Draft Genome Sequences of Clinical Respiratory Isolates of *Mycobacterium goodii* Recovered in Ireland

Simone Mok,[a,b] Peter R. Flanagan,[a,b] Emma Roycroft,[a,b] Thomas R. Rogers,[a,b] Margaret M. Fitzgibbon[a,b]

aIrish Mycobacteria Reference Laboratory, St. James's Hospital, Dublin, Ireland
bDepartment of Clinical Microbiology, Trinity College Dublin, Dublin, Ireland

Simone Mok and Peter R. Flanagan contributed equally to this work. Author order is based on the fact that Simone Mok initiated the laboratory work.

ABSTRACT   Here, we describe the draft genomes of five *Mycobacterium goodii* isolates that were recovered from respiratory clinical specimens in Ireland. Currently, one complete genome and one draft genome exist publicly for *M. goodii*.

Nontuberculous mycobacteria (NTM) are ubiquitous environmental microorganisms. With over 190 species identified, they are commonly found in water systems, vegetation, and soil (1). NTM are an important cause of human infections and may cause pulmonary disease that can resemble tuberculosis (2, 3). Of the many NTM, *Mycobacterium goodii* is an emerging pathogen. First described in 1999, it is a rapidly growing organism that is associated with nosocomial infections and is related to *Mycobacterium smegmatis* (4). *M. goodii* is resistant to many forms of decontamination and sterilization and has been reported in cases of infected pacemaker sites and osteomyelitis associated with prosthetic material (3). Moreover, it is usually resistant to clarithromycin and rifampin due to overexpression of the *wag31* gene, resulting in thickening of the peptidoglycan layer, as well as the presence of the *erm* gene, which confers macrolide resistance (5). Therefore, *M. goodii* presents a unique challenge in health care settings, and potential outbreaks caused by this rare pathogen are yet to be investigated by whole-genome sequencing (WGS).

Between 2011 and 2020, five *M. goodii* isolates were recovered from bronchoalveolar lavage or sputum samples from patients throughout Ireland. In all instances, samples were processed in the Irish Mycobacteria Reference Laboratory (IMRL), where culture confirmed the genus as *Mycobacterium* using the GenoType *Mycobacterium* CM line-probe assay v2.0 (Bruker-Hain Diagnostics, Germany). Further analysis using the GenoType *Mycobacterium* AS assay v1.0 confirmed the species as *M. goodii*.

For WGS, each isolate was cultured in a Bactec MGIT 960 system according to the manufacturer's instructions (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). Prior to DNA extraction, cultures were heat inactivated for 15 min at 95°C. Following this, genomic DNA extraction was performed using the Fuji QuickGene DNA tissue kit S (Kurabo, Osaka, Japan) according to the manufacturer's guidelines. Paired-end libraries were prepared using Nextera XT DNA library preparation kits (Illumina, San Diego, CA, USA), and sequencing was performed on an Illumina MiSeq system using a 2 × 300-bp reagent kit v3 as specified by the manufacturer. Unless stated otherwise, all software used default parameters. The sequenced reads were assessed for quality and trimmed using Fastp v0.20.0 (https://github.com/OpenGene/fastp) (6). Following quality control and trimming, *de novo* assembly was performed using the SPAdes genome assembler v3.14.0 (https://github.com/ablab/spades) (7). Each genome was uploaded to the online open-source software and data platform KBase (8). The platform contains a number of analytical tools, and genome annotation was performed using RASTtk v1.073 (9). Whole-genome single-nucleotide polymorphism (SNP) analysis was performed by mapping sequencing reads to the complete genome sequence of *M. goodii*

TABLE 1 WGS metrics for *M. goodii* isolates

Isolate	No. of contigs	G+C content (%)	Contig N_{50} (bp)	Total no. of reads	Mapping coverage (×)	Genome size (bp)	Total no. of genes	No. of coding genes	No. of noncoding RNAs	No. of noncoding repeats	No. of noncoding prophages	No. of noncoding CRISPR genes[a]	BioSample accession no.	GenBank accession no.
11IE50	202	66.82	232,271	863,356	25	7,459,433	7,734	7,476	50	208	NA[b]	NA	SAMN19031642	JAHBON000000000
14IE55	181	66.99	219,682	1,530,444	31	6,998,916	7,104	6,940	51	112	1	NA	SAMN19031643	JAHBOM000000000
16IE56	222	66.89	159,038	2,102,252	43	7,406,805	7,764	7,526	74	145	1	18	SAMN19031644	JAHBOL000000000
19IE239	229	66.87	128,208	1,794,620	37	7,210,549	7,516	7,286	51	178	1	NA	SAMN19031645	JAHBOK000000000
20IE209	438	66.80	78,153	1,577,501	34	7,340,765	7,902	7,452	46	403	1	NA	SAMN19031646	JAHBOJ000000000

[a] Noncoding CRISPR genes consisted of CRISPR_array, CRISPR_repeat, and CRISPR_spacer.
[b] NA, not applicable.

X7B using Snippy v3.2 (https://github.com/tseemann/snippy), and the average coverage was assessed using Qualimap (10).

Five *M. goodii* genomes were assembled from 7,868,173 reads, and all genomes contained a number of contigs of >200 bp; all associated metrics for each sample are presented in Table 1. The mean coverage of sequencing reads mapped to the reference genome was 34×. Whole-genome SNP analysis of our clinical *M. goodii* isolates and *M. goodii* strain ST0139456 (GenBank accession number PEBB00000000) revealed an average pairwise distance of >20,000 SNPs, indicating different origins (11).

This report presents the first draft genomes of respiratory *M. goodii* isolates from Ireland. The data will serve as a useful reference for WGS studies of *M. goodii*.

Data availability. The draft genome of each isolate was deposited in the NCBI database. The BioProject number is PRJNA727929, with BioSample numbers SAMN19031642 (11IE50), SAMN19031643 (14IE55), SAMN19031644 (16IE56), SAMN19031645 (19IE239), and SAMN19031646 (20IE209). The assembled WGS data were deposited in GenBank with accession numbers JAHBON000000000 (11IE50), JAHBOM000000000 (14IE55), JAHBOL000000000 (16IE56), JAHBOK000000000 (19IE239), and JAHBOJ000000000 (20IE209).

REFERENCES

1. Ratnatunga CN, Lutzky VP, Kupz A, Doolan DL, Reid DW, Field M, Bell SC, Thomson RM, Miles JJ. 2020. The rise of non-tuberculosis mycobacterial lung disease. Front Immunol 11:95. https://doi.org/10.3389/fimmu.2020.00303.
2. Moore JE, Kruijshaar ME, Ormerod LP, Drobniewski F, Abubakar I. 2010. Increasing reports of non-tuberculous mycobacteria in England, Wales and Northern Ireland, 1995–2006. BMC Public Health 10:612. https://doi.org/10.1186/1471-2458-10-612.
3. Diaz A, Ardura MI, Wang H, Antonara S, Ouellette CP. 2020. Osteomyelitis due to *Mycobacterium goodii* in an adolescent, United States. Emerg Infect Dis 26:2781–2783. https://doi.org/10.3201/eid2611.200206.
4. Ferguson DD, Gershman K, Jensen B, Arduino MJ, Yakrus MA, Cooksey RC, Srinivasan A. 2004. *Mycobacterium goodii* infections associated with surgical implants at Colorado hospital. Emerg Infect Dis 10:1868–1871. https://doi.org/10.3201/eid1010.040402.
5. Salas NM, Klein N. 2017. *Mycobacterium goodii*: an emerging nosocomial pathogen: a case report and review of the literature. Infect Dis Clin Pract (Baltim Md) 25:62–65. https://doi.org/10.1097/IPC.0000000000000428.
6. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
8. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.
9. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olsen R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365–8366. https://doi.org/10.1038/srep08365.
10. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294. https://doi.org/10.1093/bioinformatics/btv566.
11. Allam M, Joseph L, Ismail F, Said H, Ismail NA, Ismail A, Mtshali S, Mnyameni F, Goussard P, Pekeur JC, Lourens A, Omar SV. 2018. Whole-genome sequence of a *Mycobacterium goodii* isolate from a pediatric patient in South Africa. Genome Announc 6:e01478-17. https://doi.org/10.1128/genomeA.01478-17.