Vital Surveillances

Phylogenetic Analysis of Serogroup O5 *Vibrio cholerae* that Caused Successive Cholera Outbreaks — Guangdong Province, China, 2020–2021

Bixia Ke[1,2,&]; Bo Pang[3,&]; Dongmei He[1,2]; Jing Xu[1,2]; Qiuxia Chen[1,2]; Junhua Liang[1,2]; Jialiang Chen[3]; Zhenpeng Li[3]; Haijian Zhou[3]; Xiaoling Deng[1,2,#]; Biao Kan[3,#]

ABSTRACT

Introduction: Gastroenteritis caused by non-O1/non-O139 *Vibrio cholerae* exhibited an increasing trend in recent years in China. Whole genome sequence (WGS) data could play an important role both in the identification of the outbreaks and in the determination of the serogroup. Here, we present the employment of WGS data in the investigation of two outbreaks caused by non-O1/non-O139 *V. cholerae* in Guangdong, China, 2020–2021.

Methods: We obtained the whole genome sequence of 66 *V. cholerae* strains isolated in two outbreaks with next generation sequencing technology. We retrieved the publicly available WGS data of non-O1/non-O139 *V. cholerae* from public database. We used a pipeline integrated in China Pathogen Identification Net (PIN) to complete the phylogenetic analysis.

Results: Two outbreaks caused by non-O1/non-O139 *V. cholerae* were identified using WGS data. These *V. cholerae* strains were determined as serogroup O5. Type 3 and 6 secretion systems were detected in these serogroup O5 strains. These serogroup O5 strains belonged to sequence type (ST) 88.

Conclusions: Our analysis indicated the risk of non-O1/non-O139 *V. cholerae* leading to outbreaks of diarrheal diseases. The application of genomic data played an important role in the identification of the serogroup of non-O1/non-O139 *V. cholerae* in the lack of antiserum, which gave an example of the application of genome data in disease surveillance and public health emergency response.

Vibrio cholerae consists of more than 200 serogroups. The classification of serogroups is based on the O antigen of the lipopolysaccharide (LPS) (*1*). The classical method of serogroup determination is based on the immune agglutination reaction between the O antigen and the corresponding specific antiserum. The molecular mechanisms of different serogroups are based on the variation in structure of O-antigen polysaccharide (O-PS) coding sequence (*2*). Therefore, the phenotype of O-antigen is correlated with the molecular type of O-PS coding sequence. Till now, only serogroup O1 and O139 *V. cholerae* caused cholera epidemics and pandemics (*3*). *V. cholerae* does not belong to serogroup O1 and O139 and is designated as "non-O1/non-O139" *V. cholerae*. Usually, these non-O1/non-O139 *V. cholerae* only cause sporadic infections and seldomly cause outbreaks (*4*). Several kinds of toxins, such as a heat-stable toxin, cholera toxin, and other enterotoxins, have been detected in the non-O1/non-O139 *V. cholerae* that caused an outbreak. Except for the toxins, secretion systems, for example type 3 secretion system (T3SS) and type 6 secretion system, have been detected in some *V. cholerae* strains that caused cholera outbreaks (*5*).

In China, toxigenic serogroup O1 and O139 *V. cholerae* strains were rarely isolated after 2010 (*6*). In contrast, sporadic cholera cases even small scale of outbreaks caused by non-O1/non-O139 *V. cholerae* were reported from time to time (*4*). Here we report successive cholera outbreaks caused by non-toxin-producing serogroup O5 *V. cholerae* in 2020 and 2021 in Guangdong Province, China.

METHODS

Acquiring Specimens

Stools or anal swabs from patients and asymptomatic persons were collected. Environment samples in the kitchen and environmental water were collected. Oral consent was obtained from each eligible patient.

Strain Culture and DNA Extraction

All 66 strains sequenced in the study were recovered on Luria-Bertani (LB) agar (in a −80 °C freezer). A single colony was transferred to LB broth and was incubated at 37 °C with shaking at 200 revolutions per minute (RPM). Genomic DNA was extracted from overnight cultures with a Qiagen kit (QIAamp Mini Kit, Hilden, Germany) according to the manufacturer's instructions.

Library Preparation

A total amount of 0.2 μg DNA per sample was used as input material for the DNA library preparations. Sequencing library was generated using NEB Next® Ultra™ DNA Library Prep Kit for Illumina (NEB, Ipswich, USA) following the manufacturer's recommendations and index codes were added to each sample. Briefly, genomic DNA sample was fragmented by sonication to a size of 350 bp. Then DNA fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further polymerase chain reaction (PCR) amplification. After PCR products were purified by AMPure XP system (Beckman Coulter, Beverly, USA), DNA concentration was measured by Qubit® 3.0 Flurometer (Invitrogen, Carlsbad, USA), and libraries were analyzed for size distribution by NGS3K/Caliper and quantified by real-time PCR (3 nM).

Clustering & Sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using Illumina PE Cluster Kit (Illumina, San Diego, USA) according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced on Illumina NovaSeq6000 platform and 150 bp paired-end reads were generated.

Genome Assembly and Identification of Single Nucleotide Variants (SNVs)

Short reads were assembled *de novo* into contigs and scaffolds and a whole-genome alignment for all of the genomes investigated was built in this study. The whole-genome sequence of N16961 was used as the reference to call SNVs.

RESULTS

Description of the Two Outbreaks

Gastroenteritis outbreaks were reported in Qingyuan City, Guangdong Province in 2020 and 2021. The gastroenteritis outbreak in 2020 occurred on September 9 and lasted for 18 days in a vocational-technical college. A total of 137 cases were reported. In these cases, 97.7% of them complained of mild diarrheal and 62.3% complained of abdominal pain. A total of 303 samples from patients, asymptomatic persons, environment samples in the kitchen, and environmental water were collected and sent to the laboratory for screening of norovirus, rotavirus, *Salmonella*, pathogenic *E. coli*, Shigella, *Campylobacter jejuni*, *Vibrio parahaemolyticus*, and *Vibrio cholerae*. *V. cholerae* strains were isolated from 32 samples. In 2021, another outbreak occurred during the period from September 28 to September 30 in a vocational-technical college, which is about 2 kilometers away from the former one in Qingyuan City. A total of 79 patients in this college were reported to have diarrhea. A total of 47 anal swabs of the case-patients were collected and sent to the diagnostic laboratory for detection of norovirus, rotavirus, *Salmonella*, pathogenic *E. coli*, *Shigella*, *Campylobacter jejuni*, *Vibrio parahaemolyticus*, and *Vibrio cholerae*. Of the 47 samples, 34 anal swabs of the case-patients were positive for *V. cholerae* by PCR assay. All of the strains did not agglutinate with either O1 or O139 specific antiserum. Therefore, we determined that these strains were non-O1/non-O139 *V. cholerae*. In both of these two outbreaks, water has been suspected as the infection source; however, no *V. cholerae* was isolated from water samples.

Phylogenetic Analysis of *V. cholerae* Isolated in the Two Outbreaks

A maximum likelihood (ML) tree was reconstructed based on the SNVs identified in the non-repetitive and non-recombinant core-genome of 40 publicly available representative *V. cholerae* and 66 *V. cholerae* isolated in these two outbreaks using IQ-TREE. It showed that 97% (64/66) of the *V. cholerae* isolated in these two outbreaks formed a tight cluster and the left two were far away from this cluster and differed from each other (Figure 1A). In order to investigate the detailed relationship among *V. cholerae* in these two outbreaks, we built an ML tree based on the SNVs in the non-repetitive and non-recombinant core-genome of the 64 *V. cholerae*. The *V. cholerae* isolated in 2021 formed 2 clusters that included 25 and 8 strains, respectively, and the left one was in a cluster including another 6 *V. cholerae* isolated in 2020. We also detected the

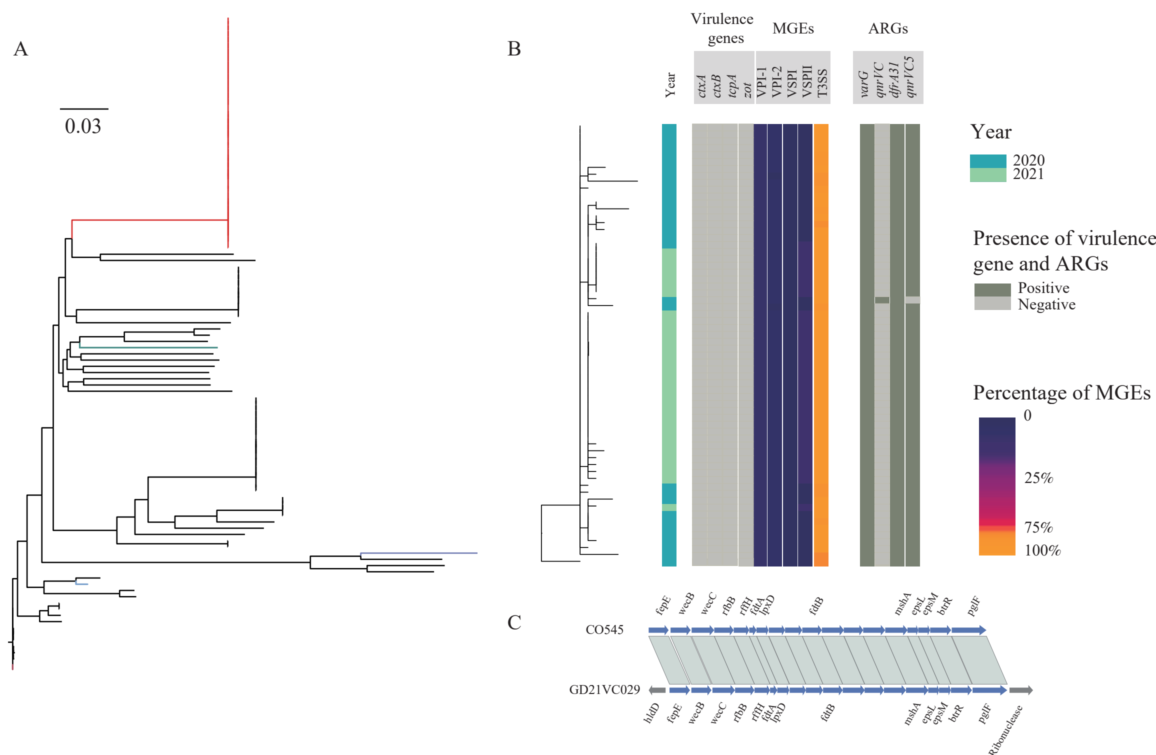

FIGURE 1. Phylogenetic analysis and molecular characteristics of serogroup O5 *V. cholerae* isolated in two outbreaks (A) Phylogenetic tree of non-O1/non-O139 *V. cholerae*. (B) Phylogenetic tree of serogroup O5 *V. cholerae*. (C) Illustration of the O-PS biosynthesis gene locus in *V. cholerae* isolated in these two outbreaks.

Notes: Maximum likelihood (ML) tree of 66 non-O1/non-O139 *V. cholerae* isolated in the 2 outbreaks, and the publicly available representative *V. cholerae* was constructed based on the SNVs in the core genome. Those isolated in this study were in red. This ML tree was constructed based on the SNVs in the non-repetitive and non-recombinant core-genome of 106 non-O1/non-O139 *V. cholerae*. O1 El Tor strain N16961 was used as reference. The ML tree was constructed on the SNVs identified in the non-repetitive, non-recombinant core-genome of the 66 most closely-related serogroup O5 *V. cholerae*. The presence and absence of virulence genes and mobile genetic elements usually detected in pandemic *V. cholerae* and antibiotic resistance genes were annotated on the right of the ML tree. Comparison between O-PS biosynthesis gene locus in serogroup O5 *V. cholerae* CO545 and that in a representative strain which was isolated in the outbreak in 2021.

Abbreviations: MGEs=mobile genetic element; ARGs=antibiotic resistance genes; SNVs=single nucleiotide variation.

presence of the virulence genes and pathogenic islands that were common in the pandemic clones of *V. cholerae*. It showed that none of the cholera toxin genes (*ctxA* and *ctxB*), the CTXΦ, the Vibrio pathogenic island (VPI)1 and 2, and the Vibrio seventh pandemic island I and II (VSPI and VSPII) were detected in all of the 64 *V. cholerae* (Figure 1B). However, we detected T3SS in all of the 64 *V. cholerae* (Figure 1B). Moreover, we detected *qnrVC5*, *dfrA31*, and *varG* in almost all of the 64 strains except that *qnrVC*, instead of *qnrVC5*, was detected in 1 strain isolated in 2020. BLAST analysis indicated that *qnrVC5* and *dfrA31* were located on an integron. In order to determine the serogroup, we tried to agglutinate these strains with O1 and O139 antigenic serum, it turned out that these strains were not O1 or O139 serogroup. The PCR result gave the same results. As we could not get the antigenic serum of other serogroups of *V. cholerae*, we extracted the O-PS sequence of these strains and did BLAST analysis against GenBank. The results showed that the 64 *V. cholerae* in the same tight cluster in Figure 1A showed best hits to the O-PS sequence of a serogroup O5 *V. cholerae* (Figure 1C). Therefore, we concluded that these 64 strains were serogroup O5. As to the left 2 *V. cholerae*, we could not determine their serogroups. We also determined that the ST of these serogroup O5 *V. cholerae* was ST88, and we could not get the ST of the left 2 non-O5 genomes as no information was available in the database.

DISCUSSION

One important characteristic of the 2 cholera outbreaks was that they were reported in 2 successive

years in almost the same place at almost the same time. The genomic analysis indicated that those strains isolated in two outbreaks were closely related. *V. cholerae* is a natural habitat of estuary water, especially in places rich in plankton (*1*). These pieces of information indicated that an unknown contaminated source existed. Therefore, enhanced surveillance, which focuses on food and environmental water, should be carried out next year.

Till now, the determination of serogroup of *V. cholerae* mainly relies on the agglutination reaction. However, due to the poor availability and cost of serogroup-specific serum, it is difficult to get information on the serogroup of *V. cholerae* other than O1 and O139. As the price of whole-genome sequencing decreased rapidly in these years, it is affordable to carry out WGS of the representative strains that caused outbreaks (*7*). Moreover, the component and structure of the O-PS coding sequences determined the various serogroups (*2*). Therefore, it is possible to establish a correlation between the serogroups and the O-PS sequences. A similar study of *Salmonella* has shown consistency between traditional serovar typing and STs (*8*). This could be achieved by the WGS of a reference collection of representative strains of different serogroups of *V. cholerae*. Ultimately, a molecular-serogroup scheme could replace the existing serogroup typing scheme.

The reported cholera cases decreased dramatically after the year 2010 in China, which is an indicator that cholera has been under control in this country (*6*). The cholera cases caused by pathogenic O1 and O139 *V. cholerae* were rare since the year 2010; however, those caused by non-toxigenic O1 and O139 or non-O1/non-O139 *V. cholerae* were reported occasionally. For example, cases caused by non-O1/O139 *V. cholerae* were reported every year since 2012 in Guangdong (data not shown). Therefore, the strategy for cholera prevention and control in China nowadays should be adjusted and attention should also be paid to non-toxigenic O1 and O139 or non-O1/non-O139 *V. cholerae*, particularly those that carry T3SS and/or T6SS (*9*). As the non-toxigenic *V. cholerae*, especially those carrying the TCP gene cluster and those carrying pre-CTXΦ, could be converted to toxigenic strains by transduction, more attention should also be paid to such strains.

Conflicts of interest: No conflicts of interest.

Funding: Guangdong Collaborative Innovation Center for Prevention and Control of Cross-border Transmission of Infectious Diseases (Project No. 2018B020207013).

doi: 10.46234/ccdcw2022.061

Corresponding authors: Xiaoling Deng, dengxiaoling@cdcp.org.cn; Biao Kan, kanbiao@icdc.cn.

[1] Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, Guangdong, China; [2] Guangdong Workstation for Emerging Infectious Disease Control and Prevention, Guangzhou, Guangdong, China; [3] State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China.
[&] Joint first authors.

Submitted: January 11, 2022; Accepted: March 22, 2022

REFERENCES

1. Kaper JB, Morris JG Jr, Levine MM. Cholera. Clin Microbiol Rev 1995;8(1):48 – 86. http://dx.doi.org/10.1128/CMR.8.1.48.
2. Chatterjee SN, Chaudhuri K. Lipopolysaccharides of *Vibrio cholera*: II. Genetics of biosynthesis. Biochim Biophys Acta 2004;1690(2):93 – 109. http://dx.doi.org/10.1016/j.bbadis.2004.06.006.
3. Pang B, Yan MY, Cui ZG, Ye XF, Diao BW, Ren YH, et al. Genetic diversity of toxigenic and nontoxigenic *Vibrio cholerae* serogroups O1 and O139 revealed by array-based comparative genomic hybridization. J Bacteriol 2007;189(13):4837 – 49. http://dx.doi.org/10.1128/JB.01959-06.
4. Xiao Y, Li BS, Xiao N, Liu X, Tan HL, He DM, et al. Drug resistance and molecular characteristics of non-O1/O139 *Vibrio cholerae* isolated from diarrhea cases in Guangdong province between 2011 and 2013. Chin J Public Health 2015;31(1):32 – 5. http://dx.doi.org/10.11847/zgggws2015-31-01-10. (In Chinese).
5. Arteaga M, Velasco J, Rodriguez S, Vidal M, Arellano C, Silva F, et al. Genomic characterization of the non-O1/non-O139 *Vibrio cholerae* strain that caused a gastroenteritis outbreak in Santiago, Chile, 2018. Microb Genom 2020;6(3):e000340. http://dx.doi.org/10.1099/mgen.0.000340.
6. Pang B, Kan B. Cholera control and prevention in the seventh pandemic in China. Dis Surveill 2021;36(9):869 – 72. http://dx.doi.org/10.3784/jbjc.202109190528. (In Chinese).
7. Kan B, Zhou HJ, Du PC, Zhang W, Lu X, Qin T, et al. Transforming bacterial disease surveillance and investigation using whole-genome sequence to probe the trace. Front Med 2018;12(1):23 – 33. http://dx.doi.org/10.1007/s11684-017-0607-7.
8. Achtman M, Wain J, Weill FX, Nair S, Zhou ZM, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog 2012;8(6):e1002776. http://dx.doi.org/10.1371/journal.ppat.1002776.
9. Kan B. New problems in monitoring of cholera. Infect Dis Inf 2006;19(1):16 – 7. http://dx.doi.org/10.3969/j.issn.1007-8134.2006.01.007. (In Chinese).