

## Phylogenetic Analysis Revealed the Dissemination of Closely Related Epidemic *Vibrio cholerae* O1 Isolates in Laos, Thailand, and Vietnam

Masatomo Morita,<sup>1,9</sup> Kazuhisa Okada,<sup>2,3</sup> Tetsu Yamashiro,<sup>4,5</sup> Tsuyoshi Sekizuka,<sup>6</sup> Amonrattana Roobthaisong,<sup>2</sup> Warawan Wongboot,<sup>2,7</sup> Siriporn Chantaroj,<sup>7</sup> Nguyen Dong Tu,<sup>8</sup> Phonepadith Xangsayath,<sup>9</sup> Noikaseumsy Sithivong,<sup>9</sup> Khambai Noilath,<sup>9</sup> Arounapha Vongdouangchanh,<sup>9</sup> Makoto Kuroda,<sup>6</sup> Shigeeyuki Hamada,<sup>3</sup> Hidemasa Izumiya,<sup>1</sup> and Makoto Ohnishi<sup>1</sup>

<sup>1</sup>Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan,

<sup>2</sup>Thailand-Japan Research Collaboration Center on Emerging and Re-emerging Infections, Nonthaburi, Thailand, <sup>3</sup>Research Institute for Microbial Diseases, Osaka University, Osaka, Japan, <sup>4</sup>Department of Bacteriology, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan, <sup>5</sup>Nagasaki University, Nagasaki, Japan, <sup>6</sup>Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan, <sup>7</sup>National Institute of Health, Department of Medical Sciences, Ministry Public Health, Nonthaburi, Thailand, <sup>8</sup>Bacteriology Department, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam, <sup>9</sup>National Center for Laboratory and Epidemiology, Vientiane, Lao People's Democratic Republic

We performed whole-genome sequencing of *Vibrio cholerae* O1 isolates from Laos, Thailand, and Vietnam, where cholera outbreaks occurred, to determine their genetic lineages. Core genome phylogenetic analysis revealed that the isolates located in same lineage without regional clusters, which suggests that closely related strains circulated in Southeast Asia.

**Keywords.** phylogeny; Southeast Asia; *Vibrio cholerae* O1; whole-genome sequencing.

Cholera is an acute, watery diarrheal disease caused by a toxigenic strain of the *Vibrio cholerae* O1 and O139 serogroups that produces a cholera toxin. It remains an important public health issue in countries where access to safe water and adequate sanitation is limited [1]. The annual global burden of cholera was estimated at 2.9 million cases in 69 countries where cholera is endemic, with 95 000 deaths [2]. In Southeast Asia, a total of 3543 cholera cases were reported in Lao People's Democratic Republic (PDR), Thailand, and Vietnam in 2007 [3], and cholera outbreaks were also reported in each country during this same period [4–8]. In addition, intranational molecular epidemiology revealed the temporal and geographical distributions of *V cholerae* serotype O1 Ogawa harboring the classical

cholera toxin B subunit gene (*ctxB1*) in these 3 countries [7–9]. Descriptive analysis of cholera outbreaks also showed that *V cholerae* O1 Ogawa was distributed in Southeast Asia [10]. It was suggested that the epidemic *V cholerae* O1 clone caused an international outbreak; however, there has been no molecular epidemiological evidence to support this claim to date. Thus, the objective of the present study was to use whole-genome sequencing to determine the genetic lineages of *V cholerae* strains found in Lao PDR, Thailand, and Vietnam.

### METHODS

We collected a total of 214 *V cholerae* O1 isolates from Lao PDR, Thailand, and Vietnam from 1998 to 2011 to infer their phylogenetic relationships (Supplementary Table 1). Genomic deoxyribonucleic acid (DNA) was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). Genomic libraries were prepared by the Nextera XT DNA Library Preparation Kit (Illumina) and sequenced on the HiSeq 2500 (Illumina) sequencer or MiSeq (Illumina) sequencer. After genome assembly was performed using SPAdes version 3.13.1 with the --careful option and a read coverage cutoff value of 10 [11], single-nucleotide variants (SNVs) were identified with Snippy version 4.3.6 (<https://github.com/tseemann/snippy>) using the *V cholerae* O1 El Tor N16961 genome (DDBJ/ENA/GenBank accession numbers LT907989 and LT907990) as a reference. We excluded the SNVs in recombinogenic regions detected using Gubbins version 2.3.4 [12], along with the SNVs in the repeat and prophage regions of the N16961 genome, which were identified using NUCmer [13] and PHAST [14] for core-genome phylogeny. A maximum likelihood phylogenetic tree was generated using RAxML version 8.2.0 [15] with 1000 bootstrap iterations. Global collection of 1302 genome sequences from public database were also included (Supplementary Table 2).

### Patient Consent Statement

The requirement for ethical approval and written informed consent was waived due to the use of unidentifiable epidemiological and microbiological data.

### RESULTS

A total of 5479 SNVs were identified in the nonrepetitive, nonrecombinogenic genome for a maximum likelihood phylogeny. All strains, except for M66-2 for the outgroup, were categorized in the seventh pandemic El Tor clade. Thirteen strains from Lao PDR in 1998 and 2000 were clustered but genetically distinct from the 201 strains isolated after 2007. The 201 strains belonged to 2 lineages, which mainly consisted of Asian isolates (Asian lineage 1 and 2). Asian lineage 1 contained 174 strains

Received 22 July 2020; editorial decision 8 October 2020; accepted 13 October 2020.

Correspondence: M. Morita, PhD, Department of Bacteriology I, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku, Tokyo 162-8640, Japan ([mmorita@niid.go.jp](mailto:mmorita@niid.go.jp)).

### Open Forum Infectious Diseases® 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)  
DOI: 10.1093/ofid/ofaa492

from 3 countries. In contrast, only 27 Thai strains were included in Asian lineage 2. Twenty-four of the 27 Thai strains in Asian lineage 2 were from Tak province, which borders Myanmar, and the isolates nearest to these were Indian strains. The 174 strains in Asian lineage 1 were further divided by the year of isolation: the 119 strains before 2008 and the 55 strains after 2009 formed different clusters. As the cholera outbreaks occurred during 2007 in Lao PDR, Thailand, and Vietnam, the genetically related 119 strains before 2008 seemed to be the outbreak strains of each country (Figure 1).

Therefore, we conducted another phylogenetic analysis and recalculated the pairwise SNV differences among the 119 strains. The median pairwise SNV distance among the 119 strains was 4 (range from 0 to 104), and 9 strains (Lao PDR [n = 1], Thailand [n = 6], Vietnam [n = 2]) showed identical SNV patterns. Relatively long branches were generated from 2 strains, LAO149 and VNM084, and the median pairwise SNV distance of LAO149 and VNM084 was 44 (range from 42 to 104) and 63 (range from 61 to 104), respectively. It is known that hypermutators with mutations in the DNA mismatch repair genes create a long branch on phylogenetic trees [16]. In the present study, we searched for mutations in DNA mismatch repair genes (*mutS*, *mutL*, *mutH*, and *uvrD*) in the draft genome sequences using BLAST+version 2.6.0 with an in-house Perl script with N16961 as a reference [17]. A nonsynonymous mutation, A1487G in *mutS*, was identified in VNM084, which would result in a Y496C amino acid substitution, and a 1220-base pair (bp) deletion was identified in LAO149 that involved 535 bp upstream from the 3'

end of *mutS*. When these 2 strains with hypermutator phenotype were excluded from the analysis, the maximal pairwise SNV distance was 18 and a more closely related phylogenetic cluster was observed (Supplementary Figure 1). The pairwise SNV differences were comparable to those for the Yemen outbreak strains (range, 0–13) [18].

## DISCUSSION

In the present study, the 214 strains from Lao PDR, Thailand, and Vietnam were assigned to a population structure of the global *V cholerae* O1 collection based on the core genome SNVs for the outbreak investigation. Recent global transmission of *V cholerae* O1 was associated with the strains of Asian lineage 2 involved in the 2 catastrophic cholera outbreaks in Haiti and Yemen [18, 19]. However, Asian lineage 1 was the lineage common to the strains from the 3 countries in this study, and the strains formed 2 distinct clusters based on the isolation year. The country had no effect on the clustering. One limitation of the present study was that our analysis did not include isolates from other countries of Southeast Asia, especially countries of the Indochina Peninsula (Cambodia, Malaysia, and Myanmar), during the same period. Nevertheless, our results indicate that the isolates in Asian lineage 1 circulated in Southeast Asia with continuous mutations on the genome. Moreover, the 119 strains from 2007 and 2008 that were associated with the outbreaks in the 3 countries were clustered, with the SNV distance ranging from 0 to 104. A large number of SNVs were derived from hypermutators, and 9 strains with identical SNV profiles were detected among the outbreak strains.

## CONCLUSIONS

These results imply mutual transmission of epidemic *V cholerae* O1 among countries and/or a common source of infection. Nevertheless, closely related strains caused cholera outbreaks in the 3 neighboring countries during the same period.

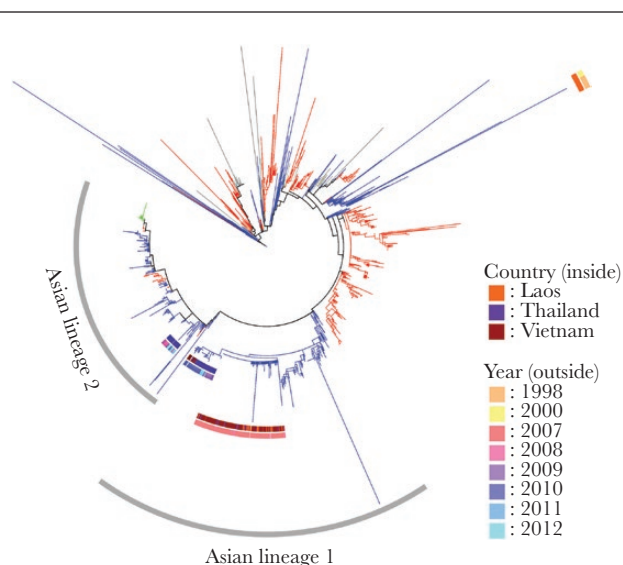
## Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

## Acknowledgments

We thank Kanako Oba and Yu Takizawa for technical assistance.

**Author contributions.** M. O. conceived the project. M. M., H. I., and M. O. designed the study. K. O., A. R., W. W., and S. C. participated in the whole-genome sequencing of isolates in Thailand. T. Y. and N. D. T. participated in the whole-genome sequencing of isolates in Vietnam. P. X., N. S., K. N., and A. V. participated in the whole-genome sequencing of isolates in Lao People's Democratic Republic (PDR). T. S. and M. M. performed the bioinformatics analysis supervised by K. N. and M. O. M. M., H. I., and M. O. contributed to the interpretation of the results. S. H. commented on the study. M. M. prepared the manuscript and figures. All authors read and approved the manuscript.



**Figure 1.** Maximum likelihood phylogeny of 1515 *Vibrio cholerae* strains. M66-2 was used as the outgroup. Branch color indicates the geographical region or country of origin for each strain (Asia, blue; Africa, red; Haiti, yellow; Yemen, green; others, gray). Colored boxes on the outside tree indicate clusters of strains in this study with their information (country and year). Asian lineages are covered with gray half circles.

**Financial support.** This research was funded by the Japan Agency for Medical Research and Development (AMED), under Grant Number JP20fk0108139, and National Center for Laboratory and Epidemiology, Ministry of Health, Lao PDR.

**Potential conflicts of interest.** All authors: no reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

## References

1. Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. *Lancet* **2004**; 363:223–33.
2. Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis* **2015**; 9:e0003832.
3. World Health Organization. Cholera, 2007. *Wkly Epidemiol Rec* **2008**; 83:269–83.
4. Nguyen BM, Lee JH, Cuong NT, et al. Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor biotype strain producing classical cholera toxin B in Vietnam in 2007 to 2008. *J Clin Microbiol* **2009**; 47:1568–71.
5. Lenglet A, Khamphongphane B, Thebvongsa P, et al. A cholera epidemic in Sekong Province, Lao People's Democratic Republic, December 2007–January 2008. *Jpn J Infect Dis* **2010**; 63:204–7.
6. Okada K, Chantaroj S, Roobthaisong A, et al. A cholera outbreak of the *Vibrio cholerae* O1 El Tor variant carrying classical CtxB in northeastern Thailand in 2007. *Am J Trop Med Hyg* **2010**; 82:875–8.
7. Sithivong N, Izumiya H, Munnalath K, et al. Cholera outbreak, Laos, 2007. *Emerg Infect Dis* **2010**; 16:745–6.
8. Okada K, Roobthaisong A, Nakagawa I, et al. Genotypic and PFGE/MLVA analyses of *Vibrio cholerae* O1: geographical spread and temporal changes during the 2007–2010 cholera outbreaks in Thailand. *PLoS One* **2012**; 7:e30863.
9. Nguyen DT, Ngo TC, Le TH, et al. Molecular epidemiology of *Vibrio cholerae* O1 in northern Vietnam (2007–2009), using multilocus variable-number tandem repeat analysis. *J Med Microbiol* **2016**; 65:1007–12.
10. Mahapatra T, Mahapatra S, Babu GR, et al. Cholera outbreaks in South and Southeast Asia: descriptive analysis, 2003–2012. *Jpn J Infect Dis* **2014**; 67:145–56.
11. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012**; 19:455–77.
12. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **2015**; 43:e15.
13. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **2002**; 30:2478–83.
14. Zhou Y, Liang Y, Lynch KH, et al. PHAST: a fast phage search tool. *Nucleic Acids Res* **2011**; 39:W347–52.
15. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**; 22:2688–90.
16. Didelot X, Pang B, Zhou Z, et al. The role of China in the global spread of the current cholera pandemic. *PLoS Genet* **2015**; 11:e1005072.
17. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**; 10:421.
18. Weill FX, Domman D, Njamkepo E, et al. Genomic insights into the 2016–2017 cholera epidemic in Yemen. *Nature* **2019**; 565:230–3.
19. Mutreja A, Kim DW, Thomson NR, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **2011**; 477:462–5.