Complete Genome Sequence of *Corynebacterium pseudotuberculosis* Strain Cp267, Isolated from a Llama

Thiago Lopes,[a] Artur Silva,[a] Rommel Thiago,[a] Adriana Carneiro,[a] Fernanda Alves Dorella,[b] Flavia Souza Rocha,[b] Anderson Rodrigues dos Santos,[b] Alex Ranieri Jerônimo Lima,[a] Luis Carlos Guimarães,[b] Eudes G. V. Barbosa,[b] Dayana Ribeiro,[b] Karina Kelly Fiaux,[b] Carlos Augusto Almeida Diniz,[b] Vinicius Augusto Carvalho de Abreu,[b] Sintia Silva de Almeida,[b] Syed Shah Hassan,[b] Amjad Ali,[b] Syeda Marriam Bakhtiar,[b] Flávia Figueira Aburjaile,[b] Anne Cybelle Pinto,[b] Siomar de Castro Soares,[b] Ulisses de Padua Pereira,[c] Maria Paula C. Schneider,[a] Anderson Miyoshi,[b] Judy Edman,[d] Sharon Spier,[d] and Vasco Azevedo[b]

Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Pará, Brazil[a]; Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil[b]; Departamento de Medicina Veterinária, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil[c]; and Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, California, USA[d]

In this work we report the genome of *Corynebacterium pseudotuberculosis* strain 267, isolated from a llama. This pathogen is of great veterinary and economic importance, as it is the cause of caseous lymphadenitis in several livestock species around the world and causes significant losses due to the high cost of treatment.

*C*orynebacterium pseudotuberculosis is an immobile, nonsporulated, facultative intracellular bacteria that causes caseous lymphadenitis (CL), a common disease in sheep and goats. CL is associated with pectoral abscesses in horses and granulomatous lymphadenitis in the camelid family (alpacas and llamas), and it is characterized by a suppurative infection of the lymph nodes and other organs (1). There are also reports of the disease in cattle and humans (6).

C. pseudotuberculosis strain Cp267, which was used in the present study, was isolated from an 11-year-old llama from northern California, specifically, from a 5-cm-diameter submandibular abscess, which yielded a pure culture of *C. pseudotuberculosis* (nitrate negative) that was morphologically similar to that obtained using the biovar ovis strain.

The genome sequencing of Cp267 was performed using the SOLiD v3 Plus platform (Applied Biosystems) with a fragment library. A total of 61,600,224 50-bp reads were produced. After sequencing, these reads were subjected to quality filtering using the qualityFilter.pl script (a homemade script), in which reads with an average Phred quality of less than 20 were removed, and error sequence correction was performed with SAET software (Life Technologies).

After the quality analysis, 41,262,623 reads with an average Phred quality score equal to or greater than 20 were used in the assembly, which generated a genome coverage corresponding to 897-fold genome coverage based on the reference genome size of 2.3 Mb of *C. pseudotuberculosis* strain FRC41 (NC_014329).

The genome of Cp267 was assembled based on the hybrid strategy using Velvet (8) and Edena (2) software. A total of 4,627 contigs were generated, with the largest, N_{50}, having 1,438 bp and the smallest contig having 82 bp. Due to the hybrid assembly methodology, the redundant contigs were removed using the Simplifier software (http://sourceforge.net/projects/simplifier).

The contigs were mapped against the reference genome (strain FRC41) using BLASTn software, and the results were analyzed using G4ALL (http://g4all.sourceforge.net/) software, to extend the contigs and identify overlaps of a minimum of 30 bp between the ends of the contigs, thus yielding larger contigs.

These contigs were later subjected to a finishing process using CLC Genomics Workbench software, with which the contigs were ordered and oriented by mapping against the reference genome, yielding a preliminary scaffold with gaps that were removed with recursive rounds of short reads mapped against the scaffold. Finally, the genome of *C. pseudotuberculosis* strain 267 was completed with 2,329,026 bp.

The following software programs were used in the automatic functional annotation of the genome: FgenesB (gene prediction; http://linux1.softberry.com/), RNAmmer (rRNA prediction) (3), tRNAscan-SE (tRNA prediction) (4), and Tandem Repeat Finder (repetitive DNA prediction) (http://tandem.bu.edu/trf/trf.html), as well as the InterProScan software (7), which integrates multiple domain and protein family databases. At the end of the automatic annotation, the data were curated manually using Artemis software (5). CLC Genomics Workbench software was used for pseudogene identification and validation.

This genome has 2,079 coding sequences (CDSs), 52,18% GC content, 4 rRNA operons, 49 tRNA genes, and 72 pseudogenes.

Nucleotide sequence accession number. The genome was deposited in the NCBI database under accession number CP003407.

ACKNOWLEDGMENTS

REFERENCES

1. **Braga WU.** 2007. Protection in alpacas against Corynebacterium pseudotuberculosis using different bacterial components. Vet. Microbiol. **119**:297–303.
2. **Hernandez D, François P, Farinelli L, Osteras M, Schrenzel J.** 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. **18**:802–809.
3. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
4. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.
5. **Rutherford K, et al.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–945.
6. **Silva A, et al.** 2011. Complete genome sequence of *Corynebacterium pseudotuberculosis* I19, a strain isolated from a cow in Israel with bovine mastitis. J. Bacteriol. **193**:323–324.
7. **Zdobnov EM, Apweiler R.** 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**:847–848.
8. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.