

Complete Genome Sequence of *Corynebacterium pseudotuberculosis* I19, a Strain Isolated from a Cow in Israel with Bovine Mastitis[∇]

Artur Silva,¹ Maria Paula C. Schneider,¹ Louise Cerdeira,¹ Maria Silvanira Barbosa,¹
Rommel Thiago J. Ramos,¹ Adriana R. Carneiro,¹ Rodrigo Santos,¹ Marília Lima,¹
Vivian D'Afonseca,² Sintia S. Almeida,² Anderson R. Santos,² Siomar C. Soares,²
Anne C. Pinto,² Amjad Ali,² Fernanda A. Dorella,² Flavia Rocha,²
Vinicius Augusto Carvalho de Abreu,² Eva Trost,³ Andreas Tauch,³
Nahum Shpigel,⁴ Anderson Miyoshi,² and Vasco Azevedo^{2*}

Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, Brazil¹; Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil²; CeBiTec, Universität Bielefeld, 33594 Bielefeld, Germany³; and the Koret School of Veterinary Medicine, Hebrew University of Jerusalem, P.O. Box 12, Rehovot 76100, Israel⁴

Received 8 October 2010/Accepted 19 October 2010

This work reports the completion and annotation of the genome sequence of *Corynebacterium pseudotuberculosis* I19, isolated from an Israeli dairy cow with severe clinical mastitis. To present the whole-genome sequence, a *de novo* assembly approach using 33 million short (25-bp) mate-paired SOLiD reads only was applied. Furthermore, the automatic, functional, and manual annotations were attained with the use of several algorithms in a multistep process.

Corynebacterium pseudotuberculosis is the etiology of common disease conditions in sheep, goats, South American camelids, and horses; however, infections in cattle and humans are sporadic and rare.

Based on nitrate reduction, *C. pseudotuberculosis* has two biovars: *C. pseudotuberculosis* bv. *equi*, infecting mainly bovines and equines, and *C. pseudotuberculosis* bv. *ovis*, infecting sheep and goats (1, 2, 8). The widespread occurrence and the economic importance of infection with this pathogen have prompted investigation of its pathogenesis. The use of whole-genome sequence analysis helps to understand the molecular and genetic bases of this bacterium's virulence. Genome sequencing of strains isolated from a human being, a goat, and a sheep was carried out by our team.

Israel is probably the only place in the world to experience large-scale outbreaks of bovine *C. pseudotuberculosis* infection. These outbreaks are also associated with cases of mastitis (9, 10). Strain I19 was isolated from a dairy cow with severe clinical mastitis in two quarters; milk samples from both quarters were positive for *C. pseudotuberculosis*. The cow was culled on the day of milk sampling. In the present research, the SOLiD system was used in sequencing the entire genome of *C. pseudotuberculosis* I19. The sequencing generated 33,368,273 mate-paired 25-nucleotide-long short reads, which is tantamount to 834,206,825 nucleotides of information, rendering a mean genome coverage depth of 321-fold given an expected genome size of 2.6 Mb. The *de novo* assembly strategy for the assembly of short reads in

this work combines De Bruijn graph and overlap-layout-consensus methods with the use of a reference genome as a basis for orientation and ordering of the *de novo*-generated contigs (6). This strategy allowed closure of all gaps and an effective coverage of 35-fold.

The genome of *C. pseudotuberculosis* strain I19 consists of a 2,337,730-bp circular chromosome. The average G+C content of the chromosome is 52.84%. The annotation procedure involved the use of several algorithms in a multistep process. For structural annotation, the following software programs were employed: FgenesB, a gene predictor (<http://www.softberry.com>); RNAmmer, an rRNA predictor (4); tRNAscan-SE, a tRNA predictor (5); and Tandem Repeats Finder, a repetitive-DNA predictor (<http://tandem.bu.edu/trf/trf.html>). Functional annotation was performed by similarity analyses using public databases and by InterProScan analysis (11). Manual annotation was performed using Artemis (7). Identification and confirmation of putative pseudogenes in the genome were carried out using Consed. Manual analysis was performed based on the Phred quality of each base in the frameshift area (3). This analysis enabled the identification of erroneous insertions or deletions of bases in the genome information produced by the sequencing process and prevented identification of false-positive pseudogenes. The genome of *C. pseudotuberculosis* strain I19 was predicted to contain 2,124 coding sequences (CDSs), 4 rRNA operons, and 50 tRNAs, and 55 pseudogenes were found.

More detailed analysis of this genome and comparative analysis with other sequenced genomes of members of the genus and the same species will provide further insight for understanding virulence and may be useful for the development of new diagnostic methods and vaccines, contributing to the control of the different diseases caused by this pathogen.

* Corresponding author. Mailing address: Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Pampulha, CEP 31270-901, Belo Horizonte, MG, Brazil. Phone and fax: 55 31 3409 2610. E-mail: vasco@icb.ufmg.br.

[∇] Published ahead of print on 29 October 2010.

Nucleotide sequence accession numbers. The genome and annotation data for strain I19 have been deposited in the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/GenBank/>) under accession no. CP002251. Genome sequences of the strains isolated from a human being, a goat, and a sheep have been deposited in the GenBank database under accession no. CP002097.1, CP001809.1, and CP001829.1, respectively.

This research work is the result of collaboration of reputable organizations, with the support of long-standing institutions which include the Rede Paraense de Genômica e Proteômica, supported by the Fundação de Amparo a Pesquisa do Estado do Pará, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). M.P.C.S., V.A., and A.S. were supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also acknowledge support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

1. Baird, G. J., and M. C. Fontaine. 2007. *Corynebacterium pseudotuberculosis* and its role in ovine caseous lymphadenitis. *J. Comp. Pathol.* **137**:179–210.
2. Dorella, F. A., L. G. C. Pacheco, S. C. Oliveira, A. Miyoshi, and V. Azevedo. 2006. *Corynebacterium pseudotuberculosis*: microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Vet. Res.* **37**:201–218.
3. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
4. Lagesen, K., P. Hallin, E. A. Rødland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100–3108.
5. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
6. Miller, J. R., S. Koren, and G. Sutton. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* **95**:315–317.
7. Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**(10):944–945.
8. Williamson, L. H. 2001. Caseous lymphadenitis in small ruminants. *Vet. Clin. North Am. Food Anim. Pract.* **17**(2):359–371.
9. Yeruham, I., D. Elad, M. Van Ham, N. Y. Shpigel, and S. Perl. 1997. *Corynebacterium pseudotuberculosis* infection in Israeli cattle: clinical and epidemiological studies. *Vet. Rec.* **140**:423–427.
10. Yeruham, I., D. Elad, S. Friedman, and S. Perl. 2003. *Corynebacterium pseudotuberculosis* infection in Israeli dairy cattle. *Epidemiol. Infect.* **131**:947–955.
11. Zdobnov, E. M., and R. Apweiler. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**(9):847–848.