

# Whole-Genome Sequence of *Corynebacterium pseudotuberculosis* Strain 226, Isolated from the Abscess of a Goat in California

Larissa M. Dias,<sup>a</sup> Jorianne T. C. Alves,<sup>a</sup> Adonney A. O. Veras,<sup>a</sup> Rafael A. Baraúna,<sup>a</sup> Pablo H. C. G. Sá,<sup>a</sup> Sharon Spier,<sup>b</sup> Judy M. Edman,<sup>b</sup> Luis C. Guimarães,<sup>a</sup> Flávia S. Rocha,<sup>c</sup> Rommel T. J. Ramos,<sup>a</sup> Vasco Azevedo,<sup>c</sup> Artur Silva,<sup>a</sup> Adriana R. Carneiro<sup>a\*</sup>

Federal University of Pará, Center of Genomics and System Biology, Laboratory of Genomic and Bioinformatics, Belém, Pará, Brazil<sup>a</sup>; Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, California, USA<sup>b</sup>; Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil<sup>c</sup>

\* Present address: Adriana Carneiro, Instituto de Ciências Biológicas, Universidade Federal do Pará, Guamá, Belém, PA, Brazil.

***Corynebacterium pseudotuberculosis* is the etiological agent of a caseous lymphadenitis disease. Herein, we present the first complete genome sequencing of *C. pseudotuberculosis* strain 226, isolated from an abscess of the sub-iliac lymph node of a goat from California (USA). The genome contains 2,138 coding sequences (CDSs), 12 rRNAs, 49 tRNAs, and 72 pseudogenes.**

Received 8 January 2016 Accepted 11 January 2016 Published 25 February 2016

**Citation** Dias LM, Alves JTC, Veras AAO, Baraúna RA, Sá PHCG, Spier S, Edman JM, Guimarães LC, Rocha FS, Ramos RTJ, Azevedo V, Silva A, Carneiro AR. 2016. Whole-genome sequence of *Corynebacterium pseudotuberculosis* strain 226, isolated from the abscess of a goat in California. *Genome Announc* 4(1):e00038-16. doi:10.1128/genomeA.00038-16.

**Copyright** © 2016 Dias et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Adriana Carneiro, [adrianarc@ufpa.br](mailto:adrianarc@ufpa.br).

*Corynebacterium pseudotuberculosis* is a pathogenic microorganism of medical and veterinary interest and has worldwide distribution (1). The species is classified by their biochemical properties: *ovis* (negative nitrate reductase) and *equi* (positive nitrate reductase) (2). The biovar *ovis* frequently infects sheep, goats, and pigs, causing caseous lymphadenitis (CLA) (3, 4).

In goats, infection can manifest in two clinical forms: cutaneous or superficial and visceral. The first form is characterized by the formation of abscesses in the superficial lymph nodes or subcutaneous tissues (3). The other form, visceral, affects organs including the kidneys, mammary glands, uterus, liver, heart, and brain (3).

Currently, there are thirty-one complete sequences and two draft genomes of *C. pseudotuberculosis* available in the GenBank database (<http://www.ncbi.nlm.nih.gov>). Although, several genomes have been elucidated, virulence factors are not fully characterized.

*C. pseudotuberculosis* strain 226, biovar *ovis*, is the first complete genome isolated and sequenced from a goat in California. It was isolated from the sub-iliac lymph node of a 4-year-old Le-Mancha breed goat from a geographic endemic region for CLA.

The genome of *C. pseudotuberculosis* strain 226 was sequenced with the Ion Torrent PGM platform using Chip 318 and the fragment library. Sequencing processes resulted in 370,840,598 reads with an estimated genomic coverage of 158× based on the reference genome of *C. pseudotuberculosis* strain 42/02-A (accession number CP003062). The quality of raw data was analyzed with the FastQC software (<http://www.bioinformatics.babraham.ac.uk>), and the assembly was performed with MIRA 4.0.2 software (5), which yielded 27 contigs. Subsequently, the contigs were subjected to analysis with the SeqMan Pro tool of the Lasergene 11 Core Suite (<http://www.dnastar.com>) to generate a preliminary

scaffold and gaps closure proceeded through CLC Genome Workbench (<http://www.clcbio.com>).

The genome was automatically annotated using Rapid Annotations using Subsystems Technology (RAST) (6). tRNAs and rRNAs were predicted using tRNAScan-SE (7) and RNAmmer (8) software. Additionally, the protein domain and motif were identified with the Pfam (9) and InterPro (10) databases. The signal peptides cleavages sites, transmembrane helices and CRISPR repeats were predicted by the SignalP 4.1 server (11), TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), and CRISPRfinder (12). Artemis program (13) was used to manual curation of the annotation. Pathogenicity islands (PAIs) was performed by the PIPS (14) software using the genome of nonpathogenic *C. glutamicum* strain ATCC13032 (accession number BA000036). The parameters utilized were a G+C content of 1.5 standard deviations, 95% codon usage deviation, and e-value of 10<sup>-5</sup> to factors virulence.

The genome of *C. pseudotuberculosis* strain 226 contains 2,337,820 bp and with a G+C content of 52.18%. A total of 2,271 genes were predicted with 2,138 coding sequences (CDSs), 49 tRNAs, 4 cluster rRNAs, and 72 pseudogenes. We identified 151 signal peptides, 595 proteins as transmembrane helices, and 1 CRISPR-associated protein (*Cas5* gene). This genome has eight PAIs with 112 genes, of which, 37 are hypothetical proteins and can be related to the virulence of the bacterium.

**Nucleotide sequence accession number.** The genome project has been deposited in GenBank under the accession number CP010889.

## ACKNOWLEDGMENTS

This study was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do

Estado do Pará (FAPESPA), and Rede Paraense de Genômica e Proteômica.

## REFERENCES

- Arsenault J, Girard C, Dubreuil P, Daignault D, Galarneau J-R, Boisclair J, Simard C, Bélanger D. 2003. Prevalence of and carcass condemnation from maedi-visna, paratuberculosis and caseous lymphadenitis in culled sheep from Quebec, Canada. *Prev Vet Med* 59:67–81. [http://dx.doi.org/10.1016/S0167-5877\(03\)00060-6](http://dx.doi.org/10.1016/S0167-5877(03)00060-6).
- Dorella FA, Pacheco LG, Oliveira SC, Miyoshi A, Azevedo VA. 2006. *Corynebacterium pseudotuberculosis*: microbiology, biochemical properties, pathogenesis and molecular studies of virulence. *Vet Res* 37:201–218. <http://dx.doi.org/10.1051/vetres:2005056>.
- Fontaine MC, Baird GJ. 2008. Caseous lymphadenitis. *Small Rumin Res* 76:42–48. <http://dx.doi.org/10.1016/j.smallrumres.2007.12.025>.
- Oliveira M, Barroco C, Mottola C, Santos R, Lemsaddek A, Tavares L, Semedo-Lemsaddek T. 2014. First report of *Corynebacterium pseudotuberculosis* from caseous lymphadenitis lesions in Black Alentejano pig (*Sus scrofa domesticus*). *BMC Vet Res* 10:218. <http://dx.doi.org/10.1186/s12917-014-0218-3>.
- Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14:1147–1159. <http://dx.doi.org/10.1101/gr.1917404>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.
- 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. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <http://dx.doi.org/10.1093/nar/gkt1223>.
- Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848. <http://dx.doi.org/10.1093/bioinformatics/17.9.847>.
- Dyrlov Bendtsen J, Nielsen H, Von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
- Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 8:R61. <http://dx.doi.org/10.1186/gb-2007-8-4-r61>.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945. <http://dx.doi.org/10.1093/bioinformatics/16.10.944>.
- Soares SC, Abreu VA, Ramos RT, Cerdeira L, Silva A, Baumbach J, Trost E, Tauch A, Hirata R, Mattos-Guaraldi AL, Miyoshi A, Azevedo V. 2012. PIPS: pathogenicity island prediction software. *PLoS One* 7:e30848.