

Vibrio cholerae O1 experiences mild bottlenecks through the gastrointestinal tract in some but not all cholera patients

Patrick Lypaczewski,¹ Denise Chac,² Chelsea N. Dunmire,² Kristine M. Tandoc,² Fahima Chowdhury,³ Ashraf I. Khan,³ Taufiqur R. Bhuiyan,³ Jason B. Harris,^{4,5,6} Regina C. LaRocque,^{5,7} Stephen B. Calderwood,^{5,7} Edward T. Ryan,^{5,7,8} Firdausi Qadri,³ B. Jesse Shapiro,¹ Ana A. Weil^{2,9}

AUTHOR AFFILIATIONS See affiliation list on p. 8.

ABSTRACT *Vibrio cholerae* O1 causes the diarrheal disease cholera, and the small intestine is the site of active infection. During cholera, cholera toxin is secreted from *V. cholerae* and induces a massive fluid influx into the small intestine, which causes vomiting and diarrhea. Typically, *V. cholerae* genomes are sequenced from bacteria passed in stool, but rarely from vomit, a fluid that may more closely represent the site of active infection. We hypothesized that *V. cholerae* O1 population bottlenecks along the gastrointestinal tract would result in reduced genetic variation in stool compared to vomit. To test this, we sequenced *V. cholerae* genomes from 10 cholera patients with paired vomit and stool samples. Genetic diversity was low in both vomit and stool, consistent with a single infecting population rather than coinfection with divergent *V. cholerae* O1 lineages. The amount of single-nucleotide variation decreased from vomit to stool in four patients, increased in two, and remained unchanged in four. The variation in gene presence/absence decreased between vomit and stool in eight patients and increased in two. Pangenome analysis of assembled short-read sequencing demonstrated that the toxin-coregulated pilus operon more frequently contained deletions in genomes from vomit compared to stool. However, these deletions were not detected by PCR or long-read sequencing, indicating that interpreting gene presence or absence patterns from short-read data alone may be incomplete. Overall, we found that *V. cholerae* O1 isolated from stool is genetically similar to *V. cholerae* recovered from the upper intestinal tract.

IMPORTANCE *Vibrio cholerae* O1, the bacterium that causes cholera, is ingested in contaminated food or water and then colonizes the upper small intestine and is excreted in stool. Shed *V. cholerae* genomes from stool are usually studied, but *V. cholerae* isolated from vomit may be more representative of where *V. cholerae* colonizes in the upper intestinal epithelium. *V. cholerae* may experience bottlenecks, or large reductions in bacterial population sizes and genetic diversity, as it passes through the gut. Passage through the gut may select for distinct *V. cholerae* mutants that are adapted for survival and gut colonization. We did not find strong evidence for such adaptive mutations, and instead observed that passage through the gut results in modest reductions in *V. cholerae* genetic diversity, and only in some patients. These results fill a gap in our understanding of the *V. cholerae* life cycle, transmission, and evolution.

KEYWORDS cholera, whole-genome sequencing, comparative genomics, single-nucleotide variants, *Vibrio cholerae*, vomit, stool, population bottleneck, nanopore sequencing

Vibrio cholerae O1 causes the diarrheal disease cholera, and recent outbreaks are increasing in size and duration (1). In this context, genomic studies are increasingly

Editor Gaurav Sharma, Indian Institute of Technology Hyderabad, Hyderabad, India

Address correspondence to B. Jesse Shapiro, jesse.shapiro@mcgill.ca, or Ana A. Weil, anaweil@uw.edu.

Patrick Lypaczewski and Denise Chac contributed equally to this article. The order of listed authors was determined by consensus among the authors.

The authors declare no conflict of interest.

See the funding table on p. 8.

Received 4 April 2024

Accepted 13 May 2024

Published 25 June 2024

Copyright © 2024 Lypaczewski et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

conducted to gain an understanding of molecular epidemiology and evolving antimicrobial resistance. Although *V. cholerae* is a small intestinal pathogen, human clinical *V. cholerae* O1 genomes are typically sequenced from stool isolates. Gastric acidity kills many ingested *V. cholerae*; however, those that survive can then move into the small bowel (2). Here, *V. cholerae* can replicate, penetrate the mucin layer overlying the small bowel epithelium, and form microcolonies through the action of colonization factors including the toxin coregulated pilus (TCP) (3). *V. cholerae* also secrete cholera toxin that binds to intestinal epithelial cells and stimulates secretion of chloride, causing sodium and water to pass into the intestinal lumen, resulting in diarrhea, vomiting, and dehydration that may be severe. The massive fluid influx into the small intestine can overflow into the stomach and can result in vomiting; the majority of cholera patients experience vomiting during the course of illness (4). This is in contrast with other diseases in which gastric contents alone are vomited. Studies of cholera vomit demonstrate high *V. cholerae* cell counts, and the vomit pH levels approximate the small intestinal environment (5, 6).

V. cholerae O1 genomes generated from stool isolates reflect the *V. cholerae* population shed via the large intestine, but *V. cholerae* isolated from vomit may be more representative of the small intestinal *V. cholerae* population that is mediating infection. The genetic diversity of *V. cholerae* O1 could vary along the gastrointestinal (GI) tract for several reasons. First, genetic diversity could be reduced through population bottlenecks when a large infecting population is reduced to a smaller number of survivors due to bile and low gastric pH. In animal models, there is an estimated 40-fold reduction in the *V. cholerae* population from the oral inoculum compared to the site of small intestinal colonization (7). Second, directional natural selection might favor distinct strains from a mixed inoculum. Third, *de novo* mutations, gene transfer events, or gene losses might occur within a patient, some of which might confer adaptation to different niches along the GI tract.

Based on sequencing of *V. cholerae* O1 isolate genomes (8) and metagenomes (9, 10) from stool, we have previously found that coinfections by distinct strains of *V. cholerae* O1 in humans appear to be rare in the excreted population, but these could be more common in the inoculum, especially in hyperendemic areas. Similarly, the *V. cholerae* O1 population in stool from single individuals contains only minor point mutations and dozens of gene content variants. The level of genetic diversity of *V. cholerae* O1 in the upper GI tract during infection is not known.

To determine the genetic diversity of *V. cholerae* during transit through infected patients, we sequenced *V. cholerae* O1 genomes from paired vomit and stool samples from 10 cholera patients in Dhaka, Bangladesh, a region hyperendemic for cholera (Fig. 1A). We compared the single-nucleotide variants (SNVs) in these 200 genomes and examined variation in gene presence and absence. We show a modest decrease in *V. cholerae* SNVs and gene content variation between vomit and stool, suggesting that passage through the gut does not dramatically reduce genetic diversity or strain variation. We also provide evidence supporting the use of long-read sequencing technologies in accurately determining gene content variation.

METHODS

To examine within-host *V. cholerae* diversity, stool and vomit samples were collected between April 2018 and August 2019 in Dhaka City, Bangladesh, from patients aged 2–60 years with severe acute diarrhea and a stool culture positive for *V. cholerae* O1 who had no major comorbid conditions, as in prior studies (8, 9). All samples were *V. cholerae* serogroup O1 and serotype Ogawa, ascertained using slide agglutination testing using polyvalent and specific antisera, as in prior studies (11). Methods used to isolate *V. cholerae* from these vomit and stool samples have been previously described (5) (see Supplemental Methods). After incubation at 37°C for 24 hours, suspect *V. cholerae* colonies were selected and confirmed for the presence of *ctxA* and *V. cholerae* O1 *rfb* gene by PCR (12) to confirm identification. Twenty individual confirmed *V. cholerae* O1

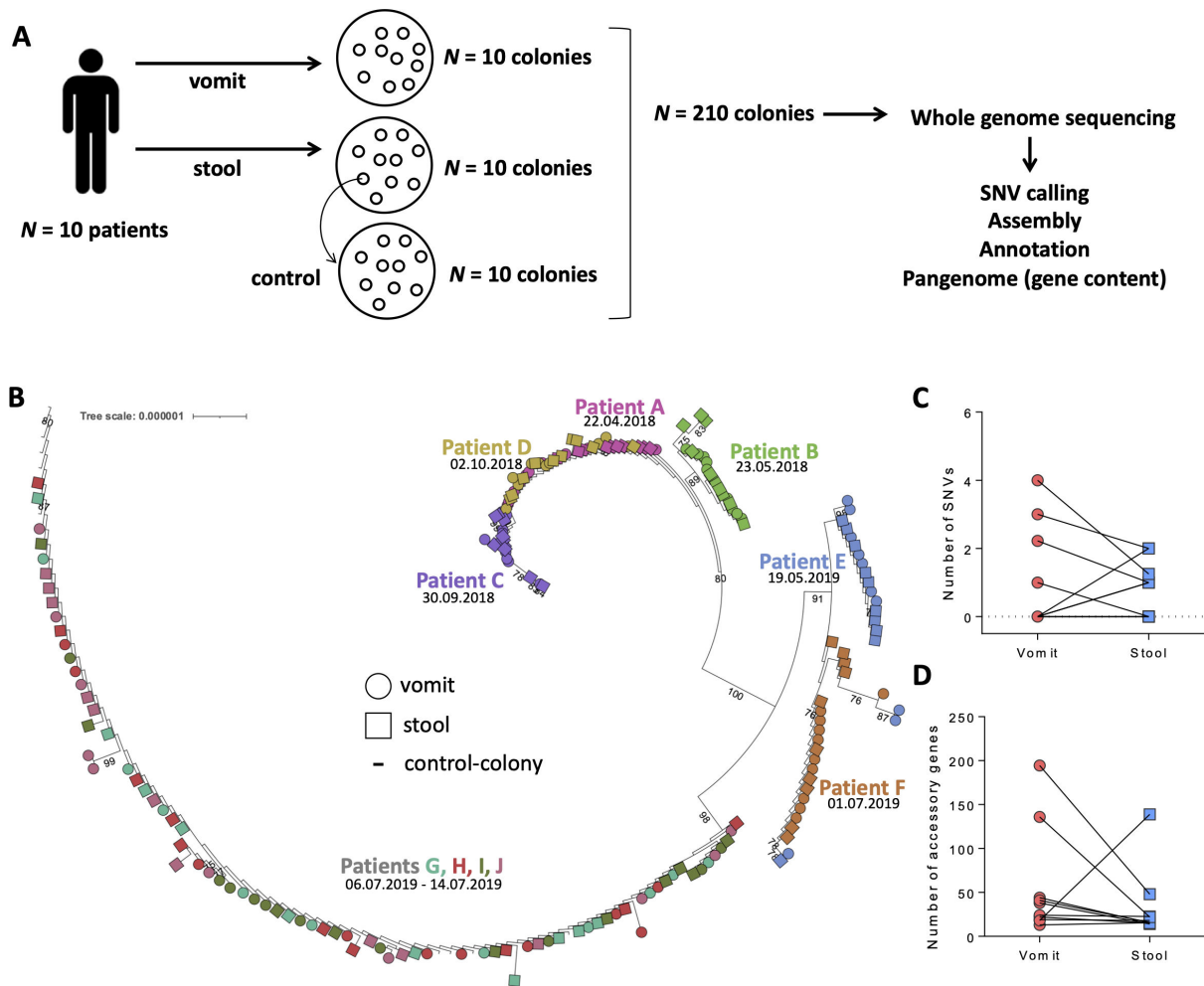


FIG 1 *V. cholerae* O1 within-patient diversity is modestly reduced in stool compared to vomit. (A) Ten patients were recruited for this study. For each patient, vomit and stool samples were plated to isolate colonies. From each vomit and stool sample, 10 colonies were used for whole-genome sequencing, in addition to 10 control colonies to control for sequencing errors. Sequence data were processed independently for single-nucleotide variant (SNV) calling and pangenome analysis (gene presence/absence) after assembly and annotations, with control colonies used to set filtering thresholds. (B) Maximum-likelihood phylogeny of 200 isolates sequenced, demonstrating that isolates cluster by patient and collection date. Ten control colony isolates are also shown. The tree is rooted on the MJ-1236 *V. cholerae* O1 reference genome (branching at the base of patient C). Samples are colored per patient. Collection dates are indicated below patient identifiers. Bootstrap values >75 are displayed. (C) Intra-patient variation based on SNVs called against the MJ-1236 reference genome across paired vomit and stool *V. cholerae* genomes, demonstrating a decrease in SNVs in four patients, an increase in three patients, and no change in three patients. The number of SNVs is normalized per 10 colonies, to account for some patients yielding only nine sequenced colonies. (D) Comparison of within-patient *V. cholerae* gene content in vomit and stool from 10 patients, showing a decrease in 8 patients and an increase in 2 patients. Only “accessory” genes that vary in their presence/absence in our data set are counted here; “core” genes common to all genomes are not included.

colonies from each patient (10 from vomit and 10 from stool) were inoculated into lysogeny broth (LB) and grown at 37°C while shaking aerobically overnight. For each colony, 1 mL of broth culture was stored at -80°C with 20% glycerol until DNA was extracted.

As a control for sequencing errors and mutations that could occur within culture rather than within patients, we selected one isolate from one cholera patient vomit sample. The glycerol stock of this isolate was streaked onto LB agar and 10 colonies were picked for whole-genome sequencing. These colonies were used in subsequent analyses as controls (Fig. 1A). DNA was extracted from each colony using standard methods (Supplemental Methods). Sequencing libraries were prepared with the Lucigen NxSeq AmpFREE kit, pooled and sequenced at the McGill Genome Centre on one lane of Illumina NovaSeq 6000 S Prime v.1.5 with paired-end 150-bp reads.

Either the *V. cholerae* O1 strain MJ-1236 (13) or a *de novo* assembly of the genome from the deeply sequenced colony control was used as a reference genome for analysis. To build a phylogeny, reads were processed using Snippy v.4.6.0 with default parameters (14), and the “snippy-core” command was used to generate a core SNV alignment. IQ-Tree v.2.2.2.7 was used to infer a maximum likelihood phylogenetic tree from this alignment (15). The TPM3u + F + I model was chosen by ModelFinder with bootstrap values determined by UFBoot (16) and rooted on the reference strain for display using iTOL (17). Demultiplexed paired-end reads were aligned to the reference genome using the Burrows-Wheeler aligner v.0.7.17 with the maximal exact match algorithm (18). The alignment files were transformed using SAMtools v.1.17 (19) to generate a pileup file. The variant calls were made using VarScan2 v.2.4.3 (20). Samples were excluded from SNV analyses if their breadth of coverage was two standard deviations or more below the median. SNVs within each patient were extracted using BCFtools v.1.13 with the command “bcftools isec -n-[#samples]”. Only SNVs at >90% frequency and read depth of >25 were included. These thresholds were established because they resulted in zero SNVs among the control colony genomes. The final list was manually inspected using Integrative Genomics Viewer v.2.16.0 (21) to remove any false positives resulting from poor mapping or instances in which one colony per group failed to be accurately called and appeared as intra-sample variation. In samples with less than 10 colonies remaining after quality filtering, the number of SNVs was normalized to the number of SNVs per 10 colonies (e.g., 1 of 9 becomes 1.1 of 10).

Isolated genomes were assembled using Unicycler v.0.4.9 (22) in Illumina-only mode. The resulting assemblies were quality controlled using checkM (23) to estimate genome completeness. Four genomes with a completeness score <100 (isolates BSC08, GSC06, HVC04, and JSC11) were removed from the pangenome analysis. Gene annotations were performed using Prokka v.1.14.5 (24) with the reference MJ-1236 proteome as an additional database using the “--proteins” additional argument to maintain consistent annotations and names. Pangenome determination was performed using Panaroo v.1.2.8 (25) using “--clean-mode sensitive” to retain the most genes found. We used the Cochran–Mantel–Haenszel test to measure systematic associations between gene presence and absence in vomit and stool, with each patient treated as an observation and the frequency of each gene’s presence in the vomit and stool derived from the Panaroo-generated pangenome table. To identify associations at the level of operons, we used the same implementation of the Cochran–Mantel–Haenszel test and counted observations at the level of operons rather than genes.

Following analysis using short-read data described above, we selected eight colonies with inferred gene content differences within the *tcp* operon for long-read re-sequencing using a MinION from Oxford Nanopore Technologies. DNA was re-extracted as above and prepared for sequencing using the Rapid Barcoding Kit (SQK-RBK004) according to manufacturer’s instructions to generate sequencing libraries. The libraries were sequenced on an R9.4.1 MinION flow cell. Raw sequencing data were basecalled and demultiplexed to FASTQ files using Guppy v.6.3.2 (Oxford Nanopore Technologies) using the model dna_r9.4.1_450bps_sup. Reads were assembled using Flye v.2.9.1 (26), and a pangenome analysis was performed as described above. The results for the *tcp* operon were manually inspected and compared to the short-read data.

Colony PCR was performed on *V. cholerae* O1 DNA extracted by boiling *V. cholerae* O1 isolates in molecular grade water at 95°C for 10 minutes. Taq 2× Master Mix (New England Biolabs) was used for the reaction, and PCR primers are listed in Table S1. PCR products were run on 1.0% agarose gels along an 1-kb ladder. Reference *V. cholerae* O1 strains were used to evaluate the *tcp* operon including PIC018, a clinical strain of *V. cholerae* O1 also isolated in Bangladesh (27) known to have an intact TCP, and a *tcpA* knockout mutant strain of a *V. cholerae* O1 clinical strain isolated in Haiti (28) gifted by Brandon Sit and Matthew Waldor.

RESULTS

From each of the 10 cholera patients, we isolated 10 *V. cholerae* colonies from vomit and 10 from stool for whole genome sequencing (Fig. 1A). We performed a phylogenetic analysis to examine relatedness of the isolates and found clustering by patient independent of sample type (Fig. 1B). The 10 control colonies grouped together, separated by very short branch lengths, indicating high-quality sequencing and low false-positive SNV identification (Fig. 1B). We observed a temporal signal in the phylogeny, with genomes isolated in 2018 and 2019 separated by a long branch with strong bootstrap support of 100. Several patients clustered together by time (e.g., A, B, C, D, and G, H, I, J), which may suggest common exposure during the same cholera outbreak (Fig. 1B). Instances in which more than one patient's isolates grouped together on the tree (e.g., patients E and F) were generally not well supported by bootstraps, making it difficult to reject a model with a single colonization event per patient.

We next focused on genetic variation within patients. Most samples yielded high breadth of coverage (median of 95% to the reference genome), and seven samples with low coverage were excluded (Fig. S1). Because we used multiple media types to isolate *V. cholerae* O1, we tested if the media type was associated with an increase or decrease in intra-sample variation and found no differences (Table S2). Within-patient diversity was generally low and consisted of low-frequency SNVs: 15 SNVs present in 1 of 10 colonies, and 4 SNVs present in 2 of 10 colonies. We therefore focused on the number of SNVs per vomit or stool sample rather than their frequencies.

If significant bottlenecks occur as *V. cholerae* O1 passed through the gut, we would expect less genetic variation in stool compared to vomit. Among the seven patients with detectable SNVs, four had fewer SNVs in stool compared to vomit, and three had more (full list of SNVs in Table S3). Three out of 10 patients had no detectable SNVs between vomit and stool isolates (Fig. 1C). Although the number of SNVs decreased from vomit to stool more often than it increased, this difference was not significant (one-sided binomial test, $P = 0.34$). To account for elements of the genome present only in our isolates and not in the MJ-1236 *V. cholerae* O1 reference strain, we repeated an identical analysis using a *de novo* genome assembly using a colony control, which yielded one additional SNV and otherwise identical results (Fig. S2; Table S4). Thus, we determined that our SNV calling was robust to the choice of reference genome.

In addition to reducing the diversity of point mutations (i.e., SNVs) in a population, bottlenecks would also be expected to reduce pangenome variation. To test this hypothesis, we compared *V. cholerae* gene content variation (e.g., presence or absence) in vomit and stool from the same patient. Based on genomes with high estimated completeness (see Supplemental Methods), we found a larger total gene content in vomit compared to stool in eight patients, and smaller in two patients (Fig. 1D; one-sided binomial test, $P = 0.055$). As previously observed (8), *V. cholerae* O1 gene content is more variable than SNVs within patients, potentially making it easier to detect a change in variation from vomit to stool. Together with the reduction in the number of SNVs from vomit to stool in more patients, these results are consistent with the hypothesis that bottlenecks may occur as *V. cholerae* O1 passes through the gut, but that bottlenecks are not evident in all patients and may produce only modest reductions in genetic variation.

We next considered the possibility that vomit and stool isolates experience different selective pressures that select for different genes in the *V. cholerae* O1 population. To identify genes potentially involved in niche adaptation, we tested for genes that varied within patients and were systematically associated with either vomit or stool across patients. We did not find any significant associations at the level of individual genes (Cochran–Mantel–Haenszel test, $P > 0.05$). Because our sample size was likely underpowered to identify gene-specific associations, and the deletion of any member of an operon was likely to disrupt the function of the entire operon, we grouped genes into annotated operons and found that all genes in the *tcp* operon were observed more often in stool than vomit (Cochran–Mantel–Haenszel test, $P = 0.002$). No other significant associations were found. The *tcp* operon encodes the toxin-coregulated pilus, a key bundle-forming

pilus factor that allows *V. cholerae* to colonize the small intestine (29, 30). Genes essential for human colonization could be stochastically lost from the *V. cholerae* O1 genome in the extra-human environment, and we could expect these loss events would be less common in isolates from stool than the vomit, because the small intestine is a strong selective filter for gut colonization.

To confirm these putative gene loss events, we aligned the raw reads for a set of eight isolates that varied in their presence or absence patterns of *tcp* genes to measure the breadth and depth of coverage of these genes. Genes with substantially reduced breadth of coverage were always called as absent from the pangenome, but the inverse was not always true (Fig. 2A and B). To further validate these apparent partial deletions, we performed a series of PCRs targeting portions of the genes (primer locations indicated on Fig. 2B). We were unable to detect any deletion within the *tcp* operon by PCR (Fig. 2C). To reconcile these conflicting data, we performed long-read re-sequencing on these eight isolates using Oxford Nanopore Technologies. Interestingly, all eight long-read re-sequenced isolates contained the entire *tcp* operon, confirming the results of the PCR analysis (Table S5). While these genes often contained frameshift mutations (Table S5), which might lead to truncated genes, the frameshifts always occurred as part of or immediately downstream of homopolymer sequences that are known to be error-prone in the long-read sequencing (31). These frameshifts are therefore likely sequencing artifacts. These validation steps suggest caution in interpreting gene presence/absence patterns from short-read data alone.

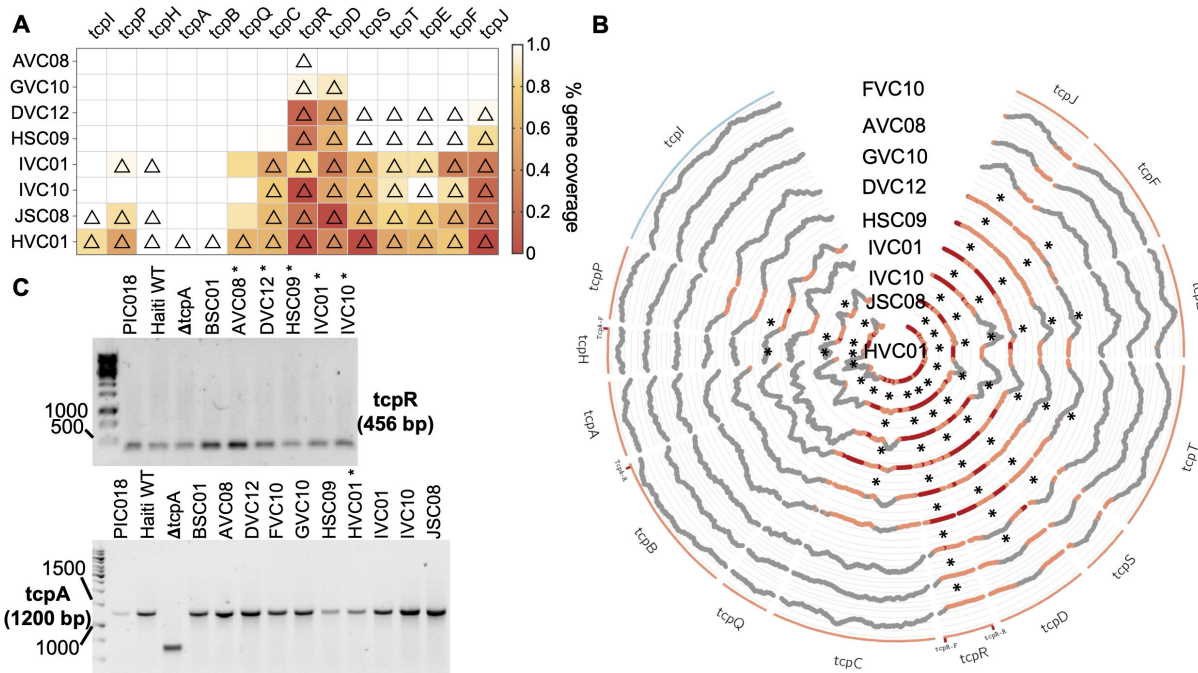


FIG 2 *tcp* genes absent from the *V. cholerae* O1 pangenome have low breadth and depth by short-read coverage and are identified as present by PCR. (A) The short-read coverage of *tcp* operon genes in eight genomes with variable presence/absence as determined by Panaroo was computed using bedtools and the raw reads aligned to the reference genome *tcp* sequences. Genes called as absent by Panaroo are marked with a triangle. The breadth of coverage of each gene is colored according to the scale shown on the right. (B) Circos plot showing coverage of *tcp* genes. Genes are colored according to the forward (blue) or reverse (orange) organization in the genome. The locations of the primers used in PCR assays are indicated on the genes as "gene-F/R." Gene coverage is plotted at each individual base and highlighted in orange at <10% and red at absolute zero. Genes called as absent by Panaroo are marked with an asterisk (*). (C) *tcp* genes assessed in study isolates using PCR. An asterisk (*) indicates expected absence of the gene according to the pangenome analysis. Clinical isolates of *V. cholerae* O1 known to have intact *tcp* and a $\Delta tcpA$ mutant were also used as controls. PCR primers are listed in Table S1. PCR products were run on 1% agarose gels with a 1-kb ladder.

DISCUSSION

V. cholerae O1 genomic studies in human disease have been based on *V. cholerae* recovered from the stool of patients with cholera, but isolates from vomit may better represent the *V. cholerae* O1 population at the site of active infection in the small intestine. Here, we examined the genetic variation between *V. cholerae* O1 recovered from the vomit versus the stool from patients with cholera in Bangladesh. We found an overall low level of genetic diversity overall. This suggests that bottlenecks between vomit and stool are not pronounced enough to reduce genetic diversity in the *V. cholerae* O1 population or that genetic diversity in the initial inoculum is so low that our sample size was insufficient to detect a difference. Additionally, vomiting typically occurs early in the course of infection (4, 32) but does not directly represent the infecting inoculum. Therefore, we cannot exclude a larger bottleneck occurring between the inoculum and the small intestinal *V. cholerae* population. Another explanation for the lack of divergence between *V. cholerae* populations from vomit and stool is that these *V. cholerae* populations are mixed during physiological processes. During high-volume fluid secretion by the small intestine, fluid may transit into large intestine and be excreted very rapidly, possibly on the order of minutes to a few hours, effectively homogenizing *V. cholerae* populations across the gut.

Our initial analysis using short-read sequencing suggested that deletions in the *tcp* operon were present more often in vomit than in stool isolates. Because vomit could include ingested environmental strains that may not encode a functional TCP, we thought it was plausible that TCP loss may be observed in vomit isolates more often than in stool isolates (33). Our results do not fully exclude the possibility that TCP may be sporadically lost by pandemic *V. cholerae* O1 strains in the environment, and these genomes would rarely be recovered from humans since they would have impaired ability to colonize the human intestine and survive transit through the GI tract. However, the TCP loss events detected using assembled short reads and pangenome analysis were not confirmed by PCR or long-read sequencing of a subset of genomes. Long-read sequencing of a larger number of genomes, matched with short-read data, could help establish if the discordant results are unique to certain genes like *tcp* or more general across the genome. The low depth and breadth of short-read coverage in many of the *tcp* genes suggests that this region of the genome may be difficult to sequence and assemble for technical reasons. It is also possible that our results could also represent within-colony variation in these genes; that is, most cells in a colony contain the intact *tcp* operon (as indicated by their detection by PCR and Nanopore sequencing), but a minority could contain deletions, detectable only by deep short-read sequencing. Such fine-scale variation in *tcp* could be a topic for future investigation. For the purposes of our study, we refrain from drawing firm conclusions regarding natural selection acting on TCP within patients, and we urge caution in interpretation of pangenome variation from short-read data alone.

While short-read Illumina sequencing is highly accurate, it seldom allows genomes to be completely assembled. In contrast, long-read sequencing produces reads with a lower individual accuracy but helps achieve closed genome assemblies with a clearer determination of a gene's presence or absence. Of note, while none of our sequenced genomes contained *tcp* deletions, they almost all contained frameshift mutations in at least one region of the *tcp* operon. These frameshifts could lead to truncated genes that might be identified as "absent" in the pangenome. However, a more likely explanation is that these frameshifts are due to sequencing errors. The frameshifts we detected always followed homopolymer repeats, which are known to be error prone in Oxford Nanopore sequencing (31). It is possible that the next generation of more accurate Nanopore flow cells (R10) combined with multiple rounds of genome polishing could resolve this issue in future studies.

In summary, we observed low within-patient diversity in *V. cholerae* O1 recovered from vomit and stool, consistent with prior studies examining only stool isolates. This indicates that *V. cholerae* O1 genomes isolated from stool are likely to represent the

population at the site of active infection. If population bottlenecks occur between the upper and lower GI tracts, they do not appear to have a large effect on *V. cholerae* O1 genetic diversity and are not universal across patients. We did identify a modest reduction in genetic diversity, particularly in pangenome diversity, in stool compared to vomit, consistent with a non-negligible role for bottlenecks, which could be explored in larger cohorts or time-series studies. Finally, we highlight that gene presence/absence observations based on short-read data should be treated with caution and confirmed by long-read sequencing or other complementary methods.

ACKNOWLEDGMENTS

We thank the patients of the International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), where these samples were collected, and the staff at the icddr,b for data entry and sample collection. We thank Matthew Waldor and Brandon Sit for the contribution of strains. The icddr,b gratefully acknowledges the government of the People's Republic of Bangladesh and Global Affairs Canada.

This work was supported by the National Institutes of Allergy and Infectious Diseases (R01AI106878 to E.T.R. and F.Q., R01AI103055 to J.B.H. and R.L., R01A1099243 to J.B.H. and F.Q., K08AI123494 to A.A.W., and T32HD007233 to D.C.), Fogarty International Center (D43TW005572 to T.R.B. and K43TW010362 to T.R.B.), the Government of the People's Republic of Bangladesh (to the icddr,b), Global Affairs Canada (to the icddr,b), the Swedish International Development Cooperation Agency (to the icddr,b), the Canadian Institutes of Health Research (project grant to B.J.S. and postdoctoral fellowship 187858 to P.L.), and the UK Department for International Development (to the icddr,b).

AUTHOR AFFILIATIONS

¹Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada

²Department of Medicine, University of Washington, Seattle, Washington, USA

³Infectious Diseases Division, International Center for Diarrheal Disease Research, Bangladesh, Dhaka

⁴Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts, USA

⁵Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts, USA

⁶Division of Global Health, Massachusetts General Hospital for Children, Boston, Massachusetts, USA

⁷Harvard Medical School, Boston, Massachusetts, USA

⁸Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA

⁹Department of Global Health, University of Washington, Seattle, Washington, USA

AUTHOR ORCIDs

Patrick Lypaczewski  <http://orcid.org/0000-0002-5550-712X>

Chelsea N. Dunmire  <http://orcid.org/0000-0003-0000-4282>

Fahima Chowdhury  <http://orcid.org/0000-0002-7082-0720>

Taufiqur R. Bhuiyan  <http://orcid.org/0000-0003-3755-4763>

Edward T. Ryan  <https://orcid.org/0009-0005-0169-394X>

B. Jesse Shapiro  <http://orcid.org/0000-0001-6819-8699>

Ana A. Weil  <http://orcid.org/0000-0002-6170-4306>

FUNDING

Funder	Grant(s)	Author(s)
HHS NIH National Institute of Allergy and Infectious Diseases (NIAID)		Denise Chac Jason B. Harris

Funder	Grant(s)	Author(s)
		Regina C. LaRocque Edward T. Ryan Firdausi Qadri Ana A. Weil
HHS NIH Fogarty International Center (FIC)		Taufiqur Bhuiyan
Ministry of Science and Technology, Government of the People's Republic of Bangladesh (MOST)		Fahima Chowdhury Ashrafal I. Khan Taufiqur Bhuiyan Firdausi Qadri
Canadian Government Global Affairs Canada (GAC)		Fahima Chowdhury Ashrafal I. Khan Taufiqur Bhuiyan Firdausi Qadri
Swedish International Development Cooperation Agency		Fahima Chowdhury Ashrafal I. Khan Taufiqur Bhuiyan Firdausi Qadri
Canadian Government Canadian Institutes of Health Research (CIHR)		Patrick Lypaczewski B. Jesse Shapiro
UK Department for International Development		Fahima Chowdhury Ashrafal I. Khan Taufiqur Bhuiyan Firdausi Qadri

DATA AVAILABILITY

The sequencing data generated for all 200 isolates and 10 colony controls were deposited in the National Center for Biotechnology Information Genbank under BioProject [PRJNA1046223](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1046223).

ETHICS APPROVAL

The ethical and research review committees of the icddr,b and the institutional review boards of Massachusetts General Hospital and the University of Washington approved the study. All adult subjects and parents/guardians of children provided written informed consent.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (Spectrum00785-24-s0001.docx). Fig. S1 to S2; Tables S1 to S5.

REFERENCES

- World Health Organization. 2023. Weekly epidemiological record, Vol. 98, p 431–452. <https://iris.who.int/handle/10665/372986>.
- Nelson EJ, Harris JB, Morris Jr JG, Calderwood SB, Camilli A. 2009. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat Rev Microbiol* 7:693–702. <https://doi.org/10.1038/nrmicro2204>
- Taylor RK, Miller VL, Furlong DB, Mekalanos JJ. 1987. Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci U S A* 84:2833–2837. <https://doi.org/10.1073/pnas.84.9.2833>
- Khan AI, Rashid MM, Islam MT, Afrad MH, Salimuzzaman M, Hegde ST, Zion MMI, Khan ZH, Shirin T, Habib ZH, Khan IA, Begum YA, Azman AS, Rahman M, Clemens JD, Flora MS, Qadri F. 2020. Epidemiology of cholera in Bangladesh: findings from nationwide hospital-based

- surveillance, 2014–2018. *Clin Infect Dis* 71:1635–1642. <https://doi.org/10.1093/cid/ciz1075>
5. Dunmire CN, Chac D, Chowdhury F, Khan AI, Bhuiyan TR, LaRocque RC, Akter A, Amin MA, Ryan ET, Qadri F, Weil AA. 2022. *Vibrio cholerae* isolation from frozen vomitus and stool samples. *J Clin Microbiol* 60:e0108422. <https://doi.org/10.1128/jcm.01084-22>
 6. Larocque RC, Harris JB, Dziejman M, Li X, Khan AI, Faruque ASG, Faruque SM, Nair GB, Ryan ET, Qadri F, Mekalanos JJ, Calderwood SB. 2005. Transcriptional profiling of *Vibrio cholerae* recovered directly from patient specimens during early and late stages of human infection. *Infect Immun* 73:4488–4493. <https://doi.org/10.1128/IAI.73.8.4488-4493.2005>
 7. Abel S, Abel zur Wiesch P, Chang H-H, Davis BM, Lipsitch M, Waldor MK. 2015. Sequence tag-based analysis of microbial population dynamics. *Nat Methods* 12:223–226. <https://doi.org/10.1038/nmeth.3253>
 8. Levade I, Terrat Y, Leducq J-B, Weil AA, Mayo-Smith LM, Chowdhury F, Khan AI, Boncy J, Buteau J, Ivers LC, Ryan ET, Charles RC, Calderwood SB, Qadri F, Harris JB, LaRocque RC, Shapiro BJ. 2017. *Vibrio cholerae* genomic diversity within and between patients. *Microb Genom* 3:e000142. <https://doi.org/10.1099/mgen.0.000142>
 9. Levade I, Khan AI, Chowdhury F, Calderwood SB, Ryan ET, Harris JB, LaRocque RC, Bhuiyan TR, Qadri F, Weil AA, Shapiro BJ. 2021. A combination of metagenomic and cultivation approaches reveals hypermutator phenotypes within *Vibrio cholerae*-infected patients. *mSystems* 6:e0088921. <https://doi.org/10.1128/mSystems.00889-21>
 10. Madi N, Cato ET, Sayeed MA, Creasy-Marrazzo A, Cuénod A, Islam K, Khabir MIU, Bhuiyan MTR, Begum YA, Freeman E, Vustepalli A, Brinkley L, Kamat M, Bailey LS, Basso KB, Qadri F, Khan AI, Shapiro BJ, Nelson EJ. 2024. Phage predation is a biomarker for disease severity and shapes pathogen genetic diversity in cholera patients. *bioRxiv*:2023.06.14.544933. <https://doi.org/10.1101/2023.06.14.544933>
 11. Qadri F, Wennerås C, Albert MJ, Hossain J, Mannoor K, Begum YA, Mohi G, Salam MA, Sack RB, Svennerholm AM. 1997. Comparison of immune responses in patients infected with *Vibrio cholerae* O139 and O1. *Infect Immun* 65:3571–3576. <https://doi.org/10.1128/iai.65.9.3571-3576.1997>
 12. Hoshino K, Yamasaki S, Mukhopadhyay AK, Chakraborty S, Basu A, Bhattacharya SK, Nair GB, Shimada T, Takeda Y. 1998. Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic *Vibrio cholerae* O1 and O139. *FEMS Immunol Med Microbiol* 20:201–207. <https://doi.org/10.1111/j.1574-695X.1998.tb01128.x>
 13. Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, Brettin TS, Bruce DC, Detter JC, Han CS, Chertkov O, Challacombe J, Huq A, Nair GB, Colwell RR. 2010. Genome sequence of hybrid *Vibrio cholerae* O1 MJ-1236, B-33, and CIR5101 and comparative genomics with *V. cholerae*. *J Bacteriol* 192:3524–3533. <https://doi.org/10.1128/JB.00040-10>
 14. Seemann T. 2015. Snippy: rapid haploid variant calling and core SNP phylogeny. <https://github.com/tseemann/snippy>.
 15. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
 16. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
 17. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
 18. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics*. <https://doi.org/10.48550/ARXIV.1303.3997>
 19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
 20. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22:568–576. <https://doi.org/10.1101/gr.129684.111>
 21. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. 2023. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* 39:btac830. <https://doi.org/10.1093/bioinformatics/btac830>
 22. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
 23. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
 24. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
 25. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 21:180. <https://doi.org/10.1186/s13059-020-02090-4>
 26. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
 27. Sayeed MA, Bufano MK, Xu P, Eckhoff G, Charles RC, Alam MM, Sultana T, Rashu MR, Berger A, Gonzalez-Escobedo G, Mandlik A, Bhuiyan TR, Leung DT, LaRocque RC, Harris JB, Calderwood SB, Qadri F, Vann WF, Kováč P, Ryan ET. 2015. A cholera conjugate vaccine containing O-specific polysaccharide (OSP) of *V. cholerae* O1 inaba and recombinant fragment of tetanus toxin heavy chain (OSP:rTTHc) induces serum, memory and lamina propria responses against OSP and is protective in mice. *PLoS Negl Trop Dis* 9:e0003881. <https://doi.org/10.1371/journal.pntd.0003881>
 28. Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK. 2011. The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364:33–42. <https://doi.org/10.1056/NEJMoa1012928>
 29. Herrington DA, Hall RH, Losonsky G, Mekalanos JJ, Taylor RK, Levine MM. 1988. Toxin, toxin-coregulated pili, and the toxR regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J Exp Med* 168:1487–1492. <https://doi.org/10.1084/jem.168.4.1487>
 30. Kaper JB, Morris JG, Levine MM. 1995. Cholera. *Clin Microbiol Rev* 8:48–86. <https://doi.org/10.1128/CMR.8.1.48>
 31. Delahaye C, Nicolas J. 2021. Sequencing DNA with nanopores: troubles and biases. *PLoS One* 16:e0257521. <https://doi.org/10.1371/journal.pone.0257521>
 32. Weil AA, Khan AI, Chowdhury F, Larocque RC, Faruque ASG, Ryan ET, Calderwood SB, Qadri F, Harris JB. 2009. Clinical outcomes in household contacts of patients with cholera in Bangladesh. *Clin Infect Dis* 49:1473–1479. <https://doi.org/10.1086/644779>
 33. Shapiro BJ, Levade I, Kovacicova G, Taylor RK, Almagro-Moreno S. 2016. Origins of pandemic *Vibrio cholerae* from environmental gene pools. *Nat Microbiol* 2:16240. <https://doi.org/10.1038/nmicrobiol.2016.240>