Complete genome sequence of an ovine ancestral strain of Mycobacterium avium subspecies paratuberculosis 6756

Franck Biet, Cyril Conde, Thierry Cochard, Fiona A Mcintoch, Marcel A Behr, John P Bannantine

HAL Id: hal-04523445
https://hal.inrae.fr/hal-04523445

Submitted on 27 Mar 2024

Complete genome sequence of an ovine ancestral strain of *Mycobacterium avium* subspecies *paratuberculosis* 6756

Franck Biet,[1] Cyril Conde,[1] Thierry Cochard,[1] Fiona A. Mcintoch,[2] Marcel A. Behr,[2] John P. Bannantine[3]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT The complete genome sequence of the most ancestral type SI strain of *Mycobacterium avium* subspecies *paratuberculosis* 6756, isolated from a sheep, was determined. The genome was sequenced using PacBio technology, yielding a genome size of 4,830,294 nucleotides with no identified plasmids.

Rationale for sequencing

*M*ycobacterium *avium* subspecies *paratuberculosis* is the causative agent of paratuberculosis, which affects ruminants worldwide (1). Sequencing efforts are still required to better understand the evolution and transmission dynamics of this pathogen (2). The main obstacle to this research lies in the fact that this mycobacterium grows slowly, which explains why very few complete genomes are available and only one for the SI type (3). Thanks to improved long-read sequencing technology, we have been able to obtain the complete assembly of type SI strain 6756.

Provenance (origin) of the isolate and culture conditions

The *M. avium* subsp. *paratuberculosis* ovine SI-type 6756 strain was isolated from a sheep in New Zealand in 1991 on Herrold's egg yolk medium with mycobactin at 37°C (4). The strain 6756 was propagated from one colony in Middlebrook 7H9 medium with 0.5% (vol/vol) glycerol, 10% (vol/vol) oleic acid-albumin-dextrose-catalase enrichment medium (Becton, Dickinson, Oxford, Oxfordshire, UK), and 2 µg/mL of mycobactin J (IdVet, France) at 37°C.

Genomic DNA purification for long-read sequencing, PacBio

D-Cycloserine (Sigma C6880-1G) was added to 15 mL of culture to obtain a final concentration of 50 µg/mL and incubated for 24 h at 37°C. Bacteria were harvested by centrifugation for 15 min at 4,000 g and then the bacterial pellet was resuspended with 160 µL of STET buffer [100 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 5% Triton X-100). One hundred microliters of Lipase (L0777-50mL, Sigma) and 20 µL of lysozyme (100 mg/mL) were added before incubating the sample at 37°C for 2 h under gentle agitation at 900 rpm (Thermomixer C, Eppendorf). Then, 25 µL of proteinase K (MagAttract HMW DNA Kit, Qiagen) was added, and the sample was incubated overnight at 56°C.

   Genomic DNA was extracted using the MagAttract HMW DNA Kit (Qiagen) for purifying high molecular weight DNA according to the manufacturer's instructions.

Genome sequencing

The genomic DNA obtained was checked by NanoDrop One (ThermoScientific) and concentrated to 59.6 ng/µL by fluorescence measurement using Qubit dsDNA BR Quantification Kit (Broad Range) and sent without size selection to the sequencing company (GENEWIZ, from Azenta Life Sciences, South Plainfield, NJ, USA) using HiFi prep to create library and a PacBio Sequel IIe sequencing platform.

High-fidelity reads with quality values greater than 20 or exhibiting an accuracy of 99% were acquired by the SMRT Link. HiFi reads below 1,000 bp were removed using NanoFilt v2.8.0 (option -l) (https://github.com/wdecoster/nanofilt) (5). Sequencing generated 406,270 reads to reach a sequencing depth of 549×, with an N_{50} length of 7,506 bp. The genome was *de novo* assembled by Flye (v2.9-b1768) with the default parameters (6) into a single circular chromosome of 4,830,294 bp, with a GC content of 69.22% rotated to start at dnaA gene using Circlator (v1.5.5) (7) fixstart command and option --genes_fa set with the K-10 dnaA gene. Genome annotation was carried out by NCBI using NCBI's Prokaryotic Genome Annotation Pipeline (8).

AUTHOR AFFILIATIONS

[1]INRAE, Université de Tours ISP, Nouzilly, France
[2]McGill University Health Centre, Montreal, Quebec, Canada
[3]USDA-Agricultural Research Service, National Animal Disease Center, Ames, Iowa, USA

AUTHOR ORCIDs

Franck Biet http://orcid.org/0000-0001-8425-6158
Marcel A. Behr http://orcid.org/0000-0003-0402-4467
John P. Bannantine http://orcid.org/0000-0002-5692-7898

DATA AVAILABILITY

The sequence was deposited in NCBI's public sequence database under the accession number CP139433 (BioProject number PRJNA1043460).

REFERENCES

1. Barkema HW, Hesselink JW, McKenna SLB, Benedictus G, Groenendaal H. 2010. Global prevalence and economics of infection with *Mycobacterium avium* subsp. *paratuberculosis* in ruminants, p 10–21. In Behr MA, Collins DM (ed), Paratuberculosis: organism, disease, control. Wallingford, Oxfordshire.
2. Barkema HW, Orsel K, Nielsen SS, Koets AP, Rutten V, Bannantine JP, Keefe GP, Kelton DF, Wells SJ, Whittington RJ, Mackintosh CG, Manning EJ, Weber MF, Heuer C, Forde TL, Ritter C, Roche S, Corbett CS, Wolf R, Griebel PJ, Kastelic JP, De Buck J. 2018. Knowledge gaps that hamper prevention and control of *Mycobacterium avium* subspecies *paratuberculosis* infection. Transbound Emerg Dis 65 Suppl 1:125–148. https://doi.org/10.1111/tbed.12723
3. Brauning R, Plain K, Gautam M, Russell T, Correa CC, Biggs P, Whittington R, Murray A, Price-Carter M. 2019. Complete genome sequence of the telford type S strain of *Mycobacterium avium* subsp. *paratuberculosis*. Microbiol Resour Announc 8:e00004-19. https://doi.org/10.1128/MRA.00004-19
4. Turenne CY, Collins DM, Alexander DC, Behr MA. 2008. *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* are independently evolved pathogenic clones of a much broader group of *M. avium* organisms. J Bacteriol 190:2479–2487. https://doi.org/10.1128/JB.01691-07
5. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
7. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0
8. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. Nucleic Acids Res 49:D1020–D1028. https://doi.org/10.1093/nar/gkaa1105