiple
540 bases were simultaneously detected at levels above the expected sequencing error rate. SNVs
541 were classified as non-synonymous, synonymous, or intergenic based on gene annotations, and
542 gene functions were recovered from the Uniprot database (The UniProt Consortium 2019) and
543 BLAST (Madden 2003). Then, similar filters to those described in (Garud et al. 2019) were
544 applied to the detected SNVs. We excluded from the analysis positions with very low or high
545 coverage value D compared to the median coverage $\overline{D}$, and positions within 100 bp of contig
546 extremities. As sites with very low coverage could result from a bias in sequencing or library
547 preparation, and sites with higher coverage could arise from mapping error or be the result of
548 repetitive region or multi-copy genes not well assembled, we masked sites in all the samples if D
549 was $< 0.3\overline{D}$ and if D was $> 3\overline{D}$ in at least two samples.
550

551 ***Mutation spectrum of hypermutator and non-mutator samples***

552 For each sample, iSNVs were categorized into six mutation types based on the chemical
553 nature of the nucleotide changes (transitions or transversions). We combined all the samples with
554 hypermutators and compared them to the mutation spectrum of the non-mutators. The mutation
555 spectrum was significantly different between the hypermutator samples and the non-
556 hypermutator samples (Chi-squared test, $P<0.01$). We then computed the mutation mean and
557 standard error of each of the six mutation types and compared the two groups (Figure 2C).

558

559 ***Bacterial replication rate estimation***

560 Replication rates were estimated with the metric iRep (index of replication), which is
561 based on the measurement of the rate of the decrease in average sequence coverage from the
562 origin to the terminus of replication. iRep values (Brown et al. 2016) were calculated by mapping
563 the sequencing reads of each sample to the *V. cholerae* MAG assembled from that sample.

564

565 ***Tests for natural selection***

566 First, we identified signals of convergent evolution in the form of nonsynonymous iSNVs
567 occurring independently in the same gene in multiple patients. To assess the significance of
568 convergent mutations, we compared their observed frequencies to expected frequencies in a
569 simple permutations model. We ran separate permutations for non-mutators (two genes with
570 convergent mutations in at least two out of eight non-mutator samples, including only one time
571 point from the patient sampled twice, and excluding the outlier patient A with a large number of
572 intergenic iSNVs) and possible hypermutators (five genes with convergent mutations in at least
573 two out of five possible hypermutator samples). In each permutation, we randomized the
574 locations of the nonsynonymous mutations, preserving the observed number of nonsynonymous
575 mutations in each sample, and the observed distribution of gene lengths. For simplicity, we
576 assumed that 2/3 of nucleotide sites in coding regions were nonsynonymous. We repeated the
577 permutations 1000 times and estimated a *P*-value as the fraction of permutations yielding greater
578 than or equal to the observed number of genes mutated in two or more samples.

579 Second, we compared natural selection at the protein level within versus between patients,
580 using the McDonald-Kreitman test (McDonald et Kreitman 1991). We again considered
581 hypermutators separately. Briefly, the four counts (Pn, Ps, Dn, Ds) of between-patient divergence

582 (D) vs. within-patient polymorphism (P), and non-synonymous (n) vs synonymous (s) mutations
583 were computed and tested for neutrality using a Fisher's exact test (FDR corrected *P*-
584 values < 0.05).
585
586 **Whole genome sequencing analyses**
587 ***Culture of Vibrio cholerae isolates***
588 We selected three of the households with asymptomatic infected contacts (households 56,
589 57, and 58) for within-patient diversity analysis using multiple *V. cholerae* colonies per
590 individual. Each index case was sampled on the day of presentation to the icddr,b, and
591 asymptomatic contacts positive for *V. cholerae* were sampled on the following day, except for
592 one contact (household 58, contact 02). This individual was only positive on day 4 following
593 presentation of the index case, and we collected samples and cultured isolates from day 4 to day
594 8. Stool samples collected from three index cases and their respective infected contacts were
595 streaked onto thiosulfate-citrate-bile salts-sucrose agar (TCBS), a medium selective for *V.*
596 *cholerae*. After overnight incubation, individual colonies were inoculated into 5 ml Luria-Bertani
597 broth and grown at 37 °C overnight. For each colony, 1 ml of broth culture was stored at −80 °C
598 with 30% glycerol until DNA extraction. We used the Qiagen DNeasy Blood and Tissue kit,
599 using 1.5 ml bacteria grown in LB media, to extract the genomic DNA. In order to obtain pure
600 gDNA templates, we performed a RNase treatment followed by a purification with the MoBio
601 PowerClean Pro DNA Clean-Up Kit.
602
603 ***Whole genome sequencing and preprocessing***
604 We prepared 48 sequencing libraries using the NEBNext Ultra II DNA library prep kit
605 (New England Biolabs) and sequenced them on the Illumina HiSeq 2500 (paired-end 125 bp)

606 platform at the Genome Québec sequencing platform (McGill University). Sequencing fastq files
607 were quality checked with FastQC, and Kraken2 was used to test for potential contamination with
608 other bacterial species (Wood, Lu, et Langmead 2019).
609
610 ***Variant calling and phylogeny***
611 We mapped the reads for each sample to the MJ-1236 reference genome and called single
612 nucleotide polymorphisms (SNPs, fixed within patients) and single nucleotide variants (SNVs,
613 variable within patients) using Snippy v.4.6.0 (Seemann 2015), with default parameters. A
614 concatenated alignment of these core variants was generated, and an unrooted phylogenic tree
615 was inferred using maximum parsimony (MP) in MEGA X (Stecher, Tamura, et Kumar 2020).
616 The percentage of replicate trees in which the associated taxa clustered together in the bootstrap
617 test (1000 replicates) are shown next to the branches. The MP tree was obtained using the
618 Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were
619 obtained by the random addition of sequences (10 replicates).
620
621 ***De novo assembly and pan genome analyses***
622 We *de novo* assembled genomes from each isolate using SPAdes v.3.11.1 on the short reads, with
623 default parameters (Bankevich et al. 2012) and used Prokka v1.5 (Seemann 2014) to annotate
624 them. We constructed a pan-genome from the resulting annotated assemblies using Roary
625 v.3.13.0 (Page et al. 2015), identifying genes present in all isolates (core genome) and genes only
626 present in some isolates (flexible genome). The flexible genome and the phylogenetic tree were
627 visualized with Phandango v.1.1.0 (Hadfield et al. 2018).
628
629
630

DATA AVAILABILITY

All metagenomic sequence data are available in NCBI GenBank under BioProject PRJNA668607, and isolate genome sequences under BioProject PRJNA668606.

FUNDING INFORMATION

This study was supported by a Canadian Institutes of Health Research Operating Grant to BJS, the icddr,b Centre for Health and Population Research, grants AI103055 (J.B.H and F.Q), AI106878 (E.T.R and F.Q.), AI058935 (E.T.R, S.B.C and F.Q.), T32A1070611976 and K08AI123494 (A.A.W.), and Emerging Global Fellowship Award TW010362 (T.R.B.), from the National Institutes of Health, and the Robert Wood Johnson Foundation Harold Amos Medical Faculty Development Program (R.C.C.).

ACKNOWLEDGEMENTS

We are grateful to the people of Dhaka where our study was undertaken; to the field, laboratory and data management staff who provided a tremendous effort to make the study successful; and to the people who provided valuable support in our study. The icddr,b gratefully acknowledges the Government of the People's Republic of Bangladesh; Global Affairs Canada (GAC); Swedish International Development Cooperation Agency (Sida) and the Department for International Development, (UKAid). We declare that we have no competing financial interest.

ETHICAL STATEMENT

The Ethical and Research Review Committees of the icddr,b and the Institutional Review Board of MGH reviewed the study. All adult subjects provided informed consent and parents/guardians of children provided informed consent. Informed consent was written.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

662 **TABLES**

Patient	Total number of iSNVs	Number of non-synonymous iSNVs	Number of synonymous iSNVs	Number of intergenic iSNVs	Mean coverage	iRep value	DNA repair and proofreading genes with NS mutation
Patient A	93	6	0	87	451.3X	3.34	-
Patient B	18	7	5	6	111.4X	1.7	-
Patient C	6	0	1	5	111.8X	1.7	-
Patient D	41	22	9	10	10X	5.43	DNA polymerase II
Patient E day 1	8	2	1	5	351X	3.25	-
Patient E day 2	21	7	1	13	258X	1.23	-
Patient F	207	133	47	27	18.2X	2.48	DNA mismatch repair endonuclease MutL Nuclease SbcCD subunit C
Patient G	16	12	3	1	7.7X	1.73	-
Patient H	32	21	11	0	98.5X	4.75	Excinuclease ABC subunit UvrB
Patient I	75	55	20	0	13X	2.79	MutT/nudix family protein
Patient J	6	1	0	5	424.6X	1.84	-
Patient K	25	13	6	6	18X	1.69	Formamidopyrimidine-DNA glycosylase mutM
Patient L	13	9	1	3	164.4X	2.67	-
Patient M	2	0	1	1	113X	2.65	-
Patient N	7	2	1	3	6.7X	2.27	-

Table 1. Within-patient *Vibrio cholerae* diversity profiles from 15 metagenomes. Mutations segregating within patients are denoted iSNVs. The number of iSNVs and mean coverage values were computed with InStrain (Olm et al. 2020) and replication rate with iRep (Brown et al. 2016).

Protein	Patient A	Patient B	Patient D	Patient E	Patient F	Patient H	Patient I	Patient K
Hemolysin (VC cytolysin)	NS (0.22)	-	-	3 NS (0.22-0.43)	-	-	-	-
2-aminoethylphosphonate ABC transporter ferric-binding protein	-	NS (0.05)	-	NS (0.05)	-	-	-	-
Peptidase B	-	-	NS (0.33)	-	-	-	NS (0.09)	-
Nuclease SbcCD subunit C	-	-	S (0.28)	-	NS (0.09)	-	-	-
C4-dicarboxylate transport sensor protein	-	-	-	-	NS (0.08)	-	NS (0.11)	-
zinc/cadmium/mercury/lead-transporting ATPase	-	-	-	-	NS (0.08)	-	-	NS (0.06)
hypothetical protein	-	-	-	-	NS (0.14)	-	-	NS (0.14)
hypothetical protein	-	-	-	-	NS (0.33)	NS (0.11)	-	-
Formamidopyrimidine-DNA glycosylase mutM	-	-	-	-	S (0.18)	-	-	NS (0.08)
Phosphoribosylformylglycinamidine synthase	-	-	-	-	-	-	NS (0.06)	S (0.08)

668

669 **Table 2. Set of genes with convergent mutations identified in more than one patient.** The
670 presence of a synonymous or non-synonymous iSNV in each gene and each patient is indicated
671 with S or NS, respectively, and the minor allele frequency is shown in parentheses. None of the
672 mutations were found at the same nucleotide or codon position. Patients containing possible or
673 likely hypermutators are underlined. Only genes and patients containing more than one mutated
674 gene are shown.

Type	Isolates	Mutation type	Nucleotide position in MJ-1236	Ref. nucleotide	Alt. nucleotide	Gene annotation	Metagenomic samples with same variant
iSNV	58.01d7C1	NS	Chr1,53054	G	A	DNA mismatch repair protein MutS	-
SNP	Households 56 and 57	S	Chr1, 198988	G	A	MSHA biogenesis protein MshQ	-
iSNV	58.01d7C1	NS	Chr1, 209665	G	A	MSHA biogenesis protein MshN	-
iSNV	56.00C4	NS	Chr1,374172	C	T	UDP-N-acetylglucosamine 4,6-dehydratase	-
SNP	Household 58	NS	Chr1,410638	G	A	Phosphopantetheine adenylyltransferase	Patients M, N
SNP	Households 56 and 57	NS	Chr1,754154	C	T	1,4-dihydroxy-2-naphthoate polyprenyltransferase	-
SNP	Household 58	S	Chr1,841538	C	T	SSU ribosomal protein S4p	Patients L, M, N
SNP	Household 58	S	Chr1,1315021	T	G	Exported zinc metalloprotease YfgC precursor	Patients L, M, N
iSNV	58.02C1	S	Chr1,1576083	C	A	Periplasmic thiol:disulfide oxidoreductase DsbB	-
SNP	Patient 58.00	NS	Chr1,1689779	A	C	Sigma-54 dependent transcriptional regulator	-
SNP	Contacts 58.01 and 58.02	NS	Chr1,2301641	G	A	Putative membrane protein	-
iSNV	58.01d7C1	NS	Chr1,1744854	C	T	Hypothetical protein	-
SNP	Contacts 58.01 and 58.02	NS	Chr1,2262202	A	G	Serine transporter	-
SNP	Households 56 and 57	NS	Chr1,2301641	C	T	LacI family DNA-binding transcriptional regulator	Patients D, J, K
iSNV	57.00C5	NS	Chr1,2509468	C	T	cyclic-di-GMP-modulating response regulator	-
iSNV	56.01C1	NS	Chr1,2588496	C	T	Amidophosphoribosyltransferase	-
iSNV	58.01d7C1	NS	Chr1,2693815	C	T	PTS system, trehalose-specific IIB component	-
SNP	Household 58	NS	Chr1,2806858	A	T	Citrate lyase alpha chain	Patients L, M, N
iSNV	56.00C1	S	Chr1,3037471	A	G	Hypothetical protein	-
SNP	Patient 58.00	NS	Chr1,3059131	C	T	DNA polymerase V (UmuC)	-
SNP	Households 56 and 57	NS	Chr1,3095039	G	A	Outer membrane protein OmpU	Patients D, F, G, I ,J, K
SNP	Contacts 58.01 and 58.02	S	Chr1,3105102	C	T	Glutamate-1-semialdehyde aminotransferase	-
iSNV	58.01d7C1	NS	Chr1,528409	C	T	Vibriolysin, extracellular zinc protease	-

675 **Table 3. Nucleotide changes identified in core genes of the *V. cholerae* isolates from index**

676 **cases (56.00, 57.00 and 58.00) and their asymptomatic contacts.** Genome position is according

677 to the MJ-1236 reference genome (CP001485.1, CP001486.1). Mutations segregating within
678 patients are denoted iSNVs; mutations fixed between patients are denoted SNP. Patient allele
679 frequency shows the allele frequency of the alternative (minor) allele. Ref=Reference allele;
680 Alt=Alternative allele. NS=non-synonymous; S=synonymous. Chr1=chromosome 1;
681 Chr2=chromosome
682

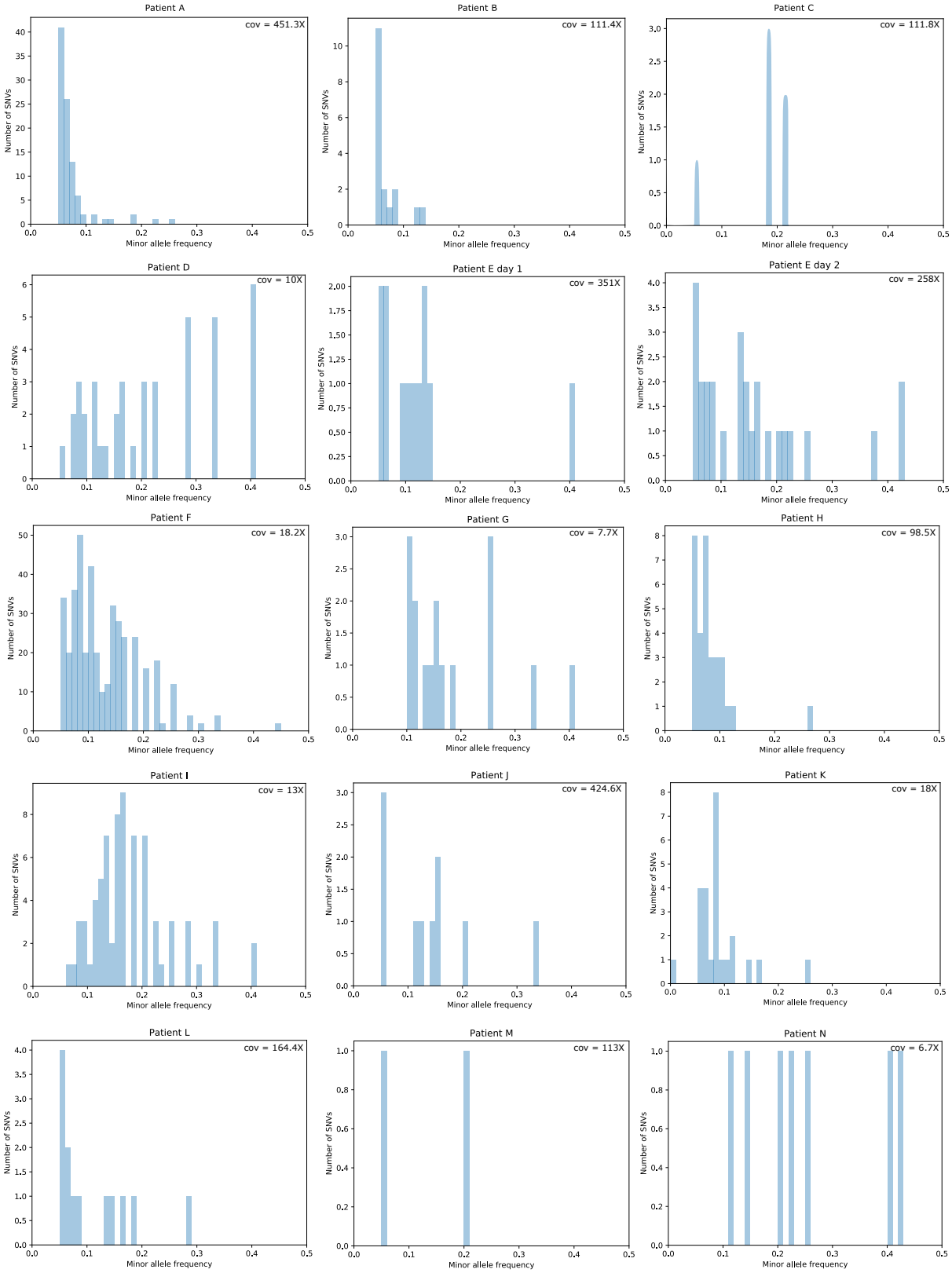
683 ## SUPPLEMENTARY FIGURES

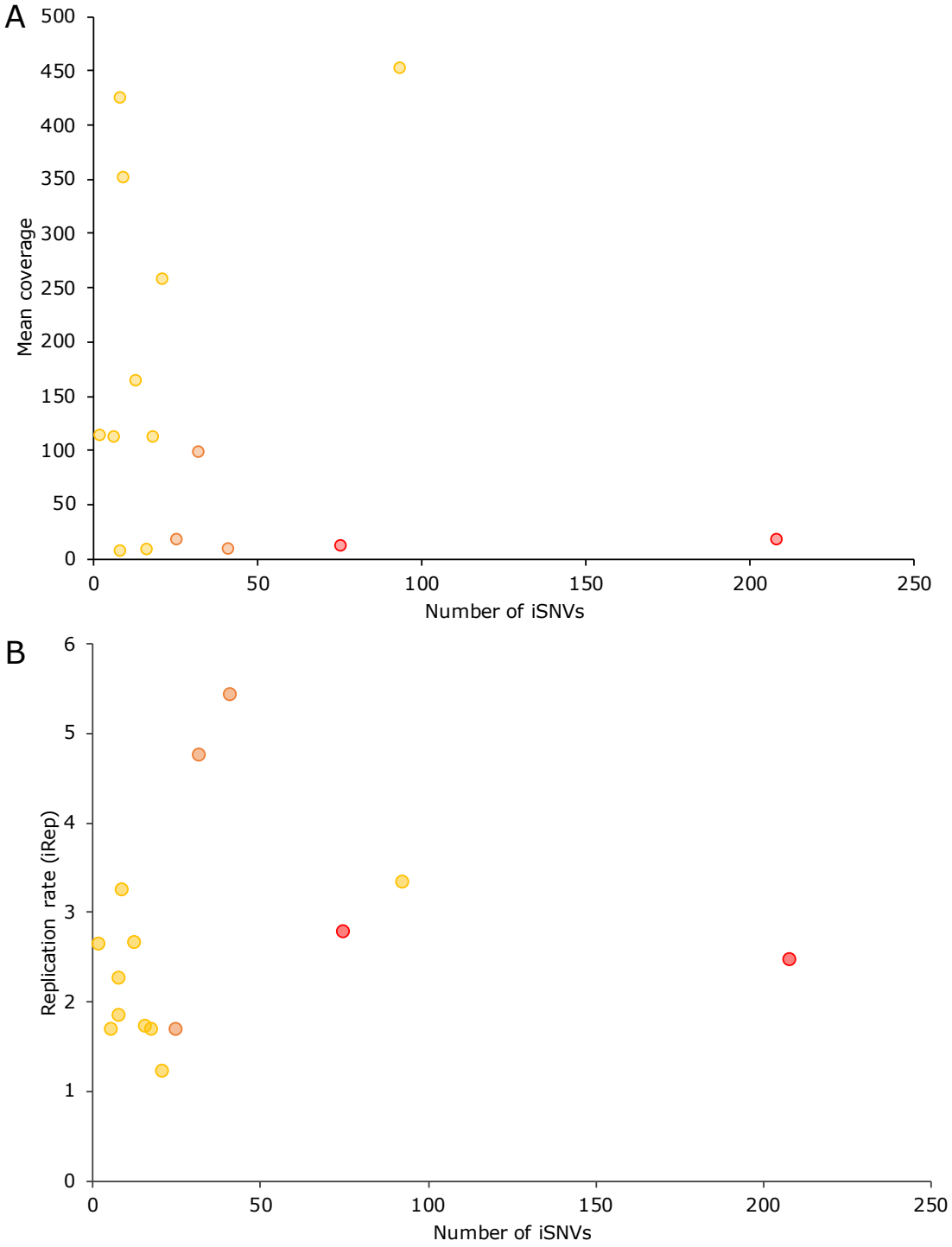
684 **Figure S1. Minor allele frequency distributions for iSNVs in 15 metagenomic samples.**

685 Allele frequencies and mean coverage values (cov) were computed with Instrain (Olm et al.

686 2020)

687


Patient A
cov = 451.3X
Patient B
cov = 111.4X
Patient C
cov = 111.8X
Patient D
cov = 10X
Patient E day 1
cov = 351X
Patient E day 2
cov = 258X
Patient F
cov = 18.2X
Patient G
cov = 7.7X
Patient H
cov = 98.5X
Patient I
cov = 13X
Patient J
cov = 424.6X
Patient K
cov = 18X
Patient L
cov = 164.4X
Patient M
cov = 113X
Patient N
cov = 6.7X
Number of SNVs
Minor allele frequency



688
689
690 **Figure S2. iSNVs numbers are not linked to depth of metagenomic read coverage (A) or**
691 **replication rate (B).** Patients with strong hypermutation phenotypes are represented in red, weak
692 hypermutation phenotypes in orange, and other samples in yellow.

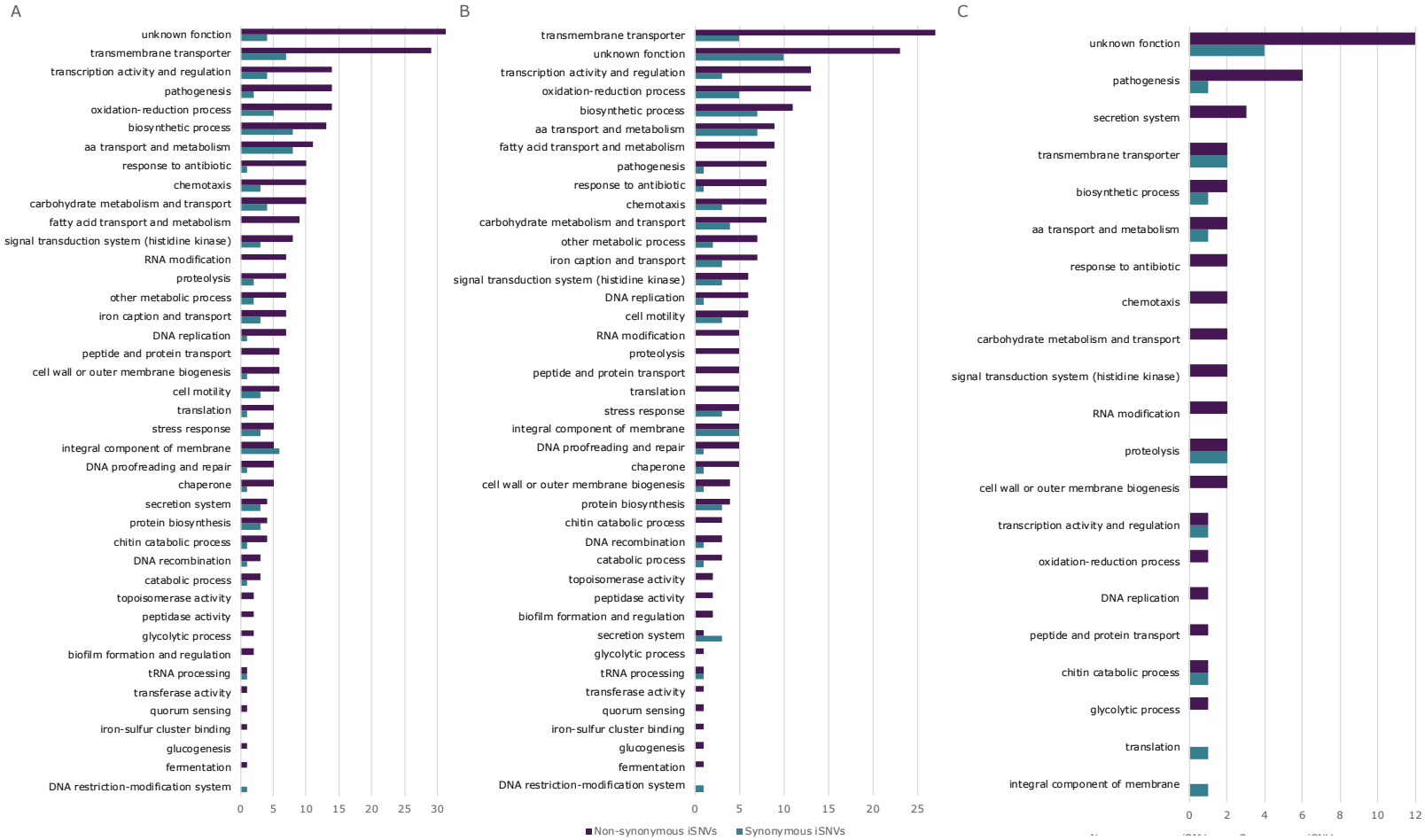


Figure S3. Functional categories of genes containing within-patient variants (iSNVs). The total number of synonymous and non-synonymous iSNVs in different functional categories was represented for all patients (A), for the 5 patients with hypermutation phenotypes (B), and for the 9 patients without hypermutation phenotypes (C). Categories are ranked in descending order of the number of non-synonymous iSNVs.

REFERENCES

Alcala A, Ramirez G, Solis A, Kim Y, Tan K, Luna O, Nguyen K, et al. 2020. « Structural and Functional Characterization of Three Type B and C Chloramphenicol Acetyltransferases from Vibrio Species ». Protein Science : A Publication of the Protein Society. Protein Sci. mars 2020. https://doi.org/10.1002/pro.3793.

Ali, Mohammad, Allyson R. Nelson, Anna Lena Lopez, et David A. Sack. 2015. « Updated Global Burden of Cholera in Endemic Countries ». Édité par Justin V. Remais. *PLOS Neglected Tropical Diseases* 9 (6): e0003832. https://doi.org/10.1371/journal.pntd.0003832.

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z Ijaz, Leo Lahti, Nicholas J Loman, Anders F Andersson, et Christopher Quince. 2014. « Binning Metagenomic Contigs by Coverage and Composition ». *Nature Methods* 11 (11): 1144–1146. https://doi.org/10.1038/nmeth.3103.

Bachmann, Nathan L., Rebecca J. Rockett, Verlaine Joy Timms, et Vitali Sintchenko. 2018. « Advances in Clinical Sample Preparation for Identification and Characterization of Bacterial Pathogens Using Metagenomics ». *Frontiers in Public Health* 6 (décembre). https://doi.org/10.3389/fpubh.2018.00363.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, Alexander S Kulikov, Valery M Lesin, et al. 2012. « SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing ». *Journal of Computational Biology* 19 (5): 455–477. https://doi.org/10.1089/cmb.2012.0021.

Brenzinger, Susanne, Lizah T. van der Aart, Gilles P. van Wezel, Jean-Marie Lacroix, Timo Glatter, et Ariane Briegel. 2019. « Structural and Proteomic Changes in Viable but Non-culturable Vibrio cholerae ». *Frontiers in Microbiology* 10 https://doi.org/10.3389/fmicb.2019.00793.

Brown, Christopher T., Matthew R. Olm, Brian C. Thomas, et Jillian F. Banfield. 2016. « Measurement of Bacterial Replication Rates in Microbial Communities ». *Nature Biotechnology* 34 (12): 1256-63. https://doi.org/10.1038/nbt.3704.

Camacho, Anton, Malika Bouhenia, Reema Alyusfi, Abdulhakeem Alkohlani, Munna Abdulla Mohammed Naji, Xavier de Radiguès, Abdinasir M Abubakar, et al. 2018. « Cholera Epidemic in Yemen, 2016–18: An Analysis of Surveillance Data ». *The Lancet Global Health* 6 (6): e680-90. https://doi.org/10.1016/S2214-109X(18)30230-4.

Chu, Nathaniel D., Sean A. Clarke, Sonia Timberlake, Martin F. Polz, Alan D. Grossman, et Eric J. Alm. 2017. « A Mobile Element in MutS Drives Hypermutation in a Marine Vibrio ». *MBio* 8 (1). https://doi.org/10.1128/mBio.02045-16.

Croucher, Nicholas J., Simon R. Harris, Christophe Fraser, Michael A. Quail, John Burton, Mark van der Linden, Lesley McGee, et al. 2011. « Rapid Pneumococcal Evolution in Response to Clinical Interventions ». *Science* 331 (6016): 430-34. https://doi.org/10.1126/science.1198545.

Darmon, Elise, Manuel A. Lopez-Vernaza, Anne C. Helness, Amanda Borking, Emily Wilson, Zubin Thacker, Laura Wardrope, et David R. F. Leach. 2007. « SbcCD Regulation and Localization in Escherichia Coli ». *Journal of Bacteriology* 189 (18): 6686-94. https://doi.org/10.1128/JB.00489-07.

Denamur, Erick, Guillaume Lecointre, Pierre Darlu, Olivier Tenaillon, Cécile Acquaviva, Chalom Sayada, Ivana Sunjevaric, et al. 2000. « Evolutionary Implications of the Frequent Horizontal Transfer of Mismatch Repair Genes ». *Cell* 103 (5): 711-21. https://doi.org/10.1016/S0092-8674(00)00175-6.

Didelot, Xavier, Christophe Fraser, Jennifer Gardy, et Caroline Colijn. 2017. « Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks ». *Molecular Biology and Evolution* 34 (4): 997-1007. https://doi.org/10.1093/molbev/msw275.

Didelot, Xavier, Bo Pang, Zhemin Zhou, Angela McCann, Peixiang Ni, Dongfang Li, Mark Achtman, et Biao Kan. 2015. « The Role of China in the Global Spread of the Current Cholera Pandemic ». *PLOS Genetics* 11 (3): e1005072. https://doi.org/10.1371/journal.pgen.1005072.

Didelot, Xavier, A Sarah Walker, Tim E Peto, Derrick W Crook, et Daniel J Wilson. 2016. « Within-Host Evolution of Bacterial Pathogens. » *Nature Publishing Group* 14 (3): 150–162. https://doi.org/10.1038/nrmicro.2015.13.

Domman, Daryl, Fahima Chowdhury, Ashraful I. Khan, Matthew J. Dorman, Ankur Mutreja, Muhammad Ikhtear Uddin, Anik Paul, et al. 2018. « Defining Endemic Cholera at Three Levels of Spatiotemporal Resolution within Bangladesh ». *Nature Genetics* 50 (7): 951-55. https://doi.org/10.1038/s41588-018-0150-8.

Eisenstark, Abraham. 2010. « Genetic Diversity among Offspring from Archived Salmonella enterica ssp. enterica Serovar Typhimurium (Demerec Collection): In Search of Survival Strategies ». *Annual Review of Microbiology* 64 (1): 277-92. https://doi.org/10.1146/annurev.micro.091208.073614.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, et Tom O. Delmont. 2015. « Anvi'o: An Advanced Analysis and Visualization Platform for 'omics Data ». *PeerJ* 3: e1319. https://doi.org/10.7717/peerj.1319.

Eren, A. Murat, Joseph H. Vineis, Hilary G. Morrison, et Mitchell L. Sogin. 2013. « A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology ». *PLOS ONE* 8 (6): e66643. https://doi.org/10.1371/journal.pone.0066643.

Foster, P. L., G. Gudmundsson, J. M. Trimarchi, H. Cai, et M. F. Goodman. 1995. « Proofreading-Defective DNA Polymerase II Increases Adaptive Mutation in Escherichia Coli ». *Proceedings of the National Academy of Sciences* 92 (17): 7951-55. https://doi.org/10.1073/pnas.92.17.7951.

Funchain, Pauline, Annie Yeung, Jean Lee Stewart, Rose Lin, Malgorzata M. Slupska, et Jeffrey

807 H. Miller. 2000. « The Consequences of Growth of a Mutator Strain of Escherichia Coli as
808 Measured by Loss of Function Among Multiple Gene Targets and Loss of Fitness ». *Genetics*
809 154 (3): 959-70.
810
811 Gardy, Jennifer L., et Nicholas J. Loman. 2018. « Towards a Genomics-Informed, Real-Time,
812 Global Pathogen Surveillance System ». *Nature Reviews Genetics* 19 (1): 9-20.
813 https://doi.org/10.1038/nrg.2017.88.
814
815 Garud, Nandita R., Benjamin H. Good, Oskar Hallatschek, et Katherine S. Pollard. 2019.
816 « Evolutionary Dynamics of Bacteria in the Gut Microbiome within and across Hosts ». *PLOS*
817 *Biology* 17 (1): e3000102. https://doi.org/10.1371/journal.pbio.3000102.
818
819 Giraud, Antoine, Ivan Matic, Olivier Tenaillon, Antonio Clara, Miroslav Radman, Michel Fons,
820 et François Taddei. 2001. « Costs and Benefits of High Mutation Rates: Adaptive Evolution of
821 Bacteria in the Mouse Gut ». *Science* 291 (5513): 2606-8.
822 https://doi.org/10.1126/science.1056421.
823
824 Hadfield, James, Nicholas J. Croucher, Richard J. Goater, Khalil Abudahab, David M. Aanensen,
825 et Simon R. Harris. 2018. « Phandango: An Interactive Viewer for Bacterial Population
826 Genomics ». *Bioinformatics* 34 (2): 292-93. https://doi.org/10.1093/bioinformatics/btx610.
827
828 Harris, Jason B., Ashraful I. Khan, Regina C. LaRocque, David J. Dorer, Fahima Chowdhury,
829 Abu S. G. Faruque, David A. Sack, Edward T. Ryan, Firdausi Qadri, et Stephen B. Calderwood.
830 2005. « Blood Group, Immunity, and Risk of Infection with Vibrio Cholerae in an Area of
831 Endemicity ». *Infection and Immunity* 73 (11): 7422-27. https://doi.org/10.1128/IAI.73.11.7422-
832 7427.2005.
833
834 Harris, Jason B., Regina C. LaRocque, Fahima Chowdhury, Ashraful I. Khan, Tanya
835 Logvinenko, Abu S. G. Faruque, Edward T. Ryan, Firdausi Qadri, et Stephen B. Calderwood.
836 2008. « Susceptibility to Vibrio cholerae Infection in a Cohort of Household Contacts of Patients
837 with Cholera in Bangladesh ». *PLoS Neglected Tropical Diseases* 2 (4).
838 https://doi.org/10.1371/journal.pntd.0000221.
839
840 Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, et Loren
841 J. Hauser. 2010. « Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site
842 Identification ». *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.
843
844 Jolivet-Gougeon, Anne, Bela Kovacs, Sandrine Le Gall-David, Hervé Le Bars, Latifa
845 Bousarghin, Martine Bonnaure-Mallet, Bernard Lobel, François Guillé, Claude-James Soussy, et
846 Peter Tenke. 2011. « Bacterial hypermutation: clinical implications ». *Journal of Medical*
847 *Microbiology,* 60 (5): 563-73. https://doi.org/10.1099/jmm.0.024083-0.
848
849 Kang, Dongwan, Feng Li, Edward S Kirton, Ashleigh Thomas, Rob S Egan, Hong An, et Zhong
850 Wang. 2019. « MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome
851 Reconstruction from Metagenome Assemblies ». Preprint. PeerJ Preprints.
852 https://doi.org/10.7287/peerj.preprints.27522v1.
853 Kendall, Emily A, Fahima Chowdhury, Yasmin Begum, Ashraful I Khan, Shan Li, James H

854 Thierer, Jason Bailey, et al. 2010. « Relatedness of Vibrio Cholerae O1/O139 Isolates from
855 Patients and Their Household Contacts, Determined by Multilocus Variable-Number Tandem-
856 Repeat Analysis. » *Journal of Bacteriology* 192 (17): 4367–4376.
857 https://doi.org/10.1128/JB.00698-10.
858
859 Kim, Daehwan, Li Song, Florian P. Breitwieser, et Steven L. Salzberg. 2016. « Centrifuge: Rapid
860 and Sensitive Classification of Metagenomic Sequences ». *Genome Research* 26 (12): 1721–
861 1729. https://doi.org/10.1101/gr.210641.116.
862
863 King, Aaron A., Edward L. Ionides, Mercedes Pascual, et Menno J. Bouma. 2008. « Inapparent
864 Infections and Cholera Dynamics ». *Nature* 454 (7206): 877-80.
865 https://doi.org/10.1038/nature07084.
866
867 Kunkel, Thomas A., et Dorothy A. Erie. 2005. « Dna mismatch repair ». *Annual Review of*
868 *Biochemistry* 74 (1): 681-710. https://doi.org/10.1146/annurev.biochem.74.082803.133243.
869
870 Labat, Francoise, Olivier Pradillon, Louis Garry, Michel Peuchmaur, Bruno Fantin, et Erick
871 Denamur. 2005. « Mutator Phenotype Confers Advantage in Escherichia Coli Chronic Urinary
872 Tract Infection Pathogenesis ». *FEMS Immunology & Medical Microbiology* 44 (3): 317-21.
873 https://doi.org/10.1016/j.femsim.2005.01.003.
874
875 Langmead, Ben, et Steven L Salzberg. 2012. « Fast Gapped-Read Alignment with Bowtie 2. »
876 *Nature Methods* 9 (4): 357–359. https://doi.org/10.1038/nmeth.1923.
877
878 Lee, Seung-Joo, Rou-Jia Sung, et Gregory L. Verdine. 2019. « Mechanism of DNA Lesion
879 Homing and Recognition by the Uvr Nucleotide Excision Repair System ». Research article.
880 Research. 2019. https://spj.sciencemag.org/research/2019/5641746/.
881
882 Levade, Inès, Morteza M. Saber, Firas Midani, Fahima Chowdhury, Ashraful I. Khan, Yasmin A.
883 Begum, Edward T. Ryan, et al. 2020. « Predicting Vibrio Cholerae Infection and Disease
884 Severity Using Metagenomics in a Prospective Cohort Study ». *BioRxiv*, février,
885 2020.02.25.960930. https://doi.org/10.1101/2020.02.25.960930.
886
887 Levade, Inès, Yves Terrat, Jean-Baptiste Leducq, Ana A. Weil, Leslie M. Mayo-Smith, Fahima
888 Chowdhury, Ashraful I. Khan, et al. 2017. « Vibrio cholerae genomic diversity within and
889 between patients ». *Microbial Genomics* 3 (12). https://doi.org/10.1099/mgen.0.000142.
890
891 Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko
892 Sadakane, Hiroshi Yamashita, et Tak-Wah Lam. 2016. « MEGAHIT v1.0: A Fast and Scalable
893 Metagenome Assembler Driven by Advanced Methodologies and Community Practices ».
894 *Methods (San Diego, Calif.)* 102: 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
895
896 Lieberman, Tami D, Kelly B Flett, Idan Yelin, Thomas R Martin, Alexander J McAdam, Gregory
897 P Priebe, et Roy Kishony. 2013. « Genetic variation of a bacterial pathogen within individuals
898 with cystic fibrosis provides a record of selective pressures ». *Nature Genetics* 46 (1): 82-87.
899 https://doi.org/10.1038/ng.2848.
900

901 Lovett, Susan T. 2011. « The DNA exonucleases of Escherichia coli ». *EcoSal Plus* 4 (2).
902 https://doi.org/10.1128/ecosalplus.4.4.7.
903
904 Lu, A.-Lien, Xianghong Li, Yesong Gu, Patrick M. Wright, et Dau-Yin Chang. 2001. « Repair of
905 Oxidative DNA Damage ». *Cell Biochemistry and Biophysics* 35 (2): 141-70.
906 https://doi.org/10.1385/CBB:35:2:141.
907
908 Madden, Tom. 2003. *The BLAST Sequence Analysis Tool*. *The NCBI Handbook [Internet]*.
909 National Center for Biotechnology Information (US).
910 https://www.ncbi.nlm.nih.gov/books/NBK21097/.
911
912 Marvig, Rasmus Lykke, Helle Krogh Johansen, Søren Molin, et Lars Jelsbak. 2013. « Genome
913 Analysis of a Transmissible Lineage of Pseudomonas Aeruginosa Reveals Pathoadaptive
914 Mutations and Distinct Evolutionary Paths of Hypermutators ». *PLOS Genetics* 9 (9): e1003741.
915 https://doi.org/10.1371/journal.pgen.1003741.
916
917 McDonald, John H., et Martin Kreitman. 1991. « Adaptive Protein Evolution at the Adh Locus in
918 Drosophila ». *Nature* 351 (6328): 652-54. https://doi.org/10.1038/351652a0.
919
920 Midani, Firas S., Ana A. Weil, Fahima Chowdhury, Yasmin A. Begum, Ashraful I. Khan, Meti
921 D. Debela, Heather K. Durand, et al. 2018. « Human Gut Microbiota Predicts Susceptibility to
922 Vibrio Cholerae Infection ». *The Journal of Infectious Diseases* 218 (4): 645-53.
923 https://doi.org/10.1093/infdis/jiy192.
924
925 Nayfach, Stephen, Beltran Rodriguez-Mueller, Nandita Garud, et Katherine S. Pollard. 2016.
926 « An Integrated Metagenomics Pipeline for Strain Profiling Reveals Novel Patterns of Bacterial
927 Transmission and Biogeography ». *Genome Research* 26 (11): 1612-25.
928 https://doi.org/10.1101/gr.201863.115.
929
930 Nelson, Eric J., Jason B. Harris, J. Glenn Morris, Stephen B. Calderwood, et Andrew Camilli.
931 2009. « Cholera transmission: the host, pathogen and bacteriophage dynamic ». *Nature Reviews*
932 *Microbiology* 7 (10): 693-702. https://doi.org/10.1038/nrmicro2204.
933
934 O'Hara, Brendan J., Zachary K. Barth, Amelia C. McKitterick, et Kimberley D. Seed. 2017. « A
935 Highly Specific Phage Defense System Is a Conserved Feature of the Vibrio Cholerae
936 Mobilome ». *PLOS Genetics* 13 (6): e1006838. https://doi.org/10.1371/journal.pgen.1006838.
937
938 Oliver, A., et A. Mena. 2010. « Bacterial Hypermutation in Cystic Fibrosis, Not Only for
939 Antibiotic Resistance ». *Clinical Microbiology and Infection* 16 (7): 798-808.
940 https://doi.org/10.1111/j.1469-0691.2010.03250.x.
941
942 Olivier, Verena, G. Kenneth Haines, Yanping Tan, et Karla J. Fullner Satchell. 2007.
943 « Hemolysin and the Multifunctional Autoprocessing RTX Toxin Are Virulence Factors during
944 Intestinal Infection of Mice with Vibrio Cholerae El Tor O1 Strains ». *Infection and Immunity* 75
945 (10): 5035-42. https://doi.org/10.1128/IAI.00506-07.
946

Olm, Matthew R., Alexander Crits-Christoph, Keith Bouma-Gregson, Brian Firek, Michael J. Morowitz, et Jillian F. Banfield. 2020. « InStrain Enables Population Genomic Analysis from Metagenomic Data and Rigorous Detection of Identical Microbial Strains ». *BioRxiv*, janvier, 2020.01.22.915579. https://doi.org/10.1101/2020.01.22.915579.

Olson, Rich, et Eric Gouaux. 2005. « Crystal Structure of the Vibrio Cholerae Cytolysin (VCC) Pro-Toxin and Its Assembly into a Heptameric Transmembrane Pore ». *Journal of Molecular Biology* 350 (5): 997-1016. https://doi.org/10.1016/j.jmb.2005.05.045.

Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T.G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, et Julian Parkhill. 2015. « Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis ». *Bioinformatics* 31 (22): 3691-93. https://doi.org/10.1093/bioinformatics/btv421.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, et Gene W. Tyson. 2015. « CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes ». *Genome Research* 25 (7): 1043-55. https://doi.org/10.1101/gr.186072.114.

Phelps, Matthew D., Lone Simonsen, et Peter K. M. Jensen. 2019. « Individual and Household Exposures Associated with Cholera Transmission in Case–Control Studies: A Systematic Review ». *Tropical Medicine & International Health* 24 (10): 1151-68. https://doi.org/10.1111/tmi.13293.

Quince, Christopher, Alan W Walker, Jared T Simpson, Nicholas J Loman, et Nicola Segata. 2017. « Shotgun Metagenomics, from Sampling to Analysis ». *Nature Biotechnology* 35 (9): 833-44. https://doi.org/10.1038/nbt.3935.

Ram, Geeta, John Chen, Krishan Kumar, Hope F. Ross, Carles Ubeda, Priyadarshan K. Damle, Kristin D. Lane, José R. Penadés, Gail E. Christie, et Richard P. Novick. 2012. « Staphylococcal Pathogenicity Island Interference with Helper Phage Reproduction Is a Paradigm of Molecular Parasitism ». *Proceedings of the National Academy of Sciences* 109 (40): 16300-305. https://doi.org/10.1073/pnas.1204615109.

Satchell, Karla J. F., Christopher J. Jones, Jennifer Wong, Jessica Queen, Shivani Agarwal, et Fitnat H. Yildiz. 2016. « Phenotypic Analysis Reveals That the 2010 Haiti Cholera Epidemic Is Linked to a Hypervirulent Strain ». *Infection and Immunity* 84 (9): 2473-81. https://doi.org/10.1128/IAI.00189-16.

Scholz, Matthias, Doyle V. Ward, Edoardo Pasolli, Thomas Tolio, Moreno Zolfo, Francesco Asnicar, Duy Tin Truong, Adrian Tett, Ardythe L. Morrow, et Nicola Segata. 2016. « Strain-Level Microbial Epidemiology and Population Genomics from Shotgun Metagenomics ». *Nature Methods* 13 (5): 435-38. https://doi.org/10.1038/nmeth.3802.

Seed, Kimberley D., Kip L. Bodi, Andrew M. Kropinski, Hans-Wolfgang Ackermann, Stephen B. Calderwood, Firdausi Qadri, et Andrew Camilli. 2011. « Evidence of a Dominant Lineage of

Vibrio Cholerae-Specific Lytic Bacteriophages Shed by Cholera Patients over a 10-Year Period in Dhaka, Bangladesh ». *MBio* 2 (1). https://doi.org/10.1128/mBio.00334-10.

Seed, Kimberley D., Shah M. Faruque, John J. Mekalanos, Stephen B. Calderwood, Firdausi Qadri, et Andrew Camilli. 2012. « Phase Variable O Antigen Biosynthetic Genes Control Expression of the Major Protective Antigen and Bacteriophage Receptor in Vibrio cholerae O1 ». Édité par Karla J. F. Satchell. *PLoS Pathogens* 8 (9): e1002917. https://doi.org/10.1371/journal.ppat.1002917.

Seed, Kimberley D., David W. Lazinski, Stephen B. Calderwood, et Andrew Camilli. 2013. « A Bacteriophage Encodes Its Own CRISPR/Cas Adaptive Response to Evade Host Innate Immunity ». *Nature* 494 (7438): 489-91. https://doi.org/10.1038/nature11927.

Seed, Kimberley D, Minmin Yen, B Jesse Shapiro, Isabelle J Hilaire, Richelle C Charles, Jessica E Teng, Louise C Ivers, Jacques Boncy, Jason B Harris, et Andrew Camilli. 2014. « Evolutionary Consequences of Intra-Patient Phage Predation on Microbial Populations. » *ELife* 3 (août): e03497. https://doi.org/10.7554/eLife.03497.

Seemann, Torsten. 2014. « Prokka: Rapid Prokaryotic Genome Annotation ». *Bioinformatics (Oxford, England)* 30 (14): 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
———. 2015. « Snippy: fast bacterial variant calling from NGS reads ». 2015. https://github.com/tseemann/snippy.

Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, et Jillian F. Banfield. 2018. « Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy ». *Nature Microbiology* 3 (7): 836-43. https://doi.org/10.1038/s41564-018-0171-1.

Stecher, Glen, Koichiro Tamura, et Sudhir Kumar. 2020. « Molecular Evolutionary Genetics Analysis (MEGA) for MacOS ». *Molecular Biology and Evolution* 37 (4): 1237-39. https://doi.org/10.1093/molbev/msz312.
The UniProt Consortium. 2019. « UniProt: A Worldwide Hub of Protein Knowledge ». *Nucleic Acids Research* 47 (D1): D506-15. https://doi.org/10.1093/nar/gky1049.
Tischler, Anna D., et Andrew Camilli. 2004. « Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation ». *Molecular microbiology* 53 (3): 857-69. https://doi.org/10.1111/j.1365-2958.2004.04155.x.

Vezzulli, L., C. Grande, G. Tassistro, I. Brettar, M. G. Höfle, R. P. A. Pereira, D. Mushi, A. Pallavicini, P. Vassallo, et C. Pruzzo. 2017. « Whole-Genome Enrichment Provides Deep Insights into Vibrio Cholerae Metagenome from an African River ». *Microbial Ecology* 73 (3): 734-38. https://doi.org/10.1007/s00248-016-0902-x.

Wang, Hui, Xiaolin Xing, Jipeng Wang, Bo Pang, Ming Liu, Jessie Larios-Valencia, Tao Liu, et al. 2018. « Hypermutation-Induced in Vivo Oxidative Stress Resistance Enhances Vibrio Cholerae Host Adaptation ». *PLOS Pathogens* 14 (10): e1007413. https://doi.org/10.1371/journal.ppat.1007413.

Weil, Ana A., Mohammad Arifuzzaman, Taufiqur R. Bhuiyan, Regina C. LaRocque, Aaron M. Harris, Emily A. Kendall, Azim Hossain, et al. 2009. « Memory T-Cell Responses to Vibrio Cholerae O1 Infection ». *Infection and Immunity* 77 (11): 5090-96. https://doi.org/10.1128/IAI.00793-09.

Weil, Ana A., Louise C. Ivers, et Jason B. Harris. 2011. « Cholera: Lessons from Haiti and Beyond ». *Current Infectious Disease Reports* 14 (1): 1-8. https://doi.org/10.1007/s11908-011-0221-9.

Weil, Ana A., Ashraful I. Khan, Fahima Chowdhury, Regina C. LaRocque, A. S. G. Faruque, Edward T. Ryan, Stephen B. Calderwood, Firdausi Qadri, et Jason B. Harris. 2009. « Clinical Outcomes in Household Contacts of Patients with Cholera in Bangladesh ». *Clinical Infectious Diseases* 49 (10): 1473-79. https://doi.org/10.1086/644779.

Weil, Ana A., et Edward T. Ryan. 2018. « Cholera: Recent Updates ». *Current Opinion in Infectious Diseases* 31 (5): 455–461. https://doi.org/10.1097/QCO.0000000000000474.
Wood, Derrick E., Jennifer Lu, et Ben Langmead. 2019. « Improved Metagenomic Analysis with Kraken 2 ». *Genome Biology* 20 (1): 257. https://doi.org/10.1186/s13059-019-1891-0.

Wu, Yu-Wei, Blake A. Simmons, et Steven W. Singer. 2016. « MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets ». *Bioinformatics* 32 (4): 605–607. https://doi.org/10.1093/bioinformatics/btv638.

Zeibell, Krystle, Sharon Aguila, Vivian Yan Shi, Andrea Chan, Hanjing Yang, et Jeffrey H. Miller. 2007. « Mutagenesis and Repair in Bacillus Anthracis: The Effect of Mutators ». *Journal of Bacteriology* 189 (6): 2331-38. https://doi.org/10.1128/JB.01656-06.

Zhao, Shijie, Tami D. Lieberman, Mathilde Poyet, Kathryn M. Kauffman, Sean M. Gibbons, Mathieu Groussin, Ramnik J. Xavier, et Eric J. Alm. 2019. « Adaptive Evolution within Gut Microbiomes of Healthy People ». *Cell Host & Microbe* 25 (5): 656-667.e8. https://doi.org/10.1016/j.chom.2019.03.007.