JB Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Genome Sequences of *Corynebacterium pseudotuberculosis* Strains 3/99-5 and 42/02-A, Isolated from Sheep in Scotland and Australia, Respectively

Florence E. Pethick,[a,b] Alex F. Lainson,[a] Raja Yaga,[a] Allen Flockhart,[a] David G. E. Smith,[a,b] Willie Donachie,[a] Louise T. Cerdeira,[c] Artur Silva,[c] Erik Bol,[c] Thiago S. Lopes,[c] Maria S. Barbosa,[c] Anne C. Pinto,[d] Anderson R. dos Santos,[d] Siomar C. Soares,[d] Sintia S. Almeida,[d] Luis C. Guimaraes,[d] Flavia F. Aburjaile,[d] Vinicius A. C. Abreu,[d] Dayana Ribeiro,[d] Karina K. Fiaux,[d] Carlos A. A. Diniz,[d] Eudes G. V. Barbosa,[d] Ulisses P. Pereira,[e] Syed S. Hassan,[d] Amjad Ali,[d] Syeda M. Bakhtiar,[d] Fernanda A. Dorella,[d] Adriana R. Carneiro,[c] Rommel T. J. Ramos,[c] Flavia S. Rocha,[d] Maria P. C. Schneider,[c] Anderson Miyoshi,[d] Vasco Azevedo,[d] and Michael C. Fontaine[a]

Moredun Research Institute, Pentlands Science Park, Bush Loan, Edinburgh, Midlothian, United Kingdom[a]; Institute of Infection, Immunity and Inflammation. College of Medical, Veterinary and Life Sciences, University of Glasgow Garscube Estate, Glasgow, United Kingdom[b]; Instituto de Ciências Biológicas, Univeridade Federal do Pará, Belém, Brazil[c]; Instituto de Ciências Biológicas, Univeridade Federal de Minas Gerais, Belo Horizonte, Brazil[d]; and Departmento de Medicina Veterinária, Universidade Federal de Lavras, Lavras, Brazil[e]

Here, we report the whole-genome sequences of two ovine-pathogenic *Corynebacterium pseudotuberculosis* isolates: strain 3/99-5, which represents the first *C. pseudotuberculosis* genome originating from the United Kingdom, and 42/02-A, the second from Australia. These genome sequences will contribute to the objective of determining the global pan-genome of this bacterium.

Corynebacterium pseudotuberculosis is responsible for several diseases in multiple host species, the most notable being caseous lymphadenitis (CLA), a chronic pyogenic disease of small ruminants (1). Previous work has revealed that ovine *C. pseudotuberculosis* strains represent a largely clonal population but that these are distinct from other strains associated with equine disease (2). To identify the core genome and other variable accessory genes contributing to host specificity, work is under way to sequence representative *C. pseudotuberculosis* strains from multiple geographic locations and host species. To facilitate this study, we have sequenced the complete genomes of two ovine *C. pseudotuberculosis* strains, isolated from natural outbreaks of CLA: 42/02-A, an isolate from Perth, Australia, and 3/99-5, from the Scottish Borders, United Kingdom.

The *C. pseudotuberculosis* 3/99-5 and 42/02-A genomes were sequenced using 454 GS-FLX and Solexa 50-bp paired-end sequencing. Reads were assembled using Velvet (8) and CABOG (Celera Assembler with the Best Overlap Graph) (5), and gaps were closed using unmapped 454 and Illumina reads.

Structural annotation was achieved using the following software: FgenesB (a gene predictor), RNAmmer (an rRNA predictor) (3), tRNA-scan-SE (a tRNA predictor) (4), and Tandem Repeat Finder (to predict repeat DNA regions) (http://tandem.bu.edu/trf/trf.html). Functional annotation was performed using InterProScan (7) analysis and homology analyses using public databases. Manual annotation was then completed using Artemis software (6).

The presence of pseudogenes within the genomes was determined using CLCBio Workbench 4.02 software. Manual analysis was also conducted based on Phred quality of each base and analysis of coverage depth at the frameshift region, allowing identification of false-positive pseudogene results.

The *C. pseudotuberculosis* 3/99-5 genome consists of a single 2,337,938-bp circular chromosome with an average G+C content of 52.18%. The genome was predicted to contain 2,142 coding sequences (CDS), four rRNA operons, and 49 tRNAs. In addition, 36 pseudogenes were also identified.

The highly similar *C. pseudotuberculosis* 42/02-A genome consists of a single 2,337,606-bp circular chromosome with an average G+C content of 52.19%. This genome was predicted to contain 2,051 coding sequences (CDS), four rRNA operons, and 49 tRNAs. In addition, 52 pseudogenes were also identified.

The sequencing of these isolates will aid comparison of genomes deriving from multiple geographical locations and host species in a wider pan-genome project. Widespread comparisons should offer insights into the organism's pathogenicity and host specificity as well as evolutionary relationships between strains originating from different geographical locations.

Nucleotide sequence accession numbers. The *C. pseudotuberculosis* 3/99-5 and 42/02-A genome sequences described in this study have been deposited in the GenBank database under accession numbers CP003152.1 and CP003062, respectively. The *C. pseudotuberculosis* 3/99-5 genome has also been deposited in the RefSeq database under accession number NC_016781.1.

ACKNOWLEDGMENTS

We acknowledge Quality Meat Scotland (QMS) and the Rural and Environment Science and Analytical Services Division (RESAS) of the Scottish Government for funding this work. This work was part of the Rede Paraense de Genômica e Proteômica, supported by FAPESPA, Núcleo Amazônico de Excelência em Genômica de Microorganismos—Pronex/CNPq/FAPESPA, CAPES, and FAPEMIG.

REFERENCES

1. **Baird GJ, Fontaine MC.** 2007. *Corynebacterium pseudotuberculosis* and its role in ovine caseous lymphadenitis. J. Comp. Pathol. 137:179–210.

Received 25 May 2012 **Accepted** 5 June 2012

Address correspondence to Michael C. Fontaine, michael.fontaine@moredun.ac.uk.

doi:10.1128/JB.00918-12

2. **Conner KM, Fontaine MC, Rudge K, Baird GJ, Donachie W.** 2007. Molecular genotyping of multinational ovine and caprine *Corynebacterium pseudotuberculosis* isolates using pulsed-field gel electrophoresis. Vet. Res. **38**:613–623.
3. **Lagesen K, et al.** 2007. RNAmmer:consistent annotation of rRNA genes in genomic sequences. 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. **Miller JR, et al.** 2008. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics **24**:2818–2824.
6. **Rutherford K, et al.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944–945.
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.