

Complete genome sequence of a new *Mycobacterium* sp. nov. from a compromised Japanese individual

Nobuhiro Asai,^{1,2} Yuriko Igarashi,³ Narimi Miyazaki,² Asami Osugi,³ Yuki Matsumoto,⁴ Shota Nakamura,⁴ Satoshi Mitarai,^{3,5} Hiroshige Mikamo^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we report the complete genome sequencing of a new *Mycobacterium* sp. nov. from a compromised Japanese individual with hematological malignancy. This will represent the essential data for future phylogenetic and comparative genome studies and will be useful for better understanding the evolution of the pathogen.

KEYWORDS *Mycobacterium nagakutense*, nagakute

We report the complete genome sequence of *Mycobacterium* sp. strain AMU20-3851, which was first clinically isolated from a human with hematological malignancy. This genome sequence report was approved by the Ethics Committee, Aichi Medical University School of Medicine (AMU 2025-182). The clinical isolate was obtained at our institute, located in Nagakute City, Aichi Prefecture, Japan. The pathogen was isolated after 90 h in blood culture (BD BACTEC FX blood culture system, Becton Dickinson, Sparks, MD, USA). The strain was grown with Ogawa growth medium at 37°C. DNA was purified with the standard phenol-chloroform method after enzymatic (100 µg/mL proteinase K with 2% SDS) treatment by bead beating (0.2 mm diameter zirconia beads) (1). The genome sequence was performed by the Kazusa DNA Research Institute (Kisarazu, Japan) on commission of PacBio Sequel II (Pacific Biosciences, USA). For PacBio Sequel II sequencing, DNA was fragmented using g-TUBE (Covaris, USA), sequencing libraries were prepared by using the PacBio SMRTbell express template prep kit 2.0 (Pacific Biosciences), and libraries between 8 and 20 kb were selected using BluePippin (Sage Science, USA). Circular Consensus Sequencing (CCS) was performed with a Sequel II (Pacific Biosciences) using the PacBio Sequel II sequencing kit v2.0 (Pacific Biosciences). High-fidelity (HiFi) reads were generated from the obtained subreads using the CCS package v6.4.0 in PacBio tools distributed via bioconda (2). Sequence quality was assessed using longQC (3).

A total of 3,267,053 subreads were obtained, from which 201,863 HiFi reads were generated. The HiFi reads were assembled into one circular contig by using the hifiasm 0.16.1-r375 (4) and rotated using the circulator v1.5.5 (5) to start at *dnaA*, whose sequence was predicted by prokka v1.14.6 (6, 7). The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline. Default software parameters were used unless otherwise noted. Average nucleotide identity (ANI) analysis was calculated using FastANI v1.34 (8, 9).

The chromosome of AMU20-3851 is 6,056,820 bp in length with 68% G + C content, encoding 5,822 predicted genes, and sequenced to an average depth of 311×. The ANI for this strain was 90.5% for *Mycobacterium diernhoferi* and 84.4% for *Mycobacterium frederiksbergense*, the raw data and the assembled genome were deposited in NCBI GenBank and SRA databases (the accession no. [GCF_043974665.1](https://www.ncbi.nlm.nih.gov/submit/seq/submit.cgi?accession=GCF_043974665.1)). This genome sequence will represent the essential data for future phylogenetic and comparative

Editor Jason E. Stajich, University of California
Riverside, Riverside, California, USA

Address correspondence to Hiroshige Mikamo,
mikamo@aichi-med-u.ac.jp.

The authors declare no conflict of interest.

Received 28 April 2025

Accepted 15 November 2025

Published 29 December 2025

Copyright © 2025 Asai 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/).

genome studies and will be useful for better understanding the evolution of the new species, *Mycobacterium* sp. strain AMU20-3851.

AUTHOR AFFILIATIONS

¹Department of Clinical Infectious Diseases, Aichi Medical University, Nagakute, Aichi, Japan

²Department of Infectious Control Diseases, Aichi Medical University, Nagakute, Aichi, Japan

³The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan

⁴Department of Infection Metagenomics, Bioinformatics Center, Research Institute for Microbial Diseases, Osaka University, Suita, Japan

⁵Department of Basic Mycobacteriology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

AUTHOR ORCID*s*

Asami Osugi  <http://orcid.org/0000-0002-1354-6040>

Satoshi Mitarai  <http://orcid.org/0000-0002-2288-077X>

Hiroshige Mikamo  <http://orcid.org/0000-0001-8876-6411>

AUTHOR CONTRIBUTIONS

Nobuhiro Asai, Conceptualization, Data curation, Supervision, Writing – original draft, Writing – review and editing | Yuriko Igarashi, Formal analysis, Investigation, Methodology, Writing – review and editing | Narimi Miyazaki, Formal analysis | Asami Osugi, Formal analysis, Investigation, Methodology | Yuki Matsumoto, Formal analysis, Investigation, Methodology | Shota Nakamura, Formal analysis, Investigation, Methodology | Satoshi Mitarai, Formal analysis, Investigation, Methodology, Supervision, Writing – review and editing | Hiroshige Mikamo, Conceptualization, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

This complete genome sequence and sequence raw data have been deposited at GenBank under the accession no. [GCF_043974665.1](https://www.ncbi.nlm.nih.gov/GenBank/ accession/GCF_043974665.1) and raw PacBio reads in the SRA database under [SRR28129243](https://www.ncbi.nlm.nih.gov/SRA/ accession/SRR28129243).

REFERENCES

1. Sambrook J, Russell DW. 2006. Extraction of bacteriophage λ DNA from large-scale cultures using proteinase K and SDS, p 3972. Cold Spring Harbor Protocols.
2. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J, BiocondaT. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat Methods 15:475–476. <https://doi.org/10.1038/s41592-018-0046-7>
3. Fukasawa Y, Ermini L, Wang H, Carty K, Cheung MS. 2020. LongQC: a quality control tool for third generation sequencing long read data. G3 (Bethesda) 10:1193–1196. <https://doi.org/10.1534/g3.119.400864>
4. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat Methods 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
5. 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>
6. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
7. Haft DH, Badretdin A, Coulouris G, DiCuccio M, Durkin AS, Jovenitti E, Li W, Mersha M, O'Neill KR, Virothaisakun J, Thibaud-Nissen F. 2024. RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes. Nucleic Acids Res 52:D762–D769. <https://doi.org/10.1093/nar/gkad988>
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>
9. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>