KAUST Academy Summer Program Project Report

Does cross-border transmission contribute to the development of drug resistance in tuberculosis (TB) in Southern Africa?

Intern: **Lama Khalid Albishri**, KAUST Academy, King Abdullah University of Science and Technology (KSA)

Project Mentor: **Dr. Andries Van Tonder**, Research Associate, Department of Veterinary Medicine, University of Cambridge, UK

Internship Organization: **University of Cambridge**

Abstract

The emergence and spread of drug-resistant tuberculosis (TB) pose a significant threat to global health [1], particularly in high-burden regions like Southern Africa [2]. Cross-border transmission can undermine national TB control programs, but its specific role in the dissemination of drug-resistant Mycobacterium tuberculosis in this region remains poorly characterized. This project utilized whole-genome sequencing (WGS) data from 325 *M. tuberculosis* isolates from Botswana [3] and Namibia [4] to investigate this phenomenon. We performed a comprehensive genomic analysis involving quality control, phylogenetic reconstruction, drug resistance profiling, and transmission network analysis [5]. Our results reveal a complex epidemiological landscape dominated by Lineage 4 and a high prevalence of multidrug-resistant TB (MDR-TB; 75.6%). Phylogenetic analysis demonstrated significant mixing of isolates from both countries, indicating a shared strain pool. Crucially, we identified a direct molecular link (≤12 SNPs) between an MDR-TB case in central Namibia and another in northern Botswana, providing concrete genomic evidence of active cross-border transmission. This study underscores the critical need for collaborative, international public health interventions and reinforces the value of genomic surveillance in tracking the spread of drug-resistant TB across national borders.

1. Introduction

Tuberculosis (TB) continues to be a leading cause of death from a single infectious agent globally[1]. The crisis is exacerbated by the rise of drug-resistant strains of Mycobacterium tuberculosis, which complicate treatment regimens and worsen patient outcomes. Southern Africa bears a disproportionate burden of the TB epidemic, with Namibia and Botswana classified as high-burden countries for TB [2], MDR-TB, and HIV-associated TB. In such a context, human mobility and cross-border transmission present a formidable challenge to TB control, potentially facilitating the spread of resistant strains and undermining the efforts of individual national programs.

While numerous studies have used whole-genome sequencing (WGS) to understand national TB transmission dynamics, few have integrated datasets from neighbouring countries to directly investigate international spread. This knowledge gap is critical, as effective public health policy requires a regional, rather than purely national, perspective. This project aimed to address this by analyzing WGS data from two published studies from Botswana and Namibia [3,4].

1.1 Objectives

1. Characterize the genetic diversity and drug resistance profiles of the isolates
2. Reconstruct phylogenetic relationships and transmission networks
3. Identify and geolocate potential cross-border transmission events between the two countries.

2. Methodology

2.1. Data Acquisition and Curation

Raw WGS read files (FASTQ) for 325 isolates (189 from Botswana, 136 from Namibia) were initially retrieved using the ***nf-core/fetchngs*** pipeline via their European Nucleotide Archive (ENA) study accessions (PRJEB83872, PRJEB54067) [3,4].

2.2. Quality Control and Read Mapping

Quality control and contamination screening were performed using ***avantonder/BacQC***. Samples with >20% of reads mapping to non-target species were removed [5]. High-quality reads were mapped to the MTBC0 reference genome using the ***nf-core/bactmap*** pipeline. Samples with less than 75% genome coverage were excluded [5], resulting in a final high-quality dataset of 316 isolates (183 from Botswana, 133 from Namibia).

2.3. Variant Calling and Phylogenetics

The bactmap pipeline was used to call core genome variants and generate consensus pseudogenomes. A maximum-likelihood phylogenetic tree was inferred from this alignment using ***IQ-TREE***.

2.4. Lineage and Drug Resistance Profiling

Lineage assignment and identification of drug resistance-conferring mutations were performed using the command-line version of ***TB-Profiler***, a pipeline that allows users to analyse *M. tuberculosis* whole genome sequencing data to predict lineage and drug resistance [5].

2.5. Transmission and Geospatial Analysis

Transmission networks were constructed by calculating pairwise SNP distances between all isolates. Networks were tested for a range of SNP thresholds from 1 to 50 to identify the first definitive cross-border transmission link, which emerged at a threshold of ≤12 SNPs. Networks were visualized using ***ggraph*** in R. The specific cross-border transmission event was geolocated and mapped using ***ggmap***. The final phylogenetic tree, annotated with metadata, was visualized using ***Microreact*** [5].

3. Results and Discussion

The analysis of 316 high-quality isolates revealed a high prevalence of multidrug-resistant TB (MDR-TB; 75.6%), underscoring the severe public health challenge in the region. Phylogenetic analysis demonstrated extensive mixing of isolates from Botswana and Namibia across all major lineages, indicating a shared strain pool with no geographical segregation **(Fig. 1)**.

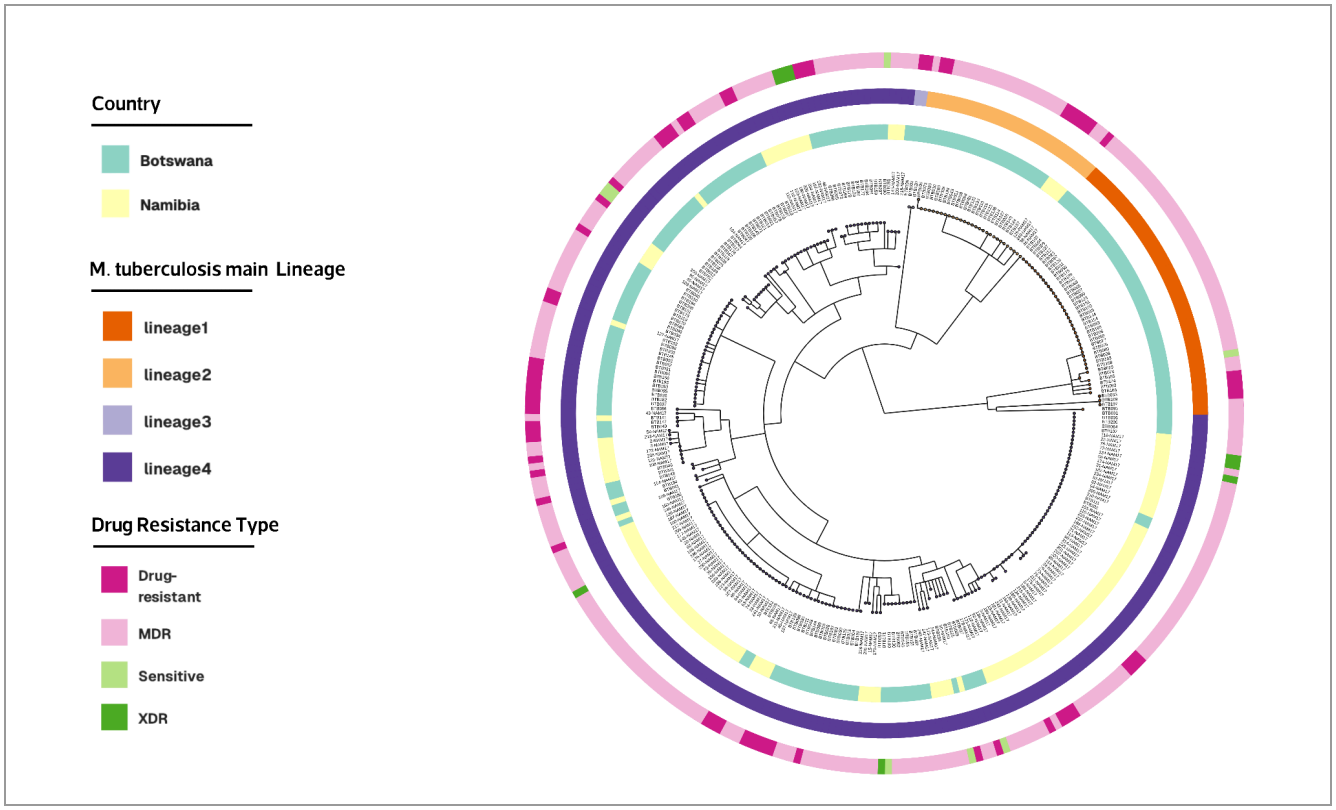


Figure 1: Maximum-Likelihood Phylogenetic Tree of 316 *M. tuberculosis* Isolates. The tree is annotated by country (Botswana, Namibia), showing extensive mixing of strains. Outer rings indicate lineage and drug resistance type. Scale bar indicates the number of nucleotide substitutions per site.

This lack of phylogenetic structure provided the first clue that cross-border transmission might be occurring. To investigate this, we tested a range of SNP thresholds from 1 to 50 to define recent transmission links. A significant cross-border link first emerged at a threshold of ≤12 SNPs **(Fig. 2)**, which was selected for further analysis.

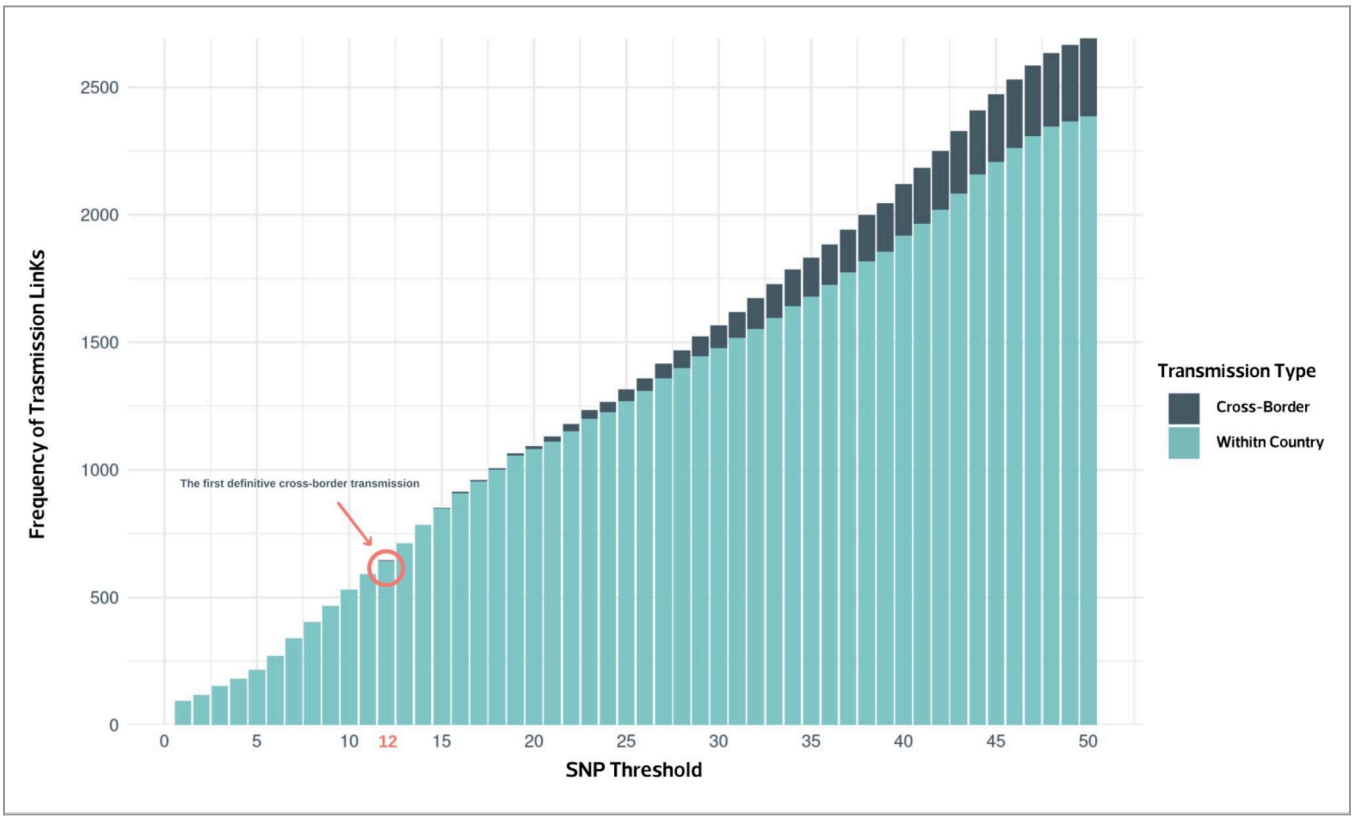


Figure 2: Transmission Link Analysis by SNP Threshold. The plot shows the number of cross-border transmission links identified at increasing SNP thresholds. The dotted vertical line indicates the ≤12 SNP threshold, where the first definitive cross-border link was identified.

Crucially, transmission network analysis at the ≤12 SNP threshold identified a direct molecular link between an MDR-TB case in central Namibia and another in northern Botswana (**Fig.3**). This short genetic distance is highly suggestive of a recent, direct transmission event, providing the first genomic evidence of cross-border transmission of MDR-TB between these two nations.

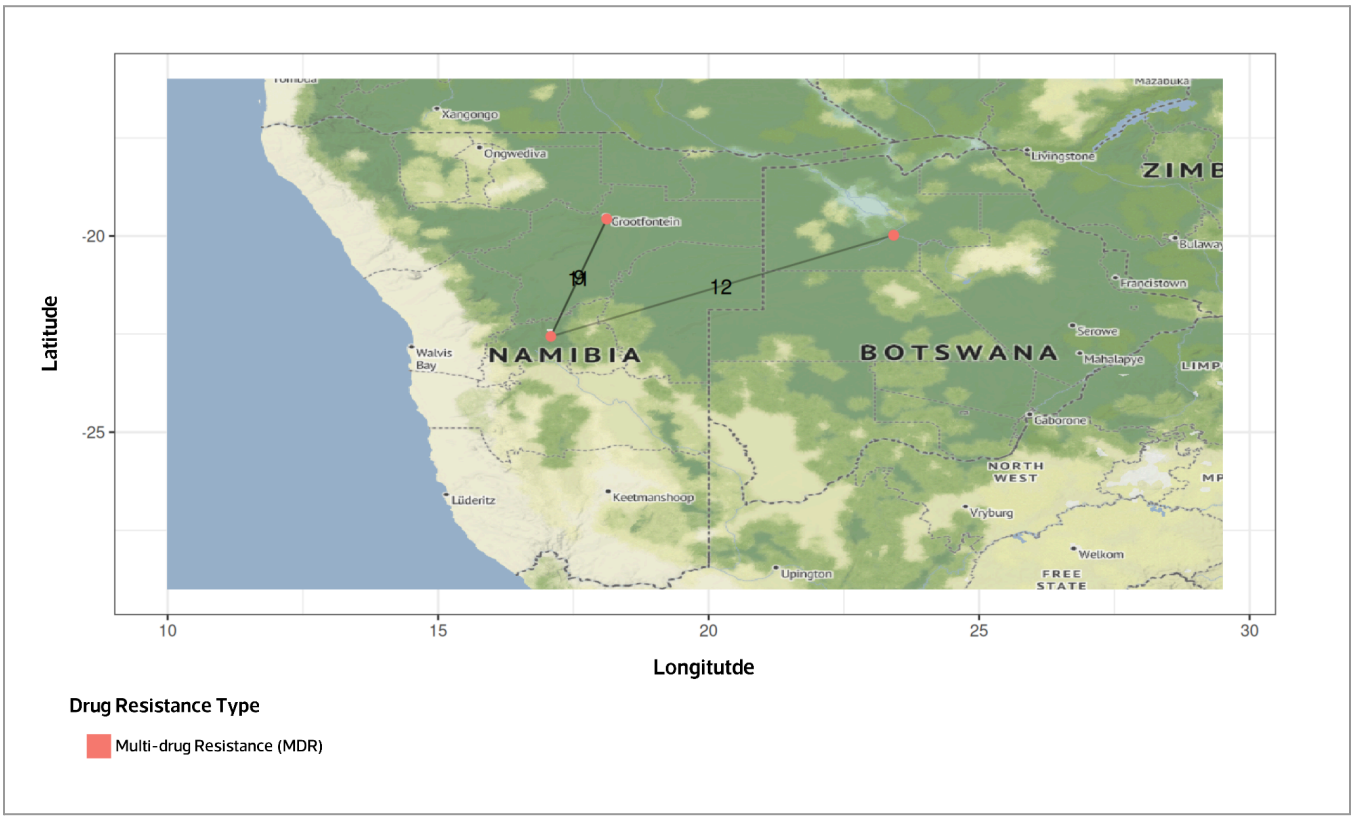


Figure 3: Evidence of a Cross-Border Transmission Event. Map highlighting the specific molecular link (≤12 SNPs) between an MDR-TB case in central Namibia and another in northern Botswana.

The finding that the transmitted strain was already multidrug-resistant confirms that cross-border transmission is not only tangible but also contributes directly to the spread of pre-existing drug resistance, representing a worst-case scenario for public health efforts. This empirically confirms that national borders are permeable to the TB epidemic and that international transmission is a tangible mechanism for the spread of drug-resistant strains.

4. Challenges and Reflections

This project encountered several significant practical challenges common in bioinformatics research. The initial automated data retrieval using nf-core/fetchngs successfully downloaded all sample files; however, a subset of these files was incomplete or corrupted, which caused downstream analysis pipelines to fail. This required manually identifying and re-downloading the affected samples to ensure data integrity, highlighting the importance of validating download completeness in automated processes. Another challenge was the lack

of critical metadata for one study in public repositories. It had to be manually curated from figures and tables in the original publication after attempts to contact authors were unsuccessful, which was time-consuming. Computational limitations, particularly local storage constraints, caused pipeline failures many times, which were overcome through several attempts to manage resources. These challenges were ultimately addressed through perseverance, adaptability, and manual effort, reinforcing the importance of these often-underestimated aspects of the research process. Overcoming these obstacles not only enabled the successful completion of the project but also provided invaluable experience in troubleshooting, debugging, and resource management that will be essential for future research endeavors.

5. Conclusion and Future Work

This study provides clear genomic evidence that cross-border transmission is a reality in the drug-resistant TB epidemic between Botswana and Namibia. The sharing of strains and the identified transmission link necessitate a paradigm shift from national to regional TB control strategies. The high prevalence of MDR-TB, coupled with evidence of international transmission, underscores the urgent need for collaborative public health interventions, including strengthened binational surveillance, data sharing agreements, and integrated contact tracing programs that operate across borders.

Future work will focus on expanding the geographical scope to incorporate data from other neighbouring countries, such as South Africa, Zimbabwe, and Zambia, to build a more comprehensive picture of transmission dynamics across Southern Africa. We also aim to incorporate detailed epidemiological data, including travel history and social contacts, to corroborate genomic inferences of transmission. Additionally, we plan to employ phylodynamic methods to infer the direction and timing of cross-border transmission events, which could further refine public health targeting. **Personally**, I plan to continue developing these research skills and apply them to public health challenges in Saudi Arabia, contributing to global efforts in combating infectious diseases through bacterial genomics.

Resources and Tools Used

- **Data Sources:** European Nucleotide Archive (ENA) study accessions (PRJEB83872, PRJEB54067).
- **Bioinformatic Pipelines:** nf-core/fetchngs, avantonder/BacQC, nf-core/bactmap, TB-Profiler
- **Phylogenetic Software:** IQ-TREE
- **Programming Languages:** Shell , R (with tidyverse, ggraph, and ggmap packages)
- **Visualization Platforms:** Microreact
- **Computational Resources:** Local high-performance desktop computer
- **Code Repository:** GitHub

Acknowledgements

I extend my deepest gratitude to my project mentor, **Dr. Andries Van Tonder**, for his exceptional guidance, expertise, and unwavering support throughout this project.

I am profoundly grateful to the program leads, **Dr. Paul Fannon, Dr. Matt Castle, and Dr. Hugo Tavares**, and all the instructors of the eight-week internship program at the **University of Cambridge** for their invaluable teachings, mentorship, and for orchestrating such a stimulating and rewarding academic experience.

My sincere thanks also go to my colleagues and fellow students for their kindness, insightful discussions, and collaboration, which greatly enriched this internship.

This incredible opportunity was made possible by the KAUST Academy, and I extend all my thanks to each and every person at KAUST Academy who helped us. Thank you for your incredible effort and for making this journey possible. Finally, I am deeply thankful to the program sponsors for their generous financial backing.

References

1. *Liebenberg, D., Gordhan, B. G., & Kana, B. D. (2022).* ***Drug resistant tuberculosis: Implications for transmission, diagnosis, and disease management. Frontiers in cellular and infection microbiology****, 12, 943545.* *https://doi.org/10.3389/fcimb.2022.943545*
2. ***WHO global lists of high burden countries for tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB),*** *Geneva: World Health Organization,2021–2025.* *https://cdn.who.int/media/docs/default-source/hq-tuberculosis/who_globalhbcliststb_2021-2025_backgrounddocument.pdf?sfvrsn=f6b854c2_9*
3. *Mogashoa, T., Loubser, J., Choga, O. T., Ngom, J. T., Choga, W. T., Mbulawa, M. B., Molefi, T., Stephen, O., Makhondo, T., Seru, K., Motshosi, P., Zuze, B., Makhema, J., Musonda, R. M., Otukile, D., Modongo, C., Kgwaadira, B. T., Fane, K., Gaseitsiwe, S., Warren, R. M., ... Streicher, E. M. (2025).* ***Whole genomic analysis uncovers high genetic diversity of rifampicin-resistant Mycobacterium tuberculosis strains in Botswana.*** *Frontiers in microbiology, 16, 1535160.* *https://doi.org/10.3389/fmicb.2025.1535160*
4. *Claassens, M., Dreyer, V., Nepolo, E., Mokomele, Q., van Rooyen, G., Ruswa, N., Günther, G., & Niemann, S. (2022).* ***Whole-Genome Sequencing for Resistance Prediction and Transmission Analysis of Mycobacterium tuberculosis Complex Strains from Namibia. Microbiology spectrum****, 10(5), e0158622.* *https://doi.org/10.1128/spectrum.01586-22*
5. *van Tonder, A., Tavares, H., Salehe, B. (2024).* ***Working with Bacterial Genomes****.* *https://cambiotraining.github.io/bacterial-genomics/*