AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

BACTERIOLOGY

Evaluation of Whole-Genome Sequencing for Identification and Typing of *Vibrio cholerae*

David R. Greig,[a] Ulf Schaefer,[a] Sophie Octavia,[b,c] Ebony Hunter,[a,d] Marie A. Chattaway,[a] Timothy J. Dallman,[a] Claire Jenkins[a]

[a]National Infection Services, Public Health England, London, United Kingdom
[b]National Public Health Laboratory, Ministry of Health, Singapore, Singapore
[c]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia
[d]School of Pharmacy & Biomolecular Sciences, University of Brighton, Sussex, United Kingdom

ABSTRACT Epidemiological and microbiological data on *Vibrio cholerae* strains isolated between April 2004 and March 2018 ($n = 836$) and held at the Public Health England culture archive were reviewed. The traditional biochemical species identification and serological typing results were compared with the genome-derived species identification and serotype for a subset of isolates ($n = 152$). Of the 836 isolates, 750 (89.7%) were from a fecal specimen, 206 (24.6%) belonged to serogroup O1, and 7 (0.8%) were serogroup O139; 792 (94.7%) isolates were from patients reporting recent travel abroad, most commonly to India ($n = 209$) and Pakistan ($n = 104$). Of the 152 *V. cholerae* isolates identified by use of kmer, 149 (98.1%) were concordant with those identified using the traditional biochemical approach. Traditional serotyping results were 100% concordant with those of the whole-genome sequencing (WGS) analysis for the identification of serogroups O1 and O139 and classical and El Tor biotypes. *ctxA* was detected in all isolates of *V. cholerae* O1 El Tor and O139 belonging to sequence type 69 (ST69) and in *V. cholerae* O1 classical variants belonging to ST73. A phylogeny of isolates belonging to ST69 from U.K. travelers clustered geographically, with isolates from India and Pakistan located on separate branches. Moving forward, WGS data from U.K. travelers will contribute to global surveillance programs and the monitoring of emerging threats to public health and the global dissemination of pathogenic lineages. At the national level, these WGS data will inform the timely reinforcement of direct public health messaging to travelers and mitigate the impact of imported infections and the associated risks to public health.

KEYWORDS *Vibrio cholerae*, global surveillance, public health, whole-genome sequencing

Cholera is an acute diarrheal disease that can kill within hours if left untreated. Patients present with the passing of voluminous rice water stools, leading to severe dehydration (1). If hydration and electrolyte therapy is not quickly initiated, symptoms can rapidly progress to hypovolemic shock, acidosis, and death. Inadequate access to clean water and sanitation facilities is a driver of transmission, and outbreaks are common among displaced populations living in overcrowded conditions (2).

The bacterial pathogen responsible for the disease is *Vibrio cholerae*. *V. cholerae* serogroups O1 and O139 are regarded as pandemic strains and harbor the *ctx* genes associated with the production of the cholera toxin (3). *ctx* has also been detected in a limited number of other serogroups (4). Serogroup O1 can be divided into two biotypes, classical and El Tor. There are over 200 different lipopolysaccharide O antigens or serogroups of *V. cholerae*. The non-O1, non-O139 serogroups are

Received 21 May 2018 **Returned for modification** 12 July 2018 **Accepted** 13 August 2018

Accepted manuscript posted online 22 August 2018

Citation Greig DR, Schaefer U, Octavia S, Hunter E, Chattaway MA, Dallman TJ, Jenkins C. 2018. Evaluation of whole-genome sequencing for identification and typing of *Vibrio cholerae*. J Clin Microbiol 56:e00831-18. https://doi.org/10.1128/JCM.00831-18.

Editor Alexander Mellmann, University Hospital Münster

Address correspondence to Claire Jenkins, claire.jenkins@phe.gov.uk.

associated with a milder form of gastroenteritis, septicemia, and other extraintestinal infections (1, 3).

Seven cholera pandemics have occurred throughout the 19th and 20th centuries. The seventh (and current) pandemic began in Indonesia in 1961 and has spread via the Bay of Bengal (the current reservoir) to Africa and South America in at least three independent but overlapping waves of transmission (5). The fifth and sixth pandemics were caused by the *V. cholerae* serogroup O1 classical biotype, while the seventh pandemic was caused by serogroup O1 biotype El Tor. In 1992, the *V. cholerae* serogroup O139 caused a large epidemic in Bangladesh and India (6); however, *V. cholerae* O1 El Tor persists as the most commonly isolated *ctx*-positive serotype/biotype. Cholera is endemic across Africa, Latin America, and Asia, resulting in a large health care burden in developing countries (7–9). The World Health Organization states that there are 1.3 to 4 million estimated cases and 21,000 to 147,000 estimated deaths annually (7).

The U.K. Standards for Microbiology Investigations Investigation of Fecal Specimens for Enteric Pathogens recommends testing fecal specimens for *V. cholerae* in cases of suspected cholera, seafood consumption, and/or recent travel (within 2 to 3 weeks) to countries where cholera is endemic (see https://www.gov.uk/government/publications/smi-b-30-investigation-of-faecal-specimens-for-enteric-pathogens). Consequently, the true incidence of domestically acquired *V. cholerae* infection in the United Kingdom is unknown, and almost all isolates of enteric origin are associated with travelers' diarrhea.

In 2015, Public Health England (PHE) implemented whole-genome sequencing (WGS) for the routine surveillance of the more common gastrointestinal pathogens, including *Escherichia coli* and *Salmonella*, *Campylobacter*, *Shigella*, and *Listeria* species (10–12). The aim of this study was to review the historical PHE data on *V. cholerae* isolates held in the PHE culture archives and to compare the results of the traditional biochemical and serological typing methods with the genome-derived species identification and subtyping data required for the public health surveillance of *V. cholerae* infection.

MATERIALS AND METHODS

Epidemiological data. All *V. cholerae* isolates from human cases in England submitted to the Gastrointestinal Bacteria Reference Unit (GBRU) by local hospital laboratories between April 2004 and March 2018 were reviewed. Patient information, including, sex, age, and recent travel, was collected from laboratory request forms upon submission and stored in the Gastro Data Warehouse (GDW), an in-house PHE database for storing and linking patient demographic and microbiological typing data. Data on symptoms were limited, stating only that the patient had either gastrointestinal symptoms or an extraintestinal infection. There were no data on the severity of symptoms or patient outcomes.

Bacterial culture and traditional biochemistry and serology. Cultures were stored on cryobeads at −40°C or in nutrient agar stabs. For each sample, one cryobead was taken and inoculated into 20 ml 3% NaCl peptone water and incubated at 37°C for 18 h, shaking at 80 rpm. Cultures were plated from either nutrient agar slopes or 3% NaCl peptone water onto blood agar, MacConkey agar with salt (NaCl 1%), thiosulfate-citrate-bile salts (TCBS) agar, and cystine lactose electrolyte-deficient (CLED) agar and incubated at 37°C for 18 h.

Identification was performed by inoculating a combination of over 40 biochemical substrates as previously described (13). Utilization of the substrate was identified by a color change or gas production within the media. The ability (or inability) of the bacteria to utilize the substrates was recorded as a positive (or negative) result and compared with that of published tables categorizing the known reactions of *Vibrio* species (13). Isolates of *V. cholerae* were agglutinated with antisera raised to O1 (Ogawa and Inaba) and O139 (Bengal) antisera to determine the serogroup. The classical and El-Tor biotypes were differentiated by the Voges-Proskauer (VP) test (classical, negative; El-Tor, positive) and hemolysis on blood agar (classical, nonhemolytic; El-Tor, hemolytic).

Whole-genome sequencing analysis. All viable cultures of *V. cholerae* submitted to the GBRU between January 2015 and March 2018 ($n = 146$) were sequenced (see Table S1 in the supplemental material). In addition, 6 isolates of *V. cholerae* from the GBRU archive and 53 noncholera *Vibrio* species, including *V. alginolyticus* ($n = 4$), *V. campbellii* ($n = 1$), *V. cincinnatiensis* ($n = 1$), *V. fluvialis* ($n = 5$), *V. mediterranei* ($n = 1$), *V. metschnikovii* ($n = 3$), *V. mimicus* ($n = 3$), *V. natriegens* ($n = 1$), *V. parahaemolyticus* ($n = 32$), and *V. vulnificus* ($n = 2$) from the GBRU archive or the National Collection of Type Cultures were also sequenced (see Table S2 in the supplementary material). Genomic DNA was extracted, fragmented, and tagged for multiplexing with Nextera XT DNA sample preparation kits (Illumina) and sequenced using the Illumina HiSeq 2500 platform at PHE. FASTQ reads were quality trimmed using Trimmomatic (v0.36) (14), with bases removed from the trailing end that fell below a PHRED score of 30. If the read length posttrimming was less than 50 bp, the read and its pair were discarded using Trimmomatic.

FASTQ reads from all sequences in this study can be found at the PHE Pathogens BioProject at the National Center for Biotechnology Information (BioProject number PRJNA438219).

An approach based on kmers (short strings of DNA of length *k*; in this method, $k = 18$) was used to confirm the identity of the sample before organism-specific algorithms were applied (see https://github.com/phe-bioinformatics/kmerid) (15). Reference genomes ($n = 1781$) in 59 bacterial genera comprising the majority of human pathogens, commensal bacteria, and common contaminants were downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria. The kmer algorithm compared each sample to representative genomes in these 59 bacterial genera and returned the most similar genome together with a similarity estimate.

Sequence type (ST) assignment was performed using a modified version of short read sequence typing (SRST) using the multilocus sequence type (MLST) database described by Tewolde et al. (16). The MOST software (for MLST) is available at https://github.com/phe-bioinformatics/MOST. Any MLST gene sequences that did not match the existing alleles were submitted to pubMLST (see https://pubmlst.org/vcholerae/) for a new allelic type assignment. Similarly, new allelic profiles were also submitted to the database for a new ST assignment.

For the isolates belonging to clonal complex 69 (CC69), high-quality Illumina reads were mapped to a SPAdes v3.5.0 *de novo* assembly of the *V. cholerae* reference genomes (GenBank accession numbers NC_002505 and NC_002506) using BWA-MEM v0.7.3 and SAMtools v1.1 (17, 18). Single nucleotide polymorphisms (SNPs) were identified using GATK v2.6.5 (19) in unified Genotyper mode. Core genome positions, defined as those present in the reference genome and in at least 80% of the isolates, that had a high-quality SNP (>90% consensus; minimum depth, 10×; mapping quality [MQ], ≥30) in at least one isolate and invariant positions were extracted using SnapperDB v0.2.5 and processed through Gubbins v2.0.0 to identify and suppress recombination within the input to RAxML v8.1.17 (20, 21).

Using the GeneFinder tool (M. Doumith, unpublished data), FASTQ reads were mapped to the virulence regulator gene *toxR* (GenBank accession number KF498634), the cholera toxin gene *ctxA* (GenBank accession number AF463401), *wbe*O1, and *wbf*O139 (GenBank accession numbers KC152957 and AB012956) encoding the somatic O antigens O1 and O139, *tcpA* classical, and *tcpA* El Tor gene sequences (GenBank accession numbers M33514 and KP187623) using Bowtie 2 (22). The best match to each target was reported with metrics, including coverage, depth, and nucleotide similarity in XML format for quality assessment. *toxR* is found in all *V. cholerae* isolates and is regarded as a marker for species identification. The *ctxA* encoding cholera toxin is most often, although not exclusively, associated with *V. cholerae* serotypes O1 and O139 and is a characteristic of these pandemic lineages (23). Variants of *tcpA* can be used to identify the classical and El Tor biotypes (23). For *in silico* predictions, only results that matched to a gene determinant at >80% nucleotide identity over >80% target gene length were accepted.

Accession number(s). Sequences were deposited in the NCBI Sequence Read Archive under accession numbers SRR7062623 to SRR7062643 for *Vibrio cholerae* isolates and ERR832415, ERR841676, ERR841677, ERR987674, ERR1173484, ERR1246938, ERR1246942, ERR1766261, ERR1766262, ERR1787513, ERR1802420, ERR1854432, SRR7232582 to SRR7232611, and SRR7637792 to SRR7637802 (see Tables S1 and S2) for noncholera *Vibrio* sp. isolates.

RESULTS

Review of the historical data. Between April 2004 and March 2018, 836 isolates of *V. cholerae* from human cases in England were submitted to the GBRU by local hospital laboratories. On average, the number of isolates per year was 60, with the lowest number of isolates being reported in 2013 ($n = 29$) and the highest number in 2007 ($n = 80$) (Fig. 1). Of the 836 isolates, 206 (24.6%) belonged to serogroup O1 and 7 belonged to serogroup O139 (0.8%), and 750 isolates were from fecal specimens, 6 were from blood cultures, 2 were from ear swabs, and 2 were from eye swabs. No clinical data were available for the remaining 76 isolates.

Gender and age data were available for 828/836 and 773/836 cases, respectively. There were 424/836 males (50.7%) and 404/836 females (48.3%), and 685/836 (81.9%) were adults (aged 16 years or older) and 88 (10.5%) were children (<16 years old). Travel data were available for 796/836 cases, of which 792 reported recent travel abroad (<7 days before the onset of symptoms). For the cases infected with *V. cholerae* non-O1, non-O139, the most common travel destinations were India ($n = 140$), Kenya ($n = 57$), Thailand ($n = 40$), and Egypt ($n = 36$). The most commonly reported destinations of cases of *V. cholerae* O1 infection were Pakistan ($n = 72$) and India ($n = 69$). The six cases of *V. cholerae* O139 infection had traveled to Thailand ($n = 2$), China, India, Jordan, and Pakistan. The four cases who stated they had not recently traveled abroad had *V. cholerae* non-O1, non-O139 isolates obtained from extraintestinal sites (blood cultures, $n = 2$; eye swab, $n = 1$; ear swab, $n = 1$).

Whole-genome sequencing. One hundred fifty-two isolates of *V. cholerae* were sequenced, including those belonging to serogroup O1 ($n = 47$), serogroup O139 ($n =$

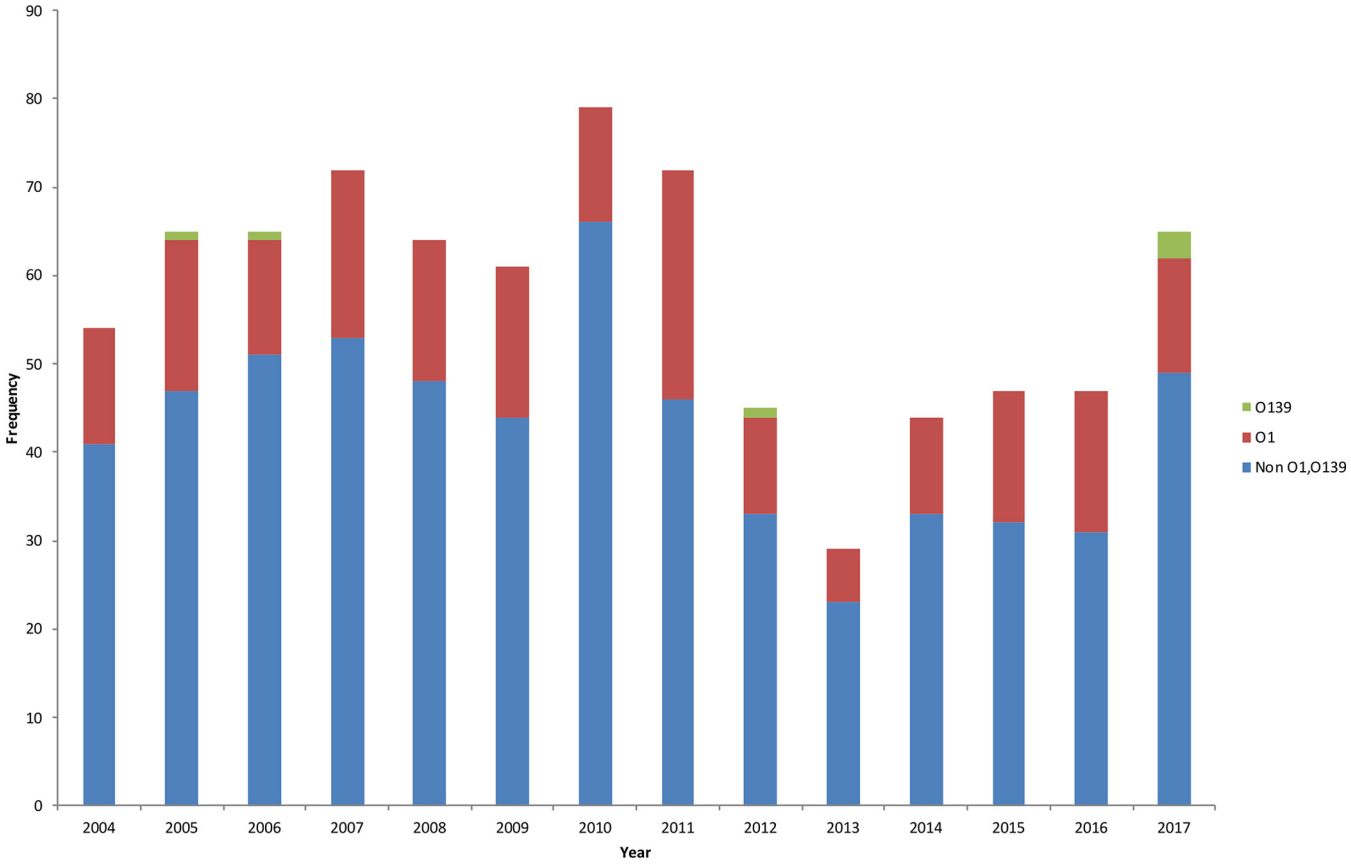

FIG 1 Number of *V. cholerae* isolates from human cases in England submitted to the GBRU by local hospital laboratories each year between March 2004 and December 2017 ($n = 836$). Blue, non-O1, non-O139 serogroups; red, serogroup O1; green, serogroup O139.

7), and those designated serogroup non-O1, non-O139 ($n = 98$) (see Table S1). One hundred thirty-eight isolates were from human cases, of which 133 were from fecal specimens from hospitalized or community cases with symptoms of gastrointestinal disease, three isolates were from ear swabs, one was from an eye swab, and one was from a blood culture from a patient with acute cholecystitis. The remaining 14 isolates were from animals ($n = 4$), food ($n = 1$), or environmental samples ($n = 3$) or were isolates from the GBRU archive ($n = 6$), including three isolates of *V. cholerae* serogroup O1 belonging to the classical biotype (see Table S1).

Kmer identification. Of the 152 *V. cholerae* isolates identified using the kmer approach, 149 (98.1%) were concordant with the traditional biochemical identification. The kmer method identified three unusual external quality assessment isolates from an obscure environmental source, previously identified as *V. cholerae*. Following further investigation, these isolates were identified as outliers in terms of genomic diversity within *V. cholerae* strains and were not well represented within the set of reference genomes in the kmer identification database. Therefore, the similarity values obtained for these isolates were lower than those for the *V. cholerae* isolates obtained from human, animal, and food specimens/samples. The kmer method provided the correct species identification for 53 isolates belonging to noncholera *Vibrio* species (see Table S2).

Use of GeneFinder for serotyping and detection of virulence genes. The *toxR* gene was detected in 144/152 isolates of *V. cholerae* (Table 1). Eight isolates identified as *V. cholerae* by traditional biochemical tests were negative for *toxR*. Further analysis of the sequence data showed the sequence coverage/similarity of *toxR* for the discrepant isolates fell just below the 80% threshold (74% and 77%) (Table 1). However, all eight isolates were identified as *V. cholerae* by the kmer approach.

TABLE 1 Summary of GeneFinder profiles and ST results

GeneFinder profile	ST(s)[a]	No. of isolates
toxR, *wbe*O1, *tcp* classical, *ctxA*	73	3
toxR, *wbe*O1, *tcp* El Tor, *ctxA*	69	34
toxR, *wbe*O1, *tcp* El Tor	75, 169, 579 (2)	4
toxR, *wbe*O1, *ctxA*	167	1
toxR, *wbe*O1	521(2), 551 (2)	4
toxR, *wbf*O139, *tcp* El Tor, *ctxA*	69	1
toxR, *wbf*O139	163, 527, 529, 544, 568	5
toxR (77%), *wbf*O139	586	1
toxR	>70 STs	94
toxR (77%)	539, 540, 541, 550, 585, 587, 600	7
	Total	152

[a]Values in parentheses are the number of isolates belonging to individual STs.

Traditional biochemistry and serotyping results were 100% concordant with the WGS analysis for the identification of O1 and O139 and classical and El Tor biotypes. Of the 37 isolates of *V. cholerae* O1 El Tor, 33 had the *wbe*O1, *tcpA* El Tor variant and *ctxA*, and 4 had the *wbe*O1, *tcpA* El Tor variant without *ctxA* (Table 1). There were five isolates that belonged to *V. cholerae* serogroup O1 that were negative for *tcpA* El Tor variant and *ctxA*, and one that was negative for *tcpA* El Tor variant but had *ctxA* (Table 1). There were only three classical variant strains in this study (Sequence Read Archive accession numbers SRR7062555, SRR7062601, and SRR7062603), all from the GBRU archive (see Table S1), as the current pandemic is caused by *V. cholerae* O1 El Tor (7–9, 24, 25). All three *V. cholerae* O1 classical strains had *wbe*O1 and the *tcpA* classical variant gene. *V. cholerae* O139 was also rare in this data set (26). Although all seven isolates of *V. cholerae* O139 had *wbf*O139, only the isolate from the GBRU archive (SRR7062531) had the *tcpA* El Tor variant and *ctxA* (Table 1; see also Table S1).

ctxA was detected in all isolates of *V. cholerae* O1 El Tor and O139 belonging to ST69 (27) and *V. cholerae* O1 classical variants belonging to ST73 (24). Four isolates of *V. cholerae* O1 were negative for *ctxA*, and the six recently obtained *V. cholerae* O139 isolates were *ctxA* negative.

Sequence typing. Sequence typing data were available for 152 isolates. The *V. cholerae* O139 El Tor-positive isolate and the 34 isolates of *V. cholerae* O1 El Tor that were *ctxA* belonged to ST69. The four *V. cholerae* O1 El Tor *ctxA*-negative isolates belonged to ST75, ST169, and ST579 ($n = 2$) and all fell within CC69. The three *V. cholerae* O1 classical isolates belonged to ST73.

The six isolates of *V. cholerae* O1 without the *tcpA* El Tor variant were ST167, ST521 ($n = 2$), ST551 ($n = 2$), and ST611. There were six isolates that had the O139 antigen but were negative for the *tcpA* El Tor variant gene and *ctxA*, and these belonged to ST163, ST527, ST529, ST544, ST568, and ST586. All *V. cholerae* O1 isolates that had the *tcpA* El Tor variant, regardless of the presence or absence of *ctxA*, belonged to the CC69 cluster, and those without the tcpA El Tor variant gene were dispersed across the population (Fig. 2).

The remaining 95 isolates of *V. cholerae* non-O1, non-O139 ($n = 95$) and *V. cholerae* O139 ($n = 3$), belonged to over 70 different STs (see Table S1). There was one major cluster of 18 isolates of *V. cholerae* non-O1, non-O139, designated CC558, comprising 15 difference STs (Fig. 2). Isolates belonging to this cluster were submitted to GBRU between 2015 from U.K. residents reporting travel to the Caribbean ($n = 6$), Southeast Asia ($n = 6$), Africa (n-2), Europe ($n = 2$), and Central America ($n = 1$) and one patient who reported recent travel but did not specify the destination (see Table S1). No clinical details or additional epidemiological information was linked to these cases.

SNP typing. As previously described, the pandemic *V. cholerae* O1 and O139 El Tor *ctxA* strains all belonged to ST69, whereas the classical biotype and *ctxA*-negative strains of *V. cholerae* O1 belonged to other STs within CC69. A phylogeny of ST69, the pandemic lineage comprising isolates from this study and sequences available in public

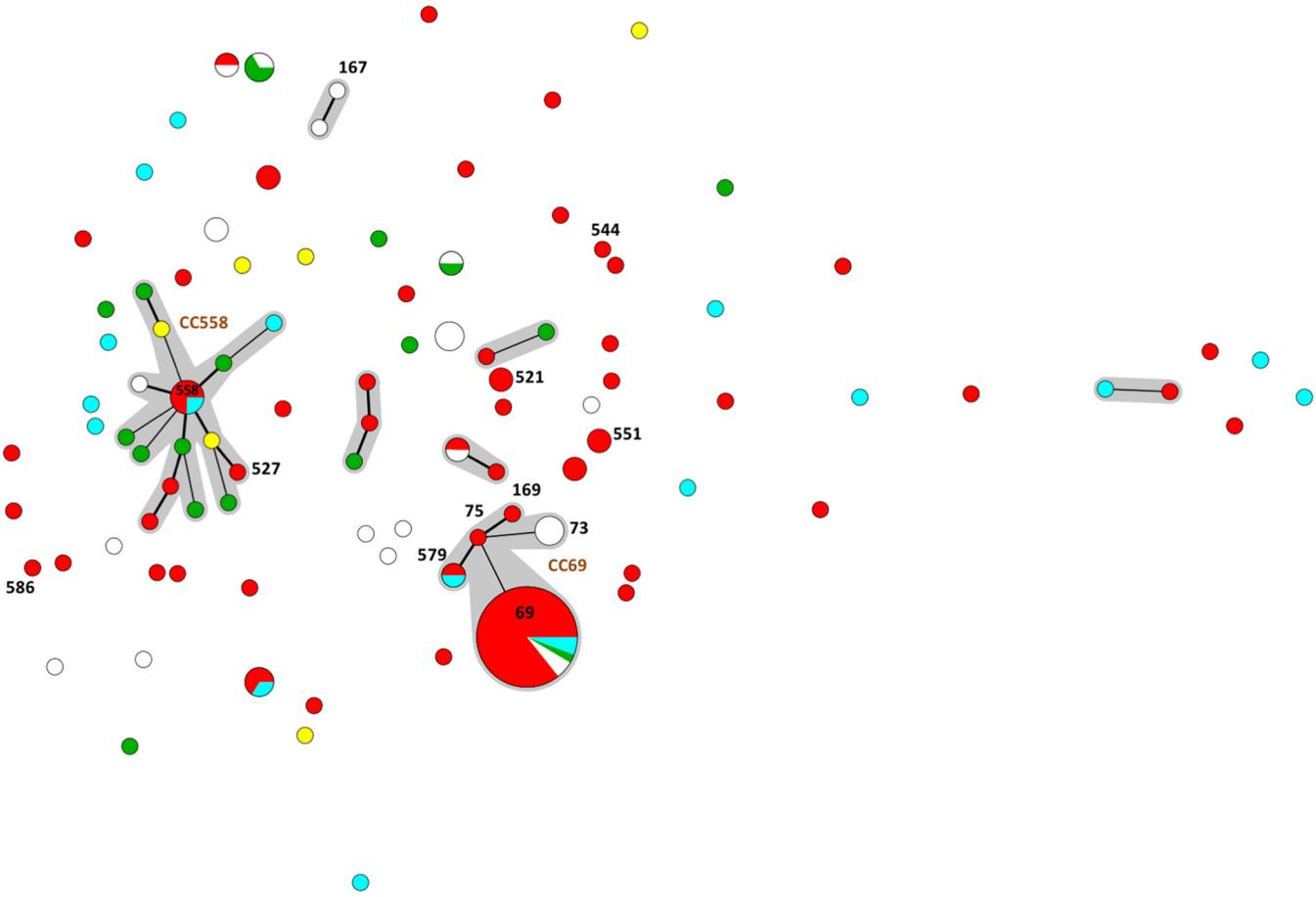

FIG 2 Minimum spanning tree illustrating the diversity in the population structure of the *V. cholerae* isolates received at PHE between January 2015 and March 2018. Clonal complexes (CC) comprising strains linked by a single locus variant (thick black line) or double locus variant (thin black line) and are shaded gray. Sequence types (ST) are shown in black. Isolates associated with cases reporting recent travel abroad are highlighted: red, Asia; blue, Africa; green, Latin America; yellow, mainland Europe; white.

databases was constructed (see Fig. S1 in the supplementary material). The isolates from U.K. travelers clustered geographically with those returning from India located on the same branch, and those reporting recent travel to Pakistan clustered on a separate branch. Further analysis based on single nucleotide polymorphisms in the core genome compared with a reference strain may be performed for outbreak detection and source attribution where the incidence of the current *V. cholerae* O1 El Tor pandemic lineage (ST69) is high (28, 29).

DISCUSSION

Historically, traditional biochemistry, biotyping, phage typing, and serology results were useful for confirming identification at the species level, typing of serogroups O1 and O139, and for identifying the classical and El Tor variants. Molecular tests for the detection of *ctxA* were not performed. In this study, the review of the historical GBRU data revealed that just less than one quarter of the *V. cholerae* isolates belonged to serogroup O1 and *V. cholerae* O139 was rarely detected (24). Due to limited resources, neither serotyping of the non-O1, non-O139 serogroups nor molecular typing of any serogroup was performed at GBRU. Therefore, prior to the implementation of WGS, it was not possible to monitor trends in emerging pathogenic lineages or to gain insight into modes of transmission for this important gastrointestinal pathogen.

Studies have shown that MLST data are an accurate, robust, reliable, high-throughput typing method that is well suited to routine public health surveillance

(11, 12, 16). For *V. cholerae*, MLST provides insight on the true evolutionary relationship between isolates, as well as a framework for fine level typing for public health surveillance (30–33). Using the STs derived from the genome data, we were able to analyze the population structure of all isolates of *V. cholerae* submitted to GBRU for the first time and identified a globally disperse clonal complex, CC558, comprising isolates of *V. cholerae* designated non-O1, non-O139. With the exception of CC69, globally disseminated clonal groups within the species *V. cholerae* are rare.

Previous studies have shown the emergence and potential spread of ST75, which belongs to the pandemic clonal complex CC69 and may pose a significant threat to public health despite testing negative for *ctxA* (34). Continued epidemiological surveillance is required to further understand the epidemic potential of *ctxA*-negative STs that are part of CC69. Currently, the correlation of ST with geography in our data set is hindered by the limited size of the data set. However, moving forward, this unprecedented level of strain discrimination available for all isolates of *V. cholerae* submitted to the GBRU will enhance our understanding of global transmission and emerging threats to public health, for the pandemic strains belonging to CC69, and the non-O1 serogroup lineages.

Global monitoring of *V. cholerae* infection is hindered by the limitations of the surveillance systems in countries where people are most at risk. The World Health Organization recommends that cholera surveillance should be part of an integrated disease surveillance system that includes feedback at the local level and information sharing at the global level. WGS linked to cases of travelers' diarrhea caused by *V. cholerae* have the potential to be a useful public health resource for global surveillance, enabling us to track the emergence and dissemination of specific lineages on a global scale (5, 8, 35). At the local level, sharing of WGS data linked to these cases could result in the timely reinforcement of direct public health messaging to travelers in order to reduce the number of imported infections and mitigate the impact of imported infections and associated risks to public health (36).

Traditional biochemistry and serotyping results were concordant with the WGS analysis for the identification of *V. cholerae* O1 and serotyping and biotyping of O1 and O139 serogroups. Moreover, using the WGS approach, species-level identification, serotyping, biotyping, determining the presence of cholera toxin, and ST and SNP typing of CC69 can all be derived from a single process workflow. WGS data may also be investigated for additional virulence factors and antimicrobial resistance determinants. The data of all *V. cholerae* genomes sequenced at PHE are publically available under the NCBI BioProject number PRJNA438219 in order to facilitate public health surveillance and monitoring of the global transmission of the pandemic lineages by the international scientific community.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JCM.00831-18.

SUPPLEMENTAL FILE 1, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 2, PDF file, 4.0 MB.

ACKNOWLEDGMENTS

This study was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections (number 109524).

Claire Jenkins and Timothy J. Dallman are affiliated with the NIHR HPRU in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with the University of East Anglia, University of Oxford, and the Quadram Institute. Claire Jenkins and Timothy J. Dallman are based at PHE.

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

REFERENCES

1. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. 2012. Cholera. Lancet 379:2466–2476. https://doi.org/10.1016/S0140-6736(12)60436-X.
2. Jutla A, Khan R, Colwell R. 2017. Natural disasters and cholera outbreaks: current understanding and future outlook. Curr Environ Health Rep 4:99–107. https://doi.org/10.1007/s40572-017-0132-5.
3. Islam MT, Alam M, Boucher Y. 2017. Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. Int Microbiol 20:106–115.
4. Crowe SJ, Newton AE, Gould LH, Parsons MB, Stroika S, Bopp CA, Freeman M, Greene K, Mahon BE. 2016. Vibriosis, not cholera: toxigenic *Vibrio cholerae* non-O1, non-O139 infections in the United States, 1984–2014. Epidemiol Infect 144:3335–3341. https://doi.org/10.1017/S0950268816001783.
5. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, Niyogi SK, Kim EJ, Ramamurthy T, Chun J, Wood JL, Clemens JD, Czerkinsky C, Nair GB, Holmgren J, Parkhill J, Dougan G. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477:462–465. https://doi.org/10.1038/nature10392.
6. Albert MJ. 1996. Epidemiology & molecular biology of *Vibrio cholerae* O139 Bengal. Indian J Med Res 104:14–27.
7. Clemens JD, Nair GB, Ahmed T, Qadri F, Holmgren J. 2017. Cholera. Lancet 390:1539–1549. https://doi.org/10.1016/S0140-6736(17)30559-7.
8. Domman D, Quilici ML, Dorman MJ, Njamkepo E, Mutreja A, Mather AE, Delgado G, Morales-Espinosa R, Grimont PAD, Lizárraga-Partida ML, Bouchier C, Aanensen DM, Kuri-Morales P, Tarr CL, Dougan G, Parkhill J, Campos J, Cravioto A, Weill FX, Thomson NR. 2017. Integrated view of *Vibrio cholerae* in the Americas. Science 358:789–793. https://doi.org/10.1126/science.aao2136.
9. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, Keddy KH, Salje H, Moore S, Mukhopadhyay AK, Bercion R, Luquero FJ, Ngandjio A, Dosso M, Monakhova E, Garin B, Bouchier C, Pazzani C, Mutreja A, Grunow R, Sidikou F, Bonte L, Breurec S, Damian M, Njanpop-Lafourcade BM, Sapriel G, Page AL, Hamze M, Henkens M, Chowdhury G, Mengel M, Koeck JL, Fournier JM, Dougan G, Grimont PAD, Parkhill J, Holt KE, Piarroux R, Ramamurthy T, Quilici ML, Thomson NR. 2017. Genomic history of the seventh pandemic of cholera in Africa. Science 358:785–789. https://doi.org/10.1126/science.aad5901.
10. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, Petrovska L, Ellis RJ, Elson R, Underwood A, Green J, Hanage WP, Jenkins C, Grant K, Wain J. 2015. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. Clin Infect Dis 61:305–12. https://doi.org/10.1093/cid/civ318.
11. Ashton PM, Nair S, Peters TM, Bale JA, Powell DG, Painset A, Tewolde R, Schaefer U, Jenkins C, Dallman TJ, de Pinna EM, Grant KA; Salmonella Whole Genome Sequencing Implementation Group. 2016. Identification of Salmonella for public health surveillance using whole genome sequencing. PeerJ 4:e1752. https://doi.org/10.7717/peerj.1752.
12. Chattaway MA, Greig DR, Gentle A, Hartman HB, Dallman TJ, Jenkins C. 2017. Whole-genome sequencing for national surveillance of *Shigella flexneri*. Front Microbiol 8:1700. https://doi.org/10.3389/fmicb.2017.01700.
13. Bergey DH, Krieg NR, Holt JG. 1984. Bergey's manual of systematic bacteriology. Williams & Wilkins, Baltimore, MD.
14. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
15. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. 2017. Identification of *Escherichia coli* and *Shigella* species from whole-genome sequences. J Clin Microbiol 55:616–623. https://doi.org/10.1128/JCM.01790-16.
16. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, Ellington M, Swift C, Green J, Underwood A. 2016. MOST: a modified MLST typing tool based on short read sequencing. PeerJ 4:e2308. https://doi.org/10.7717/peerj.2308.
17. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.
20. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
21. Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, Shaaban S, Hartman H, Myers R, Underwood A, Jenkins C, Grant K. 2018. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. Bioinformatics https://doi.org/10.1093/bioinformatics/bty212.
22. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
23. Greig DR, Hickey TJ, Boxall MD, Begum H, Gentle A, Jenkins C, Chattaway MA. 2018. A real-time multiplex PCR for the identification and typing of *Vibrio cholerae*. Diagn Microbiol Infect Dis 90:171–176. https://doi.org/10.1016/j.diagmicrobio.2017.11.017.
24. Mukhopadhyay AK, Takeda Y, Balakrish Nair G. 2014. Cholera outbreaks in the El Tor biotype era and the impact of the new El Tor variants. Curr Top Microbiol Immunol 379:17–47.
25. Siddique AK, Cash R. 2014. Cholera outbreaks in the classical biotype era. Curr Top Microbiol Immunol 379:1–16.
26. Ghosh R, Sharma NC, Halder K, Bhadra RK, Chowdhury G, Pazhani GP, Shinoda S, Mukhopadhyay AK, Nair GB, Ramamurthy T. 2016. Phenotypic and genetic heterogeneity in *Vibrio cholerae* O139 isolated from cholera cases in Delhi, India during 2001–2006. Front Microbiol 7:1250.
27. Anandan S, Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Thangamani S, Veeraraghavan B. 2017. Prevailing clone (ST69) of *Vibrio cholerae* O139 in India over 10 years. Gut Pathog 9:60. https://doi.org/10.1186/s13099-017-0210-0.
28. Ramamurthy T, Sharma NC. 2014. Cholera outbreaks in India. Curr Top Microbiol Immunol 379:49–85.
29. Shah MA, Mutreja A, Thomson N, Baker S, Parkhill J, Dougan G, Bokhari H, Wren BW. 2014. Genomic epidemiology of *Vibrio cholerae* O1 associated with floods, Pakistan, 2010. Emerg Infect Dis 20:13–20. https://doi.org/10.3201/.eid2001.130428.
30. Kotetishvili M, Stine OC, Chen Y, Kreger A, Sulakvelidze A, Sozhamannan S, Morris JG, Jr. 2003. Multilocus sequence typing has better discriminatory ability for typing *Vibrio cholerae* than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. J Clin Microbiol 41:2191–2196. https://doi.org/10.1128/JCM.41.5.2191-2196.2003.
31. Lam C, Octavia S, Reeves PR, Lan R. 2012. Multi-locus variable number tandem repeat analysis of 7th pandemic *Vibrio cholerae*. BMC Microbiol 12:82. https://doi.org/10.1186/1471-2180-12-82.
32. Octavia S, Salim A, Kurniawan J, Lam C, Leung Q, Ahsan S, Reeves PR, Nair GB, Lan R. 2013. Population structure and evolution of non-O1/non-O139 *Vibrio cholerae* by multilocus sequence typing. PLoS One 8:e65342. https://doi.org/10.1371/journal.pone.0065342.
33. Siriphap A, Leekitcharoenphon P, Kaas RS, Theethakaew C, Aarestrup FM, Sutheinkul O, Hendriksen RS. 2017. Characterization and genetic variation of *Vibrio cholerae* isolated from clinical and environmental sources in Thailand. PLoS One 12:e0169324. https://doi.org/10.1371/journal.pone.0169324.
34. Luo Y, Octavia S, Jin D, Ye J, Miao Z, Jiang T, Xia S, Lan R. 2016. US Gulf-like toxigenic O1 *Vibrio cholerae* causing sporadic cholera outbreaks in China. J Infect 72:564–72. https://doi.org/10.1016/j.jinf.2016.02.005.
35. Chowdhury FR, Nur Z, Hassan N, von Seidlein L, Dunachie S. 2017. Pandemics, pathogenicity and changing molecular epidemiology of cholera in the era of global warming. Ann Clin Microbiol Antimicrob 16:10. https://doi.org/10.1186/s12941-017-0185-1.
36. Neilson AA, Mayer CA. 2010. Cholera—recommendations for prevention in travellers. Aust Fam Physician 39:220–226.