Complete Genome Sequence of *Corynebacterium pseudotuberculosis* Cp31, Isolated from an Egyptian Buffalo


Institute of Biological Sciences, Federal University of Pará (Universidade Federal do Pará [UFPA]), Belém, Pará, Brazil; Institute of Biological Sciences, Federal University of Minas Gerais (Universidade Federal do Minas Gerais [UFMG]), Belo Horizonte, Minas Gerais, Brazil; and Faculty of Veterinary Medicine, Cairo University, Giza, Egypt

*Corynebacterium pseudotuberculosis* is of major veterinary importance because it affects many animal species, causing economically significant livestock diseases and losses. Therefore, the genomic sequencing of various lines of this organism, isolated from different hosts, will aid in the development of diagnostic methods and new prevention and treatment strategies and improve our knowledge of the biology of this microorganism. In this study, we present the genome of *C. pseudotuberculosis* Cp31, isolated from a buffalo in Egypt.

*Corynebacterium pseudotuberculosis* is a member of the CMNR group, which is composed of the *Corynebacterium*, *Mycobacterium*, *Nocardia*, and *Rhodococcus* genera (6). This microbe is a facultative intracellular Gram-positive pathogen that causes edemaform infections in various hosts. The disease caused by *C. pseudotuberculosis* is known as bovine edema (2). This microorganism is highly clonal (2) and shows evidence of dormancy or metabolic slow-down (7).

The genome of *C. pseudotuberculosis* Cp31 was isolated from a large brisket obtained from a buffalo in Egypt. The genome of *C. pseudotuberculosis* Cp31 was sequenced using a fragment library obtained according to the SOLiD version 3 protocol, and 85,530,233 50-bp reads were produced. These reads were filtered for quality control to remove reads with an average Phred quality score less than 20, resulting in 51,276,112 reads for a total of 2,563,805,600 bp. This number of reads corresponds to the SOLiD version 3 protocol, and 80,530,223 50-bp reads were produced. These reads were filtered for quality control to remove reads with an average Phred quality score less than 20, resulting in 51,276,112 reads for a total of 2,563,805,600 bp. This number of reads is equivalent to 1,114× coverage based on the 2.3-Mb genome of *C. pseudotuberculosis* FRC41 (GenBank accession number NC_014329), which was used as a reference.

The genome assembly was performed using a hybrid method (1) in which the reads were first assembled using the *C. pseudotuberculosis* FRC41 genome as a reference to obtain scaffolds, as studies have shown that *C. pseudotuberculosis* strains are highly clonal (2). The unmapped reads were then subjected to *de novo* assembly using the CLC Genomics Workbench software package, generating a set of contigs that contained gene products that were similar to those of *Corynebacterium diphtheriae*. These contigs were inserted into the genome by manual curation using the Artemis software (5). The prediction of rRNAs was performed using the FgenesB program (http://linux1.softberry.com/), which were mapped against the draft genome. The gaps were closed using three Ion Torrent sequencing runs totaling 228,643 reads, with an average size of 120 bp, which were mapped against the draft genome.

After open reading frame prediction, which was performed using the FgenesB program (http://linux1.softberry.com/), frameshifts were detected by aligning the Ion Torrent reads against the draft genome sequence by manual curation using the Artemis software (5). The prediction of rRNAs was performed using the RNAmmer program (3), which uses hidden Markov models, and tRNAs were predicted with tRNAscan-SE (4). InterProScan (8) was used to predict protein domains, motifs, and families. Noncoding RNA (ncRNA) prediction was performed using the Rfam database (http://rfam.sanger.ac.uk/), which predicted 11 ncRNAs, including a representative of the 6C class of ncRNAs that is absent in the *C. pseudotuberculosis* strain 1002 genome. Previous studies have indicated that this small RNA is conserved in *Actinobacteria* because its function is related to general dormancy or metabolic slow-down (7).

The annotation of coding sequences (CDSs) was performed in Artemis with the help of the NCBI nonredundant database. The sequence has a GC content of 52.2% and includes 2,310,587 bp, 2,171 CDSs, three rRNA operons, 49 tRNAs, and 47 pseudogenes.

**Nucleotide sequence accession number.** The genome sequence obtained in this study has been deposited in the GenBank database under accession number CP003421.

**ACKNOWLEDGMENTS**

A.S., A.R.C., A.C.P., S.C.S., A.R.S., S.S.A., L.C.G., L.T.C., F.F.A., E.G.V.B., M.P.C.S., and V.A. were supported by the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico [CNPq]). R.T.J.R. and M.S.B. acknowledge support from the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [CAPES]). This work was performed as a part of the Genomics and Proteomics Network of the state of Pará (Rede Paráense de Genômica e Proteômica [RPGP]) and was supported by the Fapesp, National Program for Academic Cooperation (Programa Nacional de Cooperação Acadêmica [PROCAD]), CAPES, and the Minas Gerais Research Foundation (Fundação de Apoio à Pesquisa do Estado de Minas Gerais [FAPEMIG]).

Received 19 September 2012 Accepted 27 September 2012

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01782-12
REFERENCES