

Revised whole genome and DNA methylome of *Mycobacterium marinum* type strain ATCC 927^T

Kirsi Savijoki,^{1,2} Paulina Deptula,³ Richard J. Roberts,⁴ Milka Hammarén,^{5,6} Jari Yli-Kauhaluoma,¹ Pekka Varmanen,² Matalena Parikka⁵

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Mycobacterium marinum*, a slow-growing Actinobacterium, typically induces tuberculosis-like disease in fish. Here, we report a new reference sequence for *M. marinum* ATCC 927^T, along with its DNA methylome. This aims to maximize the research potential of this type strain and facilitates investigations into the pathomechanisms of human tuberculosis.

KEYWORDS *Mycobacterium marinum*, tuberculosis, SMRT-PacBio, genome analysis, plasmid, DNA methylome

Mycobacterium marinum (Mmr) strain ATCC 927^T, a slow-growing Gram-positive bacterium, was originally isolated from spontaneous tuberculosis (TB) in fish (1, 2). Phylogenetic analysis based on 16S rRNA gene sequences revealed 99.3% identity to *Mycobacterium tuberculosis* (Mtb), which is the leading etiological agent of TB in humans (3, 4). Like Mtb, Mmr also survives inside its natural host (zebrafish) and triggers a positive tuberculin test during infection (5–10), making it an excellent surrogate model for Mtb.

We revised the whole-genome sequence of ATCC 927^T using PacBio SMRT, identified the methylation motifs, and matched them to corresponding restriction-modification (R-M) systems. Freeze-dried ATCC 927^T, from the American Type Culture Collection, was cultured in Middlebrook 7H9 broth containing 10% Albumin–Dextrose–Catalase (Thermo Fisher Scientific) and 0.5% glycerol for 7 days at 29°C. The genomic DNA was extracted using the MagAttract HMW DNA Kit (Qiagen, Nordic). Library preparation, quality control, sequencing, and assembly were performed by Novogene (UK) using their standard operating procedures. PacBio libraries were prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences), following the standard PacBio protocol (11) and selecting DNA fragments > 17 kb using the BluePippin system (Sage Sciences). The libraries sequenced on a PacBio Sequel II instrument using CLR mode generated a total of 334,798 reads with an average length of 77,683 bp and an N_{50} value of 120,575 bp. Default parameters were used for all the following software unless otherwise specified. Base calling and demultiplexing were performed using SMRT Link V11.1 software (11). Canu-1.6 (12) and HGAP-4 (13) were used for *de novo* assembly of the PacBio long reads, incorporating pre-processing steps for Canu and as a separate step for HGAP. Alignment of pre-assembled reads, Arrow-2.3 (14), and Circlator-1.5.3 (15) were used for error correction, polishing, and circular assembly processes, respectively. The completeness of the assemblies was assessed using BUSCO 4.0.2 (16). The resulting two scaffolds with read depths of 499× for the chromosome and 309× for the plasmid were annotated at NCBI using the Prokaryotic Genome Annotation Pipeline PGAP-6.5 (17–19). Prophage sequences were predicted using the PHASTER server (20), and SEQWARE methylome analysis at REBASE (21) was conducted as previously described (22, 23). Due to the high sequence identity (99.63%) of the plasmid to the previously published plasmid sequence of ATCC 927^T (24), the plasmid was named pMMRN. Table 1 lists all

Editor Kenneth M. Stedman, Portland State University, Portland, Oregon, USA

Address correspondence to Kirsi Savijoki, kirsi.savijoki@helsinki.fi, Paulina Deptula, deptula@food.ku.dk, or Matalena Parikka, matalena.parikka@tuni.fi.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 21 November 2023

Accepted 15 February 2024

Published 28 February 2024

[This article was published on 28 February 2024 with some author given names and surnames reversed in the byline. The byline was updated in the current version, posted on 4 March 2024.]

Copyright © 2024 Savijoki 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/).

TABLE 1 Comparison of the whole genome details of the Mmr ATCC 927^T sequenced here with the same strain sequenced earlier by Nanopore/Illumina (NI) (25) and its passage variant CCUG 20998 using PacBio SMRT (PB) (26)^a

Name of the strain	Accession no.	Size (bp)	Genome					Plasmid				Origin	
			G + C (%)	No. Genes	No. CDSs	No. pgenes ^d	No. RNAs	Accession no.	Size (bp)	No. Genes	No. tRNAs		G + C (%)
ATCC 927 ^T (NI)	NZ_AP018496.1	6,451,936	65.5	5,744	5,150	537	57	NZ_AP018497.1	127,402 ^d	130	1	63.50	25
CCUG 20998 (PB)	NZ_CP024190.1	6,453,310	65.7	5,411	5,203	152	56	NR ^b	127,576 ^c	112 ^c	1 ^c	63.96 ^c	26
ATCC 927 ^T (PB)	CP129323	6,455,215	65.7	5,502	5,300	145	57	CP129324	127,576	112	1	63.96	This study

^apgenes, pseudogenes.

^bNR, not reported.

^cDue to the lack of plasmid sequence at NCBI, the annotation was not updated from the original.

^dThe difference in length is a result of the accumulation of sequencing errors, particularly in regions of homopolymeric stretches, as determined by BLASTN 2.15.0 + nucleotide alignment.

^eFor an accurate comparison, all previously generated whole-genome sequences were obtained from the RefSeq database, where they were subjected to PGAP re-annotation.

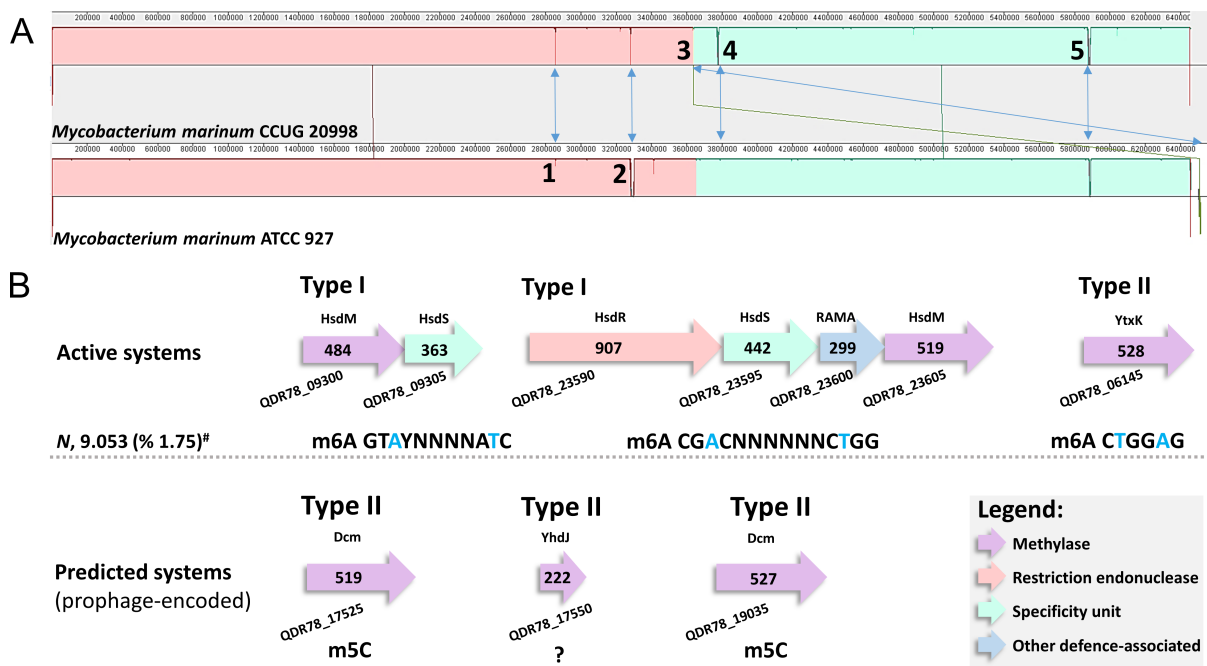


FIG 1 (A) Whole genome alignment of strain ATCC 927 to its previously sequenced passage variant CCUG 20998. ProgressiveMauve (27) was used to align the indicated genomes in a step-by-step manner for the identification of conserved regions and rearrangements between the two genomes. Significant differences are indicated with arrows and numbered as follows. 1, a truncated ABC transporter permease (QDR78_11875 versus intact CCUG20998_RS12005) due to a sequence deletion, 2, a truncated PE family protein due to a frameshift (QDR78_13350 versus CCUG20998_RS28445); pseudogenes on both genomes. 3, an insertion of six genes (CCUG20998_RS15230–CCUG20998_RS15255) coding for an IS3 family transposase, recombinase (a pseudogene), cytochrome P450, TetR/AcrR family transcriptional regulator, a frameshifted TetR/AcrR family transcriptional regulator, and a frameshifted IS256 family transposase. These genes correspond to plasmid-located genes in ATCC 927^T (QDR78_27190–QDR78_27215) coding for TetR/AcrR family transcriptional regulator, TetR/AcrR family transcriptional regulator, cytochrome P450, DUF308 domain-containing protein (a membrane-associated protein), a pseudogene recombinase family protein, and an IS3 family transposase, respectively. 4, a truncated non-ribosomal peptide synthase/polyketide synthase, which is also split into two parts: CCUG20998_RS28465 (AMP-binding protein, 322 amino acids) and CCUG20998_RS28470 (amino acid adenylation domain-containing protein, 1,468 amino acids), versus QDR78_13590 (9,908 amino acids). 5, another non-ribosomal peptide synthase/polyketide synthase: CCUG20998_RS15760 (12,024 amino acids) versus QDR78_15505 (9,880 amino acids). (B) Active and predicted R-M systems found in strain ATCC 927^T, along with matched methylation motifs detected through SMRT whole methylome analysis. The motifs GTAYNNNNATC and CTGGAG and their corresponding R-M systems are also found in the *Mycobacterium marinum* strain MMA1 (REBASE org. #46029). Other PacBio-sequenced strains contain either the first (strain 050012) or the second (*Mycobacterium marinum* strains E11 and H01) of the R-M system-motif pairs, whereas the R-M system with recognition motif CGACNNNNNNCTGG is unique, with modified adenines (indicated as T for complementary strand) marked in blue. Strain MMA1, a human-infection-associated isolate (26), also possesses an additional Type I system with recognition motif CCACNNNNNNNTCCC that is not found in other strains. No activity was detected for the prophage-encoded methyltransferases. #, numbers of detected m6A type modifications (% , number of detected m6A modifications per total number of all detected modifications).

central genomic data of ATCC 927^T in comparison to other related Mmr genomes. Figure 1 illustrates the major differences between ATCC 927^T and its passage variant CCUG 20998, also defined by PacBio SMRT (25), as well as indicates the methylation motifs with their corresponding active and predicted R-M systems. In summary, this study provides a new reference sequence for ATCC 927^T and reports the first in-depth DNA methylome analysis for this non-tuberculous bacterial model.

ACKNOWLEDGMENTS

This work was supported by H2020-MSCA-IF grant number 845658 (P.D.); Academy of Finland grant numbers 322012 and 348973 (J.Y.-K.), 325784 (P.V.), 338624 (M.H.), and 348968 (M.P.); and grants from Jane and Aatos Erkko Foundation (M.P.), Sigrid Jusélius Foundation (M.P.), and Tampere Tuberculosis Foundation (M.P. and M.H.).

Novogene Co, Ltd is greatly acknowledged for sequencing and bioinformatic analyses.

AUTHOR AFFILIATIONS

¹Drug Research Program, Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland

²Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland

³Department of Food Science (FOOD), University of Copenhagen, Frederiksberg, Denmark

⁴New England Biolabs, Ipswich, Massachusetts, USA

⁵Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

⁶European Molecular Biology Laboratory, Heidelberg, Germany

AUTHOR ORCIDs

Kirsi Savijoki  <http://orcid.org/0000-0001-5325-925X>

Paulina Deptula  <http://orcid.org/0000-0002-8854-3573>

Milka Hammarén  <http://orcid.org/0000-0001-9076-8782>

Pekka Varmanen  <http://orcid.org/0000-0002-4508-3060>

Mataleena Parikka  <http://orcid.org/0000-0001-5555-3815>

FUNDING

Funder	Grant(s)	Author(s)
EC H2020 PRIORITY 'Excellent science' H2020 Marie Skłodowska-Curie Actions (MSCA)	845658	Paulina Deptula
Research Council of Finland (AKA)	338624	Milka Hammarén
Research Council of Finland (AKA)	325784	Pekka Varmanen
Research Council of Finland (AKA)	322012	Jari Yli-Kauhaluoma
Research Council of Finland (AKA)	348973	Jari Yli-Kauhaluoma
Research Council of Finland (AKA)	348968	Mataleena Parikka
Jane ja Aatos Erkon Säätiö (J&AE)		Mataleena Parikka
Sigrid Juséliuksen Säätiö (Sigrid Jusélius Stiftelse)		Mataleena Parikka
Tampereen Tuberkuloosisäätiö (Tampere Tuberculosis Foundation)		Mataleena Parikka

AUTHOR CONTRIBUTIONS

Kirsi Savijoki, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft | Paulina Deptula, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software,

Validation, Visualization, Writing – original draft | Richard J. Roberts, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review and editing | Milka Hammarén, Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – review and editing | Jari Yli-Kauhaluoma, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Pekka Varmanen, Funding acquisition, Investigation, Project administration, Resources, Software, Supervision, Writing – original draft | Matalena Parikka, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The ATCC 927^T genome and plasmid pMMRN sequences are available at NCBI with accession numbers [CP129323](https://www.ncbi.nlm.nih.gov/nuclseq/CP129323) and [CP129324](https://www.ncbi.nlm.nih.gov/nuclseq/CP129324), respectively. The DNA methylome analysis is available at REBASE with the Org# number [27274](https://www.ncbi.nlm.nih.gov/rebase/methyl/27274). The BioProject and SRA database accession numbers are [PRJNA955429](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA955429) and [SRX22236853](https://www.ncbi.nlm.nih.gov/sra/SRX22236853).

REFERENCES

- Aronson JD. 1926. Spontaneous tuberculosis in soft water fish. *J Infect Dis* 39:315–320. <https://doi.org/10.1093/infdis/39.4.315>
- Ackleh AS, Delcambre ML, Sutton KL. 2015. A second-order high-resolution finite difference scheme for a size-structured model for the spread of *Mycobacterium marinum*. *J Biol Dyn* 9:156–187. <https://doi.org/10.1080/17513758.2014.962998>
- Akram SM, Aboobacker S. 2020. *Mycobacterium marinum* infection. In *StatPearls*. StatPearls Publishing, Treasure Island. <https://www.ncbi.nlm.nih.gov/books/NBK441883>.
- Bagcchi S. 2023. WHO's global tuberculosis report 2022. *Lancet Microbe* 4:e20. [https://doi.org/10.1016/S2666-5247\(22\)00359-7](https://doi.org/10.1016/S2666-5247(22)00359-7)
- Luukinen H, Hammarén MM, Vanha-Aho L-M, Parikka M. 2018. Modeling tuberculosis in *Mycobacterium marinum* infected adult zebrafish. *J Vis Exp* 8:58299. <https://doi.org/10.3791/58299>
- Parikka M, Hammarén MM, Harjula S-KE, Halfpenny NJA, Oksanen KE, Lahtinen MJ, Pajula ET, Iivanainen A, Pesu M, Rämetsä M. 2012. *Mycobacterium marinum* causes a latent infection that can be reactivated by gamma irradiation in adult zebrafish. *PLoS Pathog* 8:e1002944. <https://doi.org/10.1371/journal.ppat.1002944>
- Tobin DM, Ramakrishnan L. 2008. Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*. *Cell Microbiol* 10:1027–1039. <https://doi.org/10.1111/j.1462-5822.2008.01133.x>
- Takaki K, Davis JM, Winglee K, Ramakrishnan L. 2013. Evaluation of the pathogenesis and treatment of *Mycobacterium marinum* infection in zebrafish. *Nat Protoc* 8:1114–1124. <https://doi.org/10.1038/nprot.2013.068>
- Ferreras-Colino E, Contreras M, Rivalde MA, Sevilla IA, Delgado E, Domínguez L, Gortazar C, de la Fuente J. 2023. Heat-inactivated mycobacteria activate the toll-like receptor 2 and 4 pathways in the zebrafish model of tuberculosis. *EbioRxiv*. <https://doi.org/10.1101/2023.03.31.535083>
- Chen L, Liu Z, Su Y, Wang D, Yin B, Shu B, Zhang J, Zhu X, Jia C. 2017. Characterization of *Mycobacterium marinum* infections in zebrafish wounds and sinus tracts. *Wound Repair Regen* 25:536–540. <https://doi.org/10.1111/wrr.12540>
- Pacific Biosciences of California. 2014. Template preparation and sequencing guide.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>
- Lee C. 2003. Generating consensus sequences from partial order multiple sequence alignment graphs. *Bioinformatics* 19:999–1008. <https://doi.org/10.1093/bioinformatics/btg109>
- 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>
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwazdz 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>
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwazdz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–21. <https://doi.org/10.1093/nar/gkw387>
- Roberts RJ, Vincze T, Posfai J, Macelis D. 2023. REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 51:D629–D630. <https://doi.org/10.1093/nar/gkac975>
- Deptula P, Loivamaa I, Smolander O-P, Laine P, Roberts RJ, Piironen V, Paulin L, Savijoki K, Auvinen P, Varmanen P. 2019. Red-brown pigmentation of *Acidipropioni bacterium jensenii* is tied to haemolytic activity and *cyl*-like gene cluster. *Microorganisms* 7:512. <https://doi.org/10.3390/microorganisms7110512>
- Deptula P, Laine P, Paulin L, Auvinen P, Roberts RJ, Johnston CD, Varmanen P. 2020. Complete genome sequences and methylome analyses of *Cutibacterium acnes* subsp. *acnes* strains DSM 16379 and DSM 1897^T. *Microbiol Resour Announc* 9:e00705–20. <https://doi.org/10.1128/MRA.00705-20>
- Yoshida M, Fukano H, Miyamoto Y, Shibayama K, Suzuki M, Hoshino Y. 2018. Complete genome sequence of *Mycobacterium marinum* ATCC 927^T, obtained using Nanopore and Illumina sequencing technologies. *Genome Announc* 6:e00397–18. <https://doi.org/10.1128/genomeA.00397-18>

25. Das S, Pettersson BMF, Behra PRK, Mallick A, Cheramie M, Ramesh M, Shirreff L, DuCote T, Dasgupta S, Ennis DG, Kirsebom LA. 2020. Extensive genomic diversity among *Mycobacterium marinum* strains revealed by whole genome sequencing. *Sci Rep* 10:5246. <https://doi.org/10.1038/s41598-018-30152-y>
26. Manion M, Dulanto Chiang A, Pei L, Wong C-S, Khil P, Hammoud DA, Anderson M, Laidlaw E, Kuriakose S, Lisco A, Zelazny AM, Dekker JP, Sereti I, NISC Comparative Sequencing Program. 2021. Disseminated *Mycobacterium marinum* in human immunodeficiency virus unmasked by immune reconstitution inflammatory syndrome. *J Infect Dis* 224:453–457. <https://doi.org/10.1093/infdis/jiaa769>
27. Darling AE, Mau B, Perna NT. 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>