

## Complete genome sequences and multidrug resistance genotypes of nontuberculous mycobacteria isolates from the Central Tuberculosis Reference Laboratory Muhimbili Tanzania

Hortensia Nondoli<sup>1,5\*</sup>, Reuben Maghembe<sup>2,6</sup>, Winifrida Kidima<sup>3</sup>, Victor Makene<sup>1</sup>, Esther Ngada<sup>4</sup>

<sup>1</sup>Department of Molecular Biology and Biotechnology, College of Natural and Applied Sciences, University of Dar es Salaam, P. O. Box 35179, Dar es Salaam, Tanzania.

<sup>2</sup>Section of Biological and Marine Sciences, Faculty of Natural and Applied Sciences, Marian University College, P. O. Box 47, Bagamoyo, Tanzania.

<sup>3</sup>Department of Zoology and Wildlife Conservation, College of Natural and Applied Sciences, University of Dar es Salaam, P. O. Box 35160, Dar es Salaam, Tanzania.

<sup>4</sup>National Institute of Medical Research, Muhimbili Medical Research Centre, P. Box 3436, Dar es Salaam.

<sup>5</sup>Mbeya College of Health and Allied Sciences, University of Dar Es Salaam, P.O Box 608, Mbeya

<sup>6</sup>Department of Microbiology and Immunology, Faculty of Biomedical Sciences, Kampala International University-WC, Ishaka, Uganda.

### Abstract

**Background:** Nontuberculous mycobacteria (NTM) usually comprise a group of environmental bacteria, with emerging but elusive coinfection with tuberculous mycobacteria, causing pulmonary tuberculosis. Whole genome sequencing may give insight into potential antimicrobial resistance genotypes, giving clinicians and policymakers proper directions in clinical applications and management regimens.

**Methods:** WGS was performed on twenty-four gDNA isolates from archival samples at the Central Tuberculosis Reference Laboratory using the MinION Oxford Nanopore Sequencing approach. Out of twenty-four, two were confirmed to belong to the NTM group. Further analysis was done to resolve the complete genomes of two nontuberculous mycobacteria strains isolated from tuberculosis patients. We then combined phylogenomics, reference-based scaffolding and average nucleotide identity (ANI) analysis to delineate each strain's taxonomic position and corresponding features.

**Results:** Our findings reveal that the two strains fit into the genus *Mycolicibacterium*, and the closest relative is *Mycolicibacterium novocastrense*. Coupling BacAnt and CARD-based antibiotic resistance analyses revealed multidrug-resistant genotypes of diverse spectra and mechanisms. While the BC02 strain is genetically resistant to beta-lactams, macrolides and rifamycins, the BC05 strain portrays an extended drug resistance genotype encompassing beta-lactams, macrolides, polyamines, and aminoglycosides. Both strains possess a single nucleotide polymorphism (SNP) of the RNA polymerase beta-subunit (*rpoB*), representing resistance to the first-line rifampicin. Additionally, the BC05 strain genetically portrays resistance to ethambutol, isoniazid and fosfomycin through mechanisms involving target alteration through SNPs, drug inactivation and efflux.

**Conclusion:** Our findings strongly suggest the potential implication of multidrug-resistant NTM clinical isolates in the pathogenesis of pulmonary tuberculosis.

**Keywords:** Genomics; whole genome sequencing; NTM; phylogenomics; phylogenetics; antimicrobial resistance

---

<sup>1</sup> Corresponding author: Hortensia Nondoli, Email: [hnondoli@gmail.com](mailto:hnondoli@gmail.com)

## Introduction

Non-tuberculous mycobacteria (NTM) usually comprise a group of environmental *Mycobacterium* species (spp) with potential health risks considerably thought to pose a risk to pulmonary health. There are more than 160 NTM species worldwide apart from the *Mycobacterium tuberculosis* complex living in water and soil (Ratnatunga *et al.*, 2020). Previously, these organisms were considered non-dangerous to humans with immune competence and affected only immunocompromised individuals. However, evidence is increasing of their importance in pulmonary disease in both immune-competent and immunocompromised individuals (Griffith *et al.*, 2007).

In addition, the emergence of co-infection is alarming, suggesting a link between *Mycobacterium tuberculosis* (MTB) and NTM in the pathogenesis of pulmonary disease. With increased complications of pulmonary disease, evidence is accumulating pointing to the overwhelming phenomenon of antimicrobial resistance among respiratory pathogens. Draft-to-complete genomes of various *Mycobacterium tuberculosis* complex (MTBC) strains are flooding the databases. This contributes to a better understanding of the genetics underlying virulence, host-pathogen interaction features, and antimicrobial resistance patterns among MTBC strains.

However, despite this globally evident achievement, whole genome sequencing technology in the sub-Saharan African Region is in its infancy, with most infectious and noninfectious bacteria being sequenced. Except for a few MTBC sequencing studies in East African settings (Ssenooba *et al.*, 2016; Kanyerezi & Nabisubi, 2020; Katale *et al.*, 2020; Mbelele *et al.*, 2022), evidence shows that even MTBC strains are yet to be sequenced at the genomic level from East and Central Africa. This calls for attention and

emphasis on the need for whole genome sequencing of MTB and NTM strains to track the possible epidemiological trend of pulmonary disease associated with *Mycobacterium* species. In addition, even the current research interest has been skewed towards the analysis of MTBC strains, whose infection, pathogenesis, and clinical implications are elucidated to a significant level.

On the contrary, little is explored about the genomics of NTM. The increase in multidrug-resistant strains of the MTB complex has been substantially established (Al-Mutairi *et al.*, 2019; Katale *et al.* 2020; Senghore *et al.*, 2020). Evidence of co-infection between MTBC and NTM is emerging, with speculations increasing about their role in the pathogenesis of TB as well as antimicrobial resistance (Kotwal *et al.*, 2017). In Tanzania, a recent combination of microscopy and molecular markers, entailing rRNA and hsp65 gene sequences, identified over 16 NTM strains, including *M. goodnae*, *M. interjectum*, and *M. intracellulare*. *M. kumamotoense/hiberniae*, and *M. flavescens/novocastrense*, among others (Hoza *et al.* 2016).

A case study conducted in Dar es Salaam, Tanzania, identified a strain *M. yongonense* (Mnyambwa *et al.* 2018) adding to the previously identified strains. However, these results were based on partial sequences, which are of low resolution, recovering only taxonomic identities to the level of genus. This work focused on establishing whole genomic information including potential antimicrobial resistance genotypes from high throughput sequencing via Oxford Nanopore technology. Therefore, the current study reports *Mycobacterium novocastrense*UDSM-BC02 and *Mycobacterium novocastrense*UDSM BC05 as NTM strains from Tanzania with genomic characteristics potentially representing multidrug resistance and TB pathogenic potential, accounting for symptoms of pulmonary tuberculosis in TB patients.

## Materials and Methods

### Sampling and sample preparation

In this project, Ethical approval was sought from the National Health Research Review Committee and granted ethical clearance number HQ/R.84/VOLII/853. Our study involved analysis of 24 mycobacterial samples archived at the central TB reference laboratory (CTRL) in Dar es Salaam which were collected from selected cross border regions of Tanzania. The samples used for this study were received from the East African Public Health Laboratory Network (EAPHLN) project satellite (using GeneXpert MTB/RIF assay (Cepheid, USA)) hospitals namely Kibong'oto (Kilimanjaro), Mnazimmoja (Unguja), Musoma referral hospital (Mara), St Benedict Ndanda referral hospital (Mtwara), Sumbawanga referral

hospital (Rukwa) and Kigoma referral hospital (Kigoma).

In addition, the study utilized samples from the EAPHLN project non-satellite (without Xpert MTB/RIF assay) sites of Nyamagana district hospital in Mwanza, Levolosi, and St Vicent health centers in Arusha and Pwani regions respectively. These selected areas for sample collection were from the cross-border regions except for St Vicent Health center. The inclusion criteria were all tuberculosis patients with positive smear consented to participate in the study. However, presumptive tuberculosis cases not willing to participate in the study were excluded as well as tuberculosis patients who were already on anti-tuberculosis treatment.

### DNA Extraction and sequencing

The genomic DNA was extracted from twenty four heat-killed *Mycobacteria* isolates by the cetyltrimethylammonium bromide (CTAB) method as previously described (van Soolingen *et al.*, 1991). Briefly, sterile loops were used to lift bacterial isolates from LJ medium slants. The isolates were added into tubes containing 1X TE (Tris-ethylenediaminetetraacetic acid) buffer, pH 8.0, and then heat-killed in a water bath at 80°C for 20

min followed by the addition of 10 mg/ml lysozyme in each tube and incubated at 37 °C overnight. On the following day, the DNA was extracted with chloroform-isoamyl alcohol (24:1), and the pellets of genomic DNA were rehydrated in 80 µl TE and left overnight at 4 °C. The quality and quantity of gDNA were confirmed using a Qubit 2.0 fluorometer (Thermal Fisher Scientific, Waltham, MA USA). The gDNA was eventually stored at -20 °C before sequencing.

### Library preparation and sequencing

Library construction and sequencing were done using the Oxford Nanopore ligation sequencing kit (LSK 109) without fragmentation according to manufacturer's instructions. Twenty-four samples were end-prepped by adenylating the 3' end and phosphorylating the 5' end, followed by attaching barcodes for multiple sample sequencing. The end-prepped DNA was washed with a long fragment buffer to select only long fragments of

more than 3kb. Sequencing adapters were added to these long fragments of DNA and incubated for 20 minutes at room temperature. Libraries were sequenced using MinION MK1C for 48 hours generating a total of 6.45 gb reads, which were base called using Guppy as the device was sequencing and the mean read length was 2.43kb, while *de novo* assembly was accomplished with Flye (v2.8).

### Genome annotation and establishment of basic features for each sample strain

To predict the possible taxonomic placement of each strain, assembled genomes for all the strains were first annotated with Microbial Genome Atlas (MiGA) (Rodriguez-R *et al.*, 2018) and then with the comprehensive genome analysis service at PATRIC (Wattam *et al.*, 2014). Annotation was also

accomplished with Rapid Annotation by Subsystems Technology (RAST).

### Taxonomic insight: Approach from Genomic Data and 16S rRNA genes to determine possible closest strain

To predict the possible closest strains, the 16S rRNA genes were extracted using the ContEst16S algorithm (Lee *et al.*, 2017) and used for BLASTn

against the NCBI nucleotide database. Alternatively, the entire genomes from MiGA were used to identify the closest strains based on automated average nucleotide identity (ANI). Thereafter, the closest relative genome sequences were downloaded from GenBank for detailed comparison. The genomes were analyzed by the

#### **Determination of closely related type strains**

Determination of closest type strain genomes was achieved in two complementary ways: First, all user genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness (Ondov *et al.*, 2016) and, the ten type strains with the smallest MASH distances chosen per user genome. Second, an additional set of ten closely related type strains was determined via the 16S rRNA gene sequences. These were extracted from the user genomes using RNAmmer (Lagesen *et al.*, 2007) and each sequence was subsequently subjected to BLAST analysis against the 16S rRNA gene sequence of each of the currently 15679 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (according to the bitscore) for each user genome and to subsequently calculate precise distances using the Genome BLAST Distance Phylogeny approach (GBDP) under the algorithm 'coverage' and distance formula d5 (Meier-Kolthoff *et al.*, 2013). These distances were finally used to determine the 10 closest type strain genomes for each of the user genomes.

## **Results**

### **Demographic and clinical features of the patients**

Among all 24 samples isolated as Mycobacteria, 2 (8.3%) samples were identified as NTM and labeled as BC02 and BC05. The patient with a code BC02 was a male aged 23 years living in Kilimanjaro with

Type (Strain) Genome Server (TYGS), for whole genome-based taxonomic identities (Meier-Kolthoff & Göker 2019), also integrating the latest updates for the most appropriate features (Meier-Kolthoff *et al.*, 2022). The TYGS analysis was subdivided into the following steps:

#### **Scaffolding and chromosome-level assembly to deduce full genomes.**

The closest relatives identified via genome PATRIC annotation and blind phylogenetic reconstructions from TYGS were selected as reference genomes. Then contigs from each strain genome were mapped to these selected reference genomes using CONTIGuator v2.27 (Galardini *et al.*, 2015). The scaffolds from the best matches were used as a reference for further remapping of the contigs to enhance recovery. This process of reference-guided assembly was repeated four times for each strain genome and the ultimately recovered sequence was considered a chromosome, relative to the chromosome sizes of reference and other closely related strains from genome and assembly databases in the National Center for Biotechnology Information (NCBI).

#### **Genomic screening of antimicrobial resistance genes**

Antimicrobial resistance genes analysis was done using the PATRIC annotation pipeline for the identification of antimicrobial resistance genes and mechanisms (Antonopoulos *et al.*, 2019). Then the genome sequences were alternatively scanned by the comprehensive antimicrobial resistance database (CARD) (Alcock *et al.*, 2020) and, using default parameters, the BacAnt: a combinatorial pipeline for concomitant identification of antimicrobial resistance genes with integrons and transposable elements (Hua *et al.*, 2021).

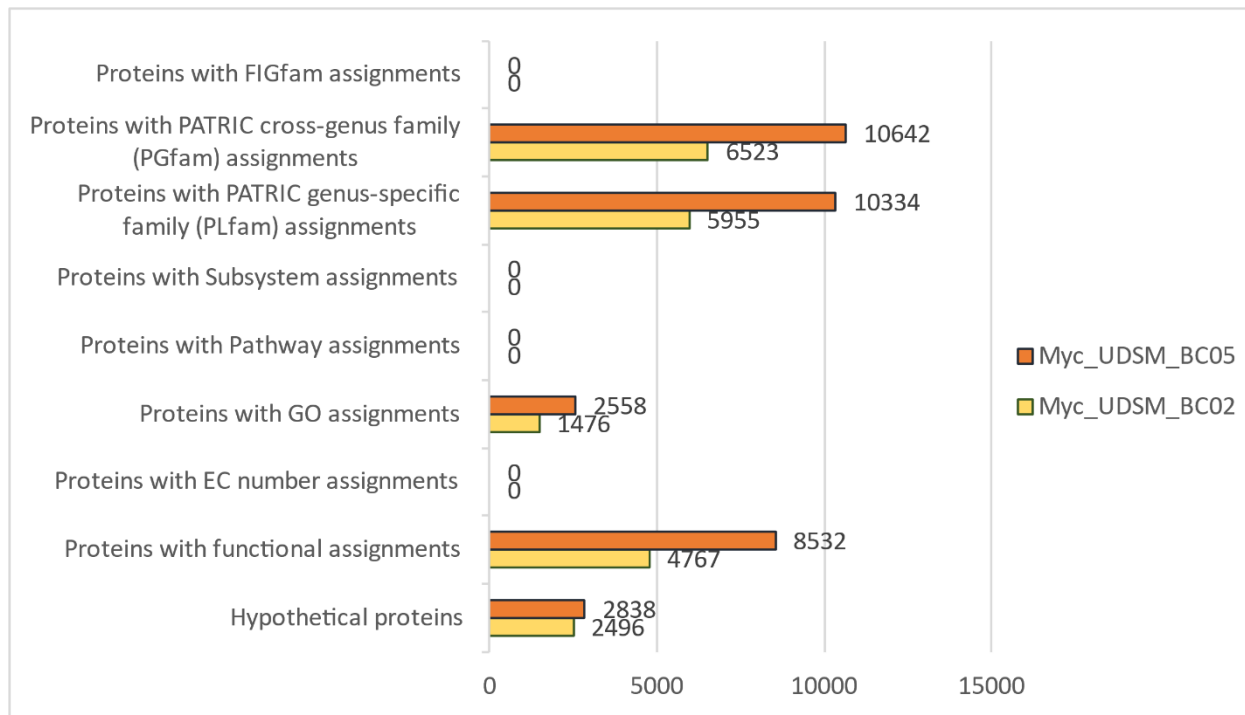
a history of TB relapse in the year 2015 and HIV negative. He was referred to a regional Hospital (Kibong'oto) after relapse with the following clinical features: coughing (60 days), fever (7 days), chest pain (7 days), fatigue (14 days), and loss of appetite (21 days). The second patient,

coded as BC05, was a male living in Arusha aged 32 years with the following clinical characteristics; HIV positive, with typical TB relapse symptoms including coughing for 58 days, fever (30 days), weight loss (53 days), breathing difficulties (30 days) and fatigue (30 days). He was previously treated with anti-Tb first-line drugs in the same facility.

### Assembly of NTM strains genomic features

The results of the assembled genome for Myc\_UDSM\_BC02 contained 108 contigs, a total of

5958338 bp (GC 66.72%), 7263 protein-coding sequences (CDS), 49 RNAs with 3 rRNA genes from PATRIC annotation. On the other hand, the Myc\_UDSM\_BC05 genome comprised 80 contigs, equivalent to 9197165 bp (GC 66.63%), 11,370 protein CDS, 90 transfer RNA (tRNA) genes, and 9 rRNA genes. Based on the annotation statistics and a comparison to other genomes in PATRIC within this same species, both genomes appeared to be of good quality. Details of the analysis, including predicted functions, and phylogenetic positions are shown (Figures 1, 2).



Figure

1. Functional characteristic features of coding sequences from the PATRIC annotation pipeline. Numbers at the end of each bar indicate the proteins that are functionally assigned according to corresponding functions and database.

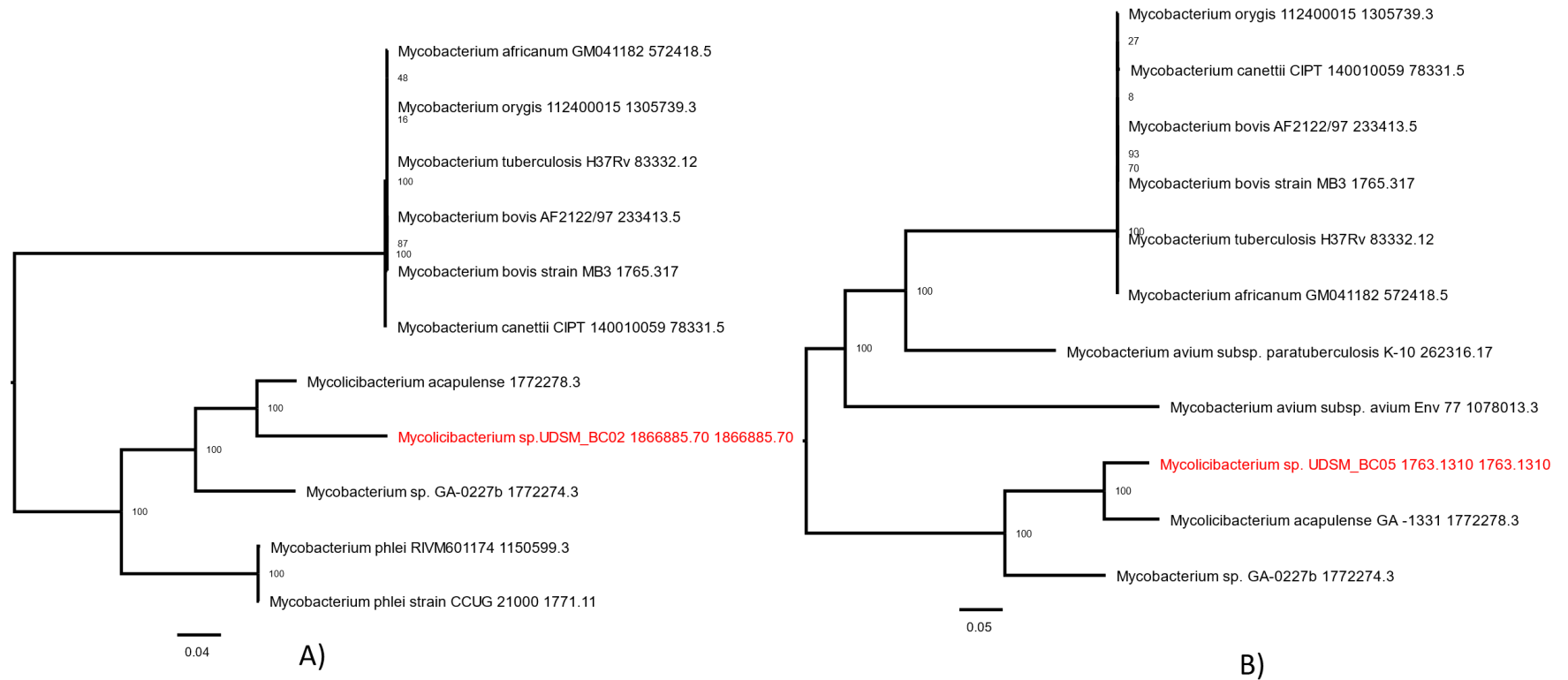
### Phylogenetic analysis of NTM strain genome from comprehensive annotation by PATRIC

Further annotation was done using PATRIC platform using reference genomes from the National Center of Biotechnology Information (NCBI) and includes them in the phylogenetic analysis as part of the comprehensive genome analysis process. The closest reference and representative genomes were identified by

Mash/MinHash (Ondov et al., 2016). PATRIC global protein families (PGFams) (Davis et al., 2016) were selected from these genomes to determine the phylogenetic placement of the study genomes. The protein sequences from the predicted families were aligned with MUSCLE (Edgar 2004), and the nucleotides for each of those sequences were mapped to the protein alignment. The joint set of amino acid and nucleotide alignments were

concatenated into a data matrix, and RaxML (Stamatakis 2014) was used to analyze this matrix, with fast bootstrapping (Stamatakis *et al.*, 2008) was used to generate the support values in the trees.

PATRIC global protein families (PGFams) (Davis *et al.* 2016) were selected from these genomes to determine the phylogenetic placement of the study genomes. The PATRIC platform retrieves high-quality and appropriate representative reference genomes from the National Center for Biotechnology Information (NCBI) and includes them in the phylogenetic analysis as part of the comprehensive genome analysis process. The closest reference and representative genomes were identified by Mash/MinHash (Ondov *et al.*, 2016). PATRIC global protein families (PGFams) (Davis *et al.*, 2016) were selected from these genomes to determine the phylogenetic placement of the study genomes. The protein sequences from the predicted families were aligned with MUSCLE (Edgar 2004), and the nucleotides for each of those sequences were mapped to the protein alignment. The joint set of amino acid and nucleotide alignments were concatenated into a data matrix, and RaxML (Stamatakis 2014) was used to analyze this matrix, with fast bootstrapping (Stamatakis *et al.*, 2008) was used to generate the support values in the trees. The initial prediction of the phylogenetic placement of each strain is indicated in Figure 2. Each of the strains forms a cluster with *Mycolicibacterium acapulense*, suggesting a close relationship between them. However, other members of the genus *Mycolicibacterium* are missing from this automated phylogenetic tree analysis. This shortfall accounts for the need to incorporate other phylogenomic approaches with the most recent and up-to-date databases.

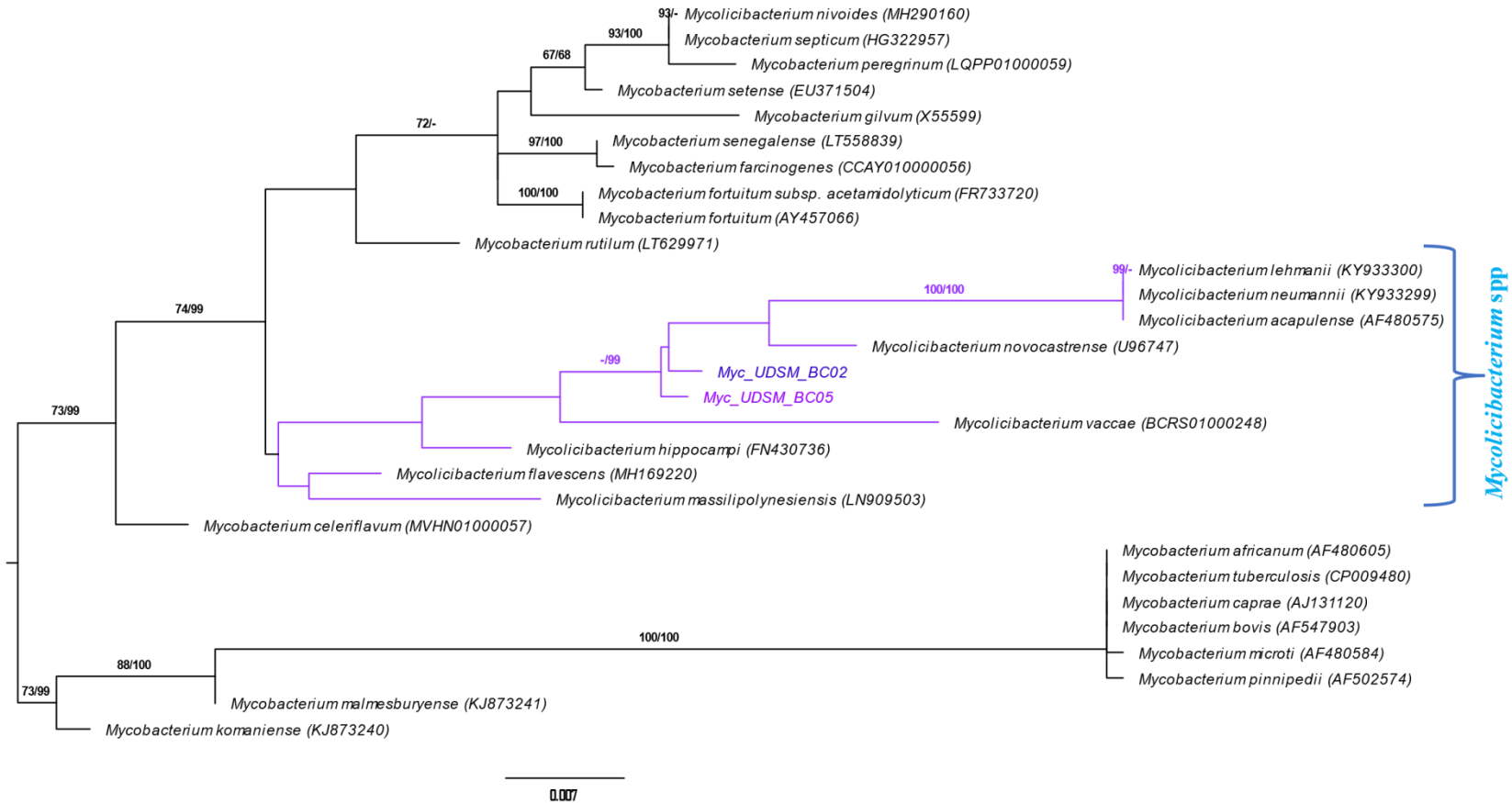


**Figure 2.** Codon phylogenetic trees generated from PATRIC annotation: A) the position of Myc\_UDSM\_BC02, B) the position of Myc\_UDSM\_BC05 based on PATRIC database search for close strains. Numbers above branches are bootstrap values for 1000 replicates automatically generated by RAxML.

### Phylogenetic position of BCo2 and BCo5 based on 16S rRNA gene sequences.

To reassess the quality of genome sequences before rRNA analysis, screening for contigs contamination was approached using ContEst16S (Lee *et al.*, 2017). The genomes were both found to be clean, i.e., devoid of intra-genus contaminant sequences. The BCo2 genome contained one 16S rRNA gene, which upon BLASTn against the NCBI database showed a percentage identity of 98.43 – 99.19 % with *Mycolicibacterium* spp exclusively *M. novocastrense* (GenBank accession no. **HM807280.1**) and *M. flavescens* (GenBank accession no. **AF174289.1**). On the other hand, the BCo5 strain contained three 16S rRNA genes. The BLASTn algorithm in ContEst16S predicted *Mycobacterium/LQIX* as the closest strain. Based on 16S rRNA phylogeny (Figure 3) annotation results revealed that the strains of the two isolates BCo2 and BCo5 were closer to each other than to the reference genomes. The results of phylogenetic position analysis of each isolate inferred from genome-based distance phylogeny (GBDP) calculated from 16S rRNA revealed that the two isolates lies within the genus *Mycolicibacterium* (Figure 4). Their possible difference is suggestive of further genomic differences and potential virulence and antibiotic resistance features, described in the next parts of this work.

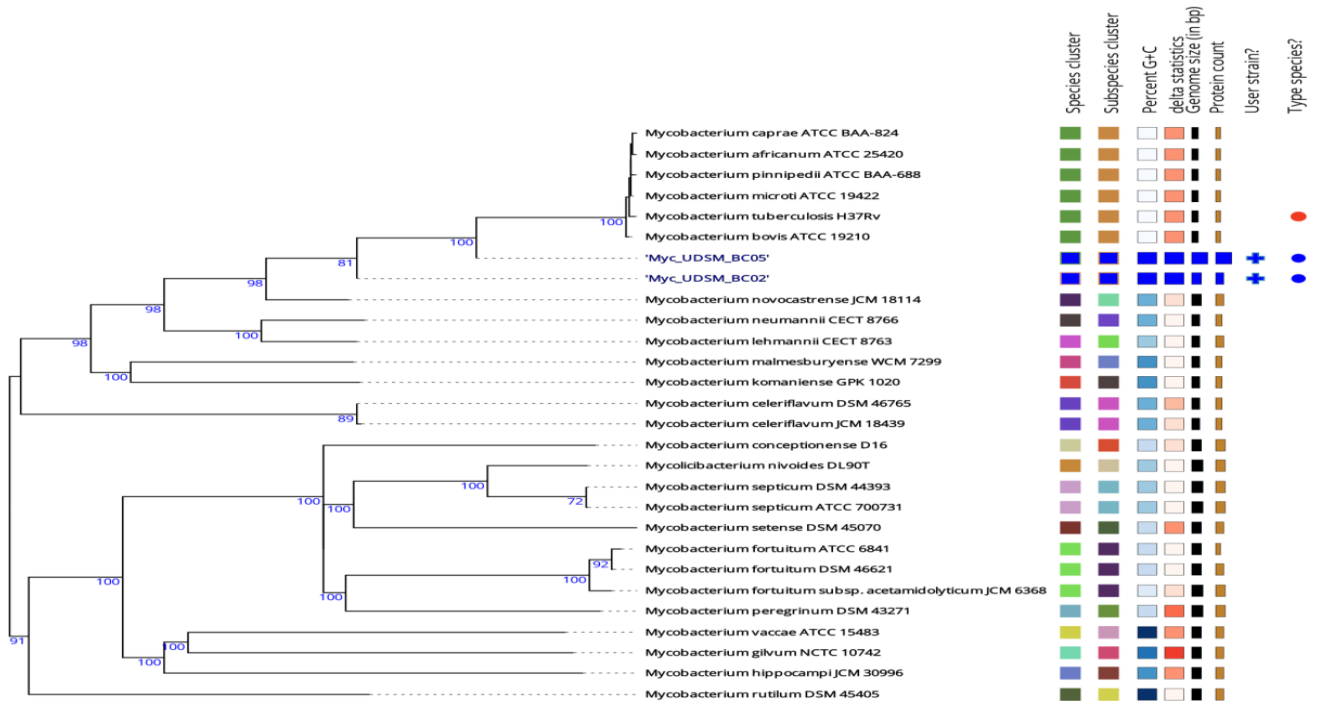




**Figure 3.** Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from 16S rRNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula  $d_5$ . The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 52.6 %. The tree was rooted at the midpoint.

### **Whole genome-based phylogenetics and taxonomic placement of BC02 and BC BC05 strains**

The two strains were further subjected to other approaches to be able to discriminate the two strains. Based on whole genome phylogenetic and taxonomic placement, the two strains BC02 and BC BC05 were first subjected to the TYGS database for fast classification (Meier-Kolthoff *et al.*, 2022). The TYGS database can retrieve the most closely and distant related strains based on several computations integrating ANI and genome-genome distance calculation (GGDC). The results of this annotation as presented in Figure 4, revealed that the two study strains form a separate clade from MTBC such as *M. tuberculosis* H37Rv, *M. africanum* ATTC 25420, and *M. caprae* ATTCC BAA-824 or *M. microti* ATTCC 19422, which are well-established strains of the MTBC.

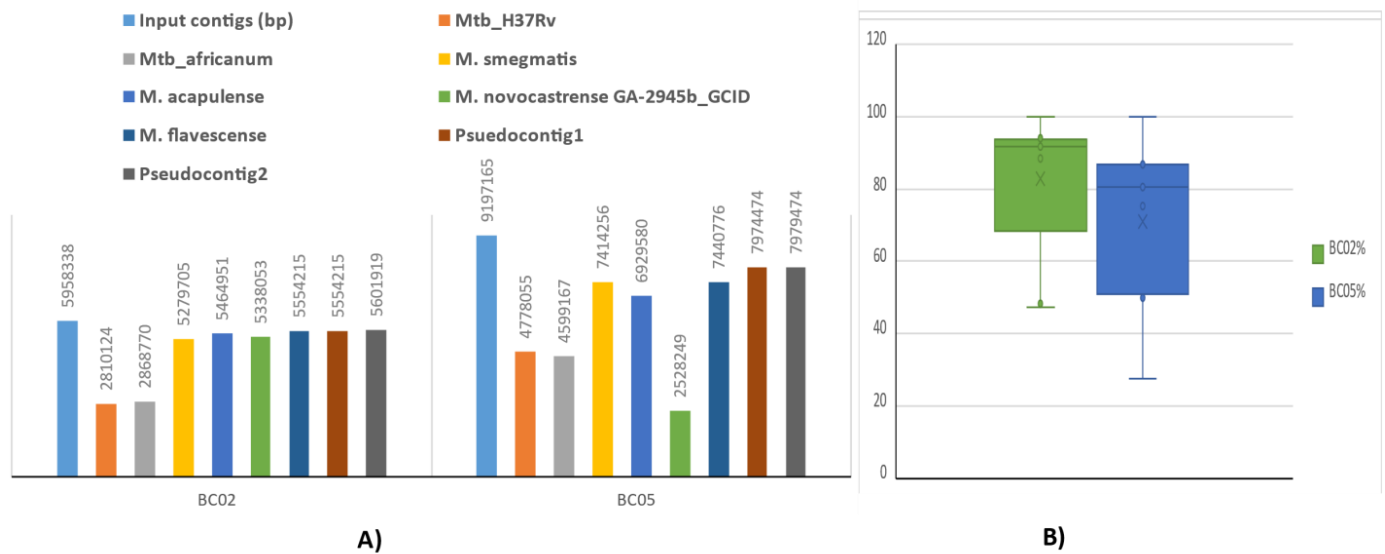


**Figure 4.** Genome-based phylogenetic tree inferred with FastME 2.1.6.1 from GBDP distances Calculated from genome sequences. The branch lengths are scaled in terms of the GBDP distance formula  $d_5$ . The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 79.7 %. The tree was rooted at the midpoint.

### Assembled chromosomes.

It is usually ideal to decrease the number of contigs as much as possible to generate a chromosomal genomic DNA and enhance the resolution of extrachromosomal sequences if any. Following the identification of phylogenetically close strains, the latter genome sequences were downloaded from the NCBI database and used as references against which the contigs for the study

genomes were mapped using CONTIGuator 2.27 (Galardini *et al.*, 2015). The scaffold size and corresponding percentage recovery generated from each mapping are indicated in Figure 5. The final recovered genome for the strain BC02 can be accessed via accession no. **CP097264** while that of BC05 is under the NCBI processing but can be shared upon request..



**Figure 5.** Chromosome recovery profiles obtained from reference-based assembly via CONTIGuator scaffolding pipeline. A) Genome size for each reference mapping, numbers above bars indicate the recovered chromosome against the corresponding reference genomes. B) Mapping percentage corresponding to each recovered genome.

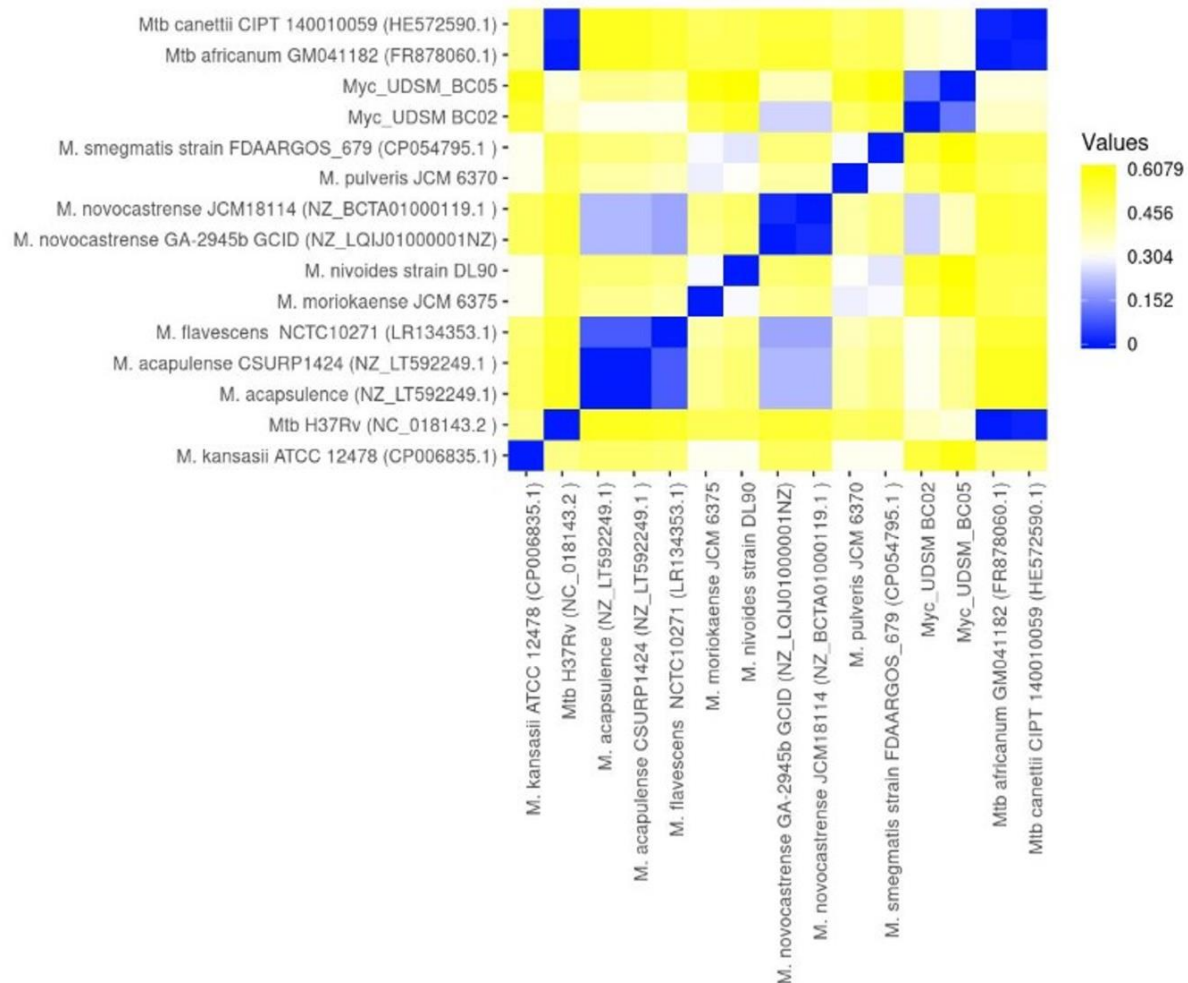
### Analysis based on Average nucleotide identity of BC02 and BC05

To enhance the power of inference, the genomes were compared with other NTM in the genus *Mycobacterium* as well as standard strains from the MTBC complex based on their average nucleotide identity. Average nucleotide identity

(ANI) was calculated using the enve-omics algorithm (Kim *et al.*, 2014). The values of ANI analysis are presented in Tables 1. From these results a distance matrix was generated and the corresponding comparative heatmap is shown in Figure 6.

Table 1. Average nucleotide identities of each strain compared to reference genome sequences

Reference sequence	BCo2		BCo5	
	ANI	SD	ANI	SD
<i>Mtb. kansasii</i> ATCC 12478 (CP006835.1)	78.38	4.45	80.17	4.12
<i>Mtb</i> H37Rv (NC_018143.2)	88.34	11.09	88.35	11.04
<i>M. acapsulence</i> (NZ_LT592249.1)	90.02	3.61	90.03	3.72
<i>M. acapulense</i> CSURP1424 (NZ_LT592249.1)	90.02	3.61	90.03	3.72
<i>M. flavescens</i> NCTC10271 (LR134353.1)	89.89	3.38	89.91	3.48
<i>M. moriokaense</i> JCM 6375	80.06	4.26	79.88	4.35
<i>M. nivoides</i> strain DL90	78.72	4.49	78.54	4.53
<i>M. novocastrense</i> GA-2945b GCID (NZ_LQIJ01000001NZ)	95.51	3.19	95.56	3.15
<i>M. novocastrense</i> JCM18114 (NZ_BCTA01000119.1)	95.63	3.36	95.67	3.38
<i>M. pulveris</i> JCM 6370	80.94	4.42	80.87	4.45
<i>M. smegmatis</i> FDAARGOS 679 (CP054795.1)	79.19	4.61	79.08	4.70
<i>Myc</i> _UDSM BCo2	100.00	0.00	99.81	0.66
<i>Myc</i> _UDSM BCo5	99.81	0.66	100.00	0.00



**Figure 6.** A heatmap representing pairwise comparison of BC02 and BC05 strains against NTM and standard MTBC strains. The correlation was inferred using Euclidean distance measurement method.

### Results from RAST annotation

To enhance precision in the description of these two genomes the assembled chromosomes were reannotated with the RAST server (Aziz *et al.*, 2008) using the RASTk pipeline (Brettin *et al.*, 2015). Default minimum gene length was used, with the minimum identity being adjusted to 70%. From this annotation, the BC02 genome (with 5,624,151 bp in size and 66.9 GC %) was found to contain 6657 CDS with 50 RNA genes and 301 subsystems. These genomic features are coherent with most members of the genus *Mycolicibacterium* reported from recent genomic studies (Sánchez *et al.*, 2019; Vatlin *et al.*, 2019).

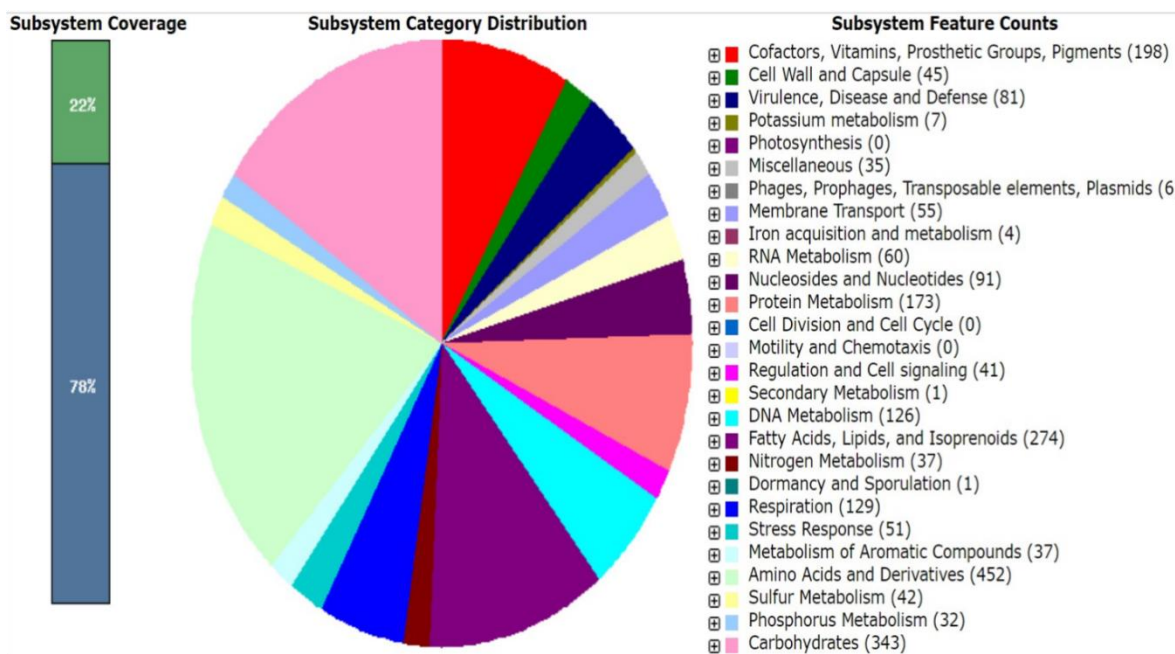
Out of the 301 subsystems (Figure 7), the richest in features is that for amino acid metabolism followed by those for carbohydrate metabolism and then fatty acids, lipids, and

isoprenoids. The most remarkable feature of *Mycolicibacterium* spp is their genomic richness in the genes for biosynthesis of membrane lipids, which have been implicated in fatty acid accumulation and biofilm formation, among other adaptive mechanisms to harsh conditions (Chen *et al.*, 2020). On the other hand, the BC05 genome 7,979,474 bp and GC 66.5%) contained 9590 CDS and 90 rRNAs with 833 subsystems with various functional categorizations indicated in Figure 8. These annotation features are consistent in terms of values with genomic size and are more coherent with *Mycolicibacterium* spp genomes from other studies (Sánchez *et al.*, 2019) than *Mycobacterium* spp (Advani *et al.*, 2019). Detailed analysis of virulence factors showed that both BC02 and BC05 strains possess genes for invasion and intracellular resistance, i.e., 51 gene features for BC02 and 519

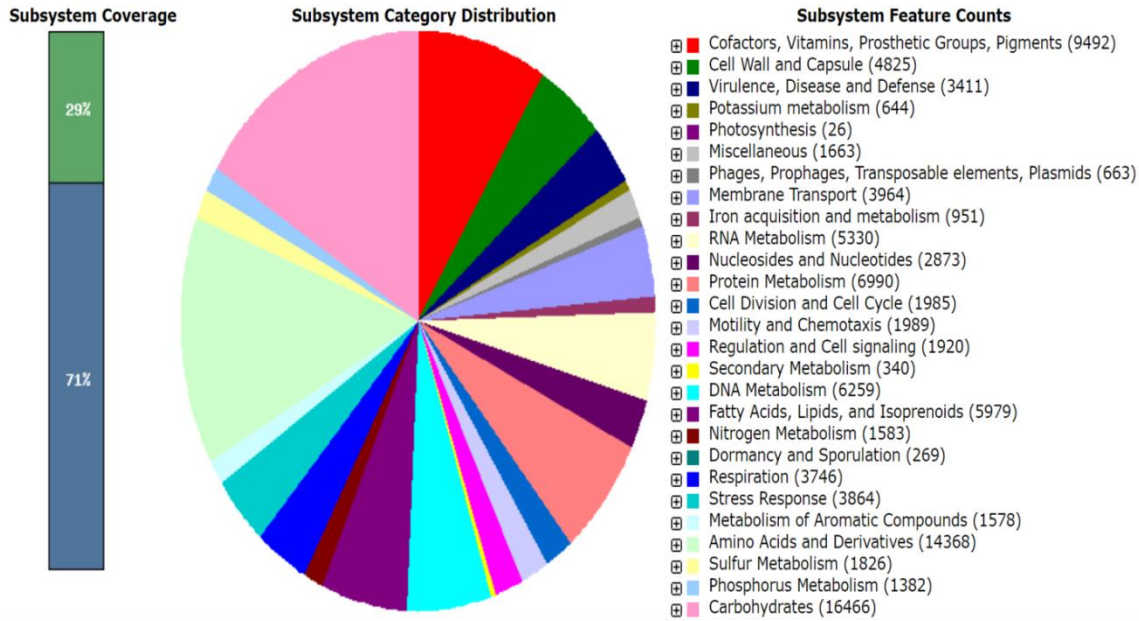
for BC05, suggesting that the strains can potentially infect and cause disease.

The BC02 genome also possesses resistance to fluoroquinolones (2 genes). All these features are indicative of their potential virulence and antibiotic resistance. In the BC05 genome, other antibiotic resistance features identified via RAST annotation include resistance to vancomycin (9 genes), tetracycline resistance, ribosome protection type (45 genes), aminoglycoside adenyl-transferases (2 genes), and tetracycline resistance, ribosome protection type, too (45

In addition, while BC02 possesses a single, potentially class A beta-lactamase gene, BC05 contains up to 105 beta-lactamase genes. genes), among others. This pattern of genotypic antimicrobial resistance is a strong indicator of multidrug resistance, characterized in the vast majority of Mycobacteria including the *Mycobacterium tuberculosis* H37Rv, *M. tuberculosis* var *africanum*, *M. tuberculosis* var *kansasii* (Katale et al., 2020; Al-Mutairi et al., 2019), and other MTBC species (Joean et al., 2020; Parthasarathy et al., 2016).



**Figure 7.** Metabolic subsystems recovered from RASTk annotation of the BC02 chromosome. The bar in the left represents the percentage of each subsystem coverage and the legend on the right indicates the counts for each subsystem feature.



**Figure 8.** Metabolic subsystems recovered from RASTk annotation of the BC05 chromosome. The bar in the left represents the percentage of each subsystem coverage and the legend on the right indicates the counts for each subsystem feature.

### Orthologous clusters

From Orthovenn2 analysis, BC02 and BC05 together with the most closely related NTM species and reference MTBC strains form 7769 clusters, 6559 orthologous clusters (at least contains two species) and 1210 single-copy gene clusters (Table 2). This indicates that most of the gene families and clusters are shared among *Mycolicibacterium* spp as contrasted to the

*Mycobacterium* representative strains. The number of singletons (genes not conforming to a cluster) is consistently dependent upon the genome size and the number of inherent CDS. It appears that, the BC02 genome (1316 singletons) has fewer singletons than the BC05. Genome (2646 singletons) Likewise, the genome of *Mycolicibacterium smegmatis* LN831039 (6847 CDS) contains 1928 singletons.

**Table 2.** Clusters of orthologous recovered from Orthovenn2 analysis and annotation. The proteome of each genome is presented against the clusters and corresponding singletons.

Species	Proteins	Clusters	Singletons
Myc_UDSM_BC02	6657	5108	1316
Myc_UDSM_BC05	9590	5950	2646
Myc_novocastrense_JCM_BCTA01000119	5336	4794	411
Myc_flavescens_NTC10271_LR134353	5698	4762	674
Myc_smegmatis_LN831039	6847	4427	1928
Mtb_africanum_GM041182	4269	4070	129
Mtb_H37Rv_NC018143	4302	4092	124

Shown in Figure 9 the members of *Mycolicibacterium* share 1537 protein families. While protein families shared between BC02 and *M. novocastrense* could not be resolved, 35 families are shared between BC05 and *M. novocastrense*.



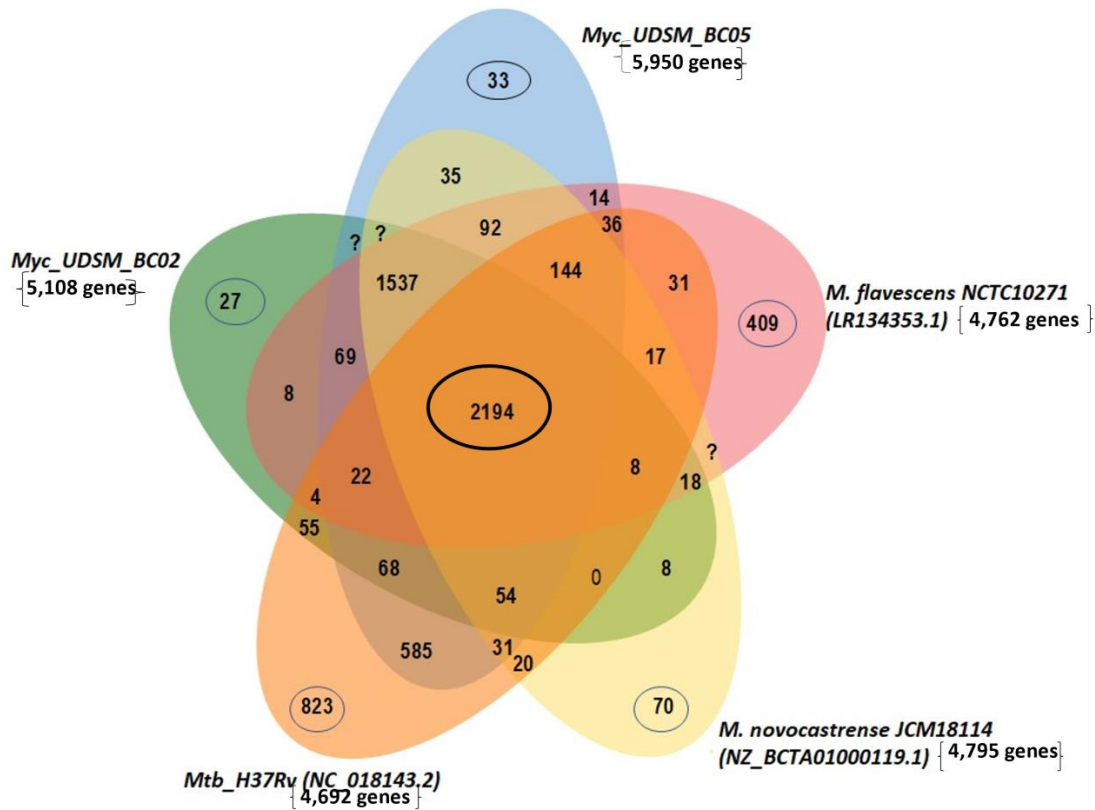


Figure 9. A Venn representation of the distribution of unique, group-specific, and core gene families among the Mycolicibacterium spp as compared with *M. tuberculosis* H37Rv. The selected closest scaffolds with at least a single one-to-one ortholog shared among the genomes were compared using

OrthoVenn2. The core genome is shown in the central circle. Each colored intersect segment represents the number of gene families shared among the respective overlapping genomes, and the outermost circled numbers represent unique gene families for individual genomes.

### Analysis of mutations associated with antimicrobial resistance genotypes in BC02 and BC05 strains of NTM to anti-TB drugs.

PATRIC database offers an opportunity to analyze genes related to antimicrobial resistance (specialty genes), including a count of virulence factors, drug targets, and antimicrobial resistance genes. As indicated in Table 3, again specialty genes for the BC05 are more than two-fold those of BC02, possibly due to larger genome size, also attributable to repetitive sequences in the BC05 genome. To better define the antimicrobial resistance potential of each strain, the FASTA files were first analyzed with BacAnt v3.3.3, which provides a database (BacAnt-database v2.0) to scan for resistance genes, insertion elements, and transposon regions from the genome sequences (Hua et al., 2021). Results from this scanning

revealed neither insertion elements nor transposons but a class A beta-lactamase (BlaA) gene flanked between 2605222 and 2605896 of the BC02 chromosomal DNA sequence, which covers 73.26% with 99.56% identity to the reference gene in the resistance database (resDB).

In the case of the BC05 chromosome, the antimicrobial resistance genotype could not be resolved with BacAnt. Thus alternative reannotation was approached using the comprehensive antimicrobial resistance database, (CARD) retaining default parameters (Alcock et al., 2020). Based on CARD analysis, nucleotide sequences undergo ORF calling to generate predicted protein sequences (Alcock et al., 2020). From this analysis (Table 3), the BC02 genome was found with three strict hits, relevant to genes conferring resistance against macrolides

and rifamycin as defined by efflux, drug inactivation, and target alteration/replacement respectively. On the other hand, results show that the strain BC05 contains one perfect and 10 strict hits for a wide range of antibiotic resistance genes (Table 5). The perfect hit presents efflux-mediated antibiotic resistance against first-line anti-TB drugs, rifamycin, and isoniazid. The rest of the hits

include efflux, and target alteration against rifamycin, macrolide, penam, disinfectants, polyamine, and fosfomycin antibiotics. These resistance genotypes in the BC05 are strongly suggestive of the emergence of a multidrug-resistant NTM strain associated with MTBC coinfection, which underlies exacerbated pulmonary tuberculosis pathogenesis.

**Table 3. Antimicrobial resistance genes for BC02 and BC05 strains as recovered from the PATRIC annotation pipeline. The annotation entails virulence factors, transporter genes and drug targets from different databases accessible to the PATRIC database.**

Specialty Genes	Source	Myc_UDSM_BC02	Myc_UDSM_BC05
Virulence Factor	PATRIC_VF	171	548
Virulence Factor	Victors	113	315
Virulence Factor	VFDB	31	76
Transporter	TCDB	51	127
Drug Target	DrugBank	38	82
Drug Target	TTD	14	35

**Table 4. Myc\_UDSM\_BCo2 antibiotic drug resistance genotypes annotated via the CARD pipelines. Models for detection criteria, antimicrobial resistance gene families, drug class and drug resistance mechanisms are included alongside percentage identity match and sequence coverage by default parameters.**

ARO term	SNP	Detection criteria	AMR gene family	Drug class	Resistance mechanism	% ID of matching region	% length of resistance sequence
mtrA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	Macrolide, penam	efflux	96.05	100
arr-1		protein homolog model	rifampin ADP-ribosyltransferase (Arr)	rifamycin	inactivation	84.51	99.3
MtbrpoB mutants conferring resistance to rifampicin	D516G, H526T, L511R	Protein variant model	rifamycin-resistant beta-subunit of RNA polymerase (rpoB)	rifamycin	Target alteration/replacement	100	52.39

**Table 5. Myc\_UDSM\_BC05 antibiotic drug resistance genotypes annotated via the CARD pipelines. Models for detection criteria, antimicrobial resistance gene families, drug class and drug resistance mechanisms are included alongside percentage identity match and sequence coverage by default parameters.**

RGI Criteria	ARO Term	SNP	Detection Criteria	AMR Gene Family	Drug Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
Perfect	efpA		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	rifamycin antibiotic, isoniazid	antibiotic efflux	100	100
Strict	qacG		protein homolog model	small multidrug resistance (SMR) antibiotic efflux pump	disinfecting agents and antiseptics	antibiotic efflux	39.81	106.54
Strict	RbpA		protein homolog model	RbpA bacterial RNA polymerase-binding protein	rifamycin antibiotic	antibiotic target protection	93.69	97.37
Strict	mtrA		protein homolog model	resistance-nodulation-cell division (RND) antibiotic efflux pump	macrolide antibiotic, penam	antibiotic efflux	96.05	100
Strict	arr-1		protein homolog model	rifampin ADP-ribosyltransferase (Arr)	rifamycin antibiotic	antibiotic inactivation	84.51	99.3
Strict	Mycobacterium tuberculosis embB mutant conferring resistance to ethambutol	E378A	protein variant model	ethambutol resistant embB	polyamine antibiotic	antibiotic target alteration	99.8	48

Strict	Mycobacterium tuberculosis embA mutant conferring resistance to ethambutol	P913S	protein variant model	ethambutol resistant embA	polyamine antibiotic	antibiotic target alteration	99.73	68.74
Strict	Mycobacterium tuberculosis embC mutant conferring resistance to ethambutol	T270I	protein variant model	ethambutol resistant embC	polyamine antibiotic	antibiotic target alteration	99.71	33
Strict	Mycobacterium tuberculosis intrinsic murA conferring resistance to fosfomycin	C117D	protein variant model	antibiotic-resistant murAtransferase	fosfomycin	antibiotic target alteration	100	34.69
Strict	Mycobacterium tuberculosis rpoB mutants conferring resistance to rifampicin	D516G, H526T, L511R	protein variant model	rifamycin-resistant beta-subunit of RNA polymerase (rpoB)	of rifamycin antibiotic	antibiotic target alteration, antibiotic target replacement	99.78	45.14
Strict	Mycobacterium tuberculosis 23S rRNA mutation conferring resistance to capreomycin	A2145G, A2045G	rRNA gene variant model	23s rRNA with mutation conferring resistance to aminoglycoside antibiotics	aminoglycoside antibiotic	antibiotic target alteration	96.28	100

## Discussion

From this study it has been observed that the genome size of BC02 is significantly smaller by approximately 3.2 Mb than that of BC05. This could generally account for a small number of CDS, therefore all the protein assignments in Figure 1 correspond with the smaller BC02 genome relative to that of BC05. The genome sizes of NTM strains are becoming known to be relatively larger compared to those of MTBC (Yoon et al. 2020). Thus, the numbers of CDS are considerably larger. This work also supports that observation, with the BC05 having a huge size and proteome.

In this study, phylogenetic analysis was achieved through PATRIC. The initial prediction of the phylogenetic placement of each strain as indicated in Figure 2 shows that each of the strains forms a cluster with *Mycolicibacterium acapulense*, suggesting a close relationship between them.

The results of the initial phylogenetic placement of the two strains (Figure 2) shows that they form cluster with *Mycobacterium acapulense* suggesting their close relationship. However, other members of the genus *Mycolicibacterium* are missing from this automated phylogenetic tree analysis. This suggests that the PATRIC phylogenetic analysis algorithms could not successfully manage to retrieve some of the most relevant reference genomes. It could also imply that the database is yet to be updated to include a wide range of the most recent genomes of the genus *Mycolicibacterium*. To overcome this shortfall, we had to incorporate other phylogenomic approaches with the most recent and up-to-date databases for proper placement.

A detailed search through the NCBI assembly database revealed that

*Mycobacterium/LQIX* correlates to the former *Mycobacterium* sp. GA-1199 (GenBank assembly accession no. **GCA\_001500045.1**), whose current best match is *Mycolicibacterium novocastrense*, with average nucleotide identity of 95.18%, query coverage of 86.65% and subject coverage of 75.67%. These findings point to the NTM, potentially *M. novocastrense* as the best match for the BC05 genome. Presented in Figure 4, the phylogenetic position of each from GBDP distances calculated from 16S rRNA lies within the genus *Mycolicibacterium*. It then follows that both Myc\_BC02 and Myc\_BC05 comprise part of the NTM group and are potentially suggestive of co-infection or super-infection with MTBC. Based on the 16S rRNA phylogeny (Figure 3), the two strains are closer to each other than to the reference strain. However, the two strains also form potentially separate clades, indicating a possible difference between them, basally rooted from *Mycolicibacterium vaccae*. Their possible difference is suggestive of further genomic differences and potential virulence and antibiotic resistance features.

The pan-genome comparative analysis provides the highest throughput and the most reliable results in genomics. In the case of phylogenetics, whole-genome-based comparison has demonstrated the potential to discriminate intraspecific differences based on multiple markers to establish various positions and lineages within and between groups (Coscolla et al. 2021; Vázquez-Chacón et al. 2021; Stephen Kanyerezi and Patricia Nabisubi 2020). The TYGS database can retrieve the most closely and distant related strains based on several computations integrating ANI and genome-genome distance calculation (GGDC). We observed that the two study strains formed a separate

clade from MTBC such as *M. tuberculosis* H37Rv, *M. africanum* ATTC 25420, and *M. caprae* ATCC BAA-824 or *M. microti* ATCC 19422, which are well-established strains of the MTBC.

The results of the phylogenies reconstructed from both the 16S rRNA gene (Figure 3) and genome-wise comparison (Figure 4) reiterate the prior observation from Figure 2 that the two strains BCo2 and BCo5 are potential members of the NTM group. A previous study based on 16S rRNA and hsp56 gene sequences from Muheza Designated District Hospital, Tanzania revealed the presence of *M. kumamotoense*, *M. scrofulaceum*/*M. avium*, *M. acapulcensis/flavescens*, *M. avium*, and *M. flavescens/novocastrense*, among others (Hoza et al. 2016). Also, another case study from Dar es Salaam exposed the presence of *M. yongonense* (Mnyambwa et al. 2018). This work confirms through genomics that *M. novocastrense/flavescens* is among the most common clinical NTM species in Tanzania, especially from Northern regions of Tanzania. However, the insufficiency of genomic data in Tanzania and East Africa hampers pan-genomic comparison within the region to establish appropriate possible linkages for tracing the potential transboundary distribution of the NTM strains.

Observations from ANI values results showed that the two strains appear closer to each other than between each and the reference strains. This is strong evidence that the strains belong to closely related taxonomic groups. The accepted species delimitation for most bacterial samples ranges from 95 – 96% (Jain et al. 2018). From the above results (Table 1) the strains BCo2 and BCo5 meet the criteria, i.e., ANI = 99.8%, genomic distance of 0.01 and the probability that the two are in the same species is 96.41%. In addition,

the strains match with *M. novocastrense* by 95.5% ANI and about 64% with *M. flavescens*.

Analysis of the mutation associated with antimicrobial resistance genotypes in BCo2 and BCo5 strains was performed on anti-TB drugs. The evolution of multidrug-resistant MTBC strains is vastly growing to integrate virtually all the established mechanisms of resistance against anti-TB drugs (Senghore et al., 2020). According to literature the most established beta-lactamase gene in *Mycobacterium* spp (particularly *M. tuberculosis* strain ATCC 25177 / H37Ra) is the class C beta-lactamase (BLaC) (Bhattacharya et al., 2021), which in the UniProt database (<https://www.uniprot.org/uniprot/A5U493>), is described as a representative of extended beta-lactamase antibiotic resistance and is known to portray extensive penicillinase, cephalosporinase, as well as carbapenamase activities. However, the BlaA gene is elusive as a beta-lactamase, considerably characterized in *Mtb var bovis*, *Mtb var canettii*, *Mtb var orygis*, and *Mycobacterium lactis* (Bhattacharya et al., 2021).

Therefore *in vitro* and *in vivo* studies are instrumental in the characterization of the antibiotic resistance mechanism for the BCo2 strain. With different coverages and percentage identities ranging from 36 – 98%, the BCo2 strain possesses the potential to resist a wide of other drugs including fosfomycin, macrolides, vancomycin, tetracycline, and gentamycin/tobramycin, among others. These forms of resistance are also common among the notorious MTBC strains from various parts (Kidenya et al., 2018; Joean et al., 2020; Katala et al., 2020). This concurrence in the genotypic multidrug resistance is a strong indication for the co-existence of

coinfections and co-pathogenesis between the less explored NTM strains and the commonly known MTBC strains such as the *Mtb* H37Rv and *Mtb var africanum*, among others reported from recent studies (Ishiekwene *et al.*, 2017; Stepanyan *et al.*, 2019).

Observation from this study attracts huge attention to the role of NTM co-pathogens as auxiliary agents in the drug resistance process. Here we demonstrate that these resistance genotypes of BCo2 and BCo5 isolates are comparative to those of MTBC strains (Parthasarathy *et al.*, 2016; Katale *et al.*, 2020). In terms of drug resistance, the BCo2 strain presents a narrow resistome spectrum, while BCo5 demonstrates a potentially broad spectrum resistome, encompassing rifamycin, isoniazid, macrolides, polyamines, as well as fosfomycin, which correlates with resistance patterns among members of the MTBC such as *M. tuberculosis* H37Rv, *M. tuberculosis africanum*, *M. tuberculosis var mungii*, *M. tuberculosis var bovis*, and *M. tuberculosis var kansasii* (Vázquez-Chacón *et al.*, 2021; Al-Mutairi *et al.*, 2019; Parthasarathy *et al.*, 2016). Until 2018, the first macrolide-binding protein was yet to be elucidated (Zhang *et al.*, 2018). The current findings underscore the possible coexistence of MTB and NTM strains with the potential for macrolide resistance.

In addition, rifamycin resistance is profoundly demonstrated in both BCo2 and BCo5 by mechanisms involving both adenosine diphosphate (ADP)-ribosyltransferase activity as well as single nucleotide polymorphism (SNP), leading to rifamycin-resistant beta-subunit of RNA polymerase (*rpoB*). These mechanisms are evident even in the extensively drug-resistant MTB *M. tuberculosis* H37Rv (Joean *et al.*, 2020). Our findings therefore suggest that molecular diagnostic techniques

integrating the principle of DNA hybridization such as those in the GeneXpert or GenoType MTBDRplus (Dorman *et al.*, 2012) kits are not specific to MTBC strains, rather they could potentially detect the BCo2 and BCo5 strains in the TB samples as the inherent assays focus on the presence of SNPs associated with rifampicin and/or isoniazid, among other drug resistances.

Although WGS provides rapid and comprehensive diagnosis of Mycobacterium organisms and their resistance mechanisms and WHO has acknowledged its potential in determining resistance TB there are challenges that has to be addressed before its implementation in low-income countries. Those challenges include high costs of equipment, training of technical staff, and expertise guidance in clinical interpretations of WGS results. The major limitation to this study was lack of funds which could enable inclusion of large number of samples for WGS but also unavailability of databanks of African origin for Mycobacteria strains.

### Conclusion

Long-read sequencing supports the recovery of nearly full genomes of strains. Comparative genomic analysis shows that the BCo2 and BCo5 strains in this study are NTM and belong to *Mycolicibacterium*. Since the samples were obtained from symptomatic relapse TB patients, this study concludes that the strains have the potential to cause pulmonary tuberculosis and probably resistant TB. While the two strains are closest to each other, they conform to *Mycolicibacterium novocastrense* as the closest relative.

With the current screening for antibiotic resistance repertoire, this study infers the existence of multidrug-resistant NTM strains whose resistance encompasses a wide range of drug classes including beta-lactam,



macrolides, polyamines, aminoglycosides, and rifamycins. Precisely, the BCo2 strain is genetically resistant to penicillins and carbapenems as well as rifamycin drugs, while the BCo5 strain is genetically resistant to most first-line anti-TB drugs including rifamycin, isoniazid, and ethambutol, among others. Phenotypic studies coupling standard antimicrobial assays with transcriptomics, proteomics, or other omics-based methods would further our understanding of the exact mechanisms of drug resistance portrayed by each strain towards effective management of pulmonary tuberculosis in Tanzania and the sub-Saharan Region in general.

#### Acknowledgements

The authors acknowledge the laboratory staff at the Central Tuberculosis Reference Laboratory Muhimbili National Hospital for sample storage and laboratory workbench for DNA extraction and processing. Department of Molecular Biology and Biotechnology Dar es Salaam University and Special thanks to Mr. Charles Kayuki for his unlimited assistance in the Nanopore sequencing work.

**Funding:** The major project was funded by the World Bank. This work was partly funded by the University of Dar es Salaam and the Swedish International Cooperation Development Agency (SIDA) via the UDSM-IMB project no. 2015-2020/2022.

#### Competing interests

The authors declare that they have no competing interest related to this article.

#### References

Ratnatunga CN, Lutzky VP, Kupz A, Doolan DL, Reid DW, Field M, Bell SC, Thomson RM and Miles JJ (2020) The Rise of Non-Tuberculosis Mycobacterial Lung

Disease. *Frontiers in Immunology* 11: 303. <https://doi.org/10.3389/fimmu.2020.00303>.

Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, Holland SM, Horsburgh R, Huitt G, Iademaro MF, Iseman M, Olivier K, Ruoss S, von Reyn CF, Wallace RJ Jr, Winthrop K; ATS Mycobacterial Diseases Subcommittee; American Thoracic Society; Infectious Disease Society of America (2007) An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med.* 2007 Feb 15;175(4):367-416. <https://doi.org/10.1164/rccm.200604-571ST>. Erratum in: *Am J Respir Crit Care Med.* 2007 Apr 1;175(7):744-5. Dosage error in article text. PMID: 17277290.

Ssenooba W, Meehan CJ, Lukoye D, Kasule GW, Musisi K, Joloba ML, Cobelens FG and de Jong BC (2016) Whole genome sequencing to complement tuberculosis drug resistance surveys in Uganda. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases.* 40: 8–16. <https://doi.org/10.1016/j.meegid.2016.02.019>.

Kanyerezi S and Nabisubi P (2020) Comparative Genomic Analysis of *Mycobacterium tuberculosis* Reveals Evolution and Genomic Instability within Uganda I Sub-Lineage." *bioRxiv*, January, 2020.10.24.353425. <https://doi.org/10.1101/2020.10.24.353425>.

- Katale BZ, Mbelele PM, Lema NA, Campino S, Mshana SE, Rweyemamu MM, Phelan JE, Keyyu JD, Majigo M, Mbugi EV, Dockrell HM, Clark TG, Matee MI and Mpagama S (2020) Whole genome sequencing of Mycobacterium tuberculosis isolates and clinical outcomes of patients treated for multidrug-resistant tuberculosis in Tanzania. *BMC Genomics*.**21**(1): 174.  
<https://doi.org/10.1186/s12864-020-6577-1>.
- Mbelele PM, Utpatel C, Sauli E, Mpolya EA, Mutayoba BK, Barilar I, Dreyer V, Merker M, Sariko ML, Swema BM, Mmbaga BT, Gratz J, Addo KK, Pletschette M, Niemann S, Houtp ER, Mpagama SG, Heysell SK (2022) Whole Genome Sequencing-Based Drug Resistance Predictions of Multidrug-Resistant Mycobacterium Tuberculosis Isolates from Tanzania. *JAC-Antimicrobial Resistance* 4 (2): dlac042.  
<https://doi.org/10.1093/jacamr/dl-ac042>.
- Al-Mutairi NM, Ahmad S, Mokaddas E, Eldeen HS and Joseph S (2019) Occurrence of disputed rpoB mutations among Mycobacterium tuberculosis isolates phenotypically susceptible to rifampicin in a country with a low incidence of multidrug-resistant tuberculosis. *BMC Infectious Diseases*.**19**(1): 3.  
<https://doi.org/10.1186/s12879-018-3638-z>.
- SenghoreM, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, Kwambana-Adams B, Kay GL, Sanogo M, Baya B, Orsega S, Doumbia S, Diallo S, de Jong BC, Pallen MJ and Antonio M (2020) Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. *Scientific Reports*.**10**(1): 327.  
<https://doi.org/10.1038/s41598-019-56001-0>.
- Kotwal A, Raghuvanshi S, Sindhwani G and Khanduri R (2017) Mycobacterium tuberculosis and nontuberculosis mycobacteria co-infection: Two cases from the sub-Himalayan region of North India in a year. *Lung India : Official Organ of Indian Chest Society*.**34**(5): 494–496.  
[https://doi.org/10.4103/lungindia.lungindia\\_108\\_17](https://doi.org/10.4103/lungindia.lungindia_108_17).
- Hoza AS, Mfinanga SGM, Rodloff AC, Moser I and König B (2016) Increased Isolation of Nontuberculous Mycobacteria among TB Suspects in Northeastern, Tanzania: Public Health and Diagnostic Implications for Control Programmes. *BMC Research Notes* 9 (February): 109.  
<https://doi.org/10.1186/s13104-016-1928-3>.
- Mnyambwa NP, Kim DJ, Ngadaya E, Chun J, Ha S, Petrucka P, Addo KK, Kazwala RR and Mfinanga SG (2018) Genome sequence of Mycobacterium yongonense RT 955-2015 isolate from a patient misdiagnosed with multidrug-resistant tuberculosis: First clinical detection in Tanzania. *International Journal of Infectious Diseases*.**71**: 82–88.  
<https://doi.org/10.1016/j.ijid.2018.04.796>.
- van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD (1991) “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." *Journal of Clinical Microbiology* 29 (11): 2578–86. <https://doi.org/10.1128/jcm.29.11.2578-2586.1991>.
- Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR and Konstantinidis KT (2018) The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research*.**46**(W1): W282–W288 <https://doi.org/10.1093/nar/gky467>.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL and Sobral BW (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*.**42**(D1): D581–D591. <https://doi.org/10.1093/nar/gkt1099>.
- Meier-Kolthoff JP and Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*.**10**(1): 2182. <https://doi.org/10.1038/s41467-019-10210-3>.
- Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL and Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research*.**50**(D1): D801–D807. <https://doi.org/10.1093/nar/gkab902>.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S and Phillippy AM (2016) Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*.**17**(1): 132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T and Ussery DW (2007) RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*.**35**(9): 3100–3108. <https://doi.org/10.1093/nar/gkm160>.
- Meier-Kolthoff JP, Auch AF, Klenk HP and Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*.**14**(1): 60. <https://doi.org/10.1186/1471-2105-14-60>.
- Galardini M, Mengoni A and Bazzicalupo M (2015) Mapping contigs using CONTIGuator. *Methods in Molecular Biology (Clifton, N.J.)*.**1231**: 163–176. [https://doi.org/10.1007/978-1-4939-1720-4\\_11](https://doi.org/10.1007/978-1-4939-1720-4_11).
- Antonopoulos DA, Assaf R, Aziz RK, Brettin T, Bun C, Conrad N, Davis JJ, Dietrich EM, Disz T, Gerdes S, Kenyon RW, Machi D, Mao C, Murphy-Olson DE, Nordberg EK, Olsen GJ, Olson R, Overbeek R, Parrello B and Yoo H (2019) PATRIC as a unique resource for studying antimicrobial resistance. *Briefings in*

- Bioinformatics*.**20**(4): 1094–1102.  
<https://doi.org/10.1093/bib/bbx083>.
- Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Boucharad M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B and McArthur AG (2020) CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*.**48**(D1): D517–D525.  
<https://doi.org/10.1093/nar/gkz935>.
- Hua X, Liang Q, Deng M, He J, Wang M, Hong W, Wu J, Lu B, Leptihn S, Yu Y and Chen H (2021) BacAnt: A Combination Annotation Server for Bacterial DNA Sequences to Identify Antibiotic Resistance Genes, Integrons, and Transposable Elements. *Frontiers in Microbiology*.**12**.  
<https://doi.org/10.3389/fmicb.2021.649969>.
- Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR and Yoo H (2016) PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Frontiers in Microbiology* **7**: 118.  
<https://doi.org/10.3389/fmicb.2016.00118>.
- Edgar RC (2004) MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* **32** (5): 1792–97.  
<https://doi.org/10.1093/nar/gkh340>.
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30** (9): 1312–1313.  
<https://doi.org/10.1093/bioinformatics/btu033>.
- Stamatakis A, Hoover P and Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*.**57**(5): 758–771.  
<https://doi.org/10.1080/10635150802429642>.
- Lee I, Chalita M, Ha SM, Na SI., Yoon SH and Chun J (2017) ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. In *International Journal of Systematic and Evolutionary Microbiology*.**67**(6): 2053–2057.  
<https://doi.org/10.1099/ijsem.0.001872>.
- Kim M, Oh HS, Park SC and Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*.**64** (Pt 2): 346–351.  
<https://doi.org/10.1099/ijls.0.059774-0>.
- Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, Meyer F, Olsen G, Olson R, Osterman A, Overbeek R, Mcneil L, Paarmann D, Paczian T, Parrello B and Zagnitko O (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*.**9**: 75.  
<https://doi.org/10.1186/1471-2164-9-75>.
- Brettin, T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R,

- Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR and Xia F (2015) RASTtk: A Modular and Extensible Implementation of the RAST Algorithm for Building Custom Annotation Pipelines and Annotating Batches of Genomes. *Scientific Reports* 5 (February): 8365–8365. <https://doi.org/10.1038/srep08365>.
- Sánchez M Blesa A, Sacristán-Horcajada E and Berenguer J (2019) Complete Genome Sequence of *Mycolicibacterium hassiacum* DSM 44199. *Microbiology Resource Announcements*.8(4): e01522-18. <https://doi.org/10.1128/MRA.01522-18>.
- Vatlin AA, Shur KV, Danilenko VN and Maslov DA (2019) Draft Genome Sequences of 12 *Mycolicibacterium smegmatis* Strains Resistant to Imidazo-Tetrazines. *Microbiology Resource Announcements*.8 (16): e00263-19. <https://doi.org/10.1128/MRA.00263-19>.
- Chen S, Teng T, Wen S, Zhang T and Huang H (2020) The aceE Involves in Mycolic Acid Synthesis and Biofilm Formation in *Mycobacterium Smegmatis*. *BMC Microbiology* 20 (1): 259. <https://doi.org/10.1186/s12866-020-01940-2>.
- Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, Yadav J, Naaz F, Ravikumar R, Buggi S, Suar M, Gupta UD, Pandey A, Chauhan DS, Tripathy SP, Gowda H and Prasad TSK (2019) Whole Genome Sequencing of *Mycobacterium tuberculosis* Clinical Isolates From India Reveals Genetic Heterogeneity and Region-Specific Variations That Might Affect Drug Susceptibility. *Frontiers in Microbiology*.10: 309.
- Joan O, Thiele T, Schütz K, Schwerk N, Sedlacek L, Kalsdorf B, Baumann U and Stoll M (2020) Multidrug-resistant *Mycobacterium tuberculosis*: A report of cosmopolitan microbial migration and an analysis of best management practices. *BMC Infectious Diseases*.20(1): 678. <https://doi.org/10.1186/s12879-020-05381-0>.
- Parthasarathy K, Sivaraj A and Sundar R (2016) Identification of new efflux pump proteins from multidrug resistant *Mycobacterium tuberculosis* and screening for peptide based efflux pump inhibitors. *International Journal of Infectious Diseases*.45: 404–405.
- Yoon JK, Kim TS, Kim JI and Yim JJ (2020) Whole genome sequencing of Nontuberculous *Mycobacterium* (NTM) isolates from sputum specimens of co-habiting patients with NTM pulmonary disease and NTM isolates from their environment. *BMC Genomics*. 23;21(1):322.
- Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, Otchere ID, Asante-Poku A, Asare P, Sánchez-Busó L, Gehre F, Sanoussi CN, Antonio M, Affolabi D, Fyfe J, Beckert P, Niemann S, Alabi AS, Grobusch MP and Brites D (2021) Phylogenomics of *Mycobacterium Africanum* Reveals a New Lineage and a Complex Evolutionary History. *Microbial Genomics*. 000477:

- Microbiology Society,  
<https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000477>.
- Vázquez-Chacón CA, Rodríguez-Gaxiola F, de J, López-Carrera CF, Cruz-Rivera M, Martínez-Guarneros A, Parra-Unda R, Arámbula-Meraz E, Fonseca-Coronado S, Vaughan G and López-Durán PA (2021) Identification of drug resistance mutations among *Mycobacterium bovis* lineages in the Americas. *PLOS Neglected Tropical Diseases*.**15** (2): e0009145.  
<https://doi.org/10.1371/journal.pntd.0009145>.
- Jain C, Rodriguez-RLM, Phillippy AM, Konstantinidis KT and Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*. **9**(1): 5114.  
<https://doi.org/10.1038/s41467-018-07641-9>.
- Bhattacharya S, Junghare V, Pandey NK, Baidya S, Agarwal H, Das N, Banerjee A, Ghosh D, Roy P, Patra HK and Hazra S (2021) Variations in the SDN Loop of Class A Beta-Lactamases: A Study of the Molecular Mechanism of BlaC (*Mycobacterium tuberculosis*) to Alter the Stability and Catalytic Activity Towards Antibiotic Resistance of MBIs. *Frontiers in Microbiology*.**12**. 710291  
<https://doi.org/10.3389/fmicb.2021.710291>.
- Kidenya BR, Mshana SE, Fitzgerald DW and Ocheretina O (2018) Genotypic drug resistance using whole-genome sequencing of *Mycobacterium tuberculosis* clinical isolates from North-western Tanzania. *Tuberculosis*.**109**: 97–101.  
<https://doi.org/10.1016/j.tube.2018.02.004>.
- Ishiekwene C, Subran M, Ghitan M, Kuhn-Basti M, Chapnick E and Lin YS (2017) Case report on pulmonary disease due to coinfection of *Mycobacterium tuberculosis* and *Mycobacterium abscessus*: Difficulty in diagnosis. *Respiratory Medicine Case Reports*.**20**: 123–124.  
<https://doi.org/10.1016/j.rmcr.2017.01.011>.
- Stepanyan IE, Zaytseva AS, Bagdasaryan TR, Larionova EE, Smirnova TG, Andreevskaya SN, Chernousova LN and Ergeshov AE (2019) Nontuberculous mycobacterial co-infection in HIV-negative patients with pulmonary tuberculosis. *European Respiratory Journal*.**54** (suppl 63): PA4644.  
<https://doi.org/10.1183/13993003.congress-2019.PA4644>.
- Zhang Q, Liu H, Liu X, Jiang D, Zhang B, Tian H, Yang C, Guddat LW, Yang H, Mi K and Rao Z 2018 Discovery of the first macrolide antibiotic binding protein in *Mycobacterium tuberculosis*: A new antibiotic resistance drug target. *Protein and Cell*.**9**(11): 971–975.  
<https://doi.org/10.1007/s13238-017-0502-7>.
- Dorman Susan E, et al (2012) Genotype MTBDRplus for Direct Detection of *Mycobacterium Tuberculosis* and Drug Resistance in Strains from Gold Miners in South Africa. *Journal of Clinical Microbiology* **50** (4): 1189–94.  
<https://doi.org/10.1128/JCM.05723-11>.

