genomeA
Journals.ASM.org

# Draft Genome Sequence of the Opportunistic Marine Pathogen *Vibrio harveyi* Strain E385

**Mingjia Yu,[a] Chunhua Ren,[b] Jinrong Qiu,[c] Peng Luo,[b] Ruyi Zhu,[b] Zhe Zhao,[b] Chaoqun Hu[b]**

Engineering Research Center of Marine Biological Resource Comprehensive Utilization, Third Institute of Oceanography, State Oceanic Administration, Xiamen, China[a]; CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China[b]; South China Institute of Environmental Sciences, MEP, Guangzhou, China[c]

***Vibrio harveyi* strain E385 was isolated from a diseased cage-cultured grouper in Daya Bay, China. Phylogenetic analysis based on the 16S rRNA gene sequence showed similarity with *V. harveyi* strain BAA-1116. We sequenced the pathogenic strain *V. harveyi* E385 and compared the genome with that of the nonpathogenic strain *V. harveyi* BAA-1116.**

**Received** 12 August 2013 **Accepted** 11 November 2013 **Published** 12 December 2013

**Citation** Yu M, Ren C, Qiu J, Luo P, Zhu R, Zhao Z, Hu C. 2013. Draft genome sequence of the opportunistic marine pathogen *Vibrio harveyi* strain E385. Genome Announc. 1(6): e00677-13. doi:10.1128/genomeA.00677-13.



Address correspondence to Zhe Zhao, zhaozhe@scsio.ac.cn, or Chaoqun Hu, hucq@scsio.ac.cn.

*Vibrio harveyi* is a Gram-negative halophilic bacterium that can be found swimming free in tropical marine waters as part of the resident microflora of marine animals. It has also been recognized as an opportunistic pathogen to many commercially farmed marine invertebrate and vertebrate species (1). *V. harveyi* strain E385, isolated from a diseased cage-cultured grouper (*Epinephelus coioides*) at the Daya Bay of China in 2009, has been identified as a member of a *V. harveyi* clade and shares a high degree of genetic similarity with *V. harveyi* strain BAA-1116. Importantly, the microorganism is a virulent strain, based on the results of artificial infection tests.

*V. harveyi* E385 was sequenced on the HiSeq 2000 sequencing platform with a paired-end 2 × 100-nucleotide (nt) procedure. A total of 24,892,942 paired-end reads were generated, with an average length of 100 bp. After adapter and quality trimming of these data, there was about 2.4 Gb of filtered or clean data remaining, which yielded almost 400-fold coverage. *De novo* assembly was performed through the CLC Genomics Workbench (2) and SeqMan (3), resulting in the production of 94 scaffolds ranging from 551 bp to 765,364 bp. Excluding the gaps, the draft genome has a total of 6,354,192 bp, with a G+C content of 44.8%.

Using the draft genome, protein-coding sequences were predicted using the Glimmer 3.02 program (4), and 5,639 predicted open reading frames (ORFs) were obtained. These predicted ORFs were annotated by a BLASTx search against the SwissProt database with a cutoff *E* value of $10^{-10}$. In total, 3,652 ORFs were significantly matched by BLASTx hits, covering 64.8% of all predicted ORFs. Furthermore, GO annotation was analyzed using the Blast2GO software (5, 6), by which GO terms were assigned to query sequences and catalogued groups were produced based on biological processes, molecular functions, and cellular components. In total, 2,857 ORFs were assigned with 14,313 GO terms. Within molecular functions, catalytic activity categories (GO:0003824) and binding (GO:0005488) were highly represented, accounting for 41.3% and 34.8% of the ORFs, respectively. Cells (GO:0005623; 45.7%) and cell parts (GO:0044464; 45.7%) were the most represented GO categories within the cellular components. As to biological processes, metabolic process (GO:0008152; 28.9%) was the most highly represented category, followed by cellular process (GO:0009987; 26.1%). Using KEGG analysis (7) with the bidirectional best hit (BBH) method, 114 pathways were mapped, including metabolic pathways (127 members, KEGG:ko01100), followed by secondary metabolite biosynthesis (64 members, KEGG:ko01110), microbial metabolism in diverse environments (37 members, KEGG:ko01120), ABC transporters (27 members, KEGG:ko02010), two-component system (26 members, KEGG:ko02020), and flagellar assembly (15 members, KEGG:ko02040). To classify the possible functions of the genes, COG annotation was performed (8), and the results showed that 177 and 99 of the predicted proteins fall into the U class (intracellular trafficking, secretion, and vesicular transport) and V class (defense mechanisms), respectively. Noncoding RNA genes were identified by tRNAscan-SE (9) and RNAmmer (10) and found 19 16S-23S-5S rRNA and 105 predicted tRNA genes in the genome. Interestingly, we also found 85 toxin genes via an online prediction tool (see http://mvirdb.llnl.gov/) (11), with the parameters of the filter for BLAST results chosen using the default settings. These results emphasize its high virulence toward marine hosts.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AYKI00000000. The version described in this paper is version AYKI01000000.

## ACKNOWLEDGMENTS

Illumina paired-end sequencing was performed at the Shanghai Biotechnology Corporation, ShanghaiBio, Shanghai, China.

This research was supported by the National Natural Science Foundation of China (no. 41276163), the Natural Science Foundation of Guangdong province (no. S2011010000232), the Project of Science and Technology New Star of Zhujiang in Guangzhou city (no. 2013J2200094), the Science and Technology Planning Project of Guangdong Province (no. 2012B020308005), and the Scientific Research Foundation of Third Institute of Oceanography, SOA (no. 2012018).

## REFERENCES


1. **Austin B.** 2010. Vibrios as causal agents of zoonoses. Vet. Microbiol. **140:**310–317.
2. **CLC bio.** 2013. Assembly cell user manual. CLC bio, Aarhus, Denmark.
3. **Swindell S, Plasterer T.** 1997. SeqMan, p 75–89. *In* Swindell S (ed), Sequence data analysis guidebook: methods in molecular biology, vol 70. Humana Press, Totowa, NJ.
4. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
5. **Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A.** 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. **36:**3420–3435.
6. **Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M.** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:**3674–3676.
7. **Altermann E, Klaenhammer TR.** 2005. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics **6:**60. doi:10.1186/1471-2164-6-60.
8. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36.
9. **Schattner P, Brooks AN, Lowe TM.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33:**W686–W689. doi:10.1093/nar/gki366.
10. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
11. **Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T.** 2007. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res. **35:**D391–D394. doi:10.1093/nar/gkl791.