AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Long-Read-Based Genome Sequences of Pandemic and Environmental *Vibrio cholerae* Strains

Noémie Matthey,[a] Natália C. Drebes Dörr,[a] Melanie Blokesch[a]

[a]Laboratory of Molecular Microbiology, Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT The bacterium *Vibrio cholerae* exhibits two distinct lifestyles, one as an aquatic bacterium and the other as the etiological agent of the pandemic human disease cholera. Here, we report closed genome sequences of two seventh pandemic *V. cholerae* O1 El Tor strains, A1552 and N16961, and the environmental strain Sa5Y.

Cholera is one of the oldest diseases known and is still a major burden for people in developing countries (1). The disease is caused by *Vibrio cholerae*, which also thrives in natural environments (2). Toxigenic strains are characterized by the presence of major virulence factors (3), while marine habitats are often dominated by nontoxigenic strains. Studying those strains helps us to understand pathogen emergence (4–8).

We sequenced three *V. cholerae* strains (A1552, N16961, and Sa5Y) using whole-genome PacBio sequencing. *V. cholerae* O1 El Tor (Inaba) strain A1552 (originally named 92A1552 [9]) was isolated by the California health authorities from a traveler returning from South America (10, 11), which links it to the Peruvian outbreak in the 1990s (12–14). First used for research in the Schoolnik laboratory at Stanford University, A1552 was rendered rifampicin resistant (9) and now represents the wild type in most laboratories, including ours. *V. cholerae* O1 El Tor strain N16961 was the first sequenced strain of this species (15). However, as a recent study suggested an inversion in the initial assembly (16), we resequenced N16961. *V. cholerae* Sa5Y is a 2004 environmental isolate from California (17).

Genomic DNA was isolated from bacteria cultured in lysogeny broth using a Qiagen genomic DNA buffer set combined with Qiagen 100/G Genomic-tips. Sequencing was performed by the Genomic Technology Facility of the University of Lausanne. DNA samples were sheared in Covaris g-TUBEs to obtain fragments with a mean length of 20 kb. The sheared DNA was used to prepare each library with the PacBio SMRTbell template prep kit 1 (Pacific Biosciences) according to the manufacturer's recommendations. The resulting library was size selected on a BluePippin system (Sage Science, Inc.) for molecules larger than 15 kb, which excluded smaller plasmids. Each library was sequenced on one single-molecule real-time (SMRT) cell with P6/C4 chemistry and MagBeads on a PacBio RS II system at a movie length of 360 min. Genome assembly was performed using the protocol RS_HGAP_Assembly.3 in SMRT Pipe 2.3.0, and circularization of the genomes was achieved using the Minimus assembler of the AMOS software package 3.1.0 using default parameters (18). The assembled genomes were annotated using Prokka 1.12 (19) (Table 1).

The stock of the A1552 strain described here was previously passed on to Kemter et al., who deposited it in the German Collection of Microorganisms and Cell Cultures (DSM 106276) concomitantly with the release of its genome sequence (20). To improve upon the automated annotation of this study, we checked the annotated gene names of all coding sequences (CDS) and manually added 1,269 commonly used gene names

Received 17 November 2018 **Accepted** 27 November 2018 **Published** 13 December 2018

Citation Matthey N, Drebes Dörr NC, Blokesch M. 2018. Long-read-based genome sequences of pandemic and environmental *Vibrio cholerae* strains. Microbiol Resour Announc 7:e01574-18. https://doi.org/10.1128/MRA.01574-18.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Melanie Blokesch, melanie.blokesch@epfl.ch.

N.M. and N.C.D.D. contributed equally to this work.

TABLE 1 Statistics on genome sequences and assemblies

Feature	A1552	N16961	Sa5Y
GenBank accession no.			
Chromosome 1	CP028894	CP028827	CP028892
Chromosome 2	CP028895	CP028828	CP028893
No. of bases	799,549,317	1,750,962,832	635,540,812
No. of reads	46,861	97,399	35,390
Mean read length (bp)	17,062	17,977	17,958
Total no. of contigs (chromosomes 1 and 2)	2	2	2
Maximum contig length (bp)	3,044,896	3,003,695	2,986,375
N_{50} (bp)	3,044,896	3,003,695	2,986,375
Contig length after circularization (bp)			
Chromosome 1	3,015,094	2,975,504	2,955,400
Chromosome 2	1,070,374	1,072,331	1,095,478
Total genome size (bp)	4,085,468	4,047,835	4,050,878
Mean coverage (×)	170	351	133
GC content (%)			
Chromosome 1	47.7	47.7	47.8
Chromosome 2	46.9	46.9	46.8

under "gene"/"gene_synonym" for CDS without/with an automatically assigned gene name. Allué-Guardia et al. also recently released an A1552 genome sequence. However, the absence of the mutation in *rpoB* conferring rifampicin resistance (RpoB[S531F]) and the presence of a streptomycin resistance-causing mutation in *rpsL* (RpsL[K88R]) (21) suggest that this isolate represents a lineage distinct from that of the more commonly used rifampicin-resistant strain A1552 described here.

Data availability. The genome sequences have been deposited in NCBI GenBank under the accession numbers CP028894 and CP028895 (A1552), CP028827 and CP028828 (N16961), and CP028892 and CP028893 (Sa5Y). The raw reads are available under SRA numbers SRX4011578, SRX4011577, and SRX4011579.

ACKNOWLEDGMENTS

We thank A. Boehm for providing strain Sa5Y, the staff of the Lausanne Genomic Technologies Facility (GTF) for sample processing and bioinformatic analysis, the GenBank staff for incorporation of the manually annotated gene names (for A1552), Ivan Mateus for addition of NC_002505/NC_002506 comparative locus tags and data uploading to NCBI, and the SRA curators for help with corrupted files.

This work was supported by EPFL intramural funding, the Swiss National Science Foundation (grant 31003A_162551), and starting (309064-VIR4ENV) and consolidator (724630-CholeraIndex) grants from the European Research Council. M.B. is a Howard Hughes Medical Institute (HHMI) international research scholar (grant 55008726). The purchase of the Pacific Biosciences RS II instrument at the University of Lausanne was financed in part by the Loterie Romande through the Fondation pour la Recherche en Médecine Génétique.

N.M., N.C.D.D., and M.B. designed the research, N.M. and N.C.D.D. performed the experiments, M.B. assigned the gene/gene_synonym names, and N.M., N.C.D.D., and M.B. wrote the manuscript.

REFERENCES

1. World Health Organization. 2018. Cholera: the forgotten pandemic. World Health Organization, Geneva, Switzerland. https://www.who.int/cholera/the-forgotten-pandemic/en/.
2. Lipp EK, Huq A, Colwell RR. 2002. Effects of global climate on infectious disease: the cholera model. Clin Microbiol Rev 15:757–770. https://doi.org/10.1128/CMR.15.4.757-770.2002.
3. Nelson EJ, Harris JB, Morris JGJ, Calderwood SB, Camilli A. 2009. Cholera transmission: the host, pathogen and bacteriophage dynamic. Nat Rev Microbiol 7:693–702. https://doi.org/10.1038/nrmicro2204.
4. Faruque SM, Asadulghani Saha MN, Alim AR, Albert MJ, Islam KM, Mekalanos JJ. 1998. Analysis of clinical and environmental strains of nontoxigenic *Vibrio cholerae* for susceptibility to CTXPhi: molecular basis for origination of new strains with epidemic potential. Infect Immun 66:5819–5825.
5. Faruque SM, Mekalanos JJ. 2003. Pathogenicity islands and phages in *Vibrio cholerae* evolution. Trends Microbiol 11:505–510. https://doi.org/10.1016/j.tim.2003.09.003.
6. Blokesch M, Schoolnik GK. 2007. Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. PLoS Pathog 3:e81. https://doi.org/10.1371/journal.ppat.0030081.
7. Shapiro BJ, Levade I, Kovacikova G, Taylor RK, Almagro-Moreno S. 2016. Origins of pandemic *Vibrio cholerae* from environmental gene

pools. Nat Microbiol 2:16240. https://doi.org/10.1038/nmicrobiol.2016.240.
8. Le Roux F, Blokesch M. 2018. Eco-evolutionary dynamics linked to horizontal gene transfer in vibrios. Annu Rev Microbiol 72:89–110. https://doi.org/10.1146/annurev-micro-090817-062148.
9. Yildiz FH, Schoolnik GK. 1998. Role of *rpoS* in stress survival and virulence of *Vibrio cholerae*. J Bacteriol 180:773–784.
10. Blokesch M. 2012. A quorum sensing-mediated switch contributes to natural transformation of *Vibrio cholerae*. Mob Genet Elements 2:224–227. https://doi.org/10.4161/mge.22284.
11. Eberhart-Phillips J, Besser RE, Tormey MP, Koo D, Feikin D, Araneta MR, Wells J, Kilman L, Rutherford GW, Griffin PM, Baron R, Mascola L. 1996. An outbreak of cholera from food served on an international aircraft. Epidemiol Infect 116:9–13. https://doi.org/10.1017/S0950268800058891.
12. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon Y-S, Kim DW, Lee J-H, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR. 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci U S A 106:15442–15447. https://doi.org/10.1073/pnas.0907787106.
13. 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.
14. Domman D, Quilici ML, Dorman MJ, Njamkepo E, Mutreja A, Mather AE, Delgado G, Morales-Espinosa R, Grimont PAD, Lizarraga-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.
15. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406:477–483. https://doi.org/10.1038/35020000.
16. Val ME, Marbouty M, de Lemos Martins F, Kennedy SP, Kemble H, Bland MJ, Possoz C, Koszul R, Skovgaard O, Mazel D. 2016. A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. Sci Adv 2:e1501914. https://doi.org/10.1126/sciadv.1501914.
17. Keymer DP, Miller MC, Schoolnik GK, Boehm AB. 2007. Genomic and phenotypic diversity of coastal *Vibrio cholerae* strains is linked to environmental factors. Appl Environ Microbiol 73:3705–3714. https://doi.org/10.1128/AEM.02736-06.
18. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8:64. https://doi.org/10.1186/1471-2105-8-64.
19. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
20. Kemter FS, Messerschmidt SJ, Schallopp N, Sobetzko P, Lang E, Bunk B, Sproer C, Teschler JK, Yildiz FH, Overmann J, Waldminghaus T. 2018. Synchronous termination of replication of the two chromosomes is an evolutionary selected feature in Vibrionaceae. PLoS Genet 14:e1007251. https://doi.org/10.1371/journal.pgen.1007251.
21. Allué-Guardia A, Echazarreta M, Koenig SSK, Klose KE, Eppinger M. 2018. Closed genome sequence of *Vibrio cholerae* O1 El Tor Inaba Strain A1552. Genome Announc 6:e00098-18. https://doi.org/10.1128/genomeA.00098-18.