



Data Article

Whole genome sequence data of *Mycolicibacterium parafortuitum* Panama NTM 1 from a dairy farm in Panama



Johanna Elizabeth Ku^{a,1}, Alejandro Llanes^{a,1},
Francisco Guizado-Batista^{b,1}, Emmanuel Hernández-Ricord^b,
Amanda Ramírez-Casanova^b, Pave Mislov-Vallarino^b,
Alexa Prescilla-Ledezma^{b,c}, Nicolas Torrales^c, Fermín Acosta^a,
Amador Goodridge^{a,*}, Nora Ortiz de Moreno^{b,*}

^a Centro de Biología Molecular y Celular de Enfermedades, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT-AIP), Ciudad del Saber, Panama

^b Universidad de Panamá, Facultad de Medicina, Departamento de Microbiología Humana, Panama

^c Rochem Biocare de Panamá, S.A., Panama

ARTICLE INFO

Article history:

Received 25 November 2024

Revised 25 January 2025

Accepted 30 January 2025

Available online 7 February 2025

Dataset link: [Mycolicibacterium parafortuitum strain:Panama NTM 1 \(Original data\)](#)

Keywords:

Mycolicibacterium parafortuitum
Nontuberculous mycobacteria
Environmental mycobacteria
Whole-genome sequencing

ABSTRACT

We report the whole genome sequence of *Mycolicibacterium parafortuitum* strain Panama NTM1, isolated from cattle feces at a dairy farm in Panama (8°08'18.1"N and 80°54'00.1"W). DNA was extracted from a pure culture of this isolate and whole-genome sequencing was performed using the Illumina MiSeq[®] platform. After *de novo* assembly, the genome has a total size of 5.92 Mbp, a GC content of 68.4 %, and 5545 annotated genes. The raw read files and genome have been deposited in the NCBI database under BioProject number PRJNA1113557.

© 2025 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC license
(<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding authors.

E-mail addresses: agoodridge@indicasat.org.pa (A. Goodridge), nora.ortiz@up.ac.pa (N. Ortiz de Moreno).

Social media: [@JkuTox](#) (J.E. Ku), [@amadorgj](#) (A. Goodridge)

¹ These authors contributed equally to this work.

Specifications Table

Subject	Biological Sciences / Omics / Genomics
Specific subject area	Environmental nontuberculous mycobacteria (NTM)
Type of data	Raw data, whole genome sequencing, genome annotation, variant calling, and comparative genomic analysis of a <i>Mycolicobacterium parafortuitum</i> strain.
Data collection	The strain was isolated from deposited cattle feces. The sample was incubated in tubes with Middlebrook 7H9 broth. Afterwards, samples were decontaminated with N-acetylcysteine + NaOH 2 % w/v, and inoculated in Lowenstein-Jensen (LJ) culture medium. Trypticase soy agar (TSA) plates were used to isolate the colonies and LJ culture medium to culture pure colonies. Bacteria DNA was extracted from a pure culture colonies using cetyltrimethylammonium bromide (CTAB) method, and whole-genome sequencing was performed using a MiSeq® Reagent Kit v3 in an Illumina MiSeq® platform. <i>De novo</i> assembly was performed using SPAdes v.3.15.3, genome annotation was performed with the Prokka pipeline v.1.1.14.6.
Data source location	GPS coordinates where the sample was collected: 8°08'18.1"N 80°54'00.1"W District/Province: Santiago, Veraguas Country: Panama
Data accessibility	Repository name: National Center for Biotechnology Information (NCBI) Data identification number: BioProject: PRJNA1113557, BioSample: SAMN41455849, Sequence Read Archive (SRA): SRR29188038, GenBank: CP157737 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1113557 https://www.ncbi.nlm.nih.gov/biosample/SAMN41455849 https://www.ncbi.nlm.nih.gov/sra/?term=SRR29188038 https://www.ncbi.nlm.nih.gov/nucleotide/CP157737.1/
Related research article	None.

1. Value of the Data

- The *M. parafortuitum* Panama NTM 1 strain was isolated for purity using trypticase soy agar (TSA) plates.
- The Panama NTM 1 genome is a reference point of an *M. parafortuitum* strain isolated in Latin America, for which raw reads and the *de novo* genome assembly are publicly available.
- The genomic sequence data of *M. parafortuitum* Panama NTM 1 will be useful for researchers working in the different genomic fields of nontuberculous mycobacteria (NTM).

2. Background

The *Mycobacterium* genus includes more than 150 aerobic bacteria species with a thick lipid cell wall that provides resistance to immune clearance, disinfectants, and laboratory dyes [1,2]. Due to their pathogenicity, the two most known species are *Mycobacterium tuberculosis* and *Mycobacterium leprae*. All other species, often referred to as nontuberculous mycobacteria (NTM), are naturally present in the environment. However, recent human cases involving NTM-causing diseases have increased interest in the characterization of these species [2,3]. A recent study using genome sequences revealed that the *Mycobacterium* genus could be divided into five clades [4]. The novel clades were designated as four new potential genera within the *Mycobacteriaceae* family, including the *Mycolicobacterium* genus [5]. *Mycobacterium parafortuitum* was discovered in Japan as a fast-growing NTM [6] and it was later suggested to belong to one of these recently recognized genera [4]. Here, we present the whole-genome sequence (WGS) of a *Mycolicobacterium parafortuitum* strain from Panama, designated as Panama NTM 1. In this article, we use the name *Mycolicobacterium* for the genus because the repository where the data is available have adopted this new classification scheme, although we acknowledge that these new genera have not been officially recognized by certain bacterial nomenclature authorities.

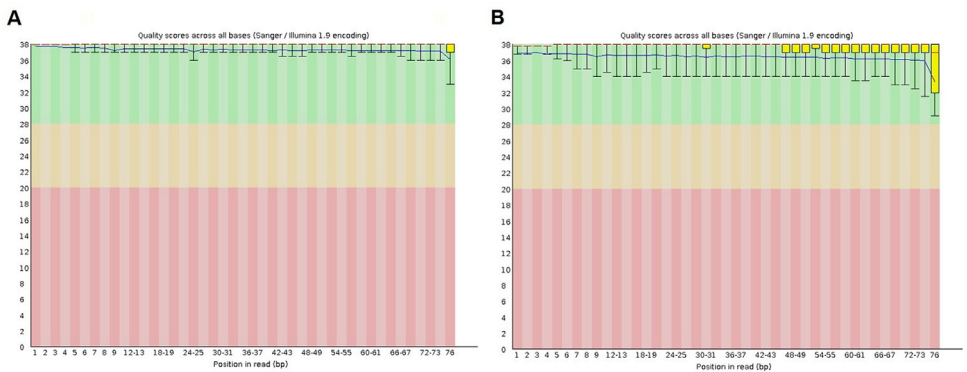


Fig. 1. Quality of read sets used for *de novo* assembly of the *M. parafortuitum* strain Panama NTM 1. A. Average quality per sequence position computed using FastQC for the quality-trimmed forward reads. B. Similar plot computed for reverse reads.

3. Data Description

We present the WGS data of the *M. parafortuitum* Panama NTM 1 strain isolated in Veraguas, Panama. Sequencing and further quality trimming of Illumina reads resulted in a dataset of 13,688,246 reads pairs with a length of 35–76 nucleotides and an average quality above Q28 (Fig. 1). Massive alignment against the non-redundant GenBank nucleotide database (Bethesda, MD, US) revealed that more than 98 % of these reads belong to *Mycolicibacterium parafortuitum*. The further assembled Panama NTM 1 genome has a total size of 5.92 Mbp, a GC content of 68.4 %, and 5545 annotated genes, of which, 5525 are protein-coding and 55 encode ribosomal and transfer RNAs. Post-assembly quality check with CheckM revealed a 99.83 % percent of genome completeness and 0.23 % of contamination, as evaluated with a set of unique marker genes for the *Mycobacteriaceae* family lineage.

A high conservation of gene content and synteny between the Panama NTM 1 genome and that of *M. parafortuitum* type strain JCM 6367 [2] can be seen in Fig. 2. However, alignment of

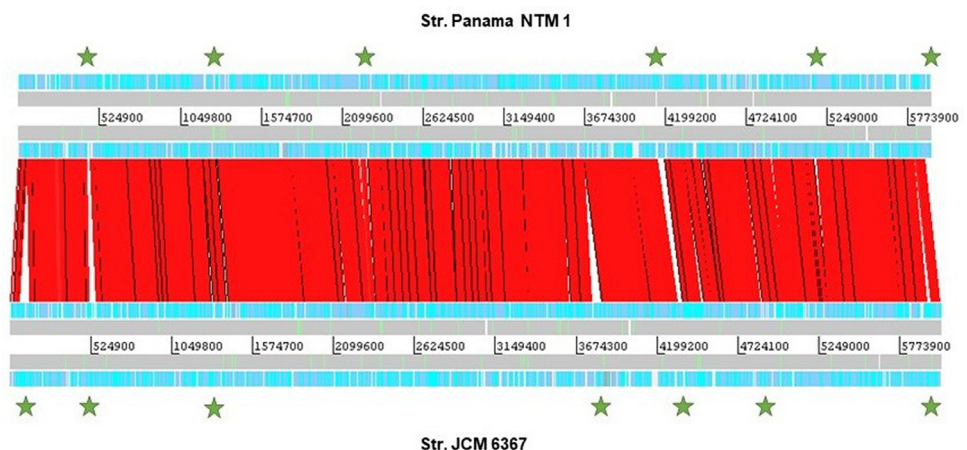


Fig. 2. Comparison of the genomes of *M. parafortuitum* strains Panama NTM 1 and JCM 6367. Regions with large-scale sequence similarity between both genomes (above 80 % on average) are connected by red bands. Protein-coding genes are represented in shades of blue, with darker blue for genes of known or predicted function and light blue for those of unknown function. Transfer RNA genes are colored green and all other RNA genes (including ribosomal RNA) are colored white. Positions of relatively large regions specific to either strain are indicated with green stars.

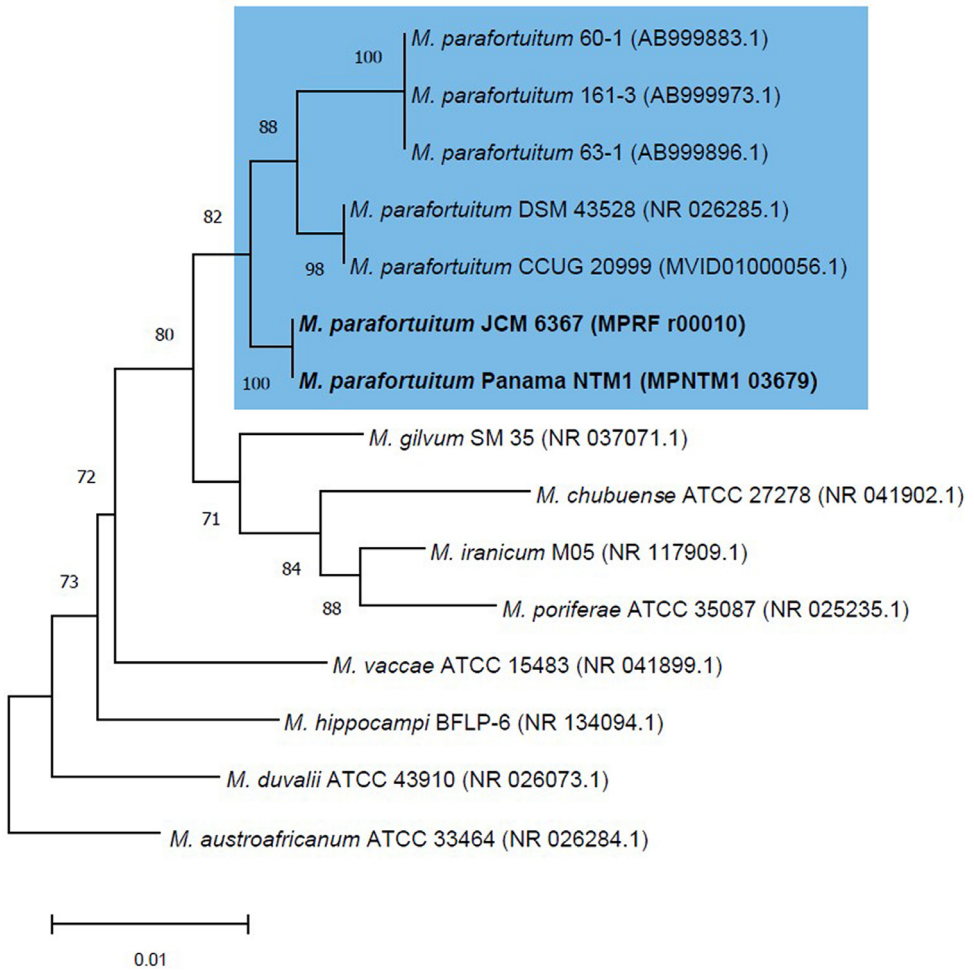


Fig. 3. Phylogenetic tree of 16S rRNA sequences of species closely related to *M. parafortuitum*. The sequences were downloaded from the GenBank database of the National Center for Biotechnology Information (NCBI) (Bethesda, MD, US), with the corresponding accession numbers and strain names indicated next to species names in the tree. Five additional 16S rRNA sequences from *M. parafortuitum* were also included in the tree for comparative purposes. Clades clustering *M. parafortuitum* sequences are shaded in blue, with the Panama NTM 1 strain and type strain JCM 6367 highlighted in bold. In the case of strains Panama NTM 1 and JCM 6367, 16S rRNA sequences were extracted from the corresponding assembled genomes. *M. austroafricanum* was used as an outgroup to root the tree, since it has been placed in a neighboring phylogenetic clade in previous studies.

the Panama NTM 1 Illumina reads to the JCM 6367 reference genome and further variant calling revealed a relatively high number of genetic variants, including 57,014 single nucleotide variants (SNVs) and 2096 insertion/deletions (indels). Of these, 52,703 SNVs were identified within protein-coding genes, the majority of which (98 %) were predicted to have a moderate or low impact on the encoded proteins. Likewise, 1737 indels appear to affect protein-coding genes, 85 % of them with predicted moderate or low functional impact.

In order to better characterize the Panama NTM 1 strain, we initially conducted a phylogenetic analysis with the 16S rRNA sequences of eight species reported to be closely related to *M. parafortuitum* in previous studies [2,4]. The close relationship of the Panama NTM 1 strain to type strain JCM 6367 and the other *M. parafortuitum* strains can also be seen in the resulting

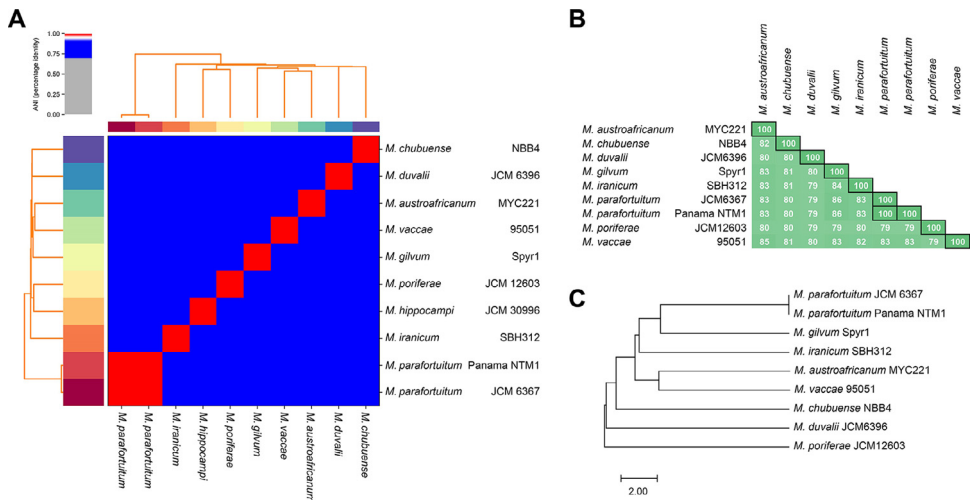


Fig. 4. Phylogenomic approaches to confirm the taxonomic position of strain Panama NTM 1. A. Average nucleotide identity (ANI) between the reference genomes for species closely related to *M. parafortuitum*. B. Average amino acid identity (AAI) calculated from the total set of protein-coding genes of selected species. C. Clustering tree built from the pairwise matrix of AAI shown in panel B. Reference genome sequences were downloaded from the GenBank database, under BioProjects PRJNA695216 (*M. austroafricanum*), PRJDB7717 (*M. hippocampi*), PRJNA53215 (*M. chubuense*), PRJDB7717 (*M. duvalii*), PRJNA28521 (*M. gilvum*), PRJNA224116 (*M. iranikum*), PRJDB7717 (*M. parafortuitum*), PRJDB7717 (*M. poriferae*) and PRJNA210723 (*M. vaccae*). *M. hippocampi* could not be included in panels B and C because its reference genome was too fragmented to perform AAI calculation in a reliable manner.

tree in Fig. 3. To further assess the taxonomic position of the Panama NTM 1 strain, we used two additional phylogenomic approaches based on average nucleotide identity (ANI) and average amino acid identity (AAI), respectively. Fig. 4A shows that the Panama NTM 1 and JCM 6367 were unambiguously clustered together on the basis of ANI. Fig. 4B y C show that these two strains share 100 % AAI and were also grouped together in a clustering tree based on this metric.

4. Experimental Design, Materials and Methods

4.1. Sample collection, bacterial culture, bacterial isolation and DNA extraction

The *M. parafortuitum* Panama NTM 1 strain was isolated from deposited cattle feces collected at a dairy farm in the district of Santiago, Veraguas province, Panama. The strain was isolated by standard microbiology methods described by Fernández de Vega et al. [7]. Briefly, the sample was incubated in tubes with Middlebrook 7H9 broth for 72 h at 28 °C, Ziehl-Neelsen stain was used to confirm the presence of acid-fast bacteria, and decontamination was performed with N-acetylcysteine + NaOH 2 % w/v (Kubica Method) [8]. Then, samples were inoculated in tubes with slanted Lowenstein-Jensen culture medium tubes in duplicate, placing one tube in the light and the other in complete darkness at 37 °C in static position for 72 h. To isolate pure colonies with the most resembling pigmentation, trypticase soy agar (TSA) plates [9] were used. The TSA plates were prepared by suspending 40 g of TSA in 1 L of distilled water, let it stand for 5 mins, mixed, heated while stirring gently, boiled for one minute, let it cool to 25 °C to adjust the pH to 7.3 +/- 0.2, autoclaved at 121 °C for 15 mins, served 20 mL in sterile petri dishes, and allowed the agar to solidify. The colonies were inoculated in the TSA plates in their previous condition (light or dark at 37 °C in static position) for 72 h. An image of the colony in TSA at 4X is available as Supplementary Material. Afterward, the Ziehl-Neelsen stain was used to confirm the presence of acid-fast bacteria, and the colonies were inoculated in the Lowenstein-Jensen medium tubes

in their previous condition (slanted medium, light or dark at 37 °C in static position) for 72 h to obtain pure colonies, which were stored at −80 °C until further molecular testing. Finally, DNA extraction was performed from the pure colonies using the cetyltrimethylammonium bromide (CTAB) method, previously described by Van Soolingen et al. [10].

4.2. Whole genome sequencing

The DNA quantity and quality were assessed by the Qubit BR Assay (Invitrogen, Thermo Scientific, USA). DNA sequences were obtained in the INDICASAT-AIP Genomics Laboratory. The sequencing library was prepared using the MiSeq Reagent Kit v3 (Illumina, USA) according to the manufacturer's instructions and sequenced in a MiSeq instrument (Illumina, USA) for 75 bp paired-end reads.

4.3. Data quality validation, genome assembly, annotation, and variant calling

Illumina reads quality was evaluated using FastQC v.0.11.9 [11] and Trimmomatic v.0.39 [12] was used to remove low-quality positions from the 3' ends. *De novo* assembly was performed using SPAdes v.3.15.3 [13], with the recommended options for MiSeq reads. Genes in the newly assembled genome draft were annotated with the Prokka pipeline v.1.14.6 [14], combining *de novo* gene prediction and transference of gene models previously annotated in the type strain JCM 6367 [2] reference genome. The two lines of evidence for annotation were manually revised and merged using Artemis and the Artemis Comparison Tool (ACT) v.18.2.0 [15]. Reads were aligned to the reference genome with BWA v.0.7.17 [16] and variant calling was further conducted from this alignment with the Genome Analysis Toolkit (GATK) v.4.2 [17]. Genome completeness and degree of contamination were estimated using CheckM v.1.2.3 [18].

4.4. Phylogenetic analysis

Phylogenetic analysis was conducted with the nucleotide sequence of the 16S ribosomal RNA (rRNA) of nine representative *Mycolicibacterium* species, including five additional sequences from *M. parafortuitum* and those from the Panama NTM1 and JCM 6367 strains. The sequences were downloaded from the GenBank database of the National Center for Biotechnology Information (NCBI) (Bethesda, MD, US). The species, strains and accession numbers are detailed in Fig. 3. These sequences were aligned using MAFFT v.7.49 [19]. MEGA v. 7.0 [20] was used to build a maximum likelihood tree with the best evolutionary model predicted by the program (HKY+G + I) and 1000 bootstrap replicates for branch support estimation. *M. austroafricanum* was used as an outgroup to root the tree since this species has been previously placed in a separate clade, when compared to the lineage leading to *M. parafortuitum* [2].

4.5. Phylogenomic analyses

Two phylogenomic approaches were used to validate the classification of the Panama NTM 1 strain. First, PyANI v.0.2.12 [21] was used to calculate the average nucleotide identity (ANI) between the reference genomes for the species included in the previous 16S rRNA phylogenetic analysis. Then, EzAAI v.1.2.3 [22] was used to build a database of protein-coding genes predicted from these reference genomes and to calculate the average amino acid identity (AAI) among these genomes. Both programs have built-in tools to cluster the corresponding samples on the basis of the comparative metric, either ANI or AAI.

Limitations

Although we did our best effort to check the quality and completeness of the newly assembled genome for the Panama NTM 1 strain, it is possible that certain regions could not be assembled completely due to the use of a short-read sequencing platform. This limitation is particularly relevant in genomic regions composed of highly repetitive sequences with relatively short repeat units, which are widely known to be difficult to assemble with short Illumina reads.

Data Availability

[Mycolicibacterium parafortuitum strain:Panama NTM 1 \(Original data\)](#) (NCBI).

Ethics Statement

The authors have read and followed the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Johanna Elizabeth Ku: Formal analysis, Data curation, Writing – original draft, Writing – review and editing. **Alejandro Llanes:** Formal analysis, Data curation, Validation, Funding acquisition, Writing – review and editing. **Francisco Guizado-Batista:** Conceptualization, Methodology, Investigation, Data Curation. **Emmanuel Hernández-Ricord:** Conceptualization, Methodology, Investigation. **Amanda Ramírez-Casanova:** Conceptualization, Methodology, Investigation. **Pave Mislov-Vallarino:** Conceptualization, Methodology, Investigation. **Alexa Prescilla-Ledezma:** Resources, Writing – Original Draft, Writing – Review & Editing, Supervision. **Nicolas Torrales:** Investigation. **Fermín Acosta:** Investigation, Funding acquisition. **Amador Goodridge:** Project administration, Funding acquisition, Writing – review and editing. **Nora Ortiz de Moreno:** Methodology, Validation, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We thank Victoria Batista, Eira Santamaría, Humberto Morris, and Jaime Chen for their support during sample isolation. We also thank Dayra Calvache and Rochem Biocare de Panamá, S.A. for financing the Illumina MiSeq® Reagent Kit v3 (150 cycles) used for DNA sequencing. A.G., F.A., A.L., and A.P. were funded by the [Sistema Nacional de Investigación \(SNI\) of Panama](#), contract numbers [022–2020](#), [072–2022](#), [043–2023](#), and [009–2023](#) respectively. F.A. was funded by the reinsertion scholarship program Grant [DDCCT no 068–2021](#) and Grant [FIED23-07 / 137-2023](#) of [SENACYT](#).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2025.111368](https://doi.org/10.1016/j.dib.2025.111368).

References

- [1] N.C. Engleberg, *Capítulo 23: micobacterias: tuberculosis y lepra*, In: *mecanismos de las enfermedades microbianas* (2013) 257–286.
- [2] Y. Matsumoto, T. Kinjo, D. Motooka, D. Nabeya, N. Jung, K. Uechi, T. Horii, T. Iida, J. Fujita, S. Nakamura, Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles, *Emerg. Microbes Infect.* 8 (2019) 1043–1053, doi:[10.1080/22221751.2019.1637702](https://doi.org/10.1080/22221751.2019.1637702).
- [3] T. Fedrizzi, C.J. Meehan, A. Grottola, E. Giacobazzi, G. Fregni Serpini, S. Tagliazucchi, A. Fabio, C. Bettua, R. Bertorelli, V. De Sanctis, F. Rumpianesi, M. Pecorari, O. Jousson, E. Tortoli, N. Segata, Genomic characterization of nontuberculous mycobacteria, *Sci. Rep.* 7 (2017) 45258, doi:[10.1038/srep45258](https://doi.org/10.1038/srep45258).
- [4] R.S. Gupta, B. Lo, J. Son, Phylogenomics and comparative genomic studies robustly support division of the genus mycobacterium into an emended genus mycobacterium and four novel genera, *Front. Microbiol.* 9 (2018), doi:[10.3389/fmicb.2018.00067](https://doi.org/10.3389/fmicb.2018.00067).
- [5] A. Oren, G. Garrity, List of new names and new combinations previously effectively, but not validly, published, *Int. J. Syst. Evol. Microbiol.* 68 (2018) 1411–1417, doi:[10.1099/ijsem.0.002711](https://doi.org/10.1099/ijsem.0.002711).
- [6] M. Tsukamura, *Mycobacterium parafortuitum*: a new species, *Microbiology* 42 (1966) 7–12, doi:[10.1099/00221287-42-1-7](https://doi.org/10.1099/00221287-42-1-7).
- [7] F.A. Fernández de Vega, J.E. Moreno, J. González Martín, J.J. Palacios Gutiérrez, *Proc. Microbiol. Clín. 9a. Micobacterias 2005* (2005) <https://www.seimc.org/ficheros/documentoscscientificos/procedimientosmicrobiologia/seimc-procedimientomicrobiologia9a.pdf/653-653>. (accessed May 16, 2024).
- [8] L. Wayne, G. Kubica, *The mycobacteria*, in: P. Sneath, N. Mair, M. Sharp, J. Holt (Eds.), *Bergey's Manual of Systematic Bacteriology*, Williams & Wilkins, Baltimore, MD, 1986, pp. 1435–1457.
- [9] M.J. Arduino, L.A. Bland, S.M. Aguero, L. Carson, M. Ridgeway, M.S. Favero, Comparison of microbiologic assay methods for hemodialysis fluids, *J. Clin. Microbiol.* 29 (1991) 592–594.
- [10] D. van Soolingen, P.W. Hermans, P.E. de Haas, D.R. Soll, J.D. van Embden, Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis, *J. Clin. Microbiol.* 29 (1991) 2578–2586, doi:[10.1128/jcm.29.11.2578-2586.1991](https://doi.org/10.1128/jcm.29.11.2578-2586.1991).
- [11] S. Andrews, FastQC: a quality control tool for high throughput sequence data., (n.d.). <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- [12] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120, doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- [13] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Pribelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477, doi:[10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- [14] T. Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics* 30 (2014) 2068–2069, doi:[10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153).
- [15] T. Carver, M. Berriman, A. Tivey, C. Patel, U. Böhme, B.G. Barrell, J. Parkhill, M.-A. Rajandream, Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database, *Bioinformatics* 24 (2008) 2672–2676, doi:[10.1093/bioinformatics/btn529](https://doi.org/10.1093/bioinformatics/btn529).
- [16] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform, *Bioinformatics* 26 (2010) 589–595, doi:[10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698).
- [17] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res.* 20 (2010) 1297–1303, doi:[10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110).
- [18] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, *Genome Res.* 25 (2015) 1043–1055, doi:[10.1101/gr.186072.114](https://doi.org/10.1101/gr.186072.114).
- [19] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.* 30 (2013) 772–780, doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- [20] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets, *Mol. Biol. Evol.* 33 (2016) 1870–1874, doi:[10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054).
- [21] L. Pritchard, R.H. Glover, S. Humphris, J.G. Elphinstone, I.K. Toth, Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens, *Anal. Methods* 8 (2015) 12–24, doi:[10.1039/C5AY02550H](https://doi.org/10.1039/C5AY02550H).
- [22] D. Kim, S. Park, J. Chun, Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity, *J. Microbiol.* 59 (2021) 476–480, doi:[10.1007/s12275-021-1154-0](https://doi.org/10.1007/s12275-021-1154-0).