Genome Sequence of *Vibrio cholerae* G4222, a South African Clinical Isolate

Wouter J. le Roux, Wai Yin Chan, Pieter De Maayer, Stephanus N. Venter

Natural Resources and the Environment, CSIR, Pretoria, South Africa; Forestry and Agriculture Biotechnology Institute and Department of Microbiology and Plant Pathology, University of Pretoria, South Africa; Center for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, South Africa

*Vibrio cholerae*, a Gram-negative pathogen autochthonous to the aquatic environment, is the causative agent of cholera. Here, we report the complete genome sequence of *V. cholerae* G4222, a clinical isolate from South Africa.

The genome was sequenced using a Roche 454 GS-FLX sequencing platform at Inqaba Biotec, South Africa. A total of 201,286 reads with an average read length of 236 bp were obtained, giving a total of 47,503,496 nucleotides and genome coverage of 11.6×. The reads were assembled into 280 contigs using Newbler assembler v2.6 (454 Life Sciences). These contigs were then scaffolded by alignment against the complete genome sequences of *V. cholerae* MJ-1236 (7) and *V. cholerae* O1 biovar El Tor strain N16961 (8) with the NCBI Genomic (NG) Aligner tool of the NCBI Genome Workbench v2.5.5. A further 38 gaps were closed by PCR amplification and Sanger sequencing. This resulted in the assembly of the *V. cholerae* G4222 genome sequence into a total of 21 contigs. Protein-coding sequence (CDS) prediction and functional annotation of the predicted proteins were done using the Rapid Annotations using Subsystems Technology (RAST) Web server (9) before manual curation was performed.

The *V. cholerae* G4222 contigs could be scaffolded into two distinct chromosomes, as is typical of *V. cholerae* strains (10).

Chromosome I consists of 14 contigs amounting to a total length of 3,139,654 bp, with an average G+C content of 47.72% and 2,809 annotated CDS. Chromosome II consists of seven contigs with a total length of 1,061,058 bp and a G+C content of 46.88%, with 1,051 CDS annotated. The chromosome sizes and G+C compositions correlate well with those of other *V. cholerae* strains (6, 8). An ~150-kb integrative conjugative element (ICE), belonging to the SXT family, is located on chromosome I of *V. cholerae* G4222 and carries the genes involved in multiple-drug resistance. Given that an African origin for SXT-related ICEs in *V. cholerae* strains has been proposed, the genome of *V. cholerae* G4222 provides further opportunity to investigate the evolution of SXT elements (11). The strain might also provide insights into the biology of South African epidemic *V. cholerae* strains.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ANNBO00000000. The version described in this paper is the first version, ANNBO01000000.

Acknowledgments

This work was supported by the National Research Foundation (NRF). We thank the National Institute for Communicable Diseases (NICD) for providing the study strain.

References


