Draft Genome Sequence of a Rare Pigmented Mycobacterium avium subsp. paratuberculosis Type C Strain

Pedro Barbosa, Célia Leão, Anabel Usié, Ana Amaro, Ana Botelho, João Inácio, Karen Stevenson, António Marcos Ramos

Centro de Biotecnologia Agrícola e Agroalimentar do Alentejo (CEBAL)/Instituto Politécnico de Beja (IPBeja), Beja, Portugal; Instituto de Ciências Agrárias e Ambientais Mediterrânicas (ICAAM), Universidade de Évora, Évora, Portugal; INIAV–Instituto Nacional de Investigação Agrária e Veterinária, Lisbon, Portugal; Moredun Research Institute, Penicuik, United Kingdom; Serviço de Desenvolvimento Agrário de São Miguel, Ponta Delgada, Açores, Portugal; School of Pharmacy and Biomolecular Sciences, University of Brighton, Brighton, United Kingdom; Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade Nova de Lisboa, Lisbon, Portugal

ABSTRACT Mycobacterium avium subsp. paratuberculosis is the causative agent of paratuberculosis. We report here the draft genome sequence of a rare pigmented M. avium subsp. paratuberculosis type C strain, comprising 58 contigs and having a genome size of 4,851,414 bp. The genome will assist in the execution of pigmentation and virulence studies on this mycobacterium.

Mycobacterium avium subsp. paratuberculosis infects a wide range of animals, including humans, and is the causative agent of paratuberculosis. We report here the draft genome sequence of a rare pigmented M. avium subsp. paratuberculosis type C strain, comprising 58 contigs and having a genome size of 4,851,414 bp. The genome will assist in the execution of pigmentation and virulence studies on this mycobacterium.

DNA was extracted from a single colony using a QIAamp DNA minikit (Qiagen). High-throughput sequence data were generated with Ion Torrent PGM (Thermo Fisher Scientific, Inc., USA), using a 400-bp sequencing kit and a 316 V1 chip. A total of 3,009,190 reads were produced, with a mean length of 299 bp. The genome assembly was created with MIRA (9) using the raw reads with preprocessing modules included in the software. The assembly contains 58 contigs, which encompass 4,851,414 bp. The largest contig obtained was 521,843 bp long and the N50 value was 188,875 bp, while the GC content was 69.3%.

Structural genome annotation was performed with Prokka (10), through Prodigal (11), yielding 4,666 protein-coding genes, 3 rRNAs, 58 tRNAs, and 1 transfer-messenger RNA. Functional annotation was executed with HMMER (12) against the eggNOG version 4.5 (13) database. The search space was reduced by selecting only hidden Markov models from actinobacteria. A total of 4,178 HMMER hits (90.8%) were detected, and 3,425 unique nonsupervised orthologous groups were found. The most common orthologs found in the genome were as follows: members of the PPE family, specific to mycobacteria.
and possibly playing a role in infection and virulence (14); S-adenosyl-L-methionine-dependent methyltransferase activity; cytochrome P450; acyl-CoA synthetase; synthase; transport proteins; and virulence factor mammalian cell entry family proteins.

A relatedness analysis was performed using 10,515 identified variants. The pigmented strain and 47 other \textit{M. avium} subsp. \textit{paratuberculosis} (15) strains were compared by mapping the reads to the K10 \textit{M. avium} subsp. \textit{paratuberculosis} reference genome (14) using BWA-MEM (16) and calling variation with GATK Haplotype Caller (17). This resulted in a dendrogram, built with SNPRelate (18), with three clearly differentiated groups for the type C, type S pigmented, and type S nonpigmented strains. The pigmented \textit{M. avium} subsp. \textit{paratuberculosis} isolate was unequivocally placed among the type C clade, confirming it to be a type C \textit{M. avium} subsp. \textit{paratuberculosis} strain. The first genome sequence of a pigmented type C \textit{M. avium} subsp. \textit{paratuberculosis} strain will assist future research focusing on pigmentation and virulence mechanisms of this bacterium.

**Accession number(s).** This whole-genome shotgun project has been deposited in GenBank under accession number MWPB00000000.

**ACKNOWLEDGMENTS**

Financial support for P.B., A.U., and A.M.R. was provided by Investigador FCT (Fundação para a Ciência e a Tecnologia) project IF/01015/2013/CP1209/CT0001. "Genetic characterization of national animal and plant resources using next-generation sequencing.” Funding was also provided by FCT projects PTDC/CVT/111634/2009, “Novas abordagens moleculares para a detecção e discriminação de membros do Complexo \textit{Mycobacterium tuberculosis} associados a tuberculose animal e avaliação do potencial zoonótico destas espécies em Portugal,” and UID/AGR/00115/2013. K.S. was funded by the Rural and Environmental Science and Analytical Services Division of the Scottish Government.

**REFERENCES**


9. Leão C, Goldstone RJ, Bryant J, McLuckie J, Inácio J, Smith DGE, Steven-...