

# Complete Genome Assemblies for Two Single-Chromosome *Vibrio cholerae* Isolates, Strains 1154-74 (Serogroup O49) and 10432-62 (Serogroup O27)

S. L. Johnson,<sup>a</sup> A. Khiani,<sup>b,c</sup> K. A. Bishop-Lilly,<sup>b,c</sup> C. Chapman,<sup>b,c</sup> M. Patel,<sup>b,c</sup> K. Verratti,<sup>b,c\*</sup> H. Teshima,<sup>a</sup> A. C. Munk,<sup>a</sup> D. C. Bruce,<sup>a</sup> C. S. Han,<sup>a</sup> G. Xie,<sup>a</sup> K. W. Davenport,<sup>a</sup> P. Chain,<sup>a</sup> S. Sozhamannan<sup>d</sup>

Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USA<sup>a</sup>; Naval Medical Research Center (NMRC)-Frederick, Fort Detrick, Maryland, USA<sup>b</sup>; Henry M. Jackson Foundation, Bethesda, Maryland, USA<sup>c</sup>; Goldbelt Raven, LLC, Frederick, Maryland, USA<sup>d</sup>

\* Present address: K. Verratti, Johns Hopkins Applied Physics Laboratory, Laurel, Maryland, USA.

**Here, we report the completed genome sequences for two non-O1/non-O139 *Vibrio cholerae* isolates. Each isolate has only a single chromosome, as opposed to the normal paradigm of two chromosomes found in all other *V. cholerae* isolates.**

Received 3 April 2015 Accepted 7 April 2015 Published 14 May 2015

**Citation** Johnson SL, Khiani A, Bishop-Lilly KA, Chapman C, Patel M, Verratti K, Teshima H, Munk AC, Bruce DC, Han CS, Xie G, Davenport KW, Chain P, Sozhamannan S. 2015. Complete genome assemblies for two single-chromosome *Vibrio cholerae* isolates, strains 1154-74 (serogroup O49) and 10432-62 (serogroup O27). *Genome Announc* 3(3): e00462-15. doi:10.1128/genomeA.00462-15.

**Copyright** © 2015 Johnson et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to S. L. Johnson, shannonj@lanl.gov.

*Vibrio cholerae* is a comma-shaped Gram-negative bacterium best known as the causative agent of cholera. Cholera represents an estimated burden of 1.4 to 4.3 million cases and 28,000 to 142,000 deaths per year worldwide (1). Historically, cholera outbreaks have been linked to *V. cholerae* O1 serogroup strains or its derivatives of the O37 and O139 serogroups. A genomic study on the 2010 Haitian cholera outbreak strains highlighted the putative role of non O1/non-O139 *V. cholerae* in causing cholera (2). In a recent study, we examined the genomic diversity of a large collection of *V. cholerae* strains belonging to different non-O1/non-O139 serogroups using whole-genome mapping (3). In that study, we found pervasive genetic and genomic structural diversity, including indels, duplications, and fusions of the usual two chromosomes. Here, we report the complete genome assemblies of two single-chromosome *V. cholerae* isolates.

Each genome was drafted and assembled using four data types: Illumina short-read, Roche 454 standard and long-insert reads, and PacBio long reads. Short-read and long-insert paired 454 data were assembled together in Newbler, and consensus sequences were computationally shredded into 2-kbp overlapping shreds. The Illumina short-read data were assembled with Velvet (4), and consensus sequences were computationally shredded into 1.5-kb overlapping shreds. The PacBio long-read data were assembled using Hierarchical Genome Assembly Process (HGAP) (5). Consensus sequences from HGAP were computationally shredded into 10-kbp overlapping pieces. All shreds were integrated using Phrap. Possible misassemblies were corrected and repeat regions verified using in-house scripts and manual editing in Consed (6–8). All genomes were assembled to finished-quality completion (9), and each assembly was annotated using an Ergatis-based (10) workflow with minor manual curation.

The finding of a single chromosome was independently verified by whole-genome mapping and pulsed-field gel electrophoresis of intact chromosomes of these two strains (3). Due to the

nonstandard topology of these genomes, we reviewed OpGen Argus-generated *in silico* optical maps for orthogonal confirmation of the single-chromosome nature of each genome.

The *V. cholerae* strain 1154-74 genome is 3.928 Mb (47.8% G+C content), and the *V. cholerae* strain 10432-62 genome is 4.077 Mb (47.7% G+C content). Annotation located 3,430 and 3,645 coding sequences, respectively, with similar gene profiles.

**Nucleotide sequence accession numbers.** Annotated genome assemblies are publicly available in NCBI under accession numbers CP010811 (*V. cholerae* 1154-74) and CP010812 (*V. cholerae* 10432-62).

## ACKNOWLEDGMENTS

Funding for this effort was provided by the Defense Threat Reduction Agency's Joint Science and Technology Office (DTRA J9-CB/JSTO).

This paper is approved by the LANL for unlimited release (LA-UR-15-21518). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

## REFERENCES

1. Ali M, Lopez AL, You YA, Kim YE, Sah B, Maskery B, Clemens J. 2012. The global burden of cholera. *Bull World Health Organ* 90:209–218. <http://dx.doi.org/10.2471/BLT.11.093427>.
2. Hasan NA, Choi SY, Eppinger M, Clark PW, Chen A, Alam M, Haley BJ, Taviani E, Hine E, Su Q, Tallon LJ, Prosper JB, Furth K, Hoq MM, Li H, Fraser-Liggett CM, Cravioto A, Huq A, Ravel J, Cebula TA, Colwell RR. 2012. Genomic diversity of 2010 Haitian cholera outbreak strains. *Proc Natl Acad Sci U S A* 109:E2010–E2017. <http://dx.doi.org/10.1073/pnas.1207359109>.
3. Chapman C, Henry M, Bishop-Lilly KA, Awosika J, Briska A, Ptashkin RN, Wagner T, Rajanna C, Tsang H, Johnson SL, Mokashi VP, Chain PS, Sozhamannan S. 2015. Scanning the landscape of genome architecture of non-O1 and non-O139 *Vibrio cholerae* by whole genome mapping reveals extensive population genetic diversity. *PLoS One* 10:e0120311. <http://dx.doi.org/10.1371/journal.pone.0120311>.
4. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read

- assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
5. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
  6. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <http://dx.doi.org/10.1101/gr.8.3.195>.
  7. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 8:175–185. <http://dx.doi.org/10.1101/gr.8.3.175>.
  8. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 8:186–194.
  9. Chain PS, Grafham DV, Fulton RS, FitzGerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium, Detter JC. 2009. Genomics. Genome project standards in a new era of sequencing. *Science* 326:236–237. <http://dx.doi.org/10.1126/science.1180614>.
  10. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q. 2010. An Ergatis-based prokaryotic genome annotation Web server. *Bioinformatics* 26:1122–1124. <http://dx.doi.org/10.1093/bioinformatics/btq090>.