



Draft Genome Sequences of Eight *Vibrio* sp. Clinical Isolates from across the United States That Form a Basal Sister Clade to *Vibrio cholerae*

Kevin Liang,^a Mohammad Tarequul Islam,^a Nora Hussain,^a Nathan S. Winkjer,^{b,d} Monica S. Im,^{b,c} Lori A. Rowe,^b Cheryl L. Tarr,^b Yan Boucher^a

^aDepartment of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

^bNational Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

^cIHRC, Inc., Atlanta, Georgia, USA

^dOak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

ABSTRACT We sequenced the genomes of eight isolates from various regions of the United States. These isolates form a monophyletic cluster clearly related to but distinct from *Vibrio cholerae*. Phylogenetic and genomic analyses suggest that they represent a basal lineage highly divergent from *Vibrio cholerae* or a novel species.

Members of the genus *Vibrio* cause a variety of human illnesses, including diarrhea, soft tissue disease, septicemia, and eye and ear infections (1). *Vibrio cholerae* has historically caused the most devastation, as toxigenic strains of this species are responsible for cholera, a diarrheal disease that can lead to death within hours. We have reported an environmental isolate, *Vibrio* sp. strain 2521-89, that forms a strongly supported monophyletic cluster with *V. cholerae* with an unusually long branch length (2). Monitoring conducted under the Cholera and Other *Vibrio* Illness Surveillance (COVIS) program (<https://www.cdc.gov/vibrio/surveillance.html>) revealed eight additional isolates that show high genome sequence similarity to *Vibrio* sp. strain 2521-89 but are genetically distinct from typical *V. cholerae* strains.

These strains were isolated on thiosulfate-citrate-bile salts-sucrose (TCBS) plates and cultured on tryptic soy agar. Initial identification was based on *rpoB* sequences following standard procedures (3). Genomic DNA was extracted using the Qiagen DNeasy blood and tissue kit according to the manufacturer's instructions. Sequencing libraries were prepared from genomic DNA using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA) and sequenced using an Illumina MiSeq or HiSeq platform (Table 1). Trimming and *de novo* assembly were completed in CLC Workbench v7 with default parameters (i.e., removing reads with less than 15 bp and assembled with a minimum contig size of 200 bp). Genome size, G+C content, and other characteristics are given in Table 1. Annotations were done using Rapid Annotations using Subsystems Technology (RAST) v2.0 (4) and the Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Average nucleotide identity (ANI) and *in silico* DNA-DNA hybridization (dDDH) values were calculated using JSpecies v1.2.1 (5) and GGDC (6), respectively. Whole-genome alignment was performed using Mugsy v1.2.3 (7). A maximum likelihood tree was constructed using RAXML v8 (8) under the GTR+GAMMA model with 1,000 bootstrap replicates. The core genome phylogeny, DDH, and ANI pairwise data were deposited in the University of Alberta Dataverse repository (<https://doi.org/10.7939/DVN/IWNRRI>).

The eight strains described here and a previously described sister strain (2) were all isolated from the environment, blood, stool, or wound infections. All strains contain the toxin transcriptional regulator gene (*toxR*) but lack the following characteristic toxigenic

Citation Liang K, Islam MT, Hussain N, Winkjer NS, Im MS, Rowe LA, Tarr CL, Boucher Y. 2019. Draft genome sequences of eight *Vibrio* sp. clinical isolates from across the United States that form a basal sister clade to *Vibrio cholerae*. Microbiol Resour Announc 8:e01473-18. <https://doi.org/10.1128/MRA.01473-18>.

Editor John J. Dennehy, Queens College
This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.
Address correspondence to Yan Boucher, yboucher@ualberta.ca.

Received 29 October 2018

Accepted 15 December 2018

Published 17 January 2019

TABLE 1 Genome features and GenBank accession numbers of the eight *Vibrio* sp. strains

| Strain name | Sequencing platform | Genome size (bp) | No. of CDS ^a | No. of rRNAs | No. of tRNAs | G+C content (%) | No. of contigs | N ₅₀ (bp) | L ₅₀ | Assembly accession no. | SRA accession no. | Source | No. of reads | Avg read length (bp) |
|-------------|---------------------|------------------|-------------------------|--------------|--------------|-----------------|----------------|----------------------|-----------------|------------------------|-------------------|-----------------|--------------|----------------------|
| 2523-88 | MiSeq | 4,181,290 | 3,843 | 7 | 70 | 47.1 | 138 | 75,895 | 16 | QKKG000000000 | SRR7962198 | NA ^b | 1,328,550 | 215.56 |
| 2015V-1076 | MiSeq | 4,050,105 | 3,637 | 3 | 65 | 47.2 | 89 | 14,5144 | 11 | QKKH000000000 | SRR7962189 | Blood | 2,857,312 | 202.9 |
| 2016V-1018 | MiSeq | 4,190,002 | 3,689 | 4 | 66 | 47.1 | 67 | 206,391 | 8 | QKKI000000000 | SRR7962195 | Wound | 1,483,569 | 212 |
| 2016V-1062 | HiSeq | 4,021,660 | 3,589 | 6 | 53 | 47.2 | 83 | 256,248 | 6 | QKKJ000000000 | SRR7962191 | Stool | 3,160,814 | 244.8 |
| 2017V-1038 | MiSeq | 4,000,833 | 3,612 | 9 | 74 | 47.1 | 116 | 117,913 | 9 | QKKK000000000 | SRR7962188 | Stool | 1,098,078 | 215.6 |
| 2017V-1070 | MiSeq | 4,062,156 | 3,602 | 9 | 55 | 47.2 | 133 | 73,382 | 18 | QKKL000000000 | SRR7962187 | Stool | 1,032,180 | 234.4 |
| 2017V-1085 | MiSeq | 4,039,328 | 3,575 | 5 | 58 | 47.2 | 109 | 104,599 | 11 | QKKM000000000 | SRR7962190 | Blood | 985,279 | 233.3 |
| 2017V-1124 | MiSeq | 4,137,761 | 3,687 | 7 | 55 | 47.1 | 80 | 164,272 | 9 | QKKN000000000 | SRR7962193 | Wound | 1,206,668 | 234.2 |

^a CDS, coding DNA sequences.^b NA, not applicable.

V. cholerae genes/genetic elements: *ctxAB*, *tcpA*, and the two *Vibrio* pathogenicity islands VPI-1 and VPI-2.

Comparative genomics and phylogenetic analyses were performed to determine if these strains could be part of a novel species. Genotypically, species are identified as having a 95% or higher average nucleotide identity (ANI) value (5) and a 70% or higher dDDH value (6) and should also form a monophyletic clade. ANI and dDDH values between the nine isolates ranged from 98% to 99% and 82% to 94%, respectively, which are higher than the corresponding species cutoffs. Although the ANI values between the nine isolates and representative *V. cholerae* strains were at the species cutoff (95%), the dDDH values ranged from 60% to 62%, putting them well below the 70% species threshold. Whole-genome phylogeny showed these isolates as a basal monophyletic clade to *V. cholerae*, and taken together with ANI and dDDH values, these isolates likely represent a close relative of *V. cholerae* or a highly divergent lineage within the species.

Data availability. The complete genome sequences were deposited in DDBJ/GenBank under the accession numbers [QKKG00000000](#) to [QKKN00000000](#). The SRA data were deposited in GenBank under the accession numbers [SRR7962187](#) to [SRR7962191](#), [SRR7962193](#), [SRR7962195](#), and [SRR7962198](#). The core genome phylogeny, ANI, and dDDH values were deposited in the University of Alberta Dataverse repository at <https://doi.org/10.7939/DVN/IWNRRI>.

ACKNOWLEDGMENTS

This work is supported by the Natural Sciences and Engineering Research Council of Canada, federal appropriations to the Centers for Disease Control and Prevention through the Advanced Molecular Detection Initiative, and Alberta Innovates–Technology Futures (AITF).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Newton A, Kendall M, Vugia DJ, Henao OL, Mahon BE. 2012. Increasing rates of vibriosis in the United States, 1996–2010: review of surveillance data from 2 systems. *Clin Infect Dis* 54:S391–395. <https://doi.org/10.1093/cid/cis243>.
2. Liang K, Orata FD, Winkler NS, Rowe LA, Tarr CL, Boucher Y. 2017. Complete genome sequence of *Vibrio* sp. strain 2521-89, a close relative of *Vibrio cholerae* isolated from lake water in New Mexico, USA. *Genome Announc* 5:e00905-17. <https://doi.org/10.1128/genomeA.00905-17>.
3. Tarr CL, Patel JS, Puhr ND, Sowers EG, Bopp CA, Strockbine NA. 2007. Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. *J Clin Microbiol* 45:134–140. <https://doi.org/10.1128/JCM.01544-06>.
4. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
5. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106:19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
6. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. <https://doi.org/10.1186/1471-2105-14-60>.
7. Angiuoli SV, Salzberg SL. 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27:334–342. <https://doi.org/10.1093/bioinformatics/btq665>.
8. 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>.