UNIVERSITÄT ZU LÜBECK

From the Research Center Borstel - Leibniz Lung Center

Priority Research Area infections

Molecular and Experimental Mycobacteriology

Director: Prof. Dr. Ulrich Schaible

MULTIDRUG RESISTANCE, POPULATION STRUCTURE, AND EPIDEMIOLOGY OF *MYCOBACTERIUM TUBERCULOSIS* COMPLEX STRAINS FROM WEST AFRICA

Dissertation for the fulfillment of requirements for the

Doctoral Degree

of the University of Lübeck

From the Department of Natural Sciences

Submitted by

Harriet Naa Afia Blankson, M.Sc.

From Accra, Ghana

Lübeck, 2024

First referee: Herr Prof. Dr. Stefan Niemann

Second referee: Herr Prof. Dr. Dennis Nurjadi

Date of oral examination: February 18th, 2025

Approved for printing: Lübeck, February 25th, 2025

Abstract

Tuberculosis (TB) remains a global public health challenge, impacting millions of individuals and placing a substantial burden on healthcare systems worldwide. The causative agents of TB, *Mycobacterium tuberculosis* complex (MTBC) strains, engage in a complex interplay of factors that influence its transmission, evolution, and the emergence of drug-resistant strains. Despite commendable progress in TB control efforts, the persistent threat of multidrug-resistant TB(MDR-TB), characterized by resistance to both isoniazid (INH) and rifampicin (RIF), has significantly complicated treatment protocols.

Recent recommendations by the World Health Organization (WHO) advocate for a new all-oral 6-month regimen to address MDR-TB. However, concerns arise with the development of resistance to fluoroquinolones (FQs) and bedaquiline (BDQ), potentially jeopardizing the regimen's efficacy. Documented outbreaks of MTBC strains, particularly of MDR strains, underscore the urgency of monitoring the transmission of MTBC strains, as it can impact diagnostics and treatment strategies in affected regions. Specific strains within MTBC lineages may exhibit unique traits, such as varying minimum inhibitory concentrations (MIC) or inherent resistance to specific drugs. Understanding the circulating MTBC strains in a country, their drug resistance profiles, and their evolutionary dynamics is imperative for effective TB management.

To close these knowledge gaps, this thesis seeks to explore the multifaceted landscape of TB through epidemiological, molecular, and clinical perspectives. The research aims to unravel the genomic intricacies of MTBC strains, by investigating the transmission dynamics of rifampicin-resistant (RR)/MDR-TB in Sierra Leone and exploring the transmission dynamics and drug resistance patterns of the sublineage 4.6.2.2 Cameroon. By shedding light on these aspects, this thesis seeks to contribute to enhanced diagnostics and more effective treatment strategies and ultimately support the global mission to eradicate TB.

In the study on RR/MDR-TB in Sierra Leone, whole genome sequencing (WGS) revealed high levels of drug resistance, with one in four strains being resistant to all first-line anti-TB drugs. While no FQ resistance was detected, five strains exhibited resistance to BDQ/clofazimine (CFZ) due to mutations in the *Rv0678* gene. The study also revealed a greater diversity of drug resistance

mutations, including borderline INH and RIF resistance mutations, which can potentially influence treatment options. The high cluster rate of over 40% indicated ongoing transmission of RR/MDR-TB strains, contributing to the burden of RR/MDR-TB in the country. Analysis of the 238 MTBC strains revealed a high diversity of strains in Sierra Leone. The presence of six major lineages (L) of MTBC strains (L1= 4%, L2 = 9%, L3= 0.8%, L4= 62%, L5= 2.9% and L6= 21%) were identified in Sierra Leone. *Mycobacterium tuberculosis* (*Mtb)* strains constitute 56%, while *Mycobacterium africanum* (*Maf*) L6 strains, account for 21% of MDR MTBC strains, which suggest a longitudinal outbreak with specific branches exhibiting resistance to multiple drugs, including BDQ/CFZ. Despite the high diversity, strains of certain sublineages 4.1.2.1. Haarlem, 4.8 mainly T and 2.2.1 Beijing Ancestral 3 and 6.3.3 West Africa 2 were implicated in the ongoing MDR transmissions.

The global population structure and phylogeography of strains of the Cameroon sublineage were also investigated using WGS. The strains were classified into eight distinct clades, with strains originating from 24 countries across Africa, Asia, Australia, and Europe. A fourth of the strains were identified as transmission strains. The strains of two clades, C.5 and C.8, exhibited high clustering rates, indicating higher transmission potential. Drug resistance was also observed, with over 10% of the strains classified as MDR. The strains remained susceptible to BDQ/CFZ, except for two strains displaying resistance.

Overall, these findings contribute to a comprehensive understanding of the transmission dynamics and drug resistance patterns of MTBC strains in Sierra Leone and the sublineage 4.6.2.2 *Cameroon*. The clades of strains that were defined, clusters, and global distribution emphasize the role of migration in the spread of these strains locally and beyond Africa. The insights gained from this research can inform and improve TB surveillance and control measures, both within Sierra Leone, West Africa, and in other parts of the world where the Cameroon sublineage strains have been identified.

Zusammenfassung

Tuberkulose (TB) ist nach wie vor eine globale Herausforderung für die öffentliche Gesundheit, von der Millionen von Menschen betroffen sind und die eine erhebliche Belastung für die Gesundheitssysteme weltweit darstellt. Die Erreger der Tuberkulose, *Mycobacterium tuberculosis* Komplex (MTBK) Stämme, stehen in einem komplexen Zusammenspiel von Faktoren, die ihre Übertragung, ihre Evolution und das Auftreten medikamentenresistente Stämme beeinflussen. Trotz lobenswerter Fortschritte bei der Tuberkulosebekämpfung hat die anhaltende Bedrohung durch mehrfach medikamentenresistente Tuberkulose (*multidrug resistant tuberculosis,* MDR-TB), die durch eine Resistenz sowohl gegen Isoniazid (INH) als auch gegen Rifampicin (RIF) gekennzeichnet ist, die Behandlungsprotokolle erheblich erschwert.

Jüngste Empfehlungen der Weltgesundheitsorganisation (WHO) sprechen sich für eine neue, rein orale 6-Monats-Therapie zur Behandlung von MDR-TB aus. Es sind jedoch Bedenken angesichts der Entwicklung von Resistenzen gegen Fluorchinolone (FQs) und Bedaquilin (BDQ) aufgekommen, die die Wirksamkeit des Regimes gefährden könnten. Dokumentierte Ausbrüche von MTBK-Stämmen, insbesondere von MDR-Stämmen, unterstreichen die Dringlichkeit der Überwachung der Übertragung von MTBK-Stämmen, da dies Auswirkungen auf die Diagnostik und Behandlungsstrategien in den betroffenen Regionen haben kann. Bestimmte Stämme innerhalb der MTBK-Linien können einzigartige Eigenschaften aufweisen, wie z.B. unterschiedliche minimale Hemmkonzentrationen (*minimum inhibitory concentration*, MIC*)* oder eine inhärente Resistenz gegen bestimmte Medikamente. Das Verständnis der in einem Land zirkulierenden MTBK-Stämme, ihrer Resistenzprofile und ihrer Entwicklungsdynamik ist für ein wirksames TB-Management unerlässlich.

Um diese Wissenslücken zu schließen, strebt diese Arbeit danach, die facettenreiche Landschaft der TB aus epidemiologischer, molekularer und klinischer Sicht zu untersuchen. Die Forschung zielt darauf ab, die genomischen Feinheiten von MTBK-Stämmen zu entschlüsseln, indem die Übertragungsdynamik von Rifampicin-resistenter (RR)/MDR-TB in Sierra Leone analysiert wird und die Übertragungsdynamik sowie die Arzneimittelresistenzmuster der Unterlinie 4.6.2.2 Cameroon untersuchtwerden. Durch die Aufklärung dieser Aspekte soll diese Arbeit zu einer

verbesserten Diagnostik und wirksameren Behandlungsstrategien beitragen und letztlich die globale Mission zur Ausrottung der TB unterstützen.

In der Studie über RR/MDR-TB in Sierra Leone ergab die Ganzgenomsequenzierung (*whole genome sequencing,* WGS) ein hohes Maß an Medikamentenresistenz, wobei einer von vier Stämmen gegen alle Erstlinien-Anti-TB-Medikamente resistent war. Während keine FQ-Resistenz festgestellt wurde, wiesen fünf Stämme aufgrund von Mutationen im *Rv0678*-Gen eine Resistenz gegen BDQ/Clofazimin (CFZ) auf. Die Studie deckte auch eine größere Vielfalt an Resistenzmutationen gegen Medikamente auf, einschließlich grenzwertiger INH- und RIF-Resistenzmutationen, die möglicherweise die Behandlungsmöglichkeiten beeinflussen. Die hohe Cluster-Rate von über 40% deutet auf eine anhaltende Übertragung von RR/MDR-TB-Stämmen hin und trägt zur Belastung durch RR/MDR-TB im Land bei. Die Analyse der 238 MTBK-Stämme ergab eine hohe Diversität der Stämme in Sierra Leone. In Sierra Leone wurde das Vorkommen von sechs Hauptlinien (L) von MTBK-Stämmen (L1= 4%, L2= 9%, L3= 0,8%, L4= 62%, L5= 2,9% und L6= 21%) festgestellt. *Mycobacterium tuberculosis* (*Mtb*) Stämme machen 56% aus, während *Mycobacterium africanum* (*Maf*) Stämme der Linie 6 21% der MDR-MTBK-Stämme ausmachen, was auf einen langristigen Ausbruch mit spezifischen Zweigen hindeutet, die Resistenz gegen mehrere Medikamente, einschließlich BDQ/CFZ, aufweisen. Trotz der hohen Diversität sind die Stämme bestimmter Unterlinien 4.1.2.1. *Haarlem*, 4.8 mainly T und 2.2.1 Beijing Ancestral 3 und 6.3.3 West Africa 2 in die laufende MDR-Übertragung involviert.

Die globale Populationsstruktur und die Phylogeographie der Stämme der Unterlinie 4.6.2.2 Cameroon wurden ebenfalls mittels WGS untersucht. Die Stämme wurden in acht verschiedene Gruppen eingeteilt, wobei die Stämme aus 24 Ländern in Afrika, Asien, Australien und Europa stammen. Ein Viertel der Stämme wurde als Übertragungsstämme identifiziert. Die Stämme von zwei Gruppen, C.5 und C.8, wiesen hohe Clusterungs raten auf, was auf ein höheres Übertragungspotenzial hindeutet. Es wurde auch Medikamentenresistenz beobachtet, wobei über 10% der Stämme als MDR eingestuft wurden. Mit Ausnahme von zwei Stämmen, die eine Resistenz aufwiesen, blieben die Stämme gegenüber BDQ/CFZ empfindlich.

Insgesamt tragen diese Ergebnisse zu einem umfassenden Verständnis der Übertragungsdynamik und der Medikamentenresistenzmuster von MTBK-Stämmen in Sierra Leone und der Unterlinie

4.6.2.2 Cameroon bei. Die Gruppen von Stämmen, die definiert wurden, Cluster und die globale Verteilung unterstreichen die Rolle der Migration bei der Verbreitung dieser Stämme vor Ort und über Afrika hinaus. Die aus dieser Untersuchung gewonnenen Erkenntnisse können die Tuberkuloseüberwachung und -bekämpfung sowohl in Sierra Leone und Westafrika als auch in anderen Teilen der Welt, in denen Stämme der Unterlinie 4.6.2.2 Cameroon identifiziert wurden, beeinflussen und verbessern.

Table of Content

List of Abbreviations

Abbreviation	Meaning
AFB	Acid fast bacilli
BCG	bacille Calmette-Guerin
BDQ	Bedaquiline
BSC	Biological safety cabinet
CC	Cameroon clades
CFZ	Clofazimine
CI	Confidence interval
CLR	C-type lectin receptor
CO_2	Carbon dioxide
COVID-19	Corona virus disease 2019
CPR	Ciprofloxacin
CR3	Complement receptor 3
CRISPRs	clustered regularly interspaced short palindromic repeats
CS	Cycloserine
CTAB	Cetyltrimethylammonium bromide
CXR	Chest X-ray
DNA	Deoxyribonucleic acid
DR	Drug resistance
DST	Phenotypic resistance testing
EDTA	Ethylenediaminetetraacetic acid
EMB	Ethambutol
EPTB	Extra pulmonary tuberculosis
ETH	Ethionamide
FcR	Fc receptors
FOB	fiberoptic bronchoscopy
G	grad
GC	Gradient centrifugation
GC content	Guanine-cytosine content
h	hour
INH	Isoniazid
IS	insertion element
KAN	Kanamycin
l	Liter
L	Lineage
LAMP	Loop-mediated isothermal amplification
LPS	long sequence polymorphisms
LTBI	Latent tuberculosis infection
Maf	*Mycobacterium africanum*
MDM	Monocyte derived macrophages

MDR	Multi-drug resistant
MGIT	Mycobacterial growth indicator tubes
MIRU-VNTR	mycobacterial interspersed repetitive unit -variable number of tandem repeats
MLT	Maximum likelihood tree
MOI	Multiplicity of infection
MOX	Moxifloxacin
MPT	Maximum parsimony tree
MST	Minimum spanning tree
MTBC	*Mycobacterium tuberculosis* complex
n	Number
NAA	Nucleic acid amplification
NGS	Next generation sequencing
NTM	Nontuberculous mycobacteria
PAS	Para-aminosalicylic acid
PGRS	polymorphic GC-rich repetitive sequences
PMN	Polymorphonuclear leukocytes
preXDR	pre-extensively drug resistant
PWD	pairwise distance
PZA	Pyrazinamide
RFLP	restriction fragment length polymorphism
RIF	Rifampicin/rifambutin
RNA	Ribonucleic acid
rtd	Root-to-tip distance
RR	rifampicin resistant
RRDR	Rifampicin resistance determining region
SDS	Sodium dodecyl sulfate
SI	Sputum indication
SM	Streptomycin
SNP	Single nucleotide polymorphism
TAG	triacylglyceride
TB	Tuberculosis
Th	T helper
TST	Tuberculin skin test
UK	United Kingdom
UN	United Nation
VNTR	variable number tandem repeats
WGS	Whole genome sequence
WHO	World health organization
XDR	Extensively drug resistant

List of Figures

List of Tables

1. Introduction

Tuberculosis (TB) is a highly contagious bacterial infection primarily affecting the lungs. It spreads through respiratory droplets and can cause various symptoms, including persistent cough, hemoptysis (cough up blood), low-grade fever, night sweats, appetite loss, malaise, exhaustion, and weight loss [1,2]. It is generally referred to as pulmonary (P) TB [2]. In addition to PTB, there is also extrapulmonary TB (EPTB), which can involve other organs such as meningitis, disseminated (miliary) illness, lymphadenitis, kidney, bone, or joint [2,3]. Untreated TB can have severe consequences and even result in death [4]. However, TB is a curable disease, and effective treatment involves a combination of several antibiotics [5]. The standard treatment regimen recommended by the World Health Organization (WHO) is an initial two-month combination of rifampicin (RIF), isoniazid (INH), ethambutol (EMB), and pyrazinamide (PZA) and then four months of RIF and INH [6]. These four are the first-line drugs for TB treatment. For drug-resistant (DR)-TB cases, the regimen must be altered to other drug combinations such as the short six or 9-month treatment or the longer regimens of 18-to-24-month duration [7].

The following sections will delve into specific aspects of TB, including epidemiology, diagnosis, drug resistance, population structure, and genotypic methods.

1.1. Epidemiology of Tuberculosis

In 2022, TB caused the deaths of an estimated 1.3 million people and remained one of the primary causes of death resulting from a single infectious agent [8]. Until 2020, there was a steady decrease in global TB incidence; however, the COVID-19 pandemic increased TB incidence [8]. The WHO reported that the incidence of new TB cases rose by 1.9% between 2021 and 2022, and about 300,000 more people fell ill with TB in 2022 compared to 2021 [8]. In some countries, there has been progress in reducing TB incidence on a global level; however, the disease continues to be a significant burden on many economies and households, particularly in the Philippines, Indonesia, Congo, Bangladesh, and India, where the incidence is above 20 per 100 000 population [9].

An important contributing factor to the TB pandemic is the emergence of drug-resistant *Mycobacterium tuberculosis* complex (MTBC) strains. The annual incidence of drug resistant (DR)-TB, specifically rifampicin-resistant (RR)/ multidrug resistant (MDR, resistance to both RIF and INH)-TB, has been stable between 2020 and 2022 [8,9], with an estimated number of 410 000 people in 2022 [8]. A projected 3.3% of all newly reported TB cases worldwide were RR/MDR-TB, while 17% of previously treated patients were RR/MDR-TB [8]. There has been an 8.5% increase in the number of individuals enrolled in RR/MDR-TB treatment in 2022 than 2021 [8]. Despite the increased number of people receiving treatment, global treatment success remained low, at 63% [8].

To tackle the issue, the WHO instituted the end TB strategy to end the epidemic. The milestones are a 95% reduction in TB deaths, a 90% reduction in TB incidence, and zero TB-affected families experiencing catastrophic costs due to TB by 2035 compared to 2015 [10]. In order to achieve their goal, the WHO is implementing three main principles: integrated, patient-centered care and prevention, policies and supportive systems, and intensified research and innovation [10,11]. These policies aim to address the needs of patients and eliminate TB from the world [12]

1.2. Tuberculosis in Sierra Leone

Sierra Leone, a West African country with a population of 8.4 million as of 2021, still faces a high burden of TB, with an incidence of 285 per 100 000 population and 7.9 per 100 000 population for MDR/ RR-TB [13]. In 2021, 2.5% of new TB cases were RR/MDR-TB, and 12% among retreatment cases [13]. The Xpert® MTB/RIF assay (Cepheid, USA) is the primary molecular diagnostic method for TB in hospitals, while smear microscopy is still used in primary care clinics [14]. However, there is a lack of in-country testing for drug susceptibility, and there is a shortage of data and studies on MDR-TB in Sierra Leone due to the absence of laboratory infrastructure and technology for disease surveillance [15]. A recent study by Kamara *et al.* (2022) showed that those who underwent DR testing and received the WHO's short treatment regimen had a better prognosis than those on the longer treatment regimen [14]. This highlights the importance of knowing the drug resistance status in determining treatment outcomes of TB.

1.3. Biology of *Mycobacterium tuberculosis* complex strains

The bacteria causing TB are of the genus *Mycobacteria* and are believed to have evolved about 40.000-70.000 years ago [16]. MTBC bacteria are Gram-positive, catalase-positive, rod-shaped bacteria with high guanine-cytosine (GC) content (62-70%) [17,18]. They are non-motile and do not form spores. Additionally, their size ranges from 0.2 - 0.6 µm wide and 1.0 -10 µm long [19,20]. The morphology of their colonies varies, with some species forming rough or smooth colonies [20].

MTBC strains are obligate aerobes and thrive in oxygen-rich tissues like the lungs [21]. They are facultative intracellular pathogens that primarily infect mononuclear phagocytes (e.g., macrophages) [22]. The organisms are slow-growing, with a generation time of 12 to 20 hours, and have a high lipid content in the cell wall [20]. Their lipid-rich walls are responsible for their classification as "acid-fast bacilli," as basic dyes cannot permeate these unless combined with phenol. As a result, once stained, these cells resist decolorization with acidified organic solvents [20,22]. MTBC strains are hydrophobic due to the configuration of their cell wall, which causes the bacteria to clump together. Detergents such as Tween® 80 are used in culture media to disperse them [23].

Strains of the MTBC, including *M. tuberculosis (Mtb) sensu stricto, M. africanum (Maf), M. bovis,* and other animal pathogens, are the main cause of TB, a condition marked by the development of tubercles and caseous necrosis in tissues [24]. Although the bacteria typically infect the lungs (PTB), there are EPTB, such as TB of the brain, kidney, and spine [25,26]. Humans primarily spread MTBC strains through respiratory pathways [23]

Not everyone infected with TB bacteria progresses to disease; the infection may remain latent (latent TB infection, LTBI) and occurs in about 90 to 95% of cases [27,28]. In LTBI, the bacteria remain dormant, preventing multiplication. However, in 5 to 10% of infected individuals, the body's immune system is unable to stop the bacteria from growing, and they develop TB disease [29].

The route of TB infection in humans is typically through the respiratory tract. Once the pathogen reaches the lungs, it moves to the alveoli, where it is phagocytosed by alveolar macrophages [30,31].

Some bacilli can survive and replicate within the macrophages due to mechanisms such as inhibition of phagosome-lysosome fusion and modulation of macrophage apoptosis [31]. When the macrophage dies, the bacilli are disseminated to the lymph nodes and bloodstream, leading to a systemic infection [32].

As a defense mechanism, the human immune system forms a granuloma (Figure 1), permitting the immune-mediated containment of the bacteria [31]. The core of the granuloma consists of a necrotic mass of fused immune cells such as macrophages, Langhans giant cells, neutrophils, epithelial cells, and mycobacteria, and a thick wall of macrophages, natural killer T cells, CD4 and CD8 cells surround them [33]. The granuloma helps to prevent the further spread of the infection. Still, if it is disrupted, it can reactivate the infection and progress to active disease, leading to dissemination to other organs in the body [33]. If left untreated, the pathophysiology of MTBC strains can result in severe pulmonary and extrapulmonary complications, potentially leading to fatal outcomes [34–36]. Therefore, it is essential to understand the pathogenesis of MTBC strains to develop effective treatment and prevention for TB.

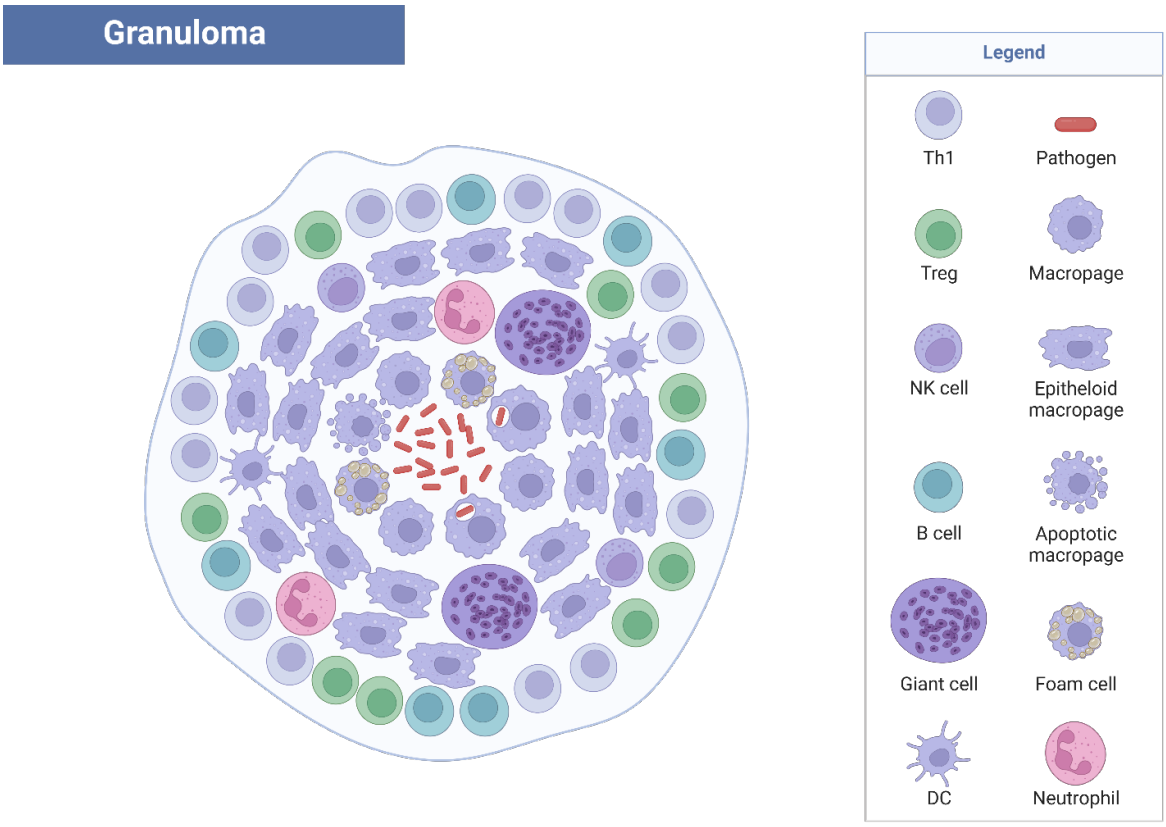


Figure 1. The structure of a granuloma induced by *Mycobacterium tuberculosis* complex strains. A granuloma is formed as a result of the body's immune response against infection by forming a cluster of immune and other cells around a pathogen. Here, strains of *Mycobacterium tuberculosis* complex are shown in the center (red rods) and are surrounded by various immune cells (B cell, dendritic cells, neutrophils, Giant cells, foam cells, apoptotic macrophages, epithelioid macrophages, natural killer cells, regulatory T cells, and type 1 T helper cells; shown in varying shapes, forms, and colors) to form a granuloma. DC = dendritic cells, Th1 = type 1 T helper, Treg = regulatory T, NK = natural killer (created with BioRender.com).

1.4. Diagnosis of Tuberculosis

Latent TB infection and the active disease differ significantly from each other. People with LTBI are asymptomatic, non-contagious, and unable to transmit the infection [37]. LTBI is diagnosed by measuring the body's immune response to antigens of the MTBC strains [37]. It is often diagnosed with the tuberculin skin test (TST); however, due to the methods' significant flaws, it has been

replaced with the interferon-gamma release assay (IGRA) [2,37,38]. About 5 to 10% of individuals infected develop active TB if untreated [27,28,39]. With an estimated 2 billion people being latently infected globally, there is a significant reservoir for future TB cases and hinders efforts toward TB control and eradication [38].

On the other hand, in active TB disease, individuals typically show signs like unexplained weight loss, loss of appetite, night sweats, fever, lethargy, chills, and protracted cough with or without blood [40]. Diagnosis of active TB involves a combination of clinical suspicion, therapy response, chest radiography, acid-fast bacilli (AFB) staining, sputum smear microscopy, mycobacteria culture, and nucleic acid amplification (NAA) testing [2,38]. Imaging such as chest X-ray aids in identifying abnormalities [2], while sputum samples are essential for microscopy and culture [30], despite limitations in sensitivity [2,38,41]. Mycobacterial cultures, with different media types, facilitate speciation and drug susceptibility testing [2,38,41].

Nucleic acid amplification assays offer rapid and specific alternatives to AFB smears and culture for MTBC identification. Resource-poor settings have a diagnostic challenge due to the limitations in available tests [39] and the upfront cost associated with the molecular test, and they end up relying on smear microscopy. This leaves a sizable proportion of cases undiscovered [42]. Even though smear-positive cases are the most contagious, missing the smear-negative cases, which account for roughly half of all cases, increases the morbidity and mortality of the illness in those patients [27,28]. This diagnostic strategy is particularly troublesome because the areas most affected by TB are also the areas most prevalent with HIV infection, which increases the risk of smear-negative TB, especially in those with advanced immunosuppression [38].

1.5. Drug Resistance

DR in TB refers to the ability of MTBC strains to grow in the presence of antibiotics frequently used for TB treatment [43]. In clinical MTBC strains, drug resistance is mainly caused by mutations or small insertions/deletions (indels) in genes coding for drug targets or converting enzymes [44]. Due to the global emergence of DR-MTBC strains, antibiotics, and other antimicrobial medicines

become ineffective, and infections become difficult or impossible to treat, resulting in a persistent infection that could be fatal [45,46]. Poor patient management, non-adherence to the recommended regimen, substandard national programs, or a combination of these are the primary causes of DR-TB [47]. Transmission is also a significant contributing factor to DR acquisition. Studies have shown a high prevalence of new MDR-TB cases among individuals who have never been treated for TB [48,49].

The WHO defines different types of DR TB [50]: INH mono-resistant (Hr), RR, MDR-TB, pre-extensively drug-resistant (pre-XDR, defined as MDR-TB plus resistance to any fluoroquinolone), and extensively drug-resistant (XDR) TB defined as pre-XDR-TB plus at least one other WHO group A drug (at the moment bedaquiline (BDQ) or linezolid) [51].

Generally, DR-TB can be diagnosed by molecular or phenotypic drug susceptibility testing assay (mDST/pDST) after confirmation of TB disease [52]. Classical pDST methods or mDST assays such as Xpert® MTB/RIF, line probe assays, sequencing technologies, or a combination of the different techniques can be used [45].

It can take weeks to get results from pDST procedures, which use cultures of MTBC inoculated with antibiotics [53]. However, without proper and timely treatment, a patient can spread the bacteria to the community [54]. Rapid drug resistance detection is essential for TB control efforts, allowing for the delivery of the proper treatment and reducing the spread of the MDR-TB strain [55]. Since most RR isolates are also INH-resistant, detecting RIF resistance may be used to discover MDR-TB [56–58]. Resistance to RIF indicates the need for second-line therapy. Currently, it is possible to quickly identify RIF resistance using genotypic methods based on detecting mutations in the hot spot region of the *rpoB* gene, which is linked to most RR MTBC strains [59,60]. Rapid detection of drug-resistant mutations from smear-positive respiratory or culture specimens is possible by combining several tests that detect genetic mutations (for example, Xpert® MTB/RIF, line probe assays, and molecular beacons) with polymerase chain reaction (PCR) or related technologies [57,60–63].

One WHO-recommended mDST assay is the Xpert® MTB/RIF assay. The assay can detect MTBC and RR strains simultaneously using real-time PCR technology. Molecular beacon technology and

Hemi-nested PCR are the basis of the Xpert® MTB/RIF [64,65]. In smear-positive cases, the Xpert® MTB/RIF has a sensitivity of 98% in RR strain detection [66] and 70% in smear-negative samples [67]. It is fully automated, which makes it simple to use with lower biohazard risk [68,69]. It has a quick turnaround time of 90 minutes from clinical specimen [70,71].

Line probe assays are alternative mDST tests that employ PCR and reverse hybridization with particular oligonucleotide probes bound to membrane strips [63]. These assays may be utilized to quickly identify *rpoB* gene mutations or detect and identify mycobacterial species [63,72,73]. Both the GenoType *Mycobacterium* (Hain Diagnostika, Nehren, Germany) and INNO-LiPA MYCOBACTERIA v2 (Innogenetics, Ghent, Belgium) are very sensitive line probe assays for the simultaneous detection and identification of mycobacteria [73]. However, all of these classical assays can cover only a limited number of targets, which is especially problematic for drugs used in current MDR-TB treatment regimens such as BDQ, or clofazimine (CFZ) [74,75]. As such, a more comprehensive method is needed such as whole genome sequencing (WGS).

Information on the nearly entire genome of a MTBC strain can be reached by WGS [76]. The characterization of the pathogens [77], drug susceptibility testing (DST) [78] genotyping [79,80], and epidemiological research on TB [81] have all been accomplished with success using this method [82]. WGS of mycobacteria can provide diagnostic and transmission information faster than routine methods (mostly involving culture) [77,83]. In a comparison study in eight laboratories across Europe and North America, WGS generated results earlier (median of 9 days), on average 21 days sooner, than the routine mycobacterial diagnostics workflow [83]. The overall cost can be further reduced by increasing the sample counts with multiplexing during the same sequence run [83]. Votintseva *et al.* (2017) published a study on using WGS for same-day diagnosis and drug resistance prediction directly from clinical specimens, demonstrating that MTBC could be identified in every DNA extract from 39 out of 40 smear-positive respiratory samples (95% to species level) [77]. However, another publication raised the issue of the possibility that human genome contamination could obscure the prediction of drug resistance-related mutation when a clinical sample is directly sequenced [84]. Although there are potential barriers to its use, such as cost, technical requirements, and standardization [85], the availability of user-friendly open-source

databases and pipelines (such as MTBseq [86], PhyResSE [87], TBProfiler [44], and Mykrobe Predictor [88]) to analyze the genomic data is likely to translate WGS from research to clinical diagnostic laboratory. The WHO recently released its catalog of resistance mutations for DR prediction, which can be used for DR prediction using WGS technology [59,89]. This catalog of DR mutations will be helpful for worldwide DR predictions and create a more uniform reference database.

Another option for detecting DR-TB is targeted next-generation sequencing (tNGS). In tNGS, specific regions or gene targets where resistance-conferring mutations occur on the MTBC genome are amplified and sequenced; as such, it has a higher depth of coverage and specificity for SNPs with mutational confidence [90,91]. With tNGS, the problem of inadequate DNA quantity from sputum and host or non-TB DNA contamination is also eliminated [90,92]. It is also less data-intensive and requires less storage [93]. Regarding feasibility in a low-resource setting, de Araujo *et al*. (2023), in their recent publication, outlined the implementation of tNGS in Namibia, an upper-middle-income country [91].

1.6. Treatment of Tuberculosis

The standard treatment of new patients with drug-susceptible TB is a 6-month regimen consisting of the first two months initiation phase with INH, RIF, PZA, and EMB and then four additional months continuation phase with INH and RIF, only [94]. This regimen has been used since 2010 and is recommended by the WHO, with a success rate of about 85% [95]. Despite the safety and efficacy of the 6-month standard regimen, many patients still find it challenging to comply due to its lengthiness [95]. Hence, the WHO has been searching for another effective but shorter regimen [94]. In a phase 3 randomized control trial in over 13 countries, two 4-month treatment regimens were tested, and one consisting of INH, rifapentine, moxifloxacin (MOX), and PZA was shown to be non-inferior to the 6-month standard regimen [96]. This shorter treatment course is now conditionally recommended for drug-susceptible pulmonary TB in individuals 12 years and above by the WHO [94].

The WHO has recommended various regimens for treating DR-TB based on the type of resistance [95]. Six months of RIF, EMB, PZA, and levofloxacin (LFX) were recommended to treat Hr-TB patients [75]. For MDR/RR-TB patients, there are several regimens that can be used based on various factors such as drug-resistance profile, prior exposure to TB medications and history, patient age, and extent of TB disease [97]. The treatment options for MDR-TB, as recommended by the WHO, are shown in Table 1. The all-oral-6-month treatment BDQ, pretomanid, linezolid, MOX (BPaLM) is the preferred regimen for MDR-TB patients, according to the WHO [50,97–99].

Table 1. Regimen options for treatment of patients with multidrug-resistant or rifampicin-resistant tuberculosis

Regimen	MDR/RR-TB fluoroquinolone susceptible	Pre-XDR-TB	XDR-TB	Extensive pulmonary TB	Age <14 years
6-month BPaLM/ BPaL	Yes (BPaLM)	Yes (BPaL)	No	Yes	No
9-month all-oral	Yes	No	No	No	Yes
Longer individualized 18-month	Yes*/ No	Yes*/No	Yes	Yes	Yes

BPaL: bedaquiline, pretomanid and linezolid, BPaLM: bedaquiline, pretonamid, linezolid and moxifloxacin, MDR/RR-TB: multidrug or rifampicin-resistant tuberculosis, XDR-TB: extensively drug-resistant tuberculosis (modified from [97]).

*When 6-month BPaLM/ BPaLM and 9-month regimens could not be used.

The 9-month all-oral regimen includes a combination of these antibiotics: bedaquiline, fluoroquinolones (levofloxacin or moxifloxacin), ethionamide or linezolid, ethambutol, isoniazid, pyrazinamide and clofazimine.

1.7. Global Population structure of *Mycobacterium tuberculosis* complex strains

MTBC strains are 99.9% similar on the nucleotide level of their genomes and have an identical 16S rRNA [100]. However, they show significant genomic diversity and can be split into seven species with phylogeographical population structure and several lineages [101,102]. MTBC strains have been classified into nine human-adapted lineages (L), L1 to L9, and several animal-adapted lineages, such as *M. bovis* and *M. caprae* (Figure 2) [103,104]. Recent studies demonstrated that MTBC strains

of the different lineages vary in phenotypical and pathobiological characteristics and their geographical distribution [105].

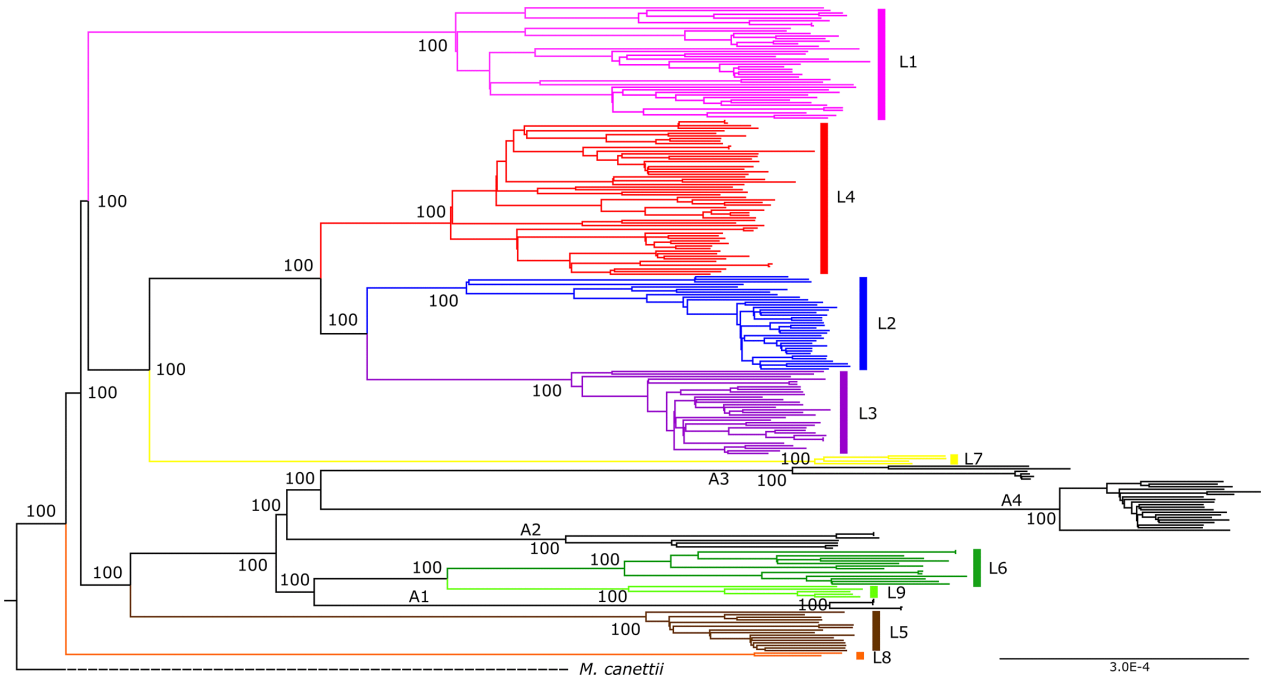


Figure 2. Maximum likelihood phylogeny of lineages of *Mycobacterium tuberculosis* complex strains. The tree was constructed based on 249 *Mycobacterium tuberculosis* complex strains rooted on *Mycobacterium canettii* as the outgroup, with support bootstrap values indicated on the nodes. The genomes were reference strains Ngabonziza *et al*. (2020) used [103]. The strains of the human-adapted lineages are color-coded from L1 to L9, and the strains of the animal-related clades are indicated with A1 to A4. The scale bar shows the number of nucleotide substitutions per site with support bootstrap values displayed on the deepest nodes (adopted from Coscolla *et al.* 2021 [106]).

MTBC strains of different lineages have been proposed to differ in geographical distribution, where generalist groups (found worldwide) and specialist groups (located in specific regions) exist [107]. Strains of lineages 2 and 4 are found worldwide and, therefore, have been suggested to be the most successful on a global level [107]. The population structure of the strains of L4 is complex, with about ten sublineage classifications having different geographical distributions, risk factors, and disease outcomes [107,108]. For example, strains of L4.1.2/Haarlem and L4.3/LAM can be

defined as generalists, while strains of L4.1.3/Ghana and L4.6.2/Cameroon are specialists proposed to be found only in specific regions of Africa and Asia [107]. Studies have suggested that strains of L4 may have a higher transmission rate than strains of other MTBC lineages [102]. Still, a lot of gaps exist in our knowledge of the pathogenesis and evolution of strains of L4, and more research is needed to fully understand the epidemiology and clinical impact of this group of MTBC strains.

Strains of the Cameroon (4.6.2.2) sublineage form a distinct group found in West and Central African countries [109–111]. They are characterized by unique genetic mutations, which set them apart from other strains of MTBC [107]. Studies have shown that the strains of the Cameroon sublineage are highly transmissible, making its recognition and understanding essential for effective TB control and management in regions where it is prevalent [109,111].

1.8. Genotyping *Mycobacterium tuberculosis* complex strains

Genotyping is a crucial tool in understanding the unique genetic makeup of organisms, including MTBC strains. Genotyping MTBC strains can provide valuable information on various aspects of the bacterium, including epidemiology, diagnosis, drug resistance, treatment options, and evolution [82,83,112]. Molecular typing tools have successfully been used for genotyping of MTBC strains in epidemiologic studies since the 1990s, leading e.g., to the discovery of population-level risk factors for transmission and the efficacy of control measures [113]. The acquired knowledge has been crucial for improving public health measures as it enabled the creation of individualized public health interventions and assessment of the efficacy of control measures [114]. The different genotyping methods have played various roles in the phylogenetic analysis and classification of MTBC strains [102,115]. It has also provided knowledge of the evolution and history of MTBC strains. For example, the discovery of *M. canetti* as the ancestor of present-day MTBC strains that have adapted to humans [116].

Genetic markers have been utilized in both molecular epidemiology and phylogeny investigations to study the transmission and population structure of MTBC strains. To be effective, these

markers must have high discriminatory power, be phylogenetically robust with low homoplasy index and convergent evolution and have a robust phylogenetic framework for the definition of biologically significant groupings [117]. The ideal DNA marker for studying the transmission dynamics of MTBC strains should be stable and polymorphic [105,117,118]. It should be easy to identify genetic polymorphisms, economical, and able to produce results quickly in a format easily transferrable between laboratories [119]. Classical genotyping methods used for studying the transmission of MTBC strains include IS*6110*-DNA-fingerprinting, spoligotyping, and Mycobacterial Interspersed Repetitive Units Variable Number Repeat (MIRU-VNTR) typing [118,120,121].

The most recent development in molecular epidemiology is the development of high-throughput, reasonably priced WGS technology [117]. While IS*6110*-RFLP, spoligotyping, and MIRU-VNTR typing only interrogate a small fraction of the genome of MTBC strains, WGS-based approaches are based on interrogating virtually the complete sequence of an organism's genome [122]. Niemann *et al.* (2009) were the first to use NGS for WGS of MTBC strains with identical classical typing data [123]. They found that the two Beijing strains investigated (one sensitive to INH, RIF, EMB, PZA, and SM and the other resistant to all five drugs) had indeed 130 SNPs and a large deletion, indicating that they were not closely linked epidemiologically [123]. In a study evaluating the potential of WGS for outbreak-related transmission chains by Roetzer *et al.* (2013), they discovered that NGS-based analysis could provide high-resolution discrimination of the outbreak strains and provided a first estimation of the variation of the genomes of MTBC strains in human-to-human recent transmission chains [124]. Walker *et al*. (2013) used WGS to identify super spreaders in MTBC strains in the United Kingdom (UK) midlands and predict the existence of undiagnosed cases [125].

WGS based on NGS is now accepted as a powerful tool for molecular epidemiological and phylogenetic studies of MTBC strains and simultaneously allows for the identification of virulence or resistance markers [126,127]. WGS provides a higher discriminatory power and, thus, has replaced classical techniques such as IS*6110*-DNA-fingerprinting, spoligotyping, and MIRU-VNTR typing [128,129]. There are still some limitations to WGS, such as the need for specialized software, the need for specialist knowledge of genome sequencing, and a limited understanding of the molecular clocks of SNPs and LSPs [130,131]. Despite these challenges, WGS has already made significant

contributions to TB diagnosis, treatment, and public health management and is expected to play an increasingly important role in the future [119,132].

1.9. Thesis objectives

Studies on the transmission, drug resistance patterns, evolution, and phylogeny of MTBC strains in Africa are sparse. To fill this knowledge gap, this dissertation aims to gain more knowledge of the properties and transmission of MTBC strains in the West African region including lineage, and sublineage classification, and correlations between phylogeny and geographical origin.

Specific objectives are:

- Investigate the determinants of the DR TB epidemiology in Sierra Leone with a focus on the prevalence of resistance and resistance mutations, and the lineage diversity and transmission rates of MTBC strains circulating in the country.
- Define the population structure and evolutionary history of MTBC strains of the Cameroon lineage in correlation with geographical occurrence and drug resistance patterns.

2. Methods

The research work in this dissertation was a combination of qualitative and quantitative methods. Both experimental and longitudinal designs were utilized, and the convenient sampling method was used to obtain the samples.

2.10. Study design and population

Sputum samples from the National TB Reference Laboratory Freetown, Sierra Leone, of individuals with RR-TB as determined by Xpert MTB/Rif and those with suspected RR-TB who had failed first-line treatment in Sierra Leone between November 2016 to March 2020 were sent to the Supranational Reference Laboratory (SRL) in Borstel for mycobacterial culture. All RR isolates based on genotypic drug susceptibility tests (gDST) were included in this analysis.

2.11. Whole Genome Sequencing and Analysis

Isolation and cultivation of mycobacterial strains were conducted at the SRL in Borstel as previously described [133]. DNA was extracted as previously described [134]. The genomes were sequenced and analyzed with bioinformatics pipelines as described in Grobbel *et al.* (2021) [135] and Merker *et al.* (2018) [136]. Briefly, genomic DNA was sequenced using Illumina Technology (NextSeq 500 or MiSeq) using the Baym protocol [137] and Nextera XT library preparation kits following the manufacturer's instructions (Illumina, San Diego, CA). The FASTQ data were analyzed with the MTBseq pipeline [86]. The reads were mapped to the H37Rv genome (GenBank ID: NC_000962.3 [138]) with BWA [139], processed with Samtools [140], and mappings were refined with the GATK 3.8 [141].

For variant detection in mapped reads, Samtools [140] derived mpileup files were filtered for minimum thresholds of at least two reads indicating a variant in both forward and reverse orientation, two reads calling the allele with at least a phred score of 4, and 5% allele frequency

for resistance determination, and for phylogeny four reads mapped in each forward and reverse direction respectively, with 75% allele frequency and at least four calls with a phred score of at least 20.

Genomic SNP positions with a reliable base call in at least 95% of the strains and covered in all strains were concatenated to a sequence alignment. SNPs within a window of 12bp from each other and those located in repetitive regions or resistance-associated genes were excluded to avoid calling SNPs related to insertions and deletions artefacts [136]. Strains were phylogenetically classified, and transmission clusters were inferred with single linkage clustering using distance cutoffs of ≤5 and ≤12 SNPs. Raw FASTQs were uploaded to the European Nucleotide Archive (ENA) (Appendix Table B.1).

2.12. Resistance analysis

Genotypic resistance predictions were based on a curated Research Center Borstel mutation catalog, as described by Grobbel *et al.* (2021) and WHOs' catalog [59,135]. Mutations linked to phenotypic drug resistance were marked as resistant, and unclear ones were classified as not resistant. When no mutation was detected, strains were defined as susceptible. Drug-resistant types were classified based on WHO classification [45].

2.13. Phylogenomic analysis

A maximum parsimony tree (MPT) was built with the software BioNumerics version 7.6.3 (Biomerieux) from the aligned sequences of concatenated SNPs. Numbers on branches indicate the number of distinct SNP positions between isolates. Using distinct SNP-sites, maximum-likelihood trees (MLT) were calculated with IQ-TREE 2, an efficient method for phylogenetic inference [142], using ModelFinder [143], an automated model selection and ultrafast approximation for phylogenetic bootstrap [144] with resampling of 1000. The MLTs were midpoint rooted with the FigTree software version 1.4.4. The trees were visualized with Interactive Tree of Life (iTOL) 5.7 [145].

2.14. Statistical analysis

Descriptive statistics (distribution frequency) were performed, and graphs were drawn with R software version 4.2.1.

2.15. DNA database screening of strains of the Cameroon sublineage

Samples from 2 cohorts were sourced. Samples from the over 14,000 sequenced isolates at the RCB were screened for strains classified as Cameroon sublineage and strains typed by MIRU-VNTR. The sequence read archive (SRA) was also screened for strains classified as belonging to the Cameroon sublineage of MTBC using SNPs from in *alaS, ethA and ethR* genes.

2.16. Phylogeny of the strains of the Cameroon sublineage

From the concatenated sequence alignments, strains were grouped based on 12 SNPs and 5 SNPs from one another. From the groupings, one strain each from a group within 5 SNPs from one another was selected for the characterization of the sublineage. Strains were then grouped based on their UN origin, RhierBAPS levels 1 and 2 [146]. RhierBAPS is an algorithm implemented in R that allows for hierarchically clustering of DNA sequences to reveal existing nested population structure [146]. The specifications for the groups were as follows: each group should have at least ten strains, then SNPs were selected based on mutations in essential genes and epitopes, with a frequency of one hundred percent (98-100%) and the mutations should be synonymous to ensure these were not under-selective pressure. The procedure for the phylogeny determination is described in the results section.

3. Results

3.1. Molecular determinants of RR/MDR-TB in Sierra Leone

The African continent bears a substantial burden of TB, accounting for 23% of global cases [8]. Amid this, drug resistance poses a significant threat to TB control [147]. This section focuses on Sierra Leone, a West African country with a high TB burden, aiming to provide valuable insights into the local challenges of DR-TB. By delving into the epidemiology, phylogeny, drug resistance patterns, and transmission dynamics of RR/MDR-TB strains in Sierra Leone, this analysis contributes to understanding the country's unique circumstances and offers essential information for targeted interventions and future research endeavors.

3.1.1. Study Demography

A total of 238 RR/MDR MTBC strains were included in the study from 2016 to 2020 (Table 2); 63% (n= 151) were from retreatment cases. Most strains were obtained from men (43%, n= 102), 17% (n= 41) were from females and 40% (n= 95) were unknown. The study population had a median age of 34 years (interquartile range 30-39) with an age range, from minors below 19 to seniors above 60 years. Among those with known HIV status, 15% (25/162) were HIV positive (Table 2).

Table 2. Main demographic characteristics of study participants

		n (RR/MDR)	Percentage (%)
Total		**238**	
Sex	Female	41	17
	Male	102	43
	Unknown	95	40
Age range	≤19	13	5.5
	20-29	50	21
	30-39	48	20.2
	40-49	37	15.5
	50-59	20	8.4
	≥ 60	8	3.4
	Unknown	62	26
Type	New	10	4.2
	Retreatment	151	63.4
	Unknown	77	32.4
HIV status	Negative	137	57.6
	Positive	25	10.5
	Unknown	76	31.9

HIV = human immunodeficiency virus, MDR=multidrug resistant, RR= rifampicin-resistant, n= number of individuals, percentages were calculated using a total of 238.

3.1.2. The population structure of *Mycobacterium tuberculosis* complex strains

WGS data analysis revealed 15,089 informative SNPs differentiating the 238 MTBC strains, which were used to calculate a maximum likelihood phylogeny based on a concatenated SNP alignment (Figure 3). Based on canonical SNP signatures [148], the 238 strains were classified into six main MTBC lineages (L 1-6) (Figure 3, Table 4). The strains of L1 comprised 4% (n= 10) and, L2 9% (n= 22) of the total number of strains. Strains of L3 and L5 comprised the lowest proportion of 0.8% (n= 2) and 2.9% (n= 7) respectively. Strains of L4 were most prominent accounting for 62% (n= 147) followed by 21% L6 strains (*Maf,* n= 50). The strains were further categorized into 23 sub-lineages as follows: 4.1.2.1 Haarlem (n= 45, 19%), 4.1 Euro-American (n= 38, 16%), 6.2.2 West

Africa 2 and 4.8 mainly T (n= 23, 9.7%), and 2.2.1 Beijing Ancestral 3 (n= 22, 9.2%), 6.3.3 West Africa 2 (n= 17, 7.1%), 4.3.3 LAM (n= 14, 5.9%), 1.1.1 EAI (n= 10, 4.2%), 4.3.4.2 LAM (n= 7, 2.9%), 4.4.1.1 S-type (n= 6, 2.5%), 6.2.3 West Africa 2 (n= 5, 2.1%), 4.1.1.1 X-type, 4.1.1.3 X-type, 4.6.2.2 Cameroon, 4.9 H37Rv-like, 5 West Africa 1, 6.3.1 West Africa 2 (n= 3, 1.3%), 3 Delhi-CAS (n= 2, 0.8%), 4.3.4.1 LAM (n= 2, 0.8%), 5.1.5 West Africa 2 (n= 2, 0.8%),, 6.1.3 West Africa 2 (n= 2, 0.8%), 5.1.3 West Africa 1 (n= 1, 0.4%), and 5.3 West Africa 1 (n= 1, 0.4%) (Figure 3, Appendix Table B.1).

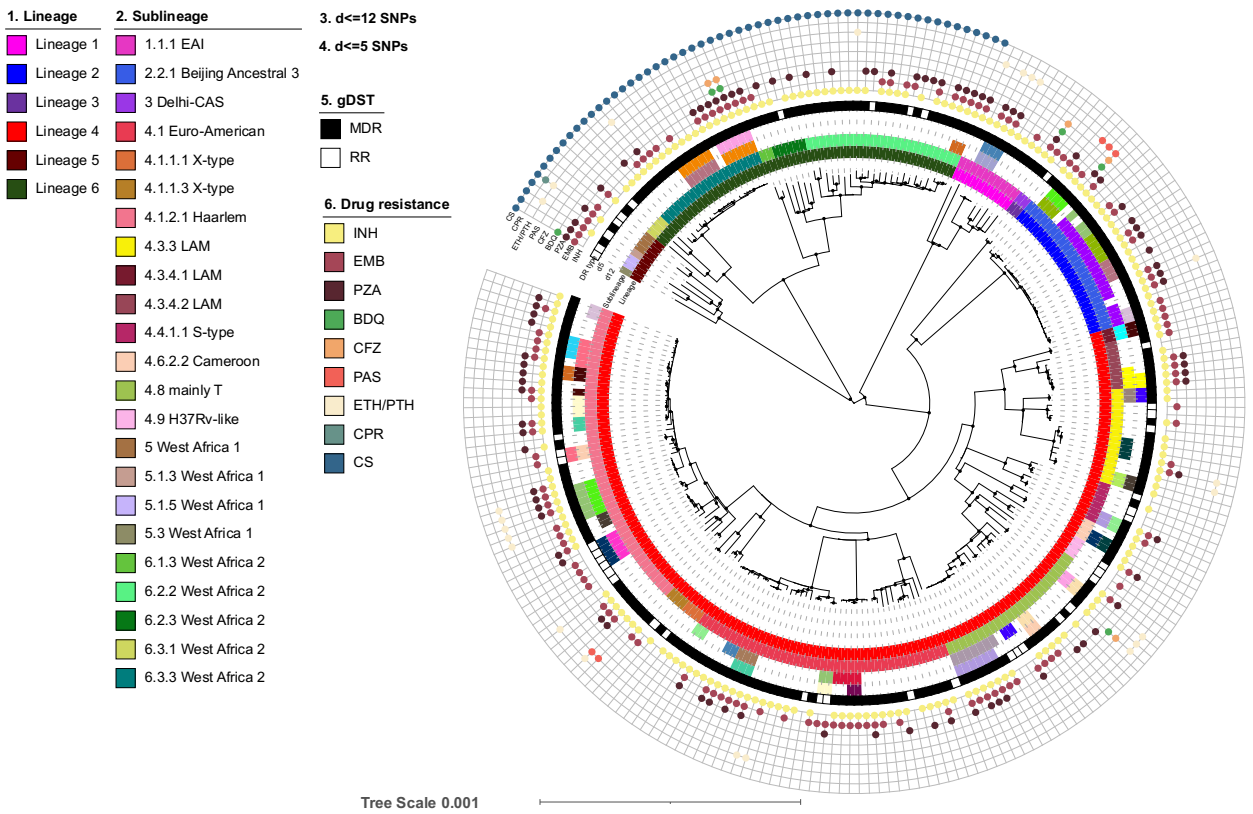


Figure 3. Phylogeny, lineage classification and drug resistance patterns of the *Mycobacterium tuberculosis* complex strains from Sierra Leone. The maximum likelihood tree was based on a concatenated single nucleotide polymorphism (SNP) alignment of 238 *Mycobacterium tuberculosis* complex (MTBC) strains. The alignment was based on 15,089 informative SNP sites. Ultrafast bootstrap values (+0.95) are indicated on the tree branches with black dots, and the circles from the inner ring to the outer show: 1. the six main MTBC lineages; 2. the MTBC sublineages; 3. and 4. show the genome clusters based on strain SNP distances of <=12 SNPs and <=5 SNPs, respectively (strains sharing the same color belong to the same cluster), 5. genotypic drug susceptibility types (gDST); 6. the colored dots indicate genotypic resistance to the respective drug. INH= isoniazid, EMB= ethambutol, PZA= pyrazinamide, PAS= para-aminosalicylic acid, CS= cycloserine, ETH= ethionamide, PTH= proteonamid, CFZ= clofazimine, BDQ= bedaquiline, CPR= capreomycin, MDR = multidrug resistance (resistance to at least isoniazid and rifampicin), RR = rifampicin resistance.

3.1.3. Drug resistance

A genotypic resistance prediction was performed based on high-confidence resistance mutations. Overall, 82% of the strains were MDR (n= 196) (Table 3), 53% (n= 126) were resistant to EMB and 39% (n= 92) to PZA (Figure 4, Appendix Table B.2 & B.3). No resistance to FQ and LIN was detected; however, five strains were resistant to BDQ/CFZ based on mutations in *Rv0678* (Table 3). A total of 61 (26%) strains were found to be resistant to all first-line drugs; three had additional resistance to BDQ/CFZ and 2 strains were also resistant to the first-line drugs with additional resistance to BDQ and streptomycin (Table 3).

Table 3. Proportion of Resistances in Antituberculosis Drug Combinations Among *Mycobacterium tuberculosis* complex Strains Grouped by Lineages

	Total (%)	**L1 (%)**	**L2 (%)**	**L3 (%)**	**L4 (%)**	**L5 (%)**	**L6 (%)**
RIF+INH	196(82.4)	10(5.1)	22(11.2)	1(0.5)	116(59.1)	2(1)	45(23)
First-line drugs	61(25.6)	5(8.2)	8(13.1)	0	31(50.8)	2(3.2)	15(24.6)
BDQ+CFZ	5(2.1)	1(20)	1(20)	0	1(20)	0	2(40)
RIF+INH+BDQ +CFZ	5(2)	1(20)	1(20)	0	1(20)	0	2(40)
First-line + BDQ + CFZ	3(1.3)	1(33.3)	0.0	0	0	0	2(66.7)
First-line + BDQ + SM	2(0.8)	0	0.0	0	0	0	2(100)

First-line drugs = rifampicin, isoniazid, ethambutol, and pyrazinamide, L= lineage, BDQ= bedaquiline, CFZ= clofazimine, SM = streptomycin, (percentages for total were based on all 238 strains, while that for lineages were from horizontal total)

Resistance patterns across strains of different lineages of MTBC in the country are visualized in Figure 4. MDR-TB was prevalent among strains of L1 and L2, with all strains being MDR (n= 10 for L1 and n= 22 for L2). Approximately 50% (n= 1) of L3 strains were MDR, while 79% (n= 116) of L4 strains, 29% (n= 2) of L5 strains, and 90% (n= 45) of L6 were MDR.

When resistance to EMB was assessed, varying degrees of resistance were observed among different lineages. For instance, in L1, 60% (n= 6) of the strains displayed resistance to EMB. Similarly, L2 exhibited a high resistance rate, with 59% (n= 13) of strains showing resistance to

EMB. The majority of L4 strains, 52% (n= 76), exhibited resistance to EMB. Furthermore, L5 strains showed a high resistance of 86% (n= 6) to EMB. And in L6, 46% (n= 23) of strains displayed resistance to EMB (Figure 4).

Resistance to PZA was observed in at least 50% of all the strains of L1 to L6, except strains of L3, which showed no resistance to PZA. Only one strain from L5 was resistant to CPR. Among the five strains resistant to BDQ/clofazimine (CFZ), one belonged to L1, one to L2, one to L4, and two to L6. Ethionamide/Prothionamide (ETH/PTH) resistance was observed in 40% (n= 4) strains of L1 strains, 5% (n= 1) of L2, 10% (n= 14) of L4, 29% (n= 2) of L5, and 4% (n= 2) of L6 strains. All strains of L5 and L6 strains were inherently resistant to cycloserine (CS). Additionally, three strains of L2 and two of L4 showed resistance to para-aminosalicylic acid (PAS) (Figure 4, Appendix Table B.1 & B.4).

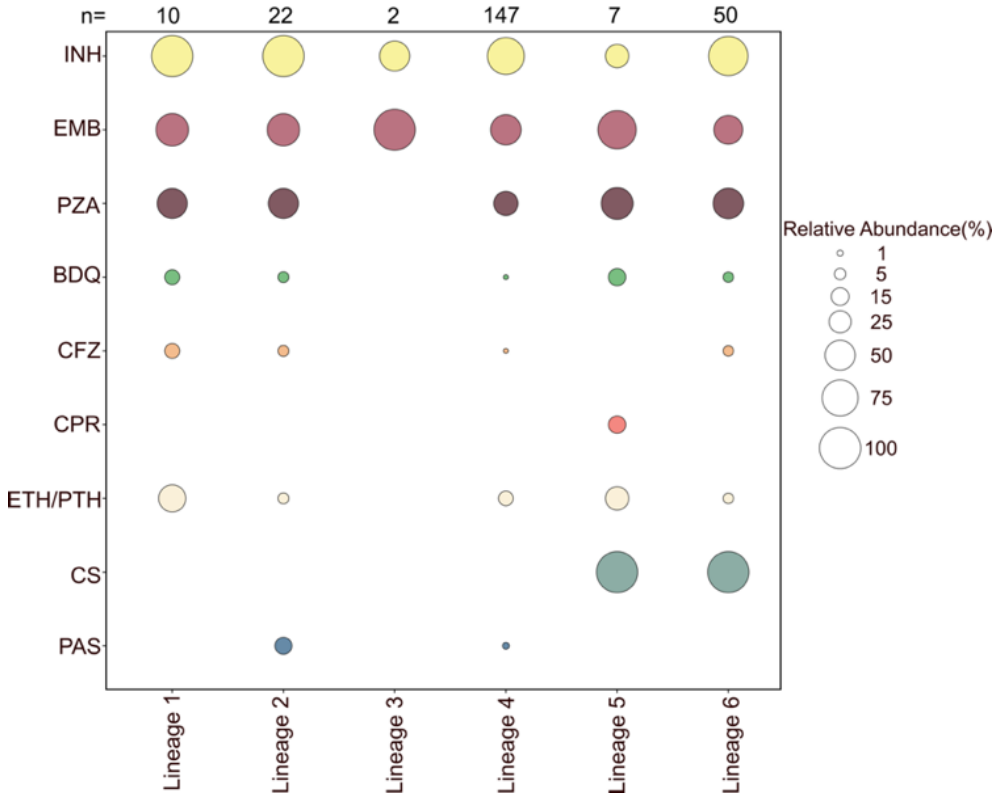


Figure 4. Drug resistance to first and second-line anti-tuberculosis drugs per lineage of *Mycobacterium tuberculosis* complex strains investigated. The bubble plot shows the relative abundance of *Mycobacterium tuberculosis* complex (MTBC) strains for each lineage that developed resistance to the corresponding drug. On the left y-axis are the abbreviated names of drugs: INH= isoniazid, EMB= ethambutol, PZA= pyrazinamide, CS= cycloserine, ETH= ethionamide, PTH= proteonamid, PAS= para-aminosalicylic acid, CFZ= clofazimine, BDQ= bedaquiline and CPR= capreomycin. On the x-axis at the bottom are the MTBC strain lineages and indicated at the top are the total number of MTBC strains per lineage. n = number of isolates, *Mycobacterium tuberculosis* complex =MTBC.

A detailed investigation was carried out on the mutation diversity related to INH and RIF resistance. All the RIF resistance-conferring mutations were found in the *rpoB* gene (Appendix Table B.6 & Figure C.1). The most prevalent mutation was *rpoB* S450L (41%, n= 98), which is part of the WHO group-1 mutations associated with resistance[59]. The *rpoB* S450L mutation was found in strains of all lineages except L2 (Appendix Table B.6). Despite this, there was a diversity of *rpoB* mutations, with 32 (13%) of the strains having one of the so-called RIF-borderline mutations (*rpoB*

D435Y, L452P, H445L/N and L430P) [149,150]. Also, three strains had the *rpoB* I491V & V170F resistance mutations (Appendix Table B.6 & Figure C.1).

Mutations linked with resistance to INH were found in the *katG* gene and the *fabG1- inhA* regulatory region (Appendix Table B.5 & C.2) [151]. The *katG* S315T resistance mutation was most prevalent, found in 106 strains (56%), and in all lineages except strains of L3 and L5. Besides, 53 strains had mutations in *fabG1- inhA/inhA,* out of which 37 acquired a second mutation in *katG* (Appendix Table B.5). Of those, the most prevalent mutation was *fabG1* 15t>c (18 strains, 9.2%), which was classified as WHO group-2 borderline mutation; 12 did not have any additional INH mutations, and four had an additional *katG* S315T mutation (Appendix Table B.5). Of note were the 12 strains with *fabG1* L203L mutations; however, 83% (n= 10) of those strains had also developed the higher level *katG* S315T resistance mutation (Appendix Table B.5).

3.1.4. Strain transmission (Clustering)

Using a maximum distance of 12 SNPs between two strains to delineate clusters, 104 (44%) strains were grouped into 31 clusters ranging in size from 2 to 16 isolates (Figure 5, Table 4). Subsequently, the cluster data was linked with resistance mutation profiles to assess the magnitude of transmission of RR/MDR strains.

Among the 10 strains belonging to L1, 3 strains formed clusters, resulting in a cluster rate of 30%. Within L2, comprising 22 strains, a significant proportion of 19 strains clustered together at a high rate of 86%. Conversely, none of the two strains of L3, formed clusters. In L4, which included 147 strains, 70 clustered, yielding a clustering rate of 48%. Conversely, none of the seven strains of L5 formed clusters. And among the 50 strains of L6, 12 strains clustered together, resulting in a clustering rate of 24% (Table 4).

Table 4. *Mycobacterium tuberculosis* complex strains proportions, clustering, and cluster rate within the lineages

Main lineages	∑ Number. of strains (%)	Clustered strains (n)	Cluster rate (%)
Lineage 1	10(4.2)	3	30
Lineage 2	22(9.2)	19	86
Lineage 3	2(0.8)	0	0
Lineage 4	147(61.8)	70	48
Lineage 5	7(2.9)	0	0
Lineage 6	50(21)	12	24

n = number, % = percentage, ∑ = total

An examination of clustering patterns within the strains based on sublineages was undertaken. Cluster sizes ranged between 2 to 16 strains. All strains of L1 belonged to one sublineage 1.1.1 EAI, and three of those strains formed one cluster, at a rate of 30% (Figure 5). Strains of 4.1 Euro-American formed five clusters, at a rate of 34%. Notably, strains of sublineage 4.1.2.1 Haarlem demonstrated the highest cluster count (n= 9), with a cluster rate of 62% (Figure 5). Furthermore, strains of 4.3.3 LAM formed three clusters at a rate of 50%, while all strains of 4.3.4.1 LAM formed a single cluster. Additionally, strains of 4.3.4.2 LAM and 4.4.1.1 S-type formed one cluster each, with cluster rates of 43% and 33%, respectively. Interestingly, among the three strains of 4.6.2.2 Cameroon, two strains formed one cluster at a rate of 67%. Similarly, strains of 6.2.2 West African 2 formed one cluster (Figure 5).

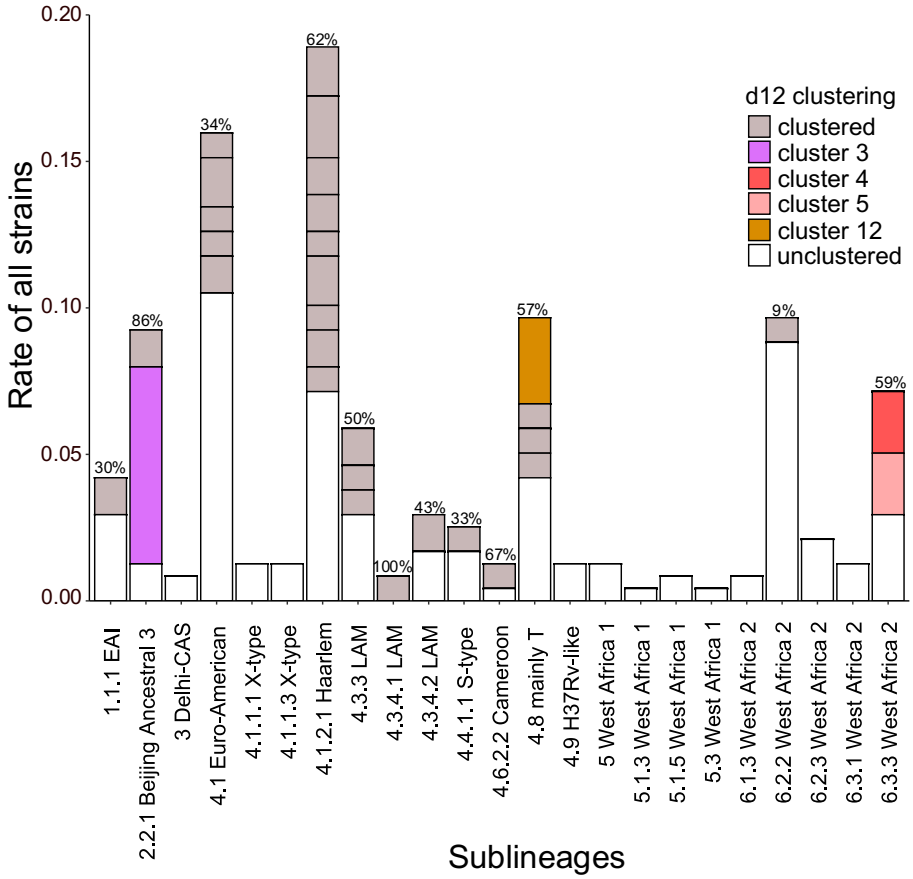
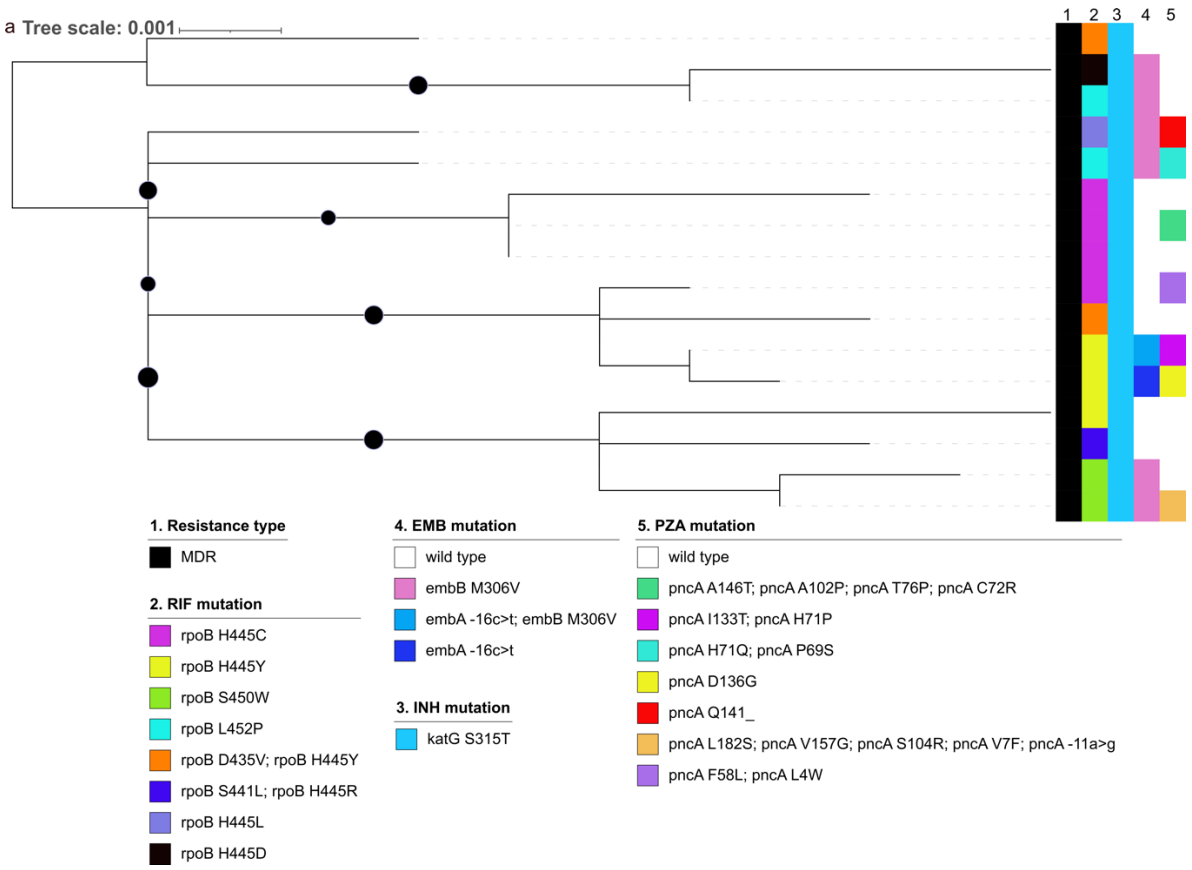


Figure 5. Clustering proportions of *Mycobacterium tuberculosis* complex strains by sublineage. The bar plot shows the distribution of *Mycobacterium tuberculosis* complex strains that clustered versus unclustered based on sublineages. The total cluster rate per sublineage is indicated on the top of the bars and distinct clusters are separated by lines. Large clusters with five or more strains are presented in specific colours, while grey represent clusters with less than five strains and white the proportion of unclustered strains.

The strains of L2 also belonged to one sublineage, 2.2.1 Beijing ancestral 3 which notably exhibited the highest cluster rate of 86%. Two clusters were formed, with one being the largest single cluster comprising 16 (73%) strains (Figure 5). This shows high transmission among the strains of 2.2.1 Beijing Ancestral 3 sublineage. Analysis of the resistance mutation profile within the largest cluster revealed that all strains within this cluster (cluster 3) were MDR and shared a *katG* S315T mutation (Figure 6a). However, variations were observed in *rpoB*, *embB*, and *pncA* resistance mutations. The strains had eight different RIF resistance mutations, with one group of four strains having the *rpoB* H445C resistance mutation and another group of four strains the H445Y

mutation. There were three strains that had two different RIF resistance mutations, two of them had developed both *rpoB* D435V and H445Y and the other *rpoB* S441L and H445R resistance mutations. Showing the transmission of an INH clone with at least seven independent developments to MDR. Meanwhile six of the strains had developed the same EMB *embB* M306V resistance mutation, and one had developed both the *embB* M306V and embA-16>t resistance mutations.

Similarly, the second-largest cluster of strains (cluster 12) formed by seven 4.8 mainly T sublineage strains was also found to be MDR and displayed a distinct resistance mutation pattern (Figure 6b). In contrast to the cluster of strains belonging to 2.2.1 Beijing ancestral 3, all the strains within this cluster shared identical resistance mutations for RIF (*rpoB* S450L), INH (*katG* S315T), and EMB (*embB* M306V), variations were only observed in PZA resistance mutations. This suggests active transmission of this MDR clone in the country (Figure 6b).



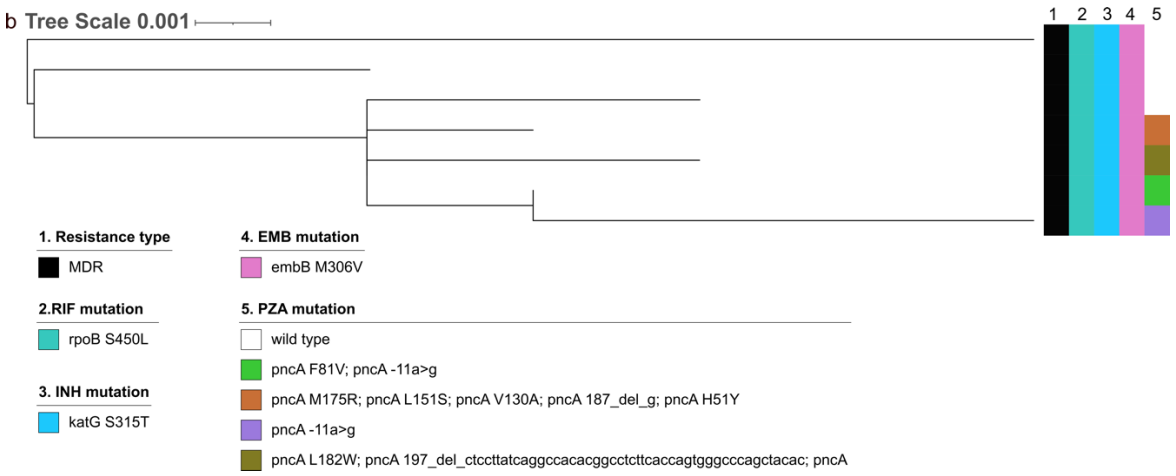


Figure 6. Phylogeny of the strains of the two largest clusters of multidrug-resistant *Mycobacterium tuberculosis* strains. Maximum likelihood trees were calculated for the *Mycobacterium tuberculosis* strains of the two largest multidrug-resistant clusters identified in the study: (a) The midpoint-rooted phylogeny of the 16 clustered *Mycobacterium tuberculosis* strains of "cluster 3" belonging to the sublineage 2.2.1 Beijing Ancestral 3. The alignment was based on 1,174 informative SNP sites. This tree shows the transmission of an isoniazid-resistant clone with multiple independent developments to multidrug resistance (MDR - resistance to at least isoniazid and rifampicin), shown by the different rifampicin resistance mutations. The strains had also developed different ethambutol and pyrazinamide resistance mutations. (b) The midpoint-rooted phylogeny of the seven clustered *Mycobacterium tuberculosis* strains of "cluster 12" belonging to sublineages 4.8 mainly T. The alignment was based on 466 informative SNP sites. In this tree, a single MDR clone, which was also resistant to ethambutol, was transmitted, shown by the same resistance mutation for all seven strains. However, the strains developed different pyrazinamide resistance mutations. 1. The genotypic drug resistance type; 2. to 5. the color-coded resistance mutations of the first-line drugs identified for the respective drugs; and the black dots on the tree branches indicate the ultrafast bootstrap values (+0.95). RIF= rifampicin, INH= isoniazid, EMB= ethambutol, PZA= pyrazinamide, MDR = multidrug resistance.

The data also revealed smaller transmission events of MDR L6 strains (Figure 7). Out of the 50 L6 strains, 12 (24%) clustered (Figure 5, Table 4). Of the 6.3.3 West Africa 2 sublineage strains, 10 out of 17 (59%) were clustered, forming two clusters (cluster 4, and cluster 5) of five strains each (Figure 5). Interestingly, while the cluster analysis separated the strains into two clusters, they were closely related in the phylogeny and shared the RIF *rpoB* D435Y mutation and INH *fabG1* -17g>t/*katG* S315T double mutation, indicating the emergence from a common MDR ancestor (Figure 7). However, only the strains of cluster 5 developed further resistance to EMB and PZA,

rendering them fully first-line resistant (Figure 7). Also, two of the strains developed BDQ/CFZ resistance due to mutations in *Rv0678* (Appendix Table B.1 & B.3).

Concluding this section, the population structure, resistance profiles, and transmission dynamics of RR/MDR MTBC strains in Sierra Leone have been comprehensively examined. These findings lay the groundwork for a deeper exploration of the evolutionary landscape within strains of a specific sublineage. The next section shifts our focus to the strains of the Cameroon sublineage, unraveling its evolutionary trajectory, characteristics, and drug resistance profiles.

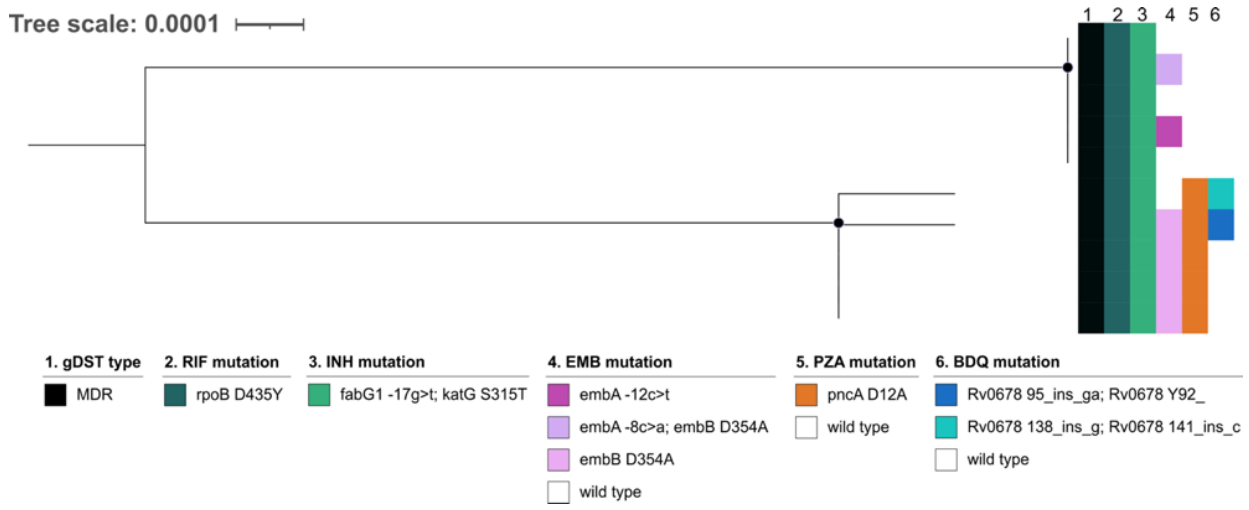


Figure 7. Phylogeny of the largest cluster of endemic multidrug resistant *Mycobacterium africanum* strains. The Mid-point rooted maximum likelihood tree was calculated for the 10 multidrug resistant (MDR, resistance to at least isoniazid and rifampicin) *Mycobacterium africanum* strains belonging to sublineage 6.3.3 West Africa 2. The alignment was based on 1,924 informative SNP sites. The tree shows a single sublineage 6.3.3 West Africa 2 MDR clone transmitting, which formed two distinct clusters (clusters 4 & 5). 1. the genotypic drug susceptibility type (gDST); 2. to 6. are the color-coded resistance mutations of the first-line drugs identified for the respective drugs; and the black dots on the tree branches indicate the ultrafast bootstrap values (+0.95). RIF= rifampicin, INH= isoniazid, EMB= ethambutol, PZA= pyrazinamide, BDQ= bedaquiline, MDR = multidrug resistance.

3.2. Lineage 4: Cameroon sublineage

This section examines the evolutionary dynamics, distinctive phylogenetic characteristics, and drug resistance profiles of MTBC strains belonging to the Cameroon sublineage. These strains are among the L4 strains causing TB and MDR-TB in West Africa [152–155]. The analysis aims to provide unique insights into the genetic intricacies and adaptive features of these strains and shed light on its relevance in the broader context of TB epidemiology. A comprehensive exploration of the strains of the Cameroon sublineage's evolutionary trajectory will provide a deeper understanding of their role in the transmission of TB and MDR-TB and contribute valuable knowledge to the ongoing efforts to combat the disease.

3.2.1. Study design

To select strains of the Cameroon sublineage (4.6.2.2) for downstream analysis, NGS data of over 14,000 inhouse (RCB, Molecular and Experimental Mycobacteriology groups) MTBC strains were screened based on SNPs specific to the sublineage [148]. Overall, 262 isolates were designated to belong to the Cameroon sublineage (Figure 8). In addition, the in-house Bionumerics (Biomerieux) database containing MIRU-VNTR profiles of 33,000 MTBC strains was screened for Cameroon sublineage. Eight hundred and sixty-two Cameroon sublineage strains were obtained, out of which 367 were selected based on the availability of the genomic DNA and maximizing geographical diversity for WGS.

To complete this collection, Cameroon sublineage specific SNPs were used to screen the online platform European Nucleotide Archive (ENA). A total of 564 NGS datasets from Cameroon sublineage strains were obtained; however, seven could not be downloaded, giving an overall total of 1,186 isolates. Duplicates and single-read sequences were excluded, and all remaining sequence datasets were run through the in-house quality check and MTBseq pipeline [86].

A total of 1,133 NGS datasets of the strains of the Cameroon sublineage were finally selected for further investigation in this study (Figure 8). The 1,133 strains were from 25 countries spanning

eight United Nations (UN) regions and 4 continents, including the major geographical areas with reported occurrences of the strains of the Cameroon sublineage (Figure 9).

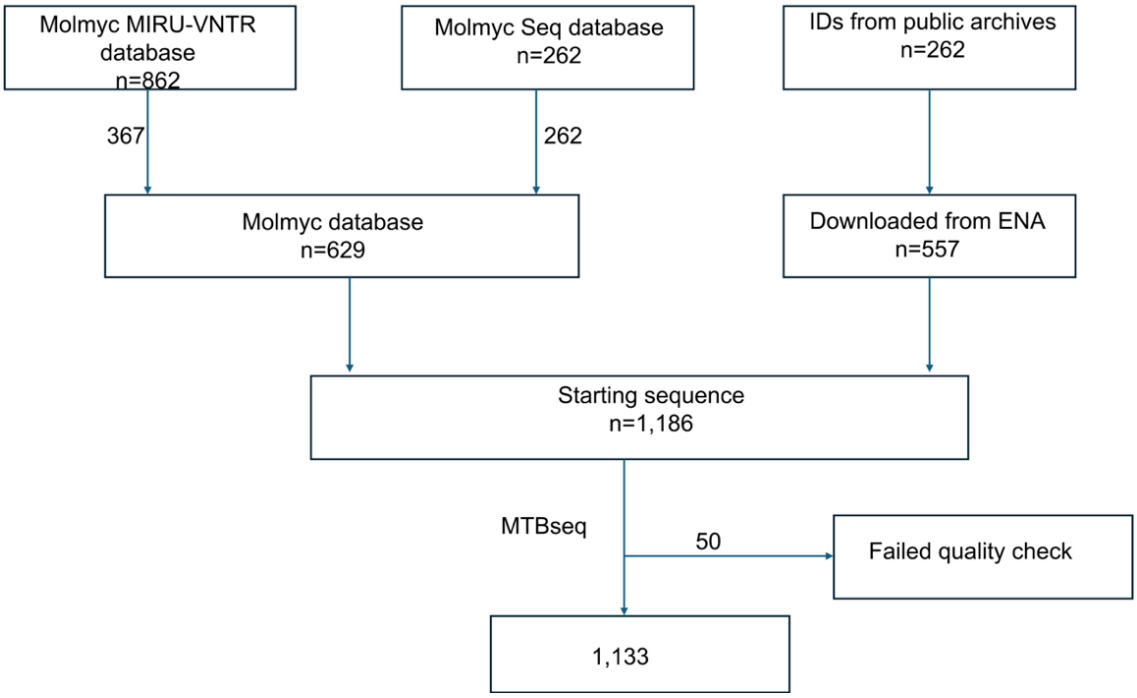


Figure 8. Workflow for selecting strains of Cameroon sublineage of Mycobacterium tuberculosis. This flow chart outlines the data sources and process for selecting the strains of the Cameroon sublineage of *Mycobacterium tuberculosis.* The chart indicates the number of strains identified in the group Molecular and Experimental Mycobacteriology's (Molmyc) Mycobacterial Interspersed Repeat Unit Variable number tandem repeat (MIRU-VNTR) database, Molmyc's sequenced strains database, and the European Nucleotide Archives (ENA). Genomic DNA samples that were available in Molmyc's MIRU-VNTR and sequenced databases were selected and sequenced, while all available sequenced data from the ENA database were downloaded. A total of 1,186 sequenced samples of strains from the Cameroon sublineage were obtained. Subsequently, the data underwent MTBseq analysis, and 50 strains that failed the quality check were excluded from the sample set (Created with BioRender.com). MIRU-VNTR= samples from Mycobacterial Interspersed Repeat Unit Variable Number Tandem Repeat database, Molmyc = Molecular and Experimental Mycobacteriology, Seq = Sequenced, IDs = Identification numbers, ENA = European Nucleotide Archives, and n = number of samples.

3.2.2. Phylogeny of the strains of the Cameroon sublineage

For this high-resolution analysis of the strains of the Cameroon sublineage, the origins, and sub-types, a MLT (Figure 9) was constructed from a concatenated sequence alignment of 1,133 strains based on 9,461 informative SNPs to give an overview of the Cameroon sublineage phylogeny. The continent and country of origin of patients were plotted on the MLT (Figure 9). Overall, the tree shows a high population diversity of Cameroon sublineage strains with several internal branches pointing towards the presence of a substructure within the Cameroon sublineage. Strikingly, a large group of strains originated from the United Kingdom. This was mainly an outbreak clone that was in transmission in the population. Most of the other strains originated from a variety of 16 African countries including: Cameroon, Côte D'Ivoire, Democratic republic of Congo, Gabon, Ghana, Guinea Bissau, Malawi, Mali, Nigeria, Republic of Niger, Senegal, Sierra Leone, Sudan, The Gambia, Togo and Zambia.

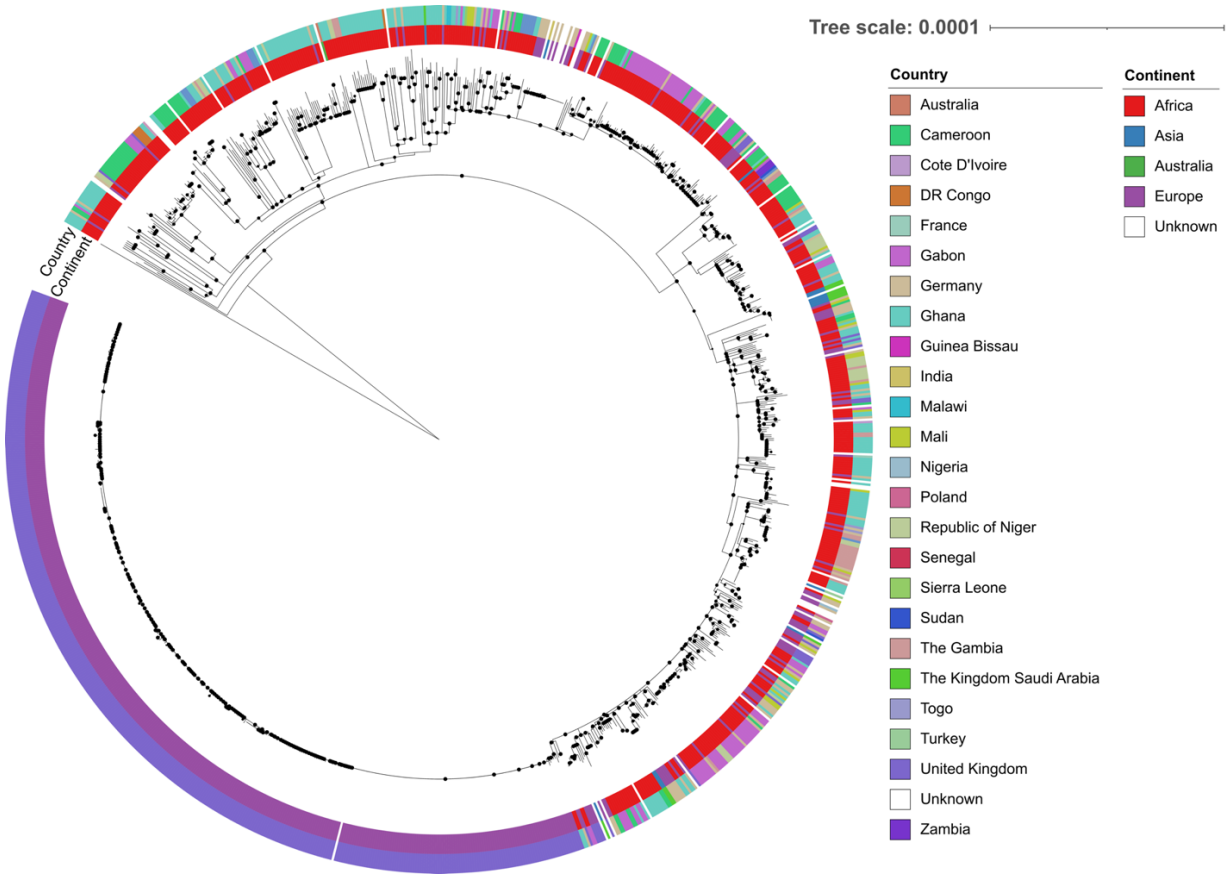


Figure 9. Phylogeny of *Mycobacterium tuberculosis* complex strains belonging to the Cameroon sublineage. The maximum likelihood tree was based on 9,461 informative SNP positions alignment of 1,133 *Mycobacterium tuberculosis* complex strains of the Cameroon sublineage. Each tip of the tree corresponds to a specific strain. Ultrafast bootstrap values (+0.95) are indicated on the tree branches with black dots. The inner ring is color-coded by the country of origin, while the outer ring is color-coded by the continent of origin of the patients.

To further investigate the phylogeny and presence of subgroups/clades within the Cameroon sublineage, first a cluster analysis was performed to enable the selection of single cluster representatives for downstream analysis. For this, a pairwise distances of 0-5 SNPs from strain to strain was used [156]. This resulted in the detection of 104 clusters, ranging in size from two to 422 strains.

For clade definition, one strain from each cluster was used, to exclude potential bias from closely related strains [156]. This resulted in a total of 444 strains that were further interrogated. For this

analysis, a concatenated sequence alignment was generated based on the NGS data of the 444 selected strains using MTBseq [86]. The alignment was based on 8,991 SNPs and was used to construct a MLT (Figure 10). To get an overview of the geographical origins of the 444 strains and to assist in the strain's clade classification, they were categorized according to UN regional designations, revealing seven primary geographical areas (Figure 10). Western and Central African regions had the highest number of strains, 233 (53%) and 107 (24%) respectively. However, there were 59 (13%) strains from Western Europe, 19 (4%) from Northern Europe, seven from Western Asia, four from Eastern African and one from North Asia and Australia (Table 5).

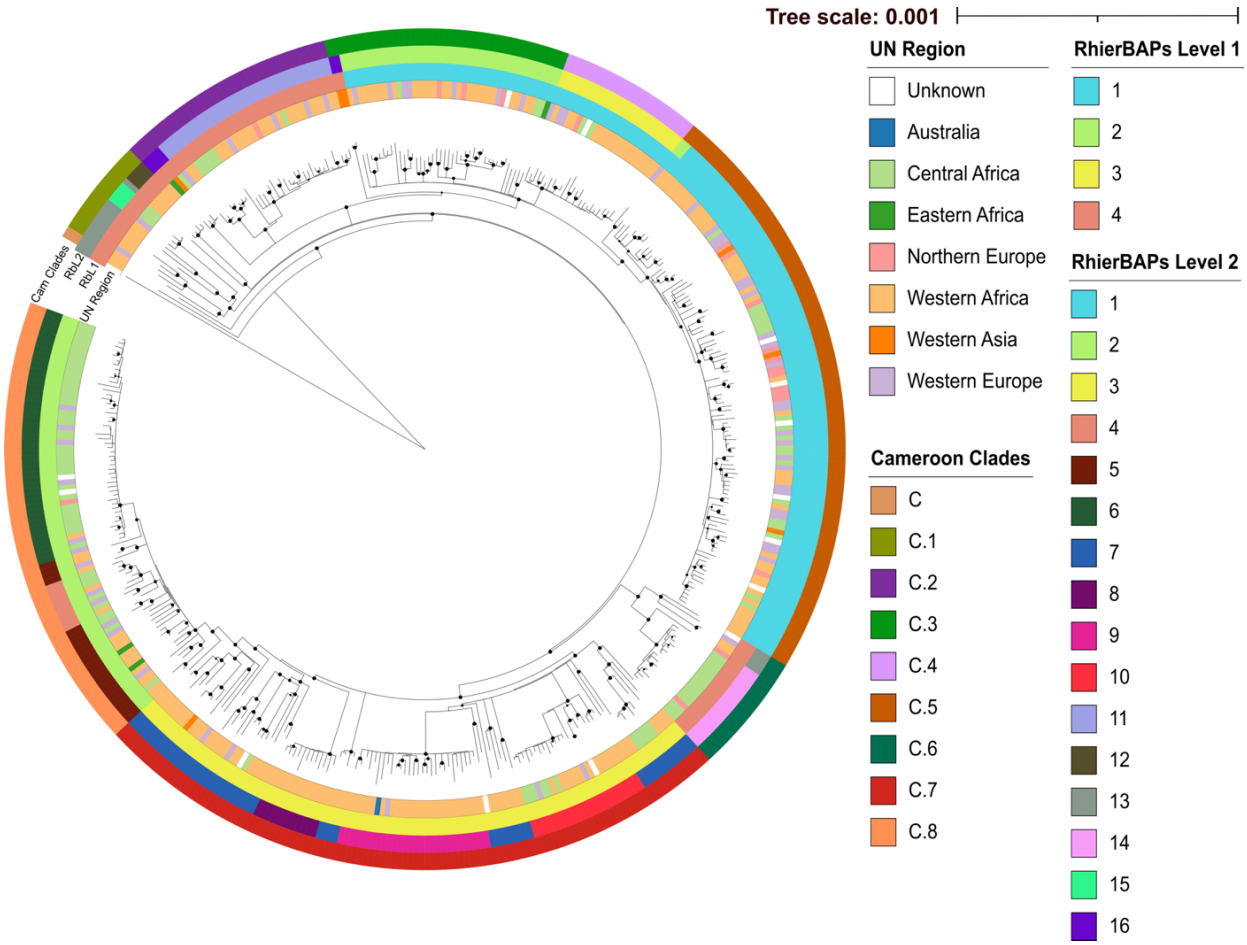


Figure 10. Phylogeny of newly defined clades of the strains of the Cameroon sublineage. The maximum likelihood tree (MLT) of 444 strains of the Cameroon sublineage was constructed based on a concatenated sequence alignment of 8,991 informative single nucleotide polymorphisms (SNP). The MLT presents the newly defined clades of the strains of the Cameroon sublineage, based on the United Nations (UN) regional designation of the origin of the infected hosts, and the RhierBAPs (Rh) level 1 and 2 nested populations of the strains. The inner-most ring is the color-coded UN-regions represented. RhierBAPs levels 1 and 2 are in the second and third rings of the phylogeny, identifying RhierBAPs levels 1 and 2. The outer ring is color-coded based on the newly defined clades of strains of the Cameroon sublineage.

The RhierBAPs model (Bayesian Analysis of Population Structure) [146] was employed to reveal nested populations in the phylogeny of the strains. The script (Appendix E) was run on two levels, with 20 initial clusters, which identified four and 16 nested populations at levels 1 and 2, respectively (Figure 10).

Table 5. United Nations regional designations of the infected host of the Cameroon strains

UN Region	**(n)**	**(%)**
Australia	1	0.2
Central Africa	107	24.1
Eastern Africa	4	0.9
North Asia	1	0.2
Northern Europe	19	4.3
Unknown	13	2.9
Western Africa	233	52.5
Western Asia	7	1.6
Western Europe	59	13.3

n= no of strains found in region, %= percentage to 444

Subsequently, the strains' clade assignments were made by combining visual inspection of the MLT, the nested populations from RhierBAPs level 1 and 2, as well as the UN regional designation of origin of the infected host, resulting in eight clades (C.1 - C.8) and one basal group (C) (Figure 10).

Signature SNPs for the strains of the clades were obtained by selecting informative markers as previously described [86]. The SNPs were selected from essential genes and mutations that led to synonymous amino acid changes; redundant markers and SNPs in drug-resistant genes were excluded. Synonymous mutations were chosen since they were less likely to be under selective pressures [148]. For each strain's clade, it was possible to select a single SNP as its marker except C.1 and C.2. For strains of clades C.1 and C.2, two SNPs were selected to define the strains that formed each group (Appendix Table D.1,Figure 10). The selected clade-specific SNPs were back tested on the 444 stains and the overall 1,133 strains of the Cameroon sublineage, leading to consistent phylogenetic assignment. The clade designation of the 444 strains is presented in Figure 10.

When the 444 strains are considered, their distribution across the clades is as follows: 17 (3.8%) strains were classified as clade C.1, followed by 39 (8.8%) as C.2, 43 strains (9.6%) as C.3, 25

(5.6%) strains as C.4, 103 (23.2%) strains as C.5, 21 (4.7%) strains as C.6, 114 (25.7%) strains to C.7, and 80 (18%) strains as C.8 (Figure 11).

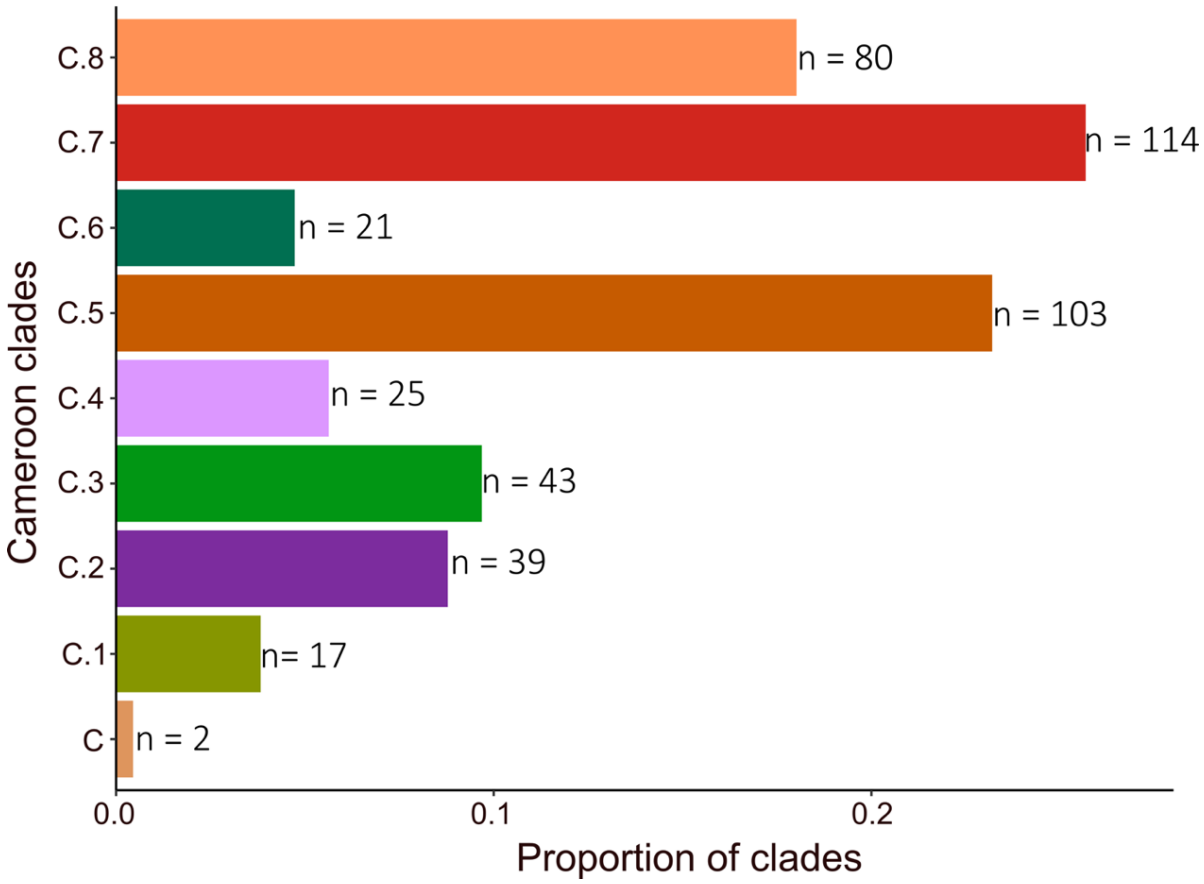


Figure 11. Distribution of the strains of the Cameroon clades. The bar chart was based on the 444 strains of the Cameroon sublineage and illustrates the proportion of the strains based on their clade designations. The total number of strains for each clade is written to the right side of the bar plot. Each clade was color-coded. n= number of strains

In the total strain population, 614 (54.2%)strains were classified to clade C.5, 155 (13.7%) to clade C.8, 145 (12.8%) to clade C.7, 69 (6.1%) to clade C.C.4, 63 (5.6%) to clade C.2, 30 (2.6%) strains to clade C.6, and 20 (1.8%) strains to clade C.1 (Figure 12).

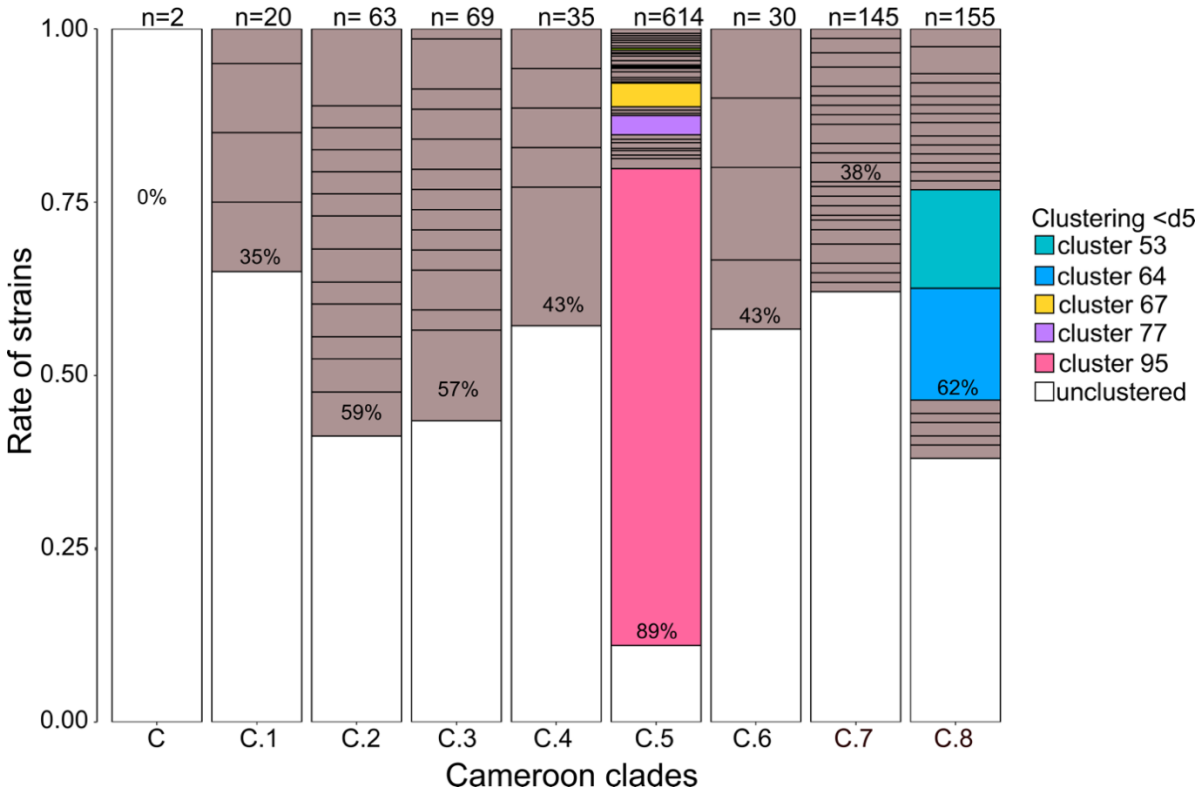


Figure 12. Clustering in the strains of the Cameroon clades of *Mycobacterium tuberculosis* complex. The bar chart illustrates the rate of transmission of the 1,133 Cameroon strains based on their clade designations. The partitioned and colored parts of the bar charts are the strains that clustered. Each partition represents a cluster of strains. The five largest clusters are brightly colored. The percentages are the cluster rate per clade. n= total number of strains per clade

The strains of the Cameroon sublineage were from four continents: Africa, Asia, Australia, and Europe. Strains of all the eight clades could be found on the African continent, while only one (C.7) was found in Australia (Appendix F, Figure 13). It was interesting to notice that strains of all except one clade could be found on the European continent. To examine potential geographical restriction among the strains, the distribution of the clades based on UN regions was investigated. More than half of the strains (52%, n= 233) were found in Western Africa, 24% (n= 107) were found in Central Africa, 13% (n= 59) in Western Europe, 4% (n= 19) in Northern Europe, and less than 3% in Australia, Eastern Africa, Northern Asia, and Western Asia (Appendix F and Table 5). The strains of clade C.8 were most prominent (43%) in Central Africa, followed by Western Europe

(20%). They were however not found in Western Asia, the region with the least diversity, with strains of only four clades. The highest strain diversity could be observed in Western Africa with strains of all clades present, followed by Central Africa and Western Europe with all but strains of clade C. Also, strains of clades C.3 and C.5 can be found in every region, and the strains of clade C.7 could only be found in four regions. However, they are most predominant (40%) in Western Africa (Appendix F, Figure 13).

When zooming to the country level, Cameroon lineage strains were concentrated in the Western and Central parts of Africa, with countries such as Cameroon, Gabon, Ghana, and Nigeria exhibiting a minimum of six clades of strains. Notably, Germany and the UK hosted strains from at least five different clades (Figure 13).

In West African countries, the distribution of strains varied. For example, in Ghana, strains of clades C.7 and C.5 constituted 45% (n= 73) and 14% (n= 24) of the Cameroon strains respectively; conversely, in Nigeria, strains of clades C.5 and C.8 comprised 35% and 23%, respectively, with no strains of clade C.7.

The strain distribution in the Central African countries was different from that of the Western African countries. In Cameroon, strains of clades C.8, C.6, and C.5 constituted 41% (n= 25), 20% (n= 12), and 12%, respectively. In Gabon, strains of clades C.8, C.5, and C.7 constituted 48% (n= 21), 27% (n= 12), and 11% of the strains of the Cameroon sublineage.

Beyond Africa, other regions displayed distinct distribution patterns. For instance, in Germany, strains of clades C.5 and C.8 constituted 40% (n= 23) and 21% (n= 12) of the Cameroon strains respectively, also in The Kingdom of Saudi Arabia, C.5 and C.3 constituted 43% and 29% of the Cameroon strains. Of note is the global prevalence of strains of clade C.5 within the Cameroon sublineage (Appendix F, Figure 13).

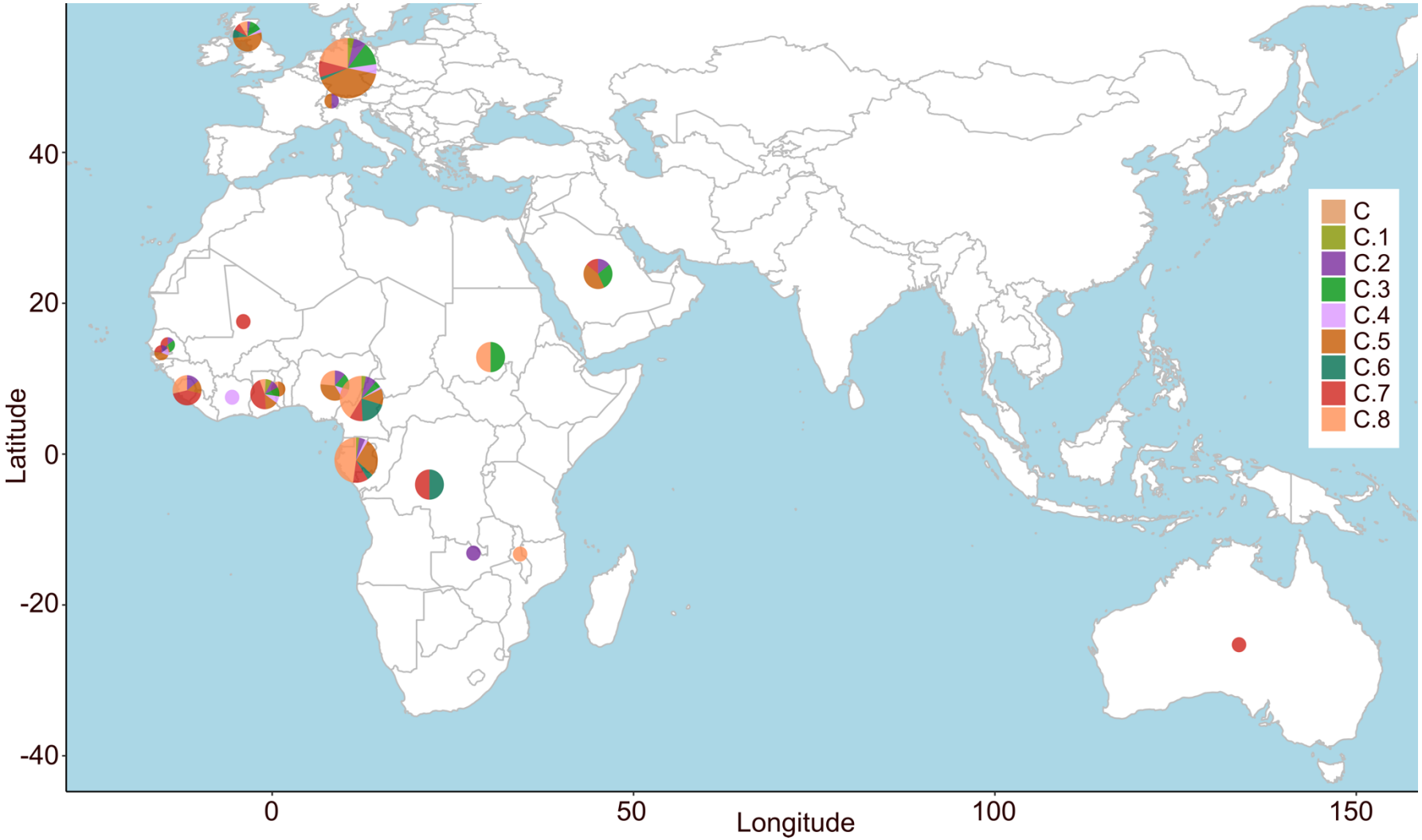


Figure 13. Global distribution of the strains of the Cameroon clades. The location of the pie chart was the country of origin of the patients of the 444 selected strains of the Cameroon sublineage. The different colors represent the clades of the strains of the Cameroon sublineage, and the proportion of their distribution based on countries is shown in the chart.

The pairwise distances (pwd) of the strains of different clades were calculated to analyse their genetic diversity (Figure 14). The highest diversity was found in strains of clade C.1, which had a median pairwise distance of 170 (mean= 151, std= 50). The strains from this clade could be found in just four countries. The lowest median pwd of 27 were in the strains of clades C.4 (mean= 27, std= 8) and C.6 (mean= 61, std= 50), which are found in six and nine countries, respectively. Strains of clades C.5 and C.8 also had low median pwd of 43 (mean= 44, std= 11) and 55 (mean= 47, std= 20) and were found in 15 and 11 countries respectively.

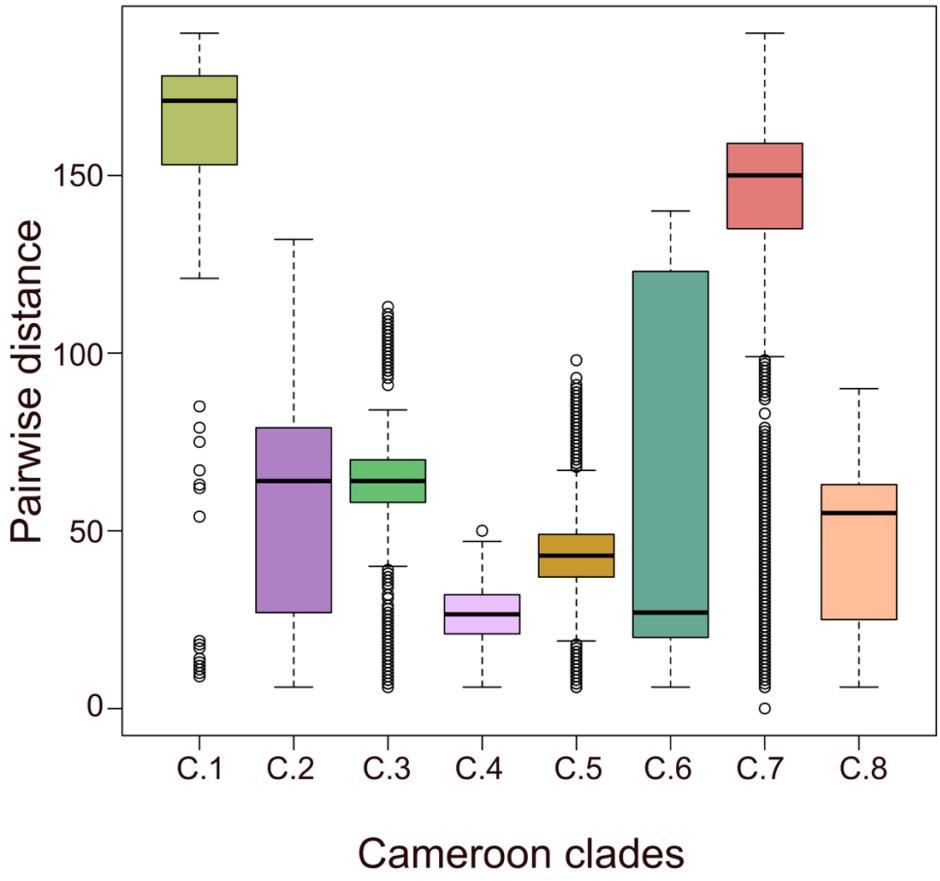


Figure 14. Genetic diversity between the strains of the Cameroon clades of *Mycobacterium tuberculosis* complex. The box plots show the pairwise distances between the strains of the clades based on the 444 strains of the Cameroon sublineage. The black horizontal line represents the median, while the black dots represent outliers. The strains of the Cameroon clades are color-coded.

3.2.3. Geographical spread of strains of different Cameroon clades

To study the evolution and transmission of the Cameroon lineage strains, patients' geographical data were mapped on the rooted ML phylogeny of the 444 strains and ancestral state reconstruction analysis was performed with PastML [157]. PastML is a software that employs a fast likelihood method to reconstruct and visualize ancestral states over evolutionary time [157]. When the continent of origin was used as a character state in PastML (Figure 15), it revealed a pattern of back-and-forth transmission of strains between Africa and Europe (Figure 15).

Notably, the ancestral reconstruction (Figure 15) indicated that strains of the Cameroon sublineage were predominantly endemic to Africa, forming a substantial cluster of 261(59%) strains. The analysis also revealed 29 independent introductions of Cameroon strains to Europe, and three to Asia. Showing that there were several small introductions of the strains to Europe, with very few transmissions to Asia. There were also two independent introductions of strains to either Africa or Europe followed by two more independent introductions, each consisting of one to four Cameroon strains, into the African continent. The group of 261 strains in Africa, were also transmitted either through Africa or Europe, after which four independent introductions of groups of strains (between one to seven) were made back to Africa, showing a clear back and forth transmission of the Cameroon strains (Figure 15). Overall, majority of the Cameroon strains never left the African continent, and they were nearly exclusively found in Africa or Europe, with many small introductions to Europe and then back to Africa.

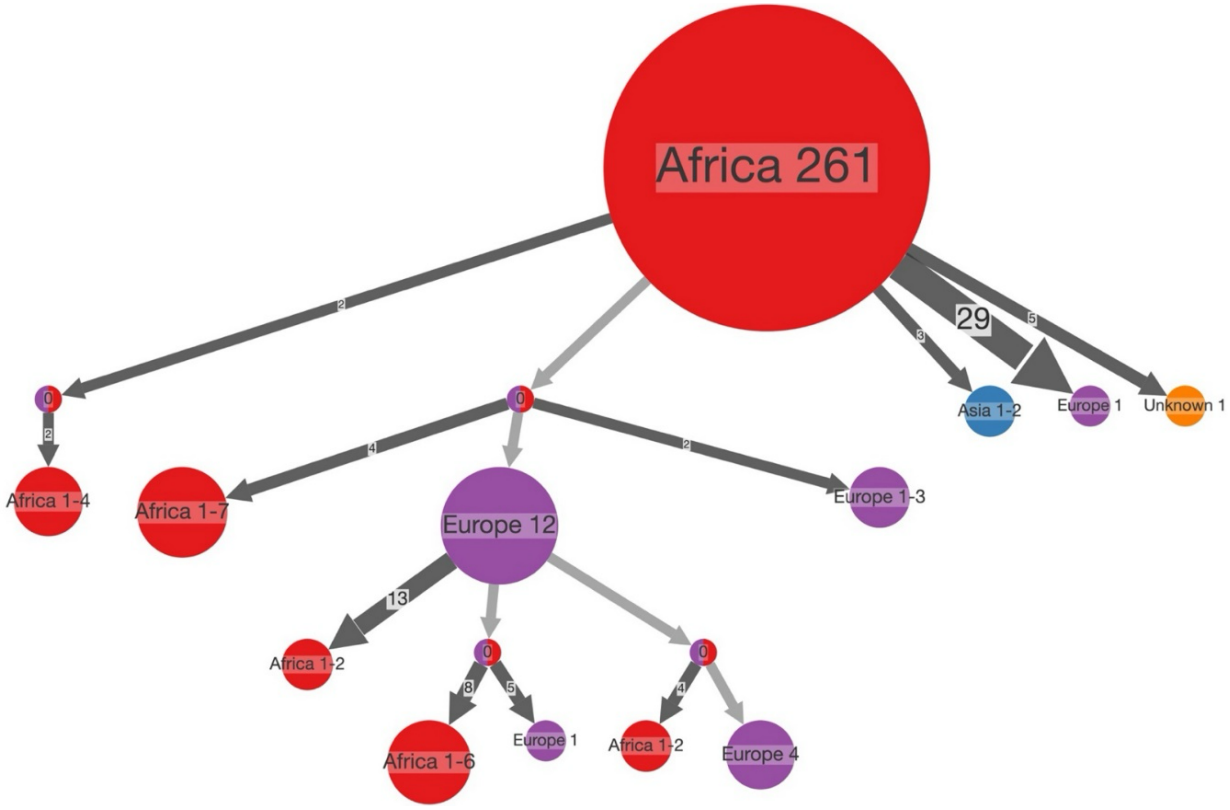


Figure 15. Ancestral reconstruction of the intercontinental transmission of strains of the Cameroon sublineage of *Mycobacterium tuberculosis* complex. The figure illustrates a compressed tree of the ancestral transmission paths of the 444 strains of the Cameroon sublineage, based on the continent of origin of the infected hosts from top to bottom. Constructed using PastML with marginal posterior probabilities approximation (MMPA) and the F81 option, the tree features color-coded continents. Nodes are circles within the tree. Nodes with multiple colors indicate ambiguity, where the program could not determine a specific continent. The arrows represent the introduction of strains to a continent, while the circles at the tips of the arrows indicate the continents where the strains were found. Vertical compression grouped together sections without continental changes, with the number of collapsed tips determining the node size. Horizontal compression merged identical subtrees (e.g., two subtrees each having one African and one European tip), forming the arrow size. A third relaxed horizontal merge was performed due to numerous continental changes, merging nodes of different sizes, with the range of merged nodes indicated. The arrow size (number on arrows) represents the frequency of strain introductions into a continent, with dark gray branches denoting two or more independent introductions.

Since most of the strains of the Cameroon sublineage never left the African continent and were found at the root of the ancestral reconstruction, an analysis of the transmission path based on

country level was conducted. A single country of origin for the strains of the Cameroon sublineage could not be identified from the ancestral reconstruction and the source country remains ambiguous (Figure 16). The countries in which the strains of the Cameroon sublineage were predominantly transmitted were Cameroon (shown with the introduction of 12 and 20 strains) and Ghana (with introduction of 20 strains, and the groups of 12 to 25 strains (Figure 16)). The strains showed regional transmission, with strains found in West Africa being transmitted among West African countries or large circulation within the country, and a similar situation in Central Africa, for instance transmission within Cameroon or Cameroon to Gabon. There were 5 independent introductions of strains from Cameroon to Gabon. Although, the ancestral reconstruction shows mainly regional transmission, there were also instances of across regional and continental transmission, where strains were introduced from Ghana to Germany and then from Germany to Gabon. Some strains were also transmitted from Ghana to Senegal, another West African country (Figure 16).

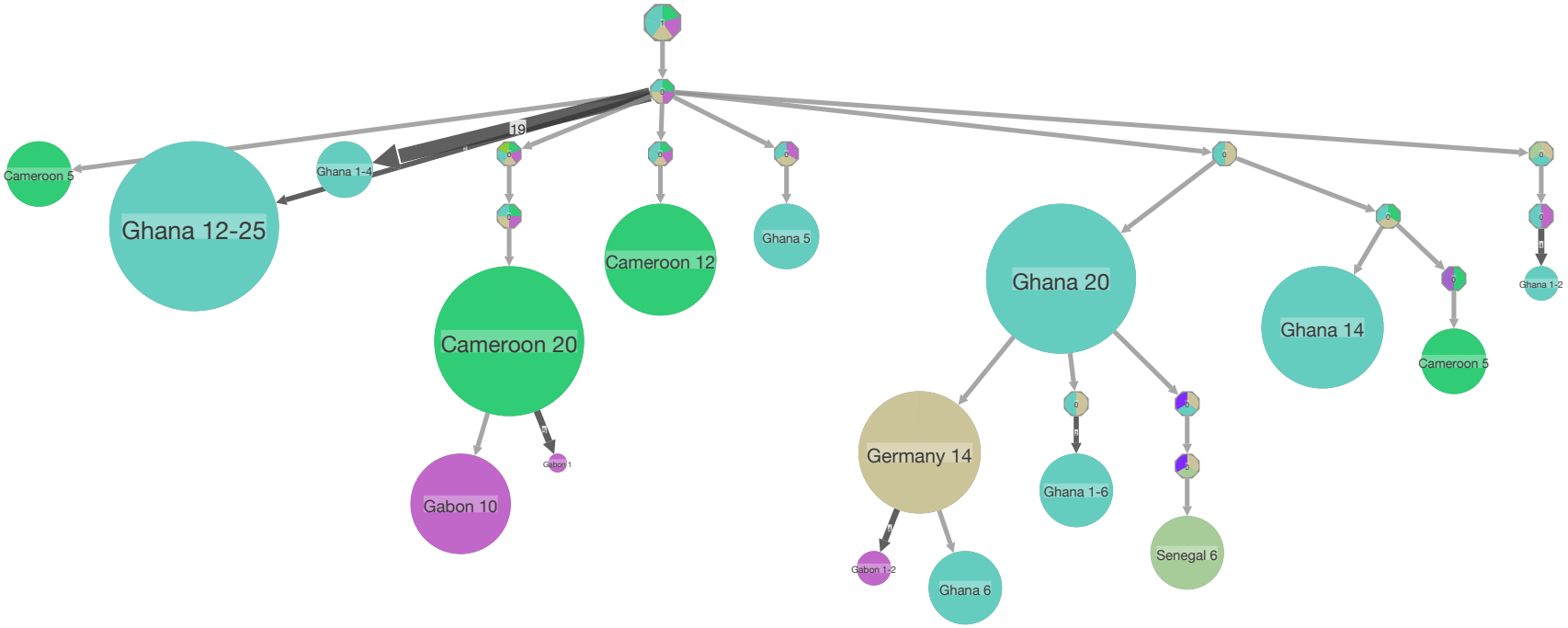


Figure 16. Ancestral reconstruction of the global transmission of Cameroon sublineage strains of *Mycobacterium tuberculosis* complex. The figure illustrates a compressed tree of the ancestral transmission paths of the 444 strains of the Cameroon sublineage, based on the country of origin from top to bottom. Constructed using PastML with marginal posterior probabilities approximation (MMPA) and the F81 option. The tree uses different shapes and colors to convey specific information. Countries are color-coded to enhance visual clarity. Nodes are either circles or hexagons within the tree. Circular nodes represent instances where the strain is linked to a single known country and hexagonal nodes indicate uncertainty in determining the exact country, showing instances where the analysis could not specify a single origin. The arrows represent the introduction of strains, while the circles at the tips of the arrows indicate the current known locations of the strains. Vertical compression grouped together sections without country changes, with the number of collapsed tips determining the node size. Horizontal compression merged identical subtrees (e.g., two subtrees each having one Ghana and one Cameroon tip), forming the branch size. A third relaxed horizontal merge was performed due to numerous country changes, merging nodes of different sizes, with the range of merged nodes indicated. The arrow size (numbers on the arrows) represents the frequency of strain introductions into a country, with dark gray branches denoting two or more independent introductions.

In order to assess the rate of recent transmission of strains within the countries, the clustering of strains was analyzed (Figure 17). The UK, which experienced an outbreak [158] of the Cameroon sublineage strains, had the highest cluster rate (97%), followed by Sierra Leone and then Gabon. Ghana and Nigeria had the lowest cluster rates of 39% and 21%. Notably, 41% of the 81 strains from Germany were clustered.

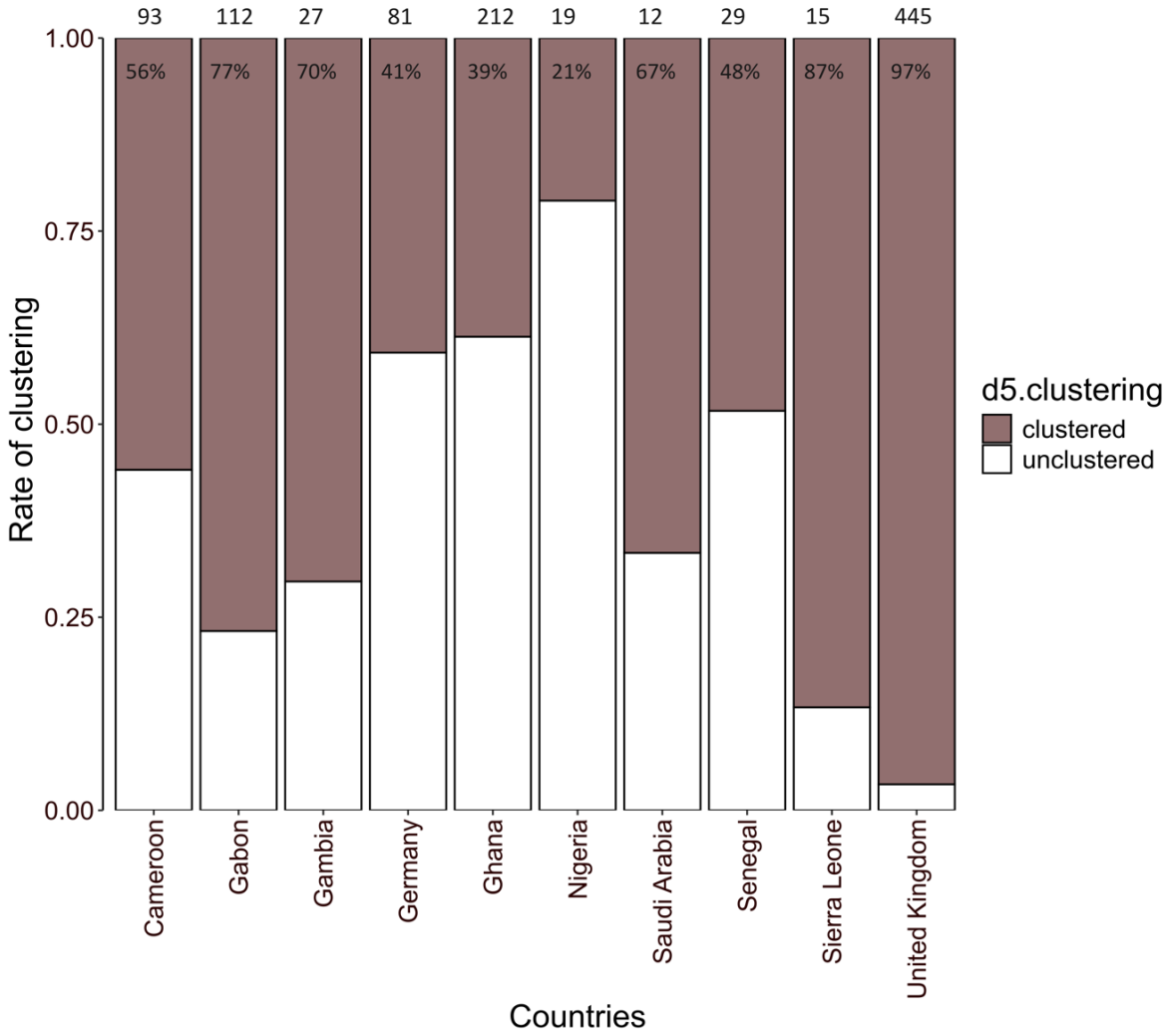


Figure 17. Transmission of Cameroon strains in countries. The bar chart displays the rate of recent transmission of the Cameroon strains in each country. Countries with less than 10 strains were not included in the analysis. Percentages in the bar chart are the clustering rate per country. The dark grey part of the bar chart are the clustered strains, while the white areas at the bottom are the unclustered strains. N= number of strains per country.

3.2.4. Clustering of strains within the clades

To have an overview of the transmission dynamics of the strains based on the newly defined Cameroon sublineages clades (CC), to identify which strain's clades are responsible for transmission of TB, the cluster rate (Figure 12) of the 1,133 strains based on the defined CC was calculated. The strains of clades (C.1 and C.7) generally had the lowest cluster rates of 35%, and 38%, respectively. The strains of clade C.5 had the highest cluster rate of 89%, the highest number of clusters (34), and the single largest cluster (cluster 95, n= 422) (Figure 12). The strains of the CC with the second most significant number of clusters were C.7, with 23 clusters. The strains of the CC with the third largest number of clusters were C.8, with 21 clusters, and had the second and third largest single clusters (Figure 12). The strain's clades with high cluster rates or large clusters signify high transmissibility of strains.

3.2.5. Major transmission events

A large number of strains of clade C.5 (cluster 95, made up of 422 strains) had clustered (Figure 18). The strains were found mainly in the UK, with a couple from Ghana. They were mainly INH mono resistant-TB (Hr-TB) strains, while 16 were MDR-TB strains [158–160]. All the strains shared the same INH *fabG1-15c>t* resistant mutation, indicating a primary transmission of a clone of the Hr-TB strain. The two strains from Ghana shared the same RIF *rpoB S450L* mutation. Meanwhile, 11 strains from the UK shared the same RIF *rpoB V170F* resistance mutation found outside of the rifampicin resistance determining region (RRDR).

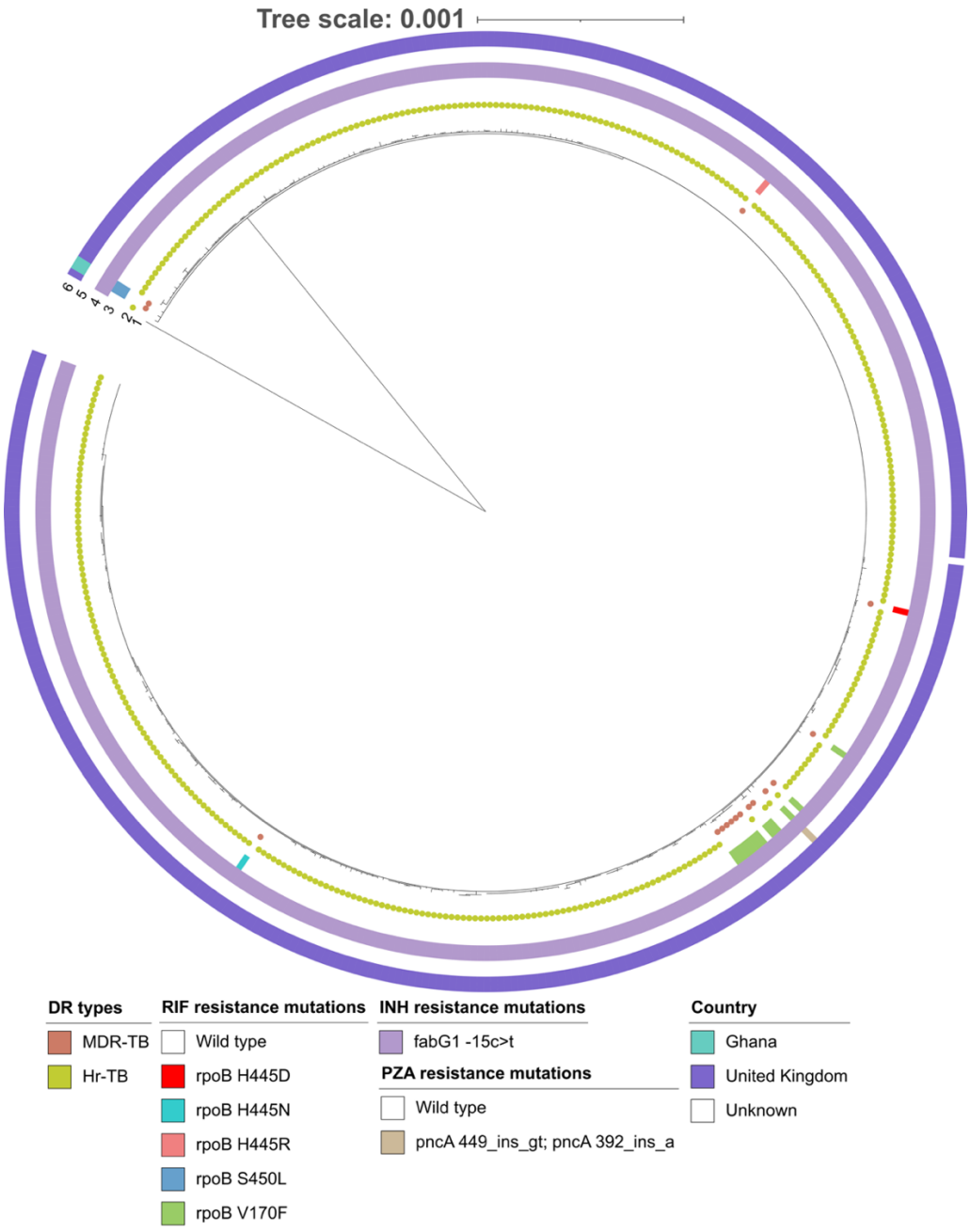
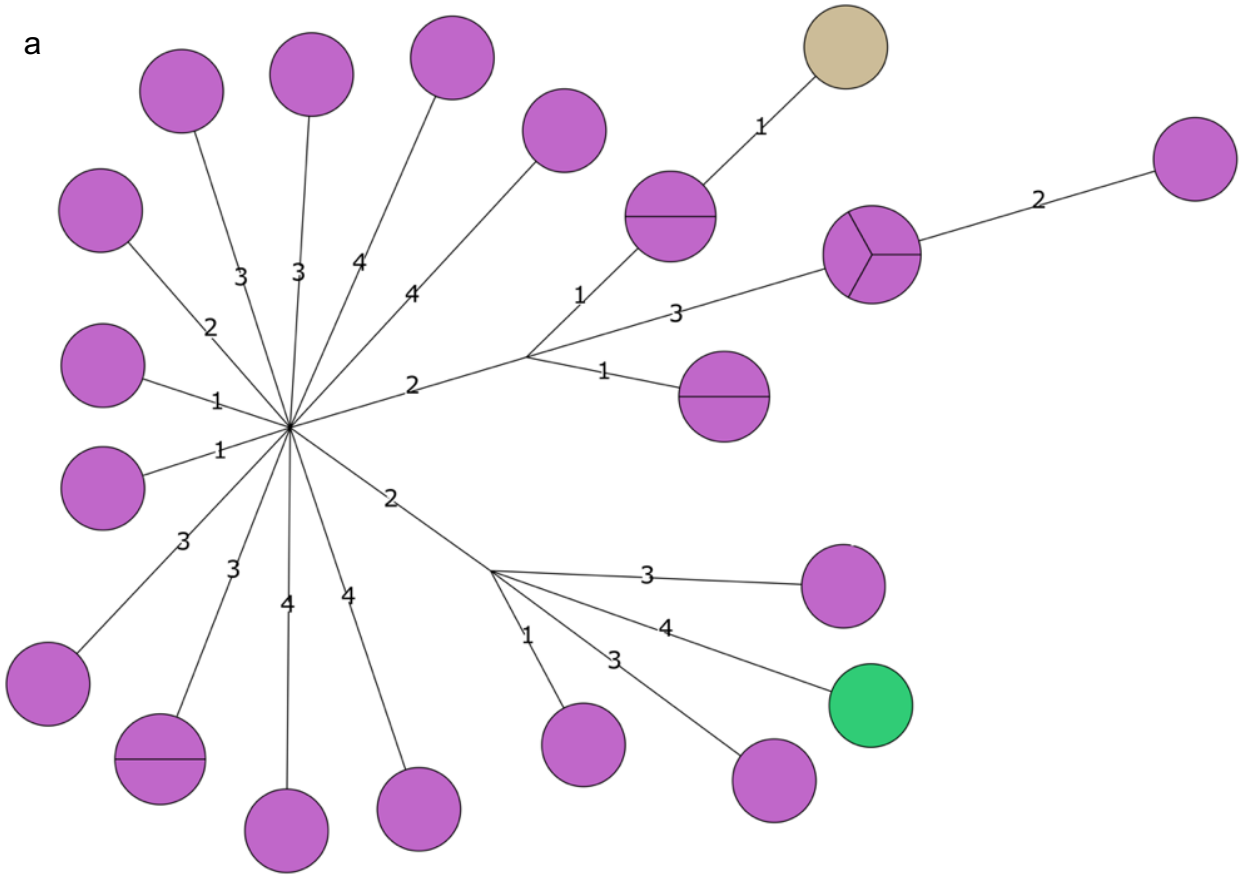


Figure 18. Phylogeny of the largest cluster of strains of the Cameroon sublineage. The Maximum likelihood phylogeny was built from a concatenated sequence alignment of 1,011 single nucleotide polymorphisms (SNP) from 422 strains (cluster 95) of clade C.5 of the Cameroon sublineage rooted on an outgroup. Indicated on the rings of the tree from inner to outer: 1. presence of multidrug-resistant strains, 2. presence of isoniazid mono-resistant strains, 3. rifampicin resistance mutations present, 4. pyrazinamide resistance mutations present, and 4. the country of origin of the infected host the strains. MDR-TB: multidrug-resistant tuberculosis (resistance to both RIF and INH); Hr-TB: isoniazid mono-resistant tuberculosis, DR: drug resistance, RIF: rifampicin, INH: isoniazid, PZA: pyrazinamide.

The second largest cluster had 25 strains belonging to clade C.8 (Figure 19a); one strain was from a patient from Cameroon, another from a patient from Germany, and the rest (23) were found in patients from Gabon, and no drug resistance was observed in any of the strains. Another significant transmission of clade C.8 strains was observed in cluster 53, which comprised of 22 strains (Figure 19b). Most strains were found in patients from Germany, with one strain each from Ghana and Guinea Bissau, these strains were also susceptible to the anti-TB drugs.



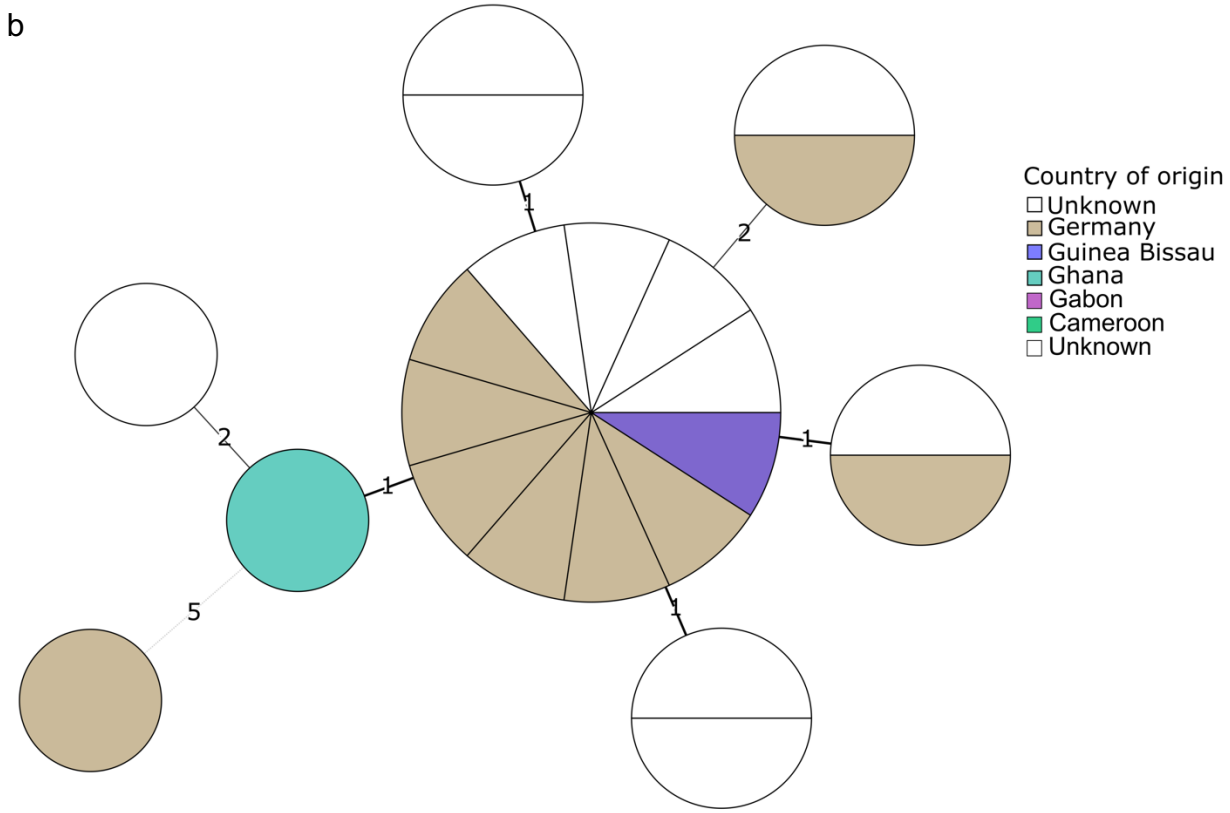


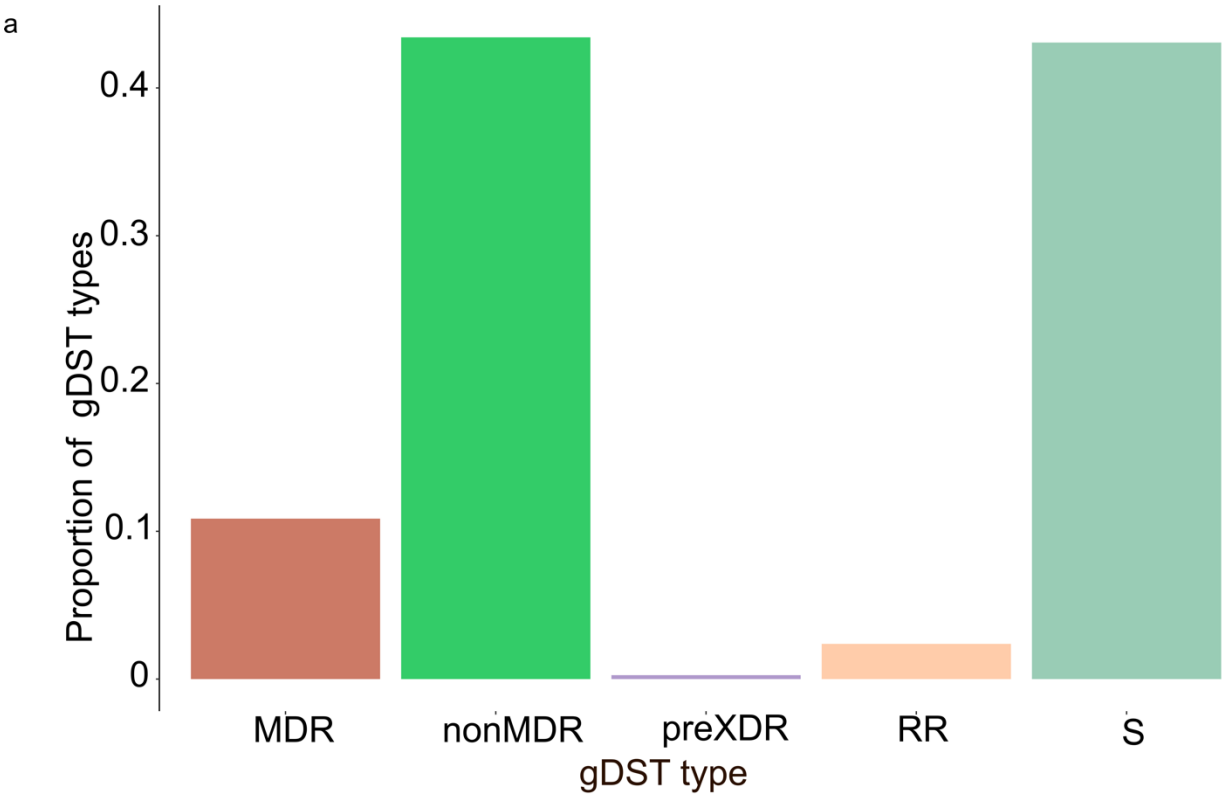
Figure 19. The phylogeny of strains of the Cameroon clades in transmission. The trees illustrate two potential outbreaks of strains of the Cameroon sublineage and the countries the strains were transmitted. A. The maximum parsimony phylogeny was constructed from 231 concatenated SNP alignment of 25 strains from clade C.5 (cluster 64). B. The maximum parsimony phylogeny was constructed from 737 concatenated SNP alignment of 22 strains from clade C.8 (cluster 53). The country of origin of the strains was color-coded. The number on the branches was the number of SNP differences from each strain. Circles represent strains; the number of parts in a circle indicates the number of strains.

3.2.6. Drug resistance

Genotypic drug resistance test was carried out on the strains to determine their drug resistance patterns. Overall 43% (n= 488) of the strains were susceptible, 2% (n= 27) were RR, 43% (n= 492) were nonMDR (resistance to drugs other than RIF and or INH), 10% (n= 123) were MDR, and 0.3% (n= 3) were preXDR (Figure 20a). Focussing on the strains' resistance to individual anti-TB drugs,

over half of the strains (51%, n= 577) were resistant to INH, 14% (n= 153) to RIF, 41% (n= 464) to ETH, 10% (n= 114) to SM and two strains to BDQ/CFZ (Table 6).

Regarding Cameroon sublineage clades (Figure 20b), all the strains of clade C were susceptible, with 23% (n= 15) of C.2 being MDR. 10% (n= 7) of C.3 and 37% (n= 13) of C.4 strains were nonMDR. The majority (70%, n= 429) of strains of clade C.5 were nonMDR, and only 4% (n= 26) were MDR. Meanwhile, 80% (n= 24) of the strains of clade C.6 were MDR, and for C.7 16% (n= 23) and 13% (n= 19) of the strains were MDR and nonMDR, respectively. Furthermore, of the 155 strains of C.8, 19% (n= 29) were classified as MDR.



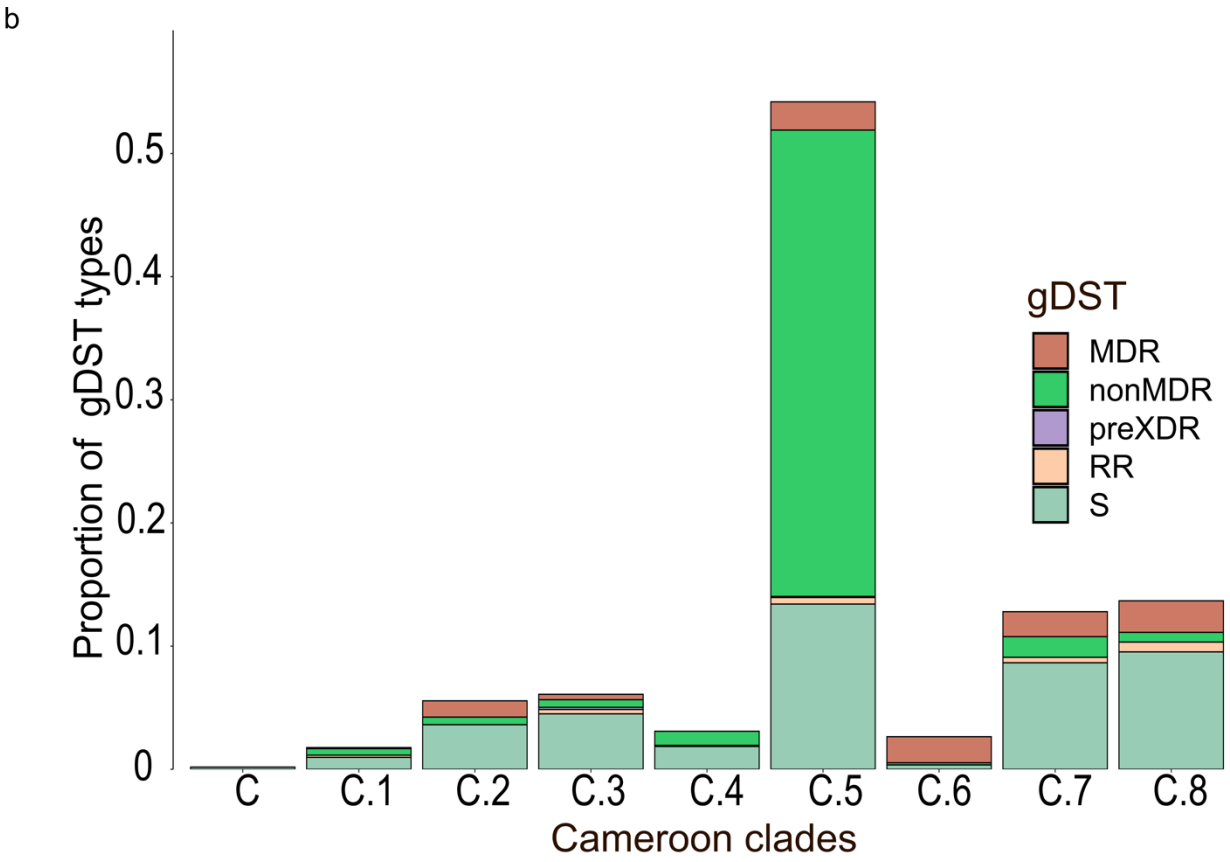


Figure 20. Distribution of drug resistances in the strains of the Cameroon sublineage of *Mycobacterium tuberculosis* complex. The bar plot illustrates the proportion of drug susceptibility types in the strains of the Cameroon sublineage. a. Proportion of drug susceptibility types in all the 1,133 Cameroon strains is illustrated. b. genotypic drug susceptibility type distribution within the 1,133 strains of the Cameroon clades is illustrated. The genotypic drug susceptibility types are color-coded. MDR: multidrug-resistant tuberculosis (resistance to both RIF and INH), nonMDR: resistance to other antituberculosis drugs other than isoniazid and rifampicin, preXDR: pre-extensively drug resistance (MDR plus resistance to any fluoroquinolone), RR: rifampicin resistant, S= susceptible, gDST = genotypic drug susceptibility type.

Table 6. Resistance of 1,133 Cameroon sublineage strains to anti-tuberculosis core drugs

Drug	n	Proportion (%)
Rifampicin	153	13.50
Isoniazid	577	50.93
Ethambutol	65	5.74
Pyrazinamide	57	5.03
Bedaquiline/ Clofazimine	2	0.18
Moxifloxacin	4	0.4
Levofloxacin	5	0.4
Ethionamide	464	41
Cycloserine	1	0.09
Streptomycin	114	10

Denominator (Total number of strains used) = 1,133; n=number; %= percent

Most of the strains of the different clades had developed resistance to at least one of the anti-TB drugs except strains of clade C (Figure 21). Strains of clades C.1 to C.8 had at least one strain that was MDR and one that was resistant to SM. One strain of C.5 and C.8 were the only ones with resistance to BDQ and CFZ. For every anti-TB drug indicated, there is at least one of the strains of clade C.5 that had developed resistance, including CS; however, strains of other clades have not developed resistance to CS (Figure 21).

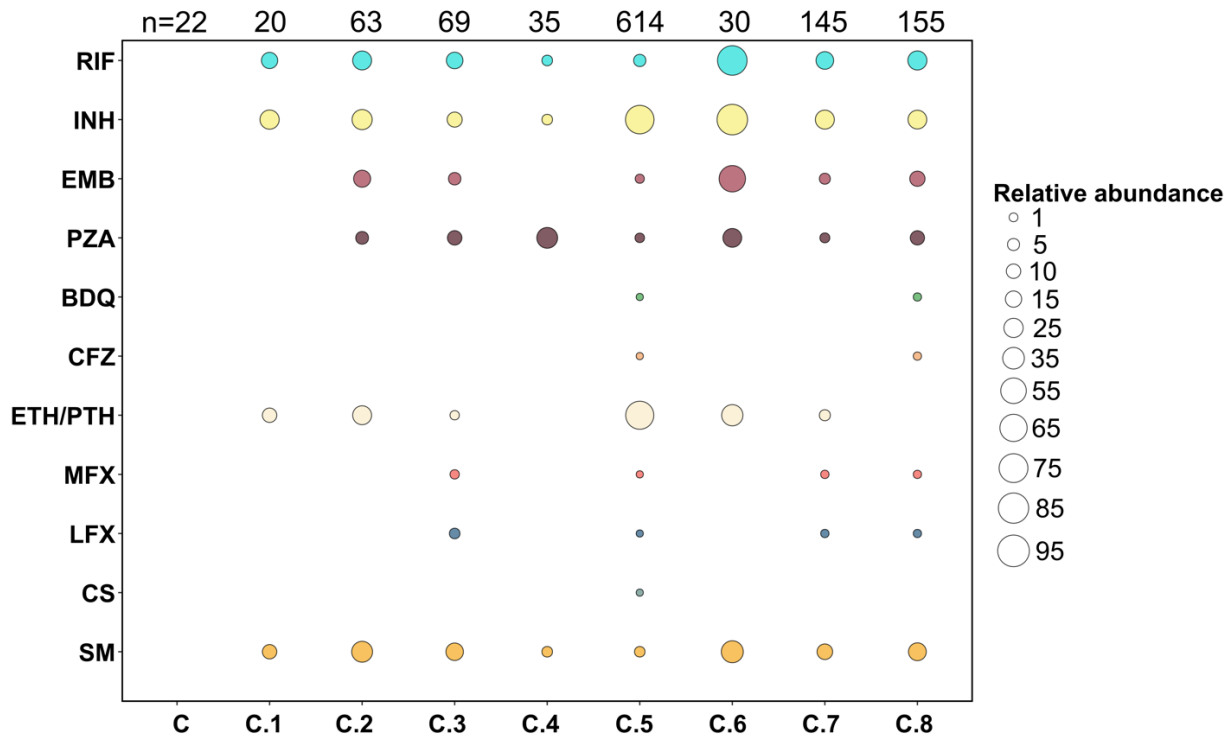


Figure 21. Resistances in the anti-tuberculosis drugs in the strains of the various Cameroon clades. The relative abundance of resistance to the anti-tuberculosis drugs based on strains of the Cameroon clade is shown in the bubble plot. The anti-tuberculosis drugs are color-coded. INH = isoniazid, EMB =ethambutol, PZA =pyrazinamide, BDQ = bedaquiline, CFZ = clofazimine, ETH/PTH = ethionamide/ prothionamide, MFX= moxifloxacin, LFX = levofloxacin, CS = cycloserine, SM = streptomycin.

Regarding RIF, 153 strains had developed resistance mutations, and these were all *rpoB* resistance mutations, which were mainly the WHO group-1 and borderline RIF-resistance mutations. Of those, the most prevalent resistance mutation was *rpoB* S450L and *rpoB* H445D, which constituted 50% (n= 76) and 11% (17) of all the RIF resistance mutations (Figure 22a, Appendix Table D.1 & E.4). These mutations were part of the group-1 mutations defined by the WHO to be associated with resistance and within the RRDR [59]. Thirteen (8.5%) strains had developed the *rpoB* V170F mutations, also a WHO group-1 resistance mutation; however, this mutation is outside the RRDR [59]. Four of the six borderline RIF-resistance mutations were identified, *rpoB* (D435Y, H445L, H445N, and L430P).

Strains of clade C.5 had a combination of 8 different RIF resistance mutations, with *rpoB* S450L as the most dominant mutation in the strains of that clade. Strains of clades C.6 and C.8 both had nine different RIF mutations, with the *rpoB* S450L again being the most dominant (Figure 22a). Meanwhile, the clade C.7 had the highest number of strains (n =25) with a RIF resistance mutation (*rpoB* S450L) (Figure 22a).

With INH, all the mutations discovered were associated with the *fabG1-inhA* operon or the *katG* gene, and 577 (51%) strains had these mutations (Figure 22b, Appendix Table D.3). The most prevalent mutation present was the *fabG1*-15c>t (75%, n= 435), a group-2 designated mutation marker for resistance [59]. There were an additional 16 strains that had developed both the *fabG1*-15c>t low-level mutation and the *katG* S315T group-1 resistance mutation. The second most prevalent mutation was the *katG* S315T which constituted 19% (n= 112) of the INH resistance mutations.

When the proportions of the INH resistance mutations were analyzed, the *fabG1* -15c>t mutation was the predominant mutation and was found in strains of clades C.5 and C.2 (Figure 22b). However, the INH mutation that dominated all the strains was the *katG* S315T, and it was found in strains of clades C.1, C.2, C.3, C.4, C.5, C.6, C.7 and C.8 (Figure 22b).

A total of 65 strains had developed EMB resistance mutation. The prevalent ones were the *embB* (M306I and M306V) which made up 31% (n= 20) and 28% (n= 18) of the mutations, respectively (Figure 22, Appendix Table D.3). For PZA, 57 strains had developed mutations; the most prevalent was *pncA* p69S, constituting 19% of the PZA mutations (Figure 22, Appendix Table D.4). Only two strains had BDQ resistance mutations, and those were in the *Rv0678* gene (Figure 22, Appendix Table D.5).

Also, strains of clades C.6 and C.8 had developed eight different EMB resistance mutations (Figure 22). However, the dominant mutations *embB* M306I and *embB* M306V had been developed by strains of clades C.2, C.5, C.6, C.7 and C.8 and for the latter, strains of clades C.3, C.5, C.6 and C.8 (Figure 22c).

With regard to PZA resistance mutations (Figure 22d), strains of clade C.8 had the highest number of resistance mutations (ten); meanwhile, clade C.4, which had 11 strains with mutation had all developed the same *pncA* P69S resistance mutation. Also, of the seven strains of C.3, six had developed the same *pncA* T47I resistance mutation. Meanwhile, except for two strains, one from clade C.5 and the other from C.8, which developed *Rv0678* (E113 and R132) resistance mutations, all the strains were susceptible to BDQ (Figure 22e).

This section provided a thorough analysis of the strains within the Cameroon sublineage of *Mycobacterium tuberculosis* complex, encompassing key aspects of phylogeny, geographical distribution, transmission dynamics, and drug resistance profiles of the strains.

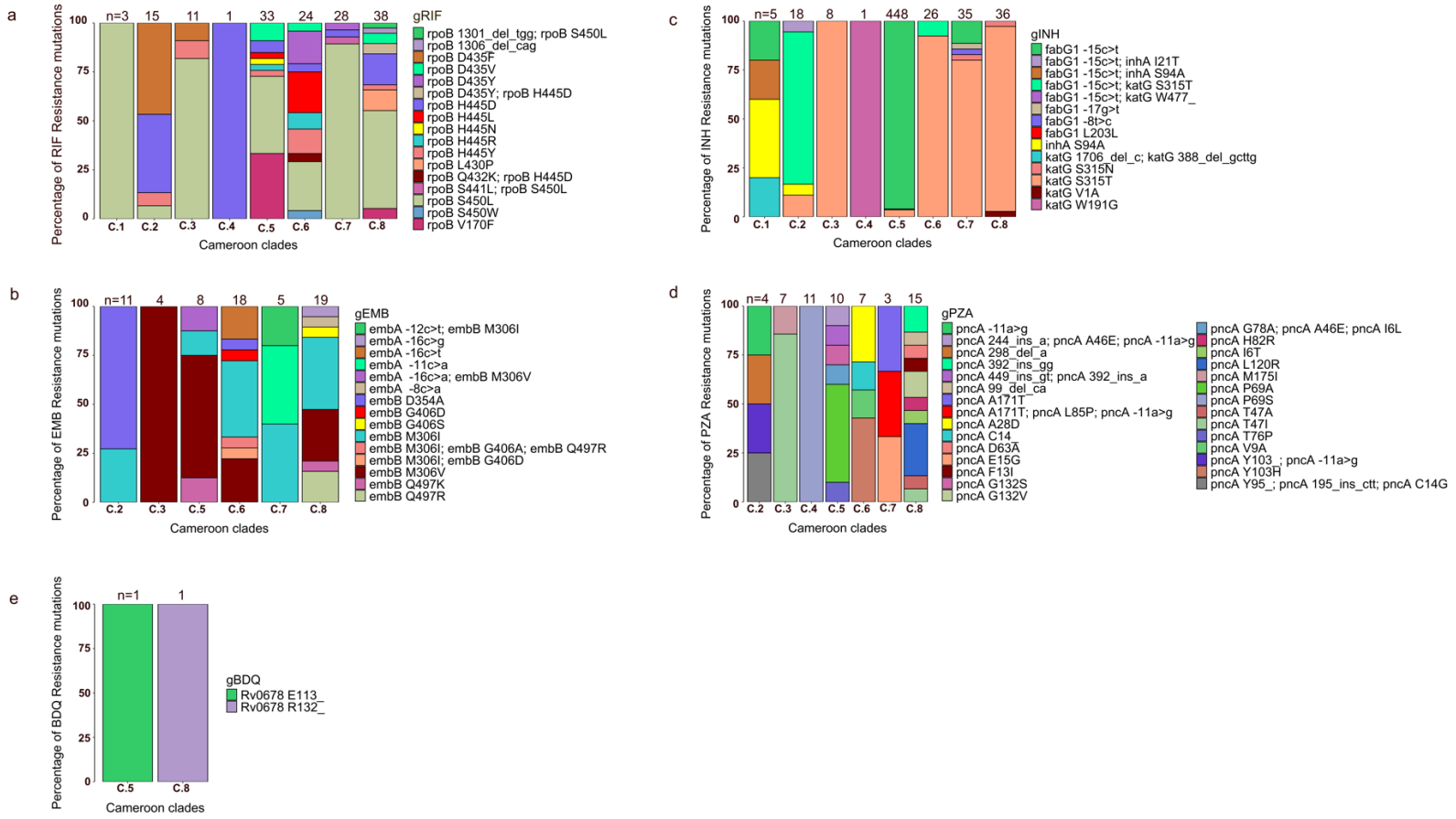


Figure 22. Distribution of resistance mutations in the strains of the Cameroon sublineage. The bar charts illustrate the proportion of resistance mutations to the anti-tuberculosis drugs in the strains of the Cameroon sublineage based on their clade designation a. Proportion of rifampicin resistance mutations is illustrated. b. Proportion of isoniazid resistance mutations is illustrated c. Proportion ethambutol resistance mutations is illustrated d. The proportion of pyrazinamide resistance mutations is illustrated. e. Proportion of bedaquiline resistance mutations is illustrated. The resistance mutations are color-coded. n = the number of strains, g = genotypic, RIF= rifampicin, INH=isoniazid, EMB= ethambutol, PZA= pyrazinamide, and BDQ = bedaquiline.

4. Discussion

The MTBC strains have afflicted humans with TB for centuries [161]. The pathogen has no environmental reservoirs, and humans serve as the natural reservoirs for MTBC, yet humanity has still been unable to eradicate the disease. The disease is highly prevalent in low- and middle-income countries such as Sierra Leone, imposing a considerable burden on many economies and households [45]. Drug resistance, especially MDR-TB, has made treating TB problematic and expensive in several areas of the world [45]. Considering how MDR-TB management has become a serious global health challenge, understanding its epidemic determinants on the regional level is crucial in developing effective control measures, as MTBC strains have developed various virulence mechanisms to evade the immune system.

Whole genome sequencing was used to analyze MTBC strains from Sierra Leone to determine the phylogeny of the strains present in the country, their transmission pattern, and drug resistance types and profile. An investigation into the evolution, transmission patterns, and drug resistance of the strains of the Cameroon sublineage was conducted. The main findings from these studies will be discussed in the following sections.

4.1. Rifampicin-resistant/multidrug-resistant-Tuberculosis in Sierra Leone

In this study WGS was employed to characterize RR/MDR TB strains from Sierra Leone. It was found that one in four strains was resistant to all first-line drugs. While no resistance to FQ was found, five strains had resistance to BDQ/CFZ, due to mutations in *Rv0678*. The overall cluster rate of more than 40%, indicated ongoing transmission of RR/MDR TB potentially contributing to the RR/MDR TB burden in the country. Strikingly, *Maf* L6 strains represented about 21% of the MDR MTBC strains investigated, forming a longitudinal outbreak with two branches, each constituting five strains. Within one of the branches, all the strains were resistant to all first-line drugs and two strains had developed additional BDQ/CFZ resistance. The data also demonstrate a high diversity of drug resistance mutations with a significant number of so-called borderline INH and RIF resistance mutations, potentially allowing high dosing for treatment.

The results of this study in Sierra Leone showed that a staggering 81% of the samples tested were positive for MDR-TB, emphasizing the critical nature of the issue. This problem extends to neighboring countries such as Liberia and Guinea, where a high incidence of DR-TB has been reported [162,163]. However, the limited number of cases being tested for DR in Liberia (43%) means that many are not receiving the appropriate treatment, exacerbating the region's DR problem [164]. Also, in a 2016 study involving eight WA countries, 6% of new cases were MDR-TB, and 35% were retreatment cases [165]. In most countries, the GeneXpert MTB/RIF is the primary diagnostic tool [166,167]. Although the GeneXpert MTB/RIF is an effective tool in diagnosing DR-TB, incorporating other molecular diagnostic methods, such as the line probe assay and the highly recommended targeted sequencing, could lead to better diagnoses and improved prognoses [168,169]. There have been instances where the GeneXpert MTB/RIF failed to detect RIF resistance or produced false positives. However, using a combination of molecular methods, particularly WGS, would allow for a more comprehensive identification of DR, including low-level resistances mutations [170–172].

The key data obtained in this study show that FQ-resistant strains are not yet circulating in the country, which is a clear difference from other African countries. For instance, in Nigeria, there was a reported FQ resistance of 13% (17/132 of RR-TB) in a study conducted between 2018 and 2020, while in Ghana, 36% (11/31) MDR-TB cases were FQ-resistant [173,174]. Also, a recent study in Namibia described a pre-XDR rate among RR/MDR strains of about 13% and a higher rate (50%) of RR/MDR MTBC strains to be resistant to all first-line drugs [150]. The difference in the rate of FQ-resistant strains is even more pronounced when compared to data from other parts of the world, such as India, where 70% of RR/MDR strains in the Mumbai Metropolitan Region were found to be FQ-resistant [175].

Given the low prevalence of BDQ resistance and the absence of FQ resistance in Sierra Leone, most RR/MDR-TB cases should be successfully treated with the WHO's 9-12 months short regimen [50]. Even the shorter all-oral-6-month regimen of BDQ, preteonamid (Pa), and linezolid (L) plus moxifloxacin (M) (in the absence of FQ resistance) (BPaLM) recently recommended by WHO is an excellent standardized treatment option for patients with RR/MDR TB in Sierra Leone [98,176].

However, FQ and/or BDQ/CFZ resistance must be monitored closely, as such resistance may rapidly reduce the new regimens' efficacy [177].

Indeed, five MTBC strains were found (2%) resistant to BDQ and CFZ due to mutations in *Rv0678*, a transcriptional repressor of the MmpS5-MmpL5 efflux pump [178,179]. Mutations in *Rv0678* are the primary driver of BDQ/CFZ resistance in clinical strains in South Africa and Moldova [180,181]. Sonnenkalb *et al.* (2023) also showed in their in-vitro and in-silico studies that mutations in *Rv0678* are the primary drivers of BDQ resistance [182]. Also, these mutations are linked to treatment failure and can potentially cause an increase in BDQ/CFZ resistances over a short timeframe [180].

Although BDQ/CFZ resistance was low, the emergence of BDQ/CFZ independent from FQL resistance is worrying. BDQ/CFZ resistance may emerge without direct drug exposure, as reported previously [172,183–185]. While the exact mechanism remains to be identified, prospective surveillance for BDQ/CFZ resistance is paramount. Given that BDQ is an essential drug in the new BPaLM regimen, research studies need to investigate mechanisms involved in selecting MTBC strains with *Rv0678* mutations.

In the investigation of MTBC strain diversity in Sierra Leone, the presence of six major MTBC strain's lineages which are responsible for RR/MDR-TB cases were observed. Among these strain lineages, strains of L4 were the most prevalent [107]. Lineage 4 strains have a wide global distribution and are known for their genotypic and phenotypic diversity [102,107,186]. Consequently, the high proportion of strains of L4 in this study aligns with previous research findings, and the associated cluster rate of 62% is consistent with those reported in other studies. Similar studies conducted in West African countries, such as Ghana, have also demonstrated high clustering rates in strains of L4, reaching up to 45% [109]. However, the specific strains' sublineages involved in transmission differ between Sierra Leone and Ghana. In Sierra Leone, the strains of sublineages 4.1.2.1 Haarlem and 4.8 mainly T were predominantly involved in transmission. In contrast, the transmission in Ghana primarily involved strains of sublineages 4.6.2.2 Cameroon and 4.1 Ghana [109]. Interestingly, the strains driving transmission in Sierra Leone were scarce or absent in Ghana. These findings suggest that the lineage and strain profiles of MTBC can vary between countries,

even within the same subregion. Therefore, it is crucial to determine the specific lineage and strain profiles for each country individually.

The strains of sublineage 4.8 mainly T exhibited significant transmission of MDR-TB. Within this sublineage, a specific MDR-TB clone was identified, characterized by resistance mutations in INH *katG* S315T, RIF *rpoB* S450L, and EMB *embB* M306V resistant mutations. Interestingly, the PZA resistance mutations varied among the strains within this cluster. The *rpoB* 450L mutation, along with the *rpoB* H445Y, are widely recognized as prominent RIF resistance mutations associated with MDR-TB [187–189]. In this study, *rpoB* S450L was the most prevalent RIF mutation, while *rpoB* H445Y ranked fourth. Considering all the strains within this cluster shared the same RIF and INH resistance mutations, it indicates that a single MDR clone of the strains of 4.8 mainly T sublineage was transmitted. This clone, with a consistent genotype, was responsible for infecting all seven individuals in the cluster.

Lineage 2 strains were the third most prevalent strains; surprisingly, they all belonged to the 2.2.1 Beijing Ancestral 3 sublineage, also showing a high cluster rate of 86%, indicating effective MDR-TB transmission. Recent studies, e.g. from Eastern European countries, have found associations between modern L2 strains such as L2.2.3 [163] and the L2.2.1 modern Beijing strains [191] with high rates of clustering and transmission [192–194]. Some studies from India and South Africa indicate that strains of ancestral Beijing lineages may also develop high drug resistance rates linked with ongoing transmission of few MDR/pre-XDR/XDR strains in a given setting [175,195]. The emergence of ancestral Beijing strains in Sierra Leone warrants further investigations towards cross-border spread in Africa, introduction by migration followed by local spread, and the overall importance of this strain type for the MDR epidemic. These strains' ancestral nature leaves the question of whether they have always been present or migrated from a different country.

The L2 strains, which belonged to the L2.2.1 Ancestral Being 3 sublineage, formed the largest cluster in this study and were also MDR. These strains shared a common resistance mutation, *katG* S315T, indicating a clonal relationship. In Hang *et al.*'s (2019) study in Vietnam, they found a similar association between the *katG* S315T mutation and L2 strains, particularly the atypical 'ancient` Beijing strains, similar to the ones observed in this study, as opposed to the modern

strains [196]. Further investigations into the resistance profiles of the strains of L2 cluster revealed that they did not share the same resistance mutations for the other first-line anti-TB drugs (RIF, PZA, EMB). While they started with a single INH clone, the strains development different resistance mutations for RIF and EMB, and PZA after transmission. Therefore, this cluster of strains does not represent a single MDR clone in terms of transmission. This could be due to poor drug compliance or inappropriate treatment regimen [197–199]. The large cluster of L2 strains with different resistance mutations suggests that the strains of L2 are highly transmissible and not because they are MDR. The different resistance mutations also highlight the need for targeted interventions and enhanced infection control measures to prevent further transmission and contain the spread of DR strains. The diverse resistance mutations within the L2 cluster showed the high level of genetic diversity among the circulating strains. This has implications for treatment strategies, as different resistant profiles may require individualized approaches and regimens. As such, comprehensive DST is needed to guide appropriate treatment decisions for patients.

In clear distinction to other geographical regions, 24% of the strains were identified to belong to *Maf*, the majority to L6 (West African 2). The higher prevalence of *Maf* L6 strains compared to L5 follows a recently published study analyzing the geographical distribution of *Maf* in different African countries, demonstrating that L6 strains are mainly found in West Africa, while L5 strains are more prominent in East-West Africa [106]. In the study, the authors also analyzed the occurrence of DR mutations in L5/L6 strains and found that L5 strains had a higher DR mutation rate than L6 strains [106]. The results here demonstrate the ability of L6 strains to transmit and account for a significant number of RR/MDR strains in a region. As such, *Maf* L6 strains should be considered potential drivers of MDR-TB in West African countries.

Of significance was that all *Maf* strains are resistant to CS due to a previously described 1-bp frameshift deletion in *ald* [200]. Cycloserine is a group B second-line antibiotic used for the treatment of TB and recommended by WHO as part of the longer regimen for MDR-TB [200–202]. This may complicate the treatment of patients infected with *Maf* MDR strains if they also gain other resistance, e.g. to FQL and BDQ/CFZ. Hence, the population structure of MTBC strains in a region

may influence the effectiveness of treatment regimens because of drug resistance mutations deeply rooted in the phylogeny [200].

In line with findings from other African countries such as Namibia, a high diversity of INH/RIF resistance mutations with a substantial proportion of strains having so-called borderline resistance mutation in *rpoB* and/or in the *inhA-fabG1* promotor region was found. These borderline resistance mutations are difficult to detect by phenotypic assays [149,150,203]. In patients infected with MTBC strains with lower-level resistance mutations, high-dose INH and/or RIF may overcome resistance, presenting a viable treatment option, especially for patients with MDR/pre-XDR/XDR-TB with advanced resistance patterns [204,205]. However, host genetic factors that lead to enhanced drug metabolisms and/or reduced bioavailability need to be investigated as they are likely to contribute to the higher frequency of strains with borderline resistance mutations in the region [206–208]. High-dose regimens can only be applied when both pathogen and host genetics/phenotypes are available.

Still, the high prevalence of MTBC strains in Sierra Leone and Namibia with RR/MDR- underlines their significance for the drug resistance TB epidemiology. Moreover, since strains with borderline INH resistance mutations can also develop higher level resistance by a second mutation [150], it is essential to have molecular methods for resistance detection to keep track of low-level resistance mutations to prevent misclassification of resistant strains and reduce treatment failure.

Three strains with resistance mutations outside the RRDR (*rpoB* I491V & V170F) were detected. These mutations are not detected by commercial molecular RR tests such as Xpert MTB/RIF assay [172,184,185]. Such strains have caused severe MDR-TB control problems in other African countries such as Eswatini [209]. Because the most widely used Xpert MTB/RIF assay does not detect these strains as resistant, patients are treated with drug-susceptible regimens resulting in high rates of treatment failures and enhanced transmission.

4.1.1. Limitations

This study was limited by the small sample size and lack of metadata to help make associations between the isolates. It was not possible to determine the relatedness of individuals infected by the 4.8 mainly T MDR-TB clone or in the large Beijing cluster, because some patients had missing data such as town/state and family history of TB. All the samples included in the study were from the national TB reference laboratory in Freetown, as such the results from the study are only based on patients who had access to the hospitals and clinics in Sierra Leone.

4.1.2. Conclusions

The data indicate that resistance to Group A, B, and C MDR-TB treatment drugs is limited in Sierra Leone. Hence, the short all-oral-6-month BPaLM or the recently proposed 6-9 month or 9-12-month regimens offer great potential to treat most MDR-TB cases in the country successfully. The data on the MTBC strains' population structure point toward the potential importance of ancestral Beijing strains for the MDR-TB epidemic in Africa. Also demonstrated is that MDR *Maf* strains contribute significantly to the MDR-TB burden in the country. As such the strains of these endemic lineages are still important and having the strains profiles could guide diagnostics.

Understanding the lineage and strain diversity within a country is important for several reasons. Firstly, it helps in characterizing the local epidemiology of TB and identifying high-transmission clusters. The information is vital for designing effective control measures and interventions tailored to the specific strains circulating in the region. Secondly, it provides insights into the potential impact of strain diversity on disease severity, treatment outcomes, and the development of drug resistance. These findings underscore the need for comprehensive studies to characterize the local strain diversity and its implications for TB efforts. By understanding the genetic diversity and transmission dynamics of DR strains, healthcare systems (in Sierra Lone and the West African region) can implement more effective strategies to prevent the spread of resistant strains, improve treatment outcomes, and ultimately reduce the burden of TB. These can include the incorporation of WGS and targeted sequencing into their routine diagnostics.

4.2. Strains of Lineage 4: Cameroon sublineage

The strains of the Cameroon sublineage contribute significantly to the burden of TB in the West African region [210]. In Ghana it was reported to be the most prevalent *Mtb* sublineage in circulation [211]. An investigation into the global population structure and phylogeography of the strains of the Cameroon sublineage was conducted using WGS. Its distribution across different regions, populations, and countries was explored, and the factors that contribute to its prevalence, as well as the role of these strains in shaping TB dynamics, was assessed. WGS data were also used to identify clades and identify transmission and drug resistance patterns within these clades.

From a cohort of 1,133 Cameroon sublineage strains, 444 were selected based on strains that were more than five SNPs from one another and used for the clade designation. These strains originating from 24 countries across seven UN regions, were categorized into eight clades and signature SNPs specific to each group were defined. The clades were defined by tree topology, UN regions of origin of the infected host, and nested populations based on the RhierBAPs algorithm. Notably, high clustering rates observed in clades C.5 and C.8 (89% and 62%, respectively) underline significant patient-to-patient transmission, suggesting robust transmission dynamics. Furthermore, a quarter of these strains were identified as part of transmission chains, including those from sampled outbreaks, which may influence the overall data interpretation. The geographical analysis shows the presence of these strains in major regions including West Africa, Central Africa and Western Europe, indicating a spread beyond traditional confines of the African continent. The study also revealed concerning level of drug resistance, with over 10% of the strains showing MDR, though most remain susceptible to newer drugs likes to BDQ/ CFZ (two strains were resistant). The findings from this study provide a detailed view of the strains of the Cameroon sublineage to improve TB surveillance not just on the African continent, but in Europe and Asia as well.

4.2.1. Strains of the Cameroon sublineage

To gain deeper insights into the geographic distribution, genetic diversity, and potential impact of the strains of the Cameroon sublineage on TB dynamics, a comprehensive exploration of its phylogeography was taken. The strains of this sublineage exhibited a widespread presence in both the Central and West African regions, as well as the Western European regions, encompassing all eight newly defined strains of clades. This suggest that the origin of the Cameroon sublineage strains may be traced back to one of these regions.

The designation of the strains within the Cameroon sublineage, as the "Cameroon family" initially stemmed from their identification in the country of Cameroon [212]. They were characterized by specific spoligotypes lacking spacers #23, #24, and #25, and were primarily found in countries across WA, including Benin, and Senegal [212]. However, the reason behind the limited distribution of these strains to West and Central Africa remained unknown [107].

It is hypothesized that the introduction and widespread implementation of the BCG-TB vaccine may have played a role in the selection of the strains belonging to the Beijing family, potentially leading to a reduction in TB cases caused by *Maf* strains and increased the prevalence of the strains of the Cameroon sublineage in the West and Central African regions [109,212–214]. Previous studies have indicated that *Maf* strains accounted for over 50 % of TB cases in these areas. As *Maf* strains share close genetic relatedness with *M. bovis* strains, the vaccine may have inadvertently favored the selection of a strain resistant to the vaccine [212].

The findings from this study, along with others, highlight the establishment of specific niches of the Cameroon sublineage strains in Ghana and Cameroon [109,212]. Certain clades, such as the C.7 in Ghana and the C.8 in Cameroon, have demonstrated distinct preferences for the specific countries. Additionally, more modern strains within the Cameroon sublineage have formed niches in countries such as the UK, Gabon, and The Gambia. These observations underscore the complex and evolving nature of the phylogeography of the Cameroon sublineage strains, indicating their ability to adapt and thrive in different geographic regions.

4.2.2. The role of migration in the spread of strains of the Cameroon sublineage

The phylogeography of the strains of the Cameroon sublineage provides valuable insights into the role of migration in the spread and dissemination of this specific strains of the lineage. The presence of the Cameroon sublineage strains in both Central and West African regions, as well as European regions, suggests the movement and migration of individuals carrying these strains across geographical boundaries.

The Cameroon sublineage strains, which are predominantly found in West and Central Africa, have shown signs of spreading to other parts of the world, including Europe [107]. Previous studies have consistently reported the prevalence of these strains in countries such as Cameroon, Ghana, Gabon, and The Gambia, corroborating the findings of Stucki *et al.* in 2016 [107,109,156]. However, this study reveals the presence of TB infections caused by the Cameroon sublineage in Western European countries, particularly Germany and The United Kingdom (UK).

The expansion of the Cameroon sublineage into the Asian and European countries can be attributed to international travel and migration [215]. With the ease and speed of modern transportation, it now takes only a day or less to travel from African countries to destinations in Asia or Europe. For example, the travel time from Ghana to the UK is approximately 6 h, while it takes around 20 h to reach China [216]. Furthermore, significant populations from Ghana and Cameroon reside in Germany, particularly in Hamburg. Also, there is a large Ghanaian community in UK, evident by the presence of their own market, Makola, similar to the one in Ghana [217,218].

The role of migration in TB transmission has been a subject of debate. Some studies suggest that TB cases in high-income countries are a result of reactivation of LTBI among foreign-born migrants. However other studies have challenged this claim [219,220]. For instance, a study conducted in São Paulo, Brazil, investigated the association between South American migrants and TB transmission, revealing a higher prevalence of the disease among Brazilians rather than migrants [221]. Similarly, a study encompassing 28 European Union countries, Iceland and Norway aimed to explore the correlation between the number of immigrants and TB notification per capita. The authors found a significantly negative correlation between the number of immigrants

and TB notifications in high-incidence countries, although there were a few exceptions. Germany, Italy, and Norway exhibited a significant positive correlation, indicating a complex relationship between immigration and TB incidence [222].

4.2.3. Transmission of strains of the Cameroon clades

The role of migration in the spread of the Cameroon sublineage provides crucial insights into understanding the transmission dynamics of its distinct clades. As these strains are introduced to new regions through human migration, it becomes essential to investigate how transmission occurs within local populations.

To explore the transmission patterns of the different strains of the clades, the study focussed on defining transmission clusters and outbreaks. Previous research has established that strains separated by five or fewer SNPs are considered outbreak strains, indicating recent transmission events [223]. In this study, the strains of the Cameroon sublineage exhibited a high clustering rate of 71%, similar to the highly transmitted Beijing lineage strains [224]. Notably, each clade (C.1 to C.8) had a significant proportion of clustered strains, suggesting ongoing transmission.

Several major TB outbreaks involving strains of the Cameroon sublineage were identified in countries such as Gabon, The Gambia, Ghana, Senegal, Germany, and the UK. While most transmission occurred within individual countries, a few transmission clusters were observed across borders. For example, strains of clade C.5 exhibited a prominent cluster with strains found in both the UK and Ghana. This could be attributed to international travel, as the UK is a frequent destination for individuals from Ghana [225,226]. However, a closer examination revealed that the strains from Ghana and the UK belonged to different branches and emerged at different times. Additionally, the strains developed different DR profiles, with strains from Ghana being MDR-TB and those from the UK being Hr-TB. This suggests the possibility of reactivation of LTBI rather than direct transmission of strains between the two countries [220]. Other studies have also reported very low to no contribution of cross-transmission of TB from migrants to the indigenous population [221,222].

Strains of clade C.5 were also involved in another transmission event, this time spanning three countries: Cameroon, Gabon, and Germany. The presence of strains from Cameroon and Germany suggests transmission through international travel. And the transmission between Gabon and Cameroon could be facilitated by their geographic proximity. The strains of clade C.5 demonstrate a high transmissibility, capable of infecting individuals not only in West and Central Africa but also in European countries.

Strains belonging to clade C.8 also showed high potential for transmissibility, with a cluster rate of 62%. The strains from the clade formed large clusters, and a transmission event involving strains from Germany, Ghana, and Guinea Bissau. International travel or migration could potentially explain the cross-country transmission observed. It is noteworthy that strains of clade C.8 consist of more recent strains, which may explain their ability to infect a broader range of individuals compared to the older clades. Similar observations have been reported in other MTBC lineage strains, where modern strains often display higher virulence and are more involved in outbreaks [227].

4.2.4. Drug resistance in strains of the Cameroon clades

The examination of drug resistance within specific clades is crucial to understanding the spread and evolution of drug resistance within the strains of the Cameroon sublineage. The high prevalence of drug resistance among strains of the Cameroon sublineage underscores the need to investigate the acquisition and maintenance of drug resistance mutations within this lineage.

In this study, it was found that a substantial proportion of the Cameroon sublineage strains exhibited drug resistance. Approximately, 10% of the strains were identified as MDR, while 43% were nonMDR, consistent with the high drug resistance observed in the Beijing strains [228] [229]. Among the different strains of clades (C.1 to C.8) within the Cameroon sublineage, strains of clades C.5, C.6, C.7, and C.8 showed the highest rate of MDR. The most predominant mutation was the *fabG1*-15c>t INH resistance mutation. This a low-level resistance mutation implying that high dose or prolonged treatment with INH might be effective with these strains [230].

Regarding RIF resistance mutations, over 10% of the strains carried mutations in the *rpoB* gene, with the most prevalent mutations being *rpoB* S450L and H445D. These findings are in line with previous studies [150,172,231]. Notably, the MDR strains of clade C.5 from the UK exhibited the *rpoB* V170F resistance mutation, which is located outside the RRDR and cannot be detected by standard commercial molecular tests [172,184,185]. This presents a challenge in the diagnosis and treatment of these strains.

In each of these clades C.2, C.3, C.5, C.6, C.7, and C.8, at least one strain had developed resistance to at least one of first-line anti-TB drugs, meaning that the standard treatment for TB may not be effective against these strains. Meanwhile at least one strain of clade C.5 was resistant to all tested anti-TB drugs in this study, making the strains from this clade of concerning from a treatment and public health perspective. Two strains, one from clade C.5 and one from C.8, demonstrated resistance to BDQ. However, these are isolated cases, indicating that most Cameroon sublineage strains are still susceptible to BDQ, thus offering potential treatment options for drug resistance [98].

One in four of the strains of clade C.2 is MDR. Particularly, one of the genetic markers for these MDR strains is a mutation in an efflux pump (EP) gene *Rv1410c.* This EP gene has been shown to be overexpressed in MDR-TB clinical strains [232] upon exposure to RIF and INH [233]. The over-expression of EP gene can reduce the intracellular concentrations of antibiotics, preventing them from reaching their target sites and resulting in drug resistance [234–236]. The differential expression of the *Rv1410c* gene between MDR and sensitive strains may have implications for MDR-TB [237]. The presence of DR strains highlights the need for continuous surveillance and monitoring of drug resistance patterns within this clade to ensure effective TB control and management.

4.2.5. Limitations

Limitations of this study include the sampling method and availability of data which may have introduced a potential bias in the results. The study was conducted at the RCB, which is a WHO supranational reference laboratory specializing in TB research. While this provided the advantage of access to strains of the Cameroon sublineage, it also limited the availability of strains to those

from regions and countries that collaborate with the RCB. Also, researchers in countries who do have the capacity for WGS may not have their samples on the ENA. Consequently, the sample may not represent the full diversity and distribution of the strains of the Cameroon sublineage globally.

Furthermore, the limitations of the initial studies from which the data was obtained also impacted the available metadata and sample information. Not all articles included sampling dates, and the country of birth information for all patients was either provided or not available. These missing or incomplete data may have affected the accuracy and representativeness of the findings.

4.2.6. Outlook

In this study, the clades of the strains of the Cameroon sublineage were successfully defined, providing valuable insights into the genetic diversity and distribution of these strains. Further studies should consider investigating the growth rate of the different clades within the Cameroon sublineage to determine if there are any variations in their replication dynamics.

Additionally, infection model studies using monocyte derived macrophages (MDM) and (polymorphonuclear) leukocytes PMNs from individuals representing the prevalent regions identified in this study could be conducted. These would enable the measurement of cytokine expression profiles, gene expression patterns, and other relevant parameters specific to each clade. By examining how the strains of different clades interact with host immune cells and how they induce immune responses, a better understanding of their adaptation to specific populations would be obtained.

4.2.7. Conclusion

In this study, the classification of strains from the Cameroon sublineage into eight distinct clades has provided important insights into their genetic diversity and distribution. While these strains are predominantly found in West and Central Africa, it is noteworthy that they are also causing disease in Western and Northern Europe. The high levels of drug resistance observed in these

strains, coupled with their susceptibility to the novel drug BDQ, present both challenges and potential treatment options.

In conclusion, the transmission dynamics within the strains of the Cameroon sublineage highlight the significant role of migration in the spread of these strains. The identification of transmission clusters and outbreaks provides valuable insights into the local and global dissemination of the strains of the Cameroon sublineage. The occurrence of cross-country transmission events emphasizes the role of international travel and migration in facilitating the spread of these strains. Furthermore, the transmissibility of specific strains of clades, such as C.5 and C.8, suggests their potential to infect individuals across different regions, including both African and European countries. These findings contribute to a better understanding of TB transmission patterns and can inform targeted interventions for effective control and prevention strategies.

5. References

1. Luies, L. & Du Preez, I. The Echo of Pulmonary Tuberculosis: Mechanisms of Clinical Symptoms and Other Disease-Induced Systemic Complications. (2020) doi:10.1128/CMR.00036-20.

2. Acharya, B. *et al.* Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. *Molecular Biology Reports* vol. 47 4065–4075 Preprint at https://doi.org/10.1007/s11033-020-05413-7 (2020).

3. Enarson, Donald. & International Union against Tuberculosis and Lung Disease. Management of tuberculosis : a guide for low income countries. 91 (2000).

4. World Health Organization. Tuberculosis. *World Health Organization Newsroom* https://www.who.int/news-room/fact-sheets/detail/tuberculosis (2023).

5. Center for Disease Control. Tuberculosis: Basic TB Facts. https://www.cdc.gov/tb/topic/basics/default.htm (2016).

6. World Health Organization. Implementing the WHO Stop TB strategy: A handbook for National Tuberculosis Control ProgrammesTreatment of tuberculosis patients. (2008).

7. World Health Organization. *Rapid Communication: Key Changes to the Treatment of Drug-Resistant Tuberculosis*. (2022).

8. World Health Organization. *Global Tuberculosis Report 2023*. https://iris.who.int/. (2023).

9. World Health Organization. Global Tuberculosis report 2022. (2022).

10. World Health Organization. *The End TB Strategy*. *Wo* (2015).

11. Villar-Hernández, R. *et al.* Tuberculosis: current challenges and beyond. *Breathe* **19**, (2023).

12. Global Tuberculosis Programme. https://www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy.

13. World Health Organization. TB profile: Sierra Leone. https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22SL%22.

14. Kamara, R. F. *et al.* Social and health factors associated with adverse treatment outcomes among people with multidrug-resistant tuberculosis in Sierra Leone: a national, retrospective cohort study. *Lancet Glob Health* **10**, e543 (2022).

15. Lakoh, S. & Yendewa, G. A. Multidrug-resistant tuberculosis in Sierra Leone. *Lancet Glob Health* **10**, e459–e460 (2022).

16. Wirth, T. *et al.* Origin, Spread and Demography of the Mycobacterium tuberculosis Complex. *PLoS Pathog* **4**, e1000160 (2008).

17. Cook, G. M. *et al.* Physiology of Mycobacteria. *Adv Microb Physiol* **55**, 81 (2009).

18. The Central TB Division. *Revised National TB Control Programme Training Manual for Mycobacterium Tuberculosis Culture and Drug Susceptibility Testing*. (New Delhi, 2009).

19. Percival, S. L., Yates, M. V., Williams, D. W., Chalmers, R. M. & Gray, N. F. *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks: Second Edition*. *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks: Second Edition* (Elsevier Ltd, 2013). doi:10.1016/C2010-0-67101-X.

20. Gunasingam, N. Morphology and Pathological Characteristics of Mycobacteria. *Mycobacterial Diseases* **0**, 1–1 (2022).

21. Oosthuizen, C. B., Reid, A. M. & Lall, N. Can Medicinal Plants Provide an Adjuvant for Tuberculosis Patients? *Medicinal Plants for Holistic Health and Well-Being* 213–253 (2018) doi:10.1016/B978-0-12-812475-8.00007-X.

22. McMurray, D. N. Mycobacteria and Nocardia. in *Laboratory Procedures in Clinical Microbiology* 379–418 (University of Texas Medical Branch at Galveston, 1996). doi:10.1007/978-1-4612-5070-8_6.

23. Antonini, J. M. Sarcoidosis☆. *Reference Module in Biomedical Sciences* (2014) doi:10.1016/B978-0-12-801238-3.05290-9.

24. Payeur, J. B. Mycobacterium. *Encyclopedia of Food Microbiology: Second Edition* 841–853 (2014) doi:10.1016/B978-0-12-384730-0.00229-9.

25. Garg, R. K., Malhotra, H. S. & Gupta, R. Spinal cord involvement in tuberculous meningitis. *Spinal Cord* vol. 53 649–657 Preprint at https://doi.org/10.1038/sc.2015.58 (2015).

26. Lee, J. Y. Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberculosis and Respiratory Diseases* vol. 78 47–55 Preprint at https://doi.org/10.4046/trd.2015.78.2.47 (2015).

27. Asghar, M. U., Mehta, S. S., Cheema, H. A., Patti, R. & Pascal, W. Sputum Smear and Culture-negative Tuberculosis with Associated Pleural Effusion: A Diagnostic Challenge. *Cureus* **10**, (2018).

28. Calis, J., Bakker, M., Elens, R., Borgdorff, M. & Harries, A. Mortality in smear-negative tuberculosis patients in Phalombe. *Malawi Med J* **14**, 13 (2002).

29. Lee, S. H. Tuberculosis Infection and Latent Tuberculosis. *Tuberc Respir Dis (Seoul)* **79**, 201 (2016).

30. Pieters, J. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. *Cell Host Microbe* **3**, 399–407 (2008).

31. Ryndak, M. B. & Laal, S. Mycobacterium tuberculosis Primary Infection and Dissemination: A Critical Role for Alveolar Epithelial Cells. *Front Cell Infect Microbiol* **9**, 299 (2019).

32. Moule, M. G. & Cirillo, J. D. Mycobacterium tuberculosis Dissemination Plays a Critical Role in Pathogenesis. *Front Cell Infect Microbiol* **10**, 65 (2020).

33. Cambier, C. J., Falkow, S. & Ramakrishnan, L. Host Evasion and Exploitation Schemes of Mycobacterium tuberculosis. *Cell* **159**, 1497–1509 (2014).

34. Vyas, J. M. Pulmonary tuberculosis. *MedlinePlus Medical Encyclopedia* https://medlineplus.gov/ency/article/000077.htm (2022).

35. Maleche, M. A. *et al.* A case report of a 16-year-old male with extrapulmonary tuberculosis presenting as multiloculated mediastinal, pleural, and paravertebral fluid collections with chest pain, Eldoret, Kenya. *Clin Case Rep* **11**, e7574 (2023).

36. Jawed, A. *et al.* Better understanding extrapulmonary tuberculosis: A scoping review of public health impact in Pakistan, Afghanistan, India, and Bangladesh. *Health Sci Rep* **6**, (2023).

37. Kiazyk, S. & Ball, T. Tuberculosis (TB): Latent tuberculosis infection: An overview. *Canada Communicable Disease Report* **43**, 62 (2017).

38. Brodie, D. & Schluger, N. W. The diagnosis of tuberculosis. *Clinics in Chest Medicine* vol. 26 247–271 Preprint at https://doi.org/10.1016/j.ccm.2005.02.012 (2005).

39. Fact Sheets | General | Latent TB Infection vs. TB Disease | TB | CDC. https://www.cdc.gov/tb/publications/factsheets/general/ltbiandactivetb.htm.

40. Luies, L. & Preez, I. du. The Echo of Pulmonary Tuberculosis: Mechanisms of Clinical Symptoms and Other Disease-Induced Systemic Complications. *Clin Microbiol Rev* **33**, 1–19 (2020).

41. Parsons, L. M. *et al.* Laboratory Diagnosis of Tuberculosis in Resource-Poor Countries: Challenges and Opportunities. *Clin Microbiol Rev* **24**, 314 (2011).

42. Desikan, P. Sputum smear microscopy in tuberculosis: Is it still relevant? *Indian J Med Res* **137**, 442 (2013).

43. Drug-Resistant TB | TB |CDC. https://www.cdc.gov/tb/topic/drtb/default.htm.

44. Coll, F. *et al.* Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* **7**, (2015).

45. World Health Organization. *GLOBAL TUBERCULOSIS REPORT 2021*. (2021).

46. Jang, J. G. & Chung, J. H. Diagnosis and treatment of multidrug-resistant tuberculosis. *Yeungnam Univ J Med* **37**, 277 (2020).

47. Beale, J. Drug-resistant tuberculosis. *J R Soc Med* **94**, 261–264 (2001).

48. Institute of Medicine (US) Forum on Drug Discovery, D. and T. & Science, R. A. of M. Transmission and Infection Control of Drug-Resistant TB. (2011).

49. Leung, E. C. C. *et al.* Transmission of multidrug-resistant and extensively drug-resistant tuberculosis in a metropolitan city. *European Respiratory Journal* **41**, 901–908 (2013).

50. World Health Organization. *WHO Consolidated Guidelines on Drug-Resistant Tuberculosis Treatment.* vol. End TB stragety (2019).

51. World Health Organization. WHO announces updated definitions of extensively drug-resistant tuberculosis. *World Health Organization* https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis (2021).

52. Yusoof, K. A. *et al.* Tuberculosis Phenotypic and Genotypic Drug Susceptibility Testing and Immunodiagnostics: A Review. *Front Immunol* **13**, (2022).

53. Amini, S. *et al.* Direct drug susceptibility testing of Mycobacterium tuberculosis using the proportional method: A multicenter study. *J Glob Antimicrob Resist* **17**, 242–244 (2019).

54. Brett, K., Dulong, C. & Severn, M. Identification of Tuberculosis: A Review of the Guidelines. *Identification of Tuberculosis: A Review of the Guidelines* (2020).

55. Otchere, I. D. *et al.* Analysis of drug resistance among difficult-to-treat tuberculosis patients in Ghana identifies several pre-XDR TB cases. *Front Microbiol* **13**, 1069292 (2023).

56. Albert, H., Trollip, A. P., Mole, R. J., Hatch, S. J. B. & Blumberg, L. Rapid indication of multidrug-resistant tuberculosis from liquid cultures using FASTPlaqueTB-RIF, a manual phage-based test. *Int J Tuberc Lung Dis* **6**, 523–528 (2002).

57. Fan, X. Y. *et al.* Rapid detection of rpoB gene mutations in rifampin-resistant Mycobacterium tuberculosis isolates in Shanghai by using the amplification refractory mutation system. *J Clin Microbiol* **41**, 993–997 (2003).

58. Fox, G. J. *et al.* Preventing the spread of multidrug-resistant tuberculosis and protecting contacts of infectious cases. *Clinical Microbiology and Infection* **23**, 147–153 (2017).

59. World Health Organization. *Catalogue of Mutations in Mycobacterium Tuberculosis Complex and Their Association with Drug Resistance*. (2021).

60. Uddin, M. K. M. *et al.* Distribution and Frequency of rpoB Mutations Detected by Xpert MTB/RIF Assay Among Beijing and Non-Beijing Rifampicin Resistant Mycobacterium tuberculosis Isolates in Bangladesh. *Infect Drug Resist* **13**, 789 (2020).

61. Cooksey, R. C., Morlock, G. P., Glickman, S. & Crawford, J. T. Evaluation of a line probe assay kit for characterization of rpoB mutations in rifampin-resistant Mycobacterium tuberculosis isolates from New York City. *J Clin Microbiol* **35**, 1281–1283 (1997).

62. Hirano, K., Abe, C. & Takahashi, M. Mutations in the rpoB Gene of Rifampin-Resistant Mycobacterium tuberculosis Strains Isolated Mostly in Asian Countries and Their Rapid Detection by Line Probe Assay. *J Clin Microbiol* **37**, 2663 (1999).

63. Mäkinen, J., Marttila, H. J., Marjamäki, M., Viljanen, M. K. & Soini, H. Comparison of Two Commercially Available DNA Line Probe Assays for Detection of Multidrug-Resistant Mycobacterium tuberculosis. *J Clin Microbiol* **44**, 350 (2006).

64. Lawn, S. D. & Nicol, M. P. Xpert® MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiol* **6**, 1067–1082 (2011).

65. Saeed, M. *et al.* GeneXpert technology: A breakthrough for the diagnosis of tuberculous pericarditis and pleuritis in less than 2 hours. *Saudi Med J* **38**, 699 (2017).

66. Helb, D. *et al.* Rapid detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol* **48**, 229–237 (2010).

67. Boehme, C. C. *et al.* Rapid Molecular Detection of Tuberculosis and Rifampicin Resistance. *New England Journal of Medicine* **363**, 1005–1020 (2010).

68. Wu, Z. *et al.* Effect of the Xpert MTB/RIF on the detection of pulmonary tuberculosis cases and rifampicin resistance in Shanghai, China. *BMC Infect Dis* **20**, 1–10 (2020).

69. Feliciano, C. S. *et al.* Xpert MTB/RIF performance to diagnose tuberculosis and rifampicin resistance in a reference centre in southern Brazil. *ERJ Open Res* **5**, (2019).

70. Boehme, C. C. *et al.* Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* **377**, 1495–1505 (2011).

71. Pinto, L. M. & Udwadia, Z. F. Xpert MTB/RIF and pulmonary tuberculosis: time to delve deeper? *Thorax* **68**, 987–988 (2013).

72. Nathavitharana, R. R. *et al.* Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal* **49**, 1601075 (2017).

73. Padilla, E. *et al.* Comparative evaluation of the new version of the INNO-LiPA Mycobacteria and genotype Mycobacterium assays for identification of Mycobacterium species from MB/BacT liquid cultures artificially inoculated with Mycobacterial strains. *J Clin Microbiol* **42**, 3083–3088 (2004).

74. Gröschel, M. I. *et al.* Pathogen-based precision medicine for drug-resistant tuberculosis. *PLoS Pathog* **14**, (2018).

75. Domínguez, J. *et al.* Clinical implications of molecular drug resistance testing for Mycobacterium tuberculosis: a 2023 TBnet/RESIST-TB consensus statement. *Lancet Infect Dis* **23**, e122–e137 (2023).

76. Yin, R., Kwoh, C. K. & Zheng, J. Whole Genome Sequencing Analysis. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* **1–3**, 176–183 (2019).

77. Votintseva, A. A. *et al.* Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. *J Clin Microbiol* **55**, 1285–1298 (2017).

78. Witney, A. A. *et al.* Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases. *J Clin Microbiol* **53**, 1473–1483 (2015).

79. Amlerova, J., Bitar, I. & Hrabak, J. Genotyping of Mycobacterium tuberculosis using whole genome sequencing. *Folia Microbiol (Praha)* **63**, 537–545 (2018).

80. Guthrie, J. L. *et al.* Genotyping and Whole-Genome Sequencing to Identify Tuberculosis Transmission to Pediatric Patients in British Columbia, Canada, 2005-2014. *J Infect Dis* **218**, 1155–1163 (2018).

81. Gautam, S. S. *et al.* Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. *PLoS One* **13**, (2018).

82. Buermans, H. P. J. & den Dunnen, J. T. Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta* **1842**, 1932–1941 (2014).

83. Pankhurst, L. J. *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med* **4**, 49–58 (2016).

84. Doughty, E. L., Sergeant, M. J., Adetifa, I., Antonio, M. & Pallen, M. J. Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* **2**, (2014).

85. Lee, R. S. & Pai, M. Real-Time Sequencing of Mycobacterium tuberculosis: Are We There Yet? *J Clin Microbiol* **55**, 1249–1254 (2017).

86. Kohl, T. A. *et al.* MTBseq: A comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. *PeerJ* **2018**, (2018).

87. Feuerriegel, S. *et al.* PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. *J Clin Microbiol* **53**, 1908–1914 (2015).

88. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* **6**, (2015).

89. World Health Organization. *Catalogue of Mutations in Mycobacterium Tuberculosis Complex and Their Association with Drug Resistance Second Edition*. (2023).

90. Dookie, N., Khan, A., Padayatchi, N. & Naidoo, K. Application of Next Generation Sequencing for Diagnosis and Clinical Management of Drug-Resistant Tuberculosis: Updates on Recent Developments in the Field. *Front Microbiol* **13**, 775030 (2022).

91. de Araujo, L. *et al.* Implementation of targeted next-generation sequencing for the diagnosis of drug-resistant tuberculosis in low-resource settings: a programmatic model, challenges, and initial outcomes. *Front Public Health* **11**, 1204064 (2023).

92. Mansoor, H. *et al.* Clinical utility of target-based next-generation sequencing for drug-resistant TB. *The International Journal of Tuberculosis and Lung Disease* **27**, 41 (2023).

93. Dilliott, A. A. *et al.* Targeted Next-generation Sequencing and Bioinformatics Pipeline to Evaluate Genetic Determinants of Constitutional Disease. *J Vis Exp* **2018**, (2018).

94. World Health Organization. *WHO Consolidated Guidelines on Tuberculosis. Module 4, Treatment : Drug-Resistant Tuberculosis Treatment.* (2020).

95. World Health Organization. *WHO Consolidated Guidelines on Tuberculosis Module 4: Treatment Drug-Susceptible Tuberculosis Treatment*. (2022).

96. Dorman, S. E. *et al.* Four-Month Rifapentine Regimens with or without Moxifloxacin for Tuberculosis. *N Engl J Med* **384**, 1705–1718 (2021).

97. World Health Organization. *Module 4: Treatment WHO Operational Handbook on Tuberculosis Drug-Resistant Tuberculosis Treatment 2022 Update*. (2022).

98. Haley, C. A. *et al.* Novel 6-Month Treatment for Drug-Resistant Tuberculosis, United States. *Emerg Infect Dis* **27**, 332 (2021).

99. Conradie, F. *et al.* Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N Engl J Med* **382**, 893–902 (2020).

100. Todd, E. C. D. Mycobacterium bovis. *Encyclopedia of Food Safety* 189–200 (2024) doi:10.1016/B978-0-12-822521-9.00064-2.

101. Yeboah-Manu, D. *et al.* Spatio-Temporal Distribution of Mycobacterium tuberculosis Complex Strains in Ghana. *PLoS One* **11**, e0161892 (2016).

102. Coscolla, M. & Gagneux, S. Consequences of genomic diversity in mycobacterium tuberculosis. *Seminars in Immunology* vol. 26 431–444 Preprint at https://doi.org/10.1016/j.smim.2014.09.012 (2014).

103. Ngabonziza, J. C. S. *et al.* A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nat Commun* **11**, (2020).

104. Brites, D. *et al.* A New Phylogenetic Framework for the Animal-Adapted Mycobacterium tuberculosis Complex. *Front Microbiol* **9**, (2018).

105. Coll, F. *et al.* A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nature Communications 2014 5:1* **5**, 1–5 (2014).

106. Coscolla, M. *et al.* Phylogenomics of mycobacterium africanum reveals a new lineage and a complex evolutionary history. *Microb Genom* **7**, 1–14 (2021).

107. Stucki, D. *et al.* Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* **48**, 1535–1543 (2016).

108. Niemann, S., Merker, M., Kohl, T. & Supply, P. Impact of Genetic Diversity on the Biology of Mycobacterium tuberculosis Complex Strains . *Microbiol Spectr* **4**, (2016).

109. Asare, P. *et al.* Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. *International Journal of Infectious Diseases* **73**, 30–42 (2018).

110. Thumamo Pokam, B. D. *et al.* A pilot study on the genetic diversity of Mycobacterium tuberculosis complex strains from tuberculosis patients in the Littoral region of Cameroon. *J Clin Tuberc Other Mycobact Dis* **21**, (2020).

111. Koro, F. K. *et al.* Genetic Structure and Drug Susceptibility Patterns of Mycobacterium tuberculosis Complex Strains Responsible of Human Pulmonary Tuberculosis in the Major Rearing Region in Cameroon. *Biomed Res Int* **2016**, (2016).

112. Shaw, B., von Bredow, B., Tsan, A., Garner, O. & Yang, S. Clinical Whole-Genome Sequencing Assay for Rapid Mycobacterium tuberculosis Complex First-Line Drug Susceptibility Testing and Phylogenetic Relatedness Analysis. *Microorganisms* **11**, 2538 (2023).

113. Foxman, B. & Riley, L. Molecular epidemiology: focus on infection. *Am J Epidemiol* **153**, 1135–1141 (2001).

114. Guernier, V., Sola, C., Brudey, K., Guégan, J.-F. & Rastogi, N. Use of cluster-graphs from spoligotyping data to study genotype similarities and a comparison of three indices to quantify recent tuberculosis transmission among culture positive cases in French Guiana during a eight year period. *BMC Infect Dis* **8**, 46 (2008).

115. Jagielski, T. *et al.* Current methods in the molecular typing of mycobacterium tuberculosis and other Mycobacteria. *Biomed Res Int* **2014**, (2014).

116. M Cristina, G. *et al.* Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis. *PLoS Pathog* **1**, 0055–0061 (2005).

117. Kato-Maeda, M., Metcalfe, J. Z. & Flores, L. Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. *Future Microbiol* **6**, 203 (2011).

118. Asare, P., Asante-Poku, A., Osei-Wusu, S., Otchere, I. D. & Yeboah-Manu, D. The Relevance of Genomic Epidemiology for Control of Tuberculosis in West Africa. *Front Public Health* **9**, 706651 (2021).

119. Kato-Maeda, M., Metcalfe, J. Z. & Flores, L. Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. *Future Microbiol* **6**, 203 (2011).

120. Allix-Béguec, C., Harmsen, D., Weniger, T., Supply, P. & Niemann, S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. *J Clin Microbiol* **46**, 2692–2699 (2008).

121. Weniger, T., Krawczyk, J., Supply, P., Niemann, S. & Harmsen, D. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Res* **38**, (2010).

122. Meehan, C. J. *et al.* Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nature Reviews Microbiology 2019 17:9* **17**, 533–545 (2019).

123. Niemann, S. *et al.* Genomic Diversity among Drug Sensitive and Multidrug Resistant Isolates of Mycobacterium tuberculosis with Identical DNA Fingerprints. *PLoS One* **4**, (2009).

124. Roetzer, A. *et al.* Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS Med* **10**, (2013).

125. Walker, T. M. *et al.* Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet Infect Dis* **13**, 137–146 (2013).

126. Desikan, S. & Narayanan, S. Genetic markers, genotyping methods & next generation sequencing in Mycobacterium tuberculosis. *Indian Journal of Medical Research* **142**, 761–774 (2015).

127. Meehan, C. J. *et al.* Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nature Reviews Microbiology 2019 17:9* **17**, 533–545 (2019).

128. Asare, P., Asante-Poku, A., Osei-Wusu, S., Otchere, I. D. & Yeboah-Manu, D. The Relevance of Genomic Epidemiology for Control of Tuberculosis in West Africa. *Front Public Health* **9**, 706651 (2021).

129. Schürch, A. C. & van Soolingen, D. DNA fingerprinting of Mycobacterium tuberculosis: From phage typing to whole-genome sequencing. *Infection, Genetics and Evolution* **12**, 602–609 (2012).

130. van Belzen, I. A. E. M., Schönhuth, A., Kemmeren, P. & Hehir-Kwa, J. Y. Structural variant detection in cancer genomes: computational challenges and perspectives for precision oncology. *npj Precision Oncology 2021 5:1* **5**, 1–11 (2021).

131. Field, M. A. Bioinformatic Challenges Detecting Genetic Variation in Precision Medicine Programs. *Front Med (Lausanne)* **9**, 806696 (2022).

132. D, M., JD, J. & DJ, S. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**, 260–261 (2009).

133. Kent, P. T. & George, K. P. *Public Health Mycobacteriology A Guide For The Level III Laboratory*. (1985).

134. Van Soolingen, D., Hermans, P. W. M., De Haas, P. E. W., Soll, D. R. & Van Embden, J. D. A. 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**, 2578–2586 (1991).

135. Grobbel, H. P. *et al.* Design of Multidrug-Resistant Tuberculosis Treatment Regimens Based on DNA Sequencing. *Clin Infect Dis* **73**, 1194–1202 (2021).

136. Merker, M. *et al.* Compensatory evolution drives multidrug-resistant tuberculosis in central Asia. *Elife* **7**, (2018).

137. Baym, M. *et al.* Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. (2015) doi:10.1371/journal.pone.0128036.

138. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537–544 (1998).

139. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754 (2009).

140. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078 (2009).

141. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

142. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

143. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6* **14**, 587–589 (2017).

144. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**, 1188–1195 (2013).

145. Letunic, I. & Bork, P. *Interactive Tree Of Life (ITOL): An Online Tool for Phylogenetic Tree Display and Annotation*. http://itol.embl.de.

146. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res* **3**, (2018).

147. Chiang, S. S. & Starke, J. R. Mycobacterium tuberculosis. *Principles and Practice of Pediatric Infectious Diseases* 790-806.e5 (2018) doi:10.1016/B978-0-323-40181-4.00134-1.

148. Coll, F. *et al.* A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nat Commun* (2014) doi:10.1038/ncomms5812.

149. Miotto, P., Cabibbe, A. M., Borroni, E., Degano, M. & Cirilloa, D. M. Role of Disputed Mutations in the rpoB Gene in Interpretation of Automated Liquid MGIT Culture Results for Rifampin Susceptibility Testing of Mycobacterium tuberculosis. *J Clin Microbiol* **56**, (2018).

150. Claassens, M. *et al.* Whole-Genome Sequencing for Resistance Prediction and Transmission Analysis of Mycobacterium tuberculosis Complex Strains from Namibia. *Microbiol Spectr* **10**, (2022).

151. Lavender, C., Globan, M., Sievers, A., Billman-Jacobe, H. & Fyfe, J. Molecular Characterization of Isoniazid-Resistant Mycobacterium tuberculosis Isolates Collected in Australia. *Antimicrob Agents Chemother* **49**, 4068 (2005).

152. Acquah, S. K. E. *et al.* Molecular epidemiology and drug susceptibility profiles of Mycobacterium tuberculosis complex isolates from Northern Ghana. *International Journal of Infectious Diseases* **109**, 294–303 (2021).

153. Kone, B. *et al.* Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in referral health centers of Bamako, Mali: What is new? *International Journal of Infectious Diseases* **117**, 204–211 (2022).

154. Bakuła, Z. *et al.* Molecular snapshot of drug-resistant Mycobacterium tuberculosis strains from the Plateau State, Nigeria. *PLoS One* **17**, (2022).

155. Senghore, M. *et al.* Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. *Scientific Reports 2020 10:1* **10**, 1–9 (2020).

156. Sven Malm, L. S. G. New Mycobacterium tuberculosisComplex Sublineage, Brazzaville, Congo. doi:10.3201/eid2303.160679.

157. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. doi:10.1093/molbev/msz131.

158. Casali, N. *et al.* Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. *PLoS Med* **13**, e1002137 (2016).

159. Hoopes, A. J., Kammerer, J. S., Harrington, T. A., Ijaz, K. & Armstrong, L. R. Isoniazid-monoresistant tuberculosis in the United States, 1993 to 2003. *Arch Intern Med* **168**, 1984–1992 (2008).

160. Ruddy, M. C. *et al.* Outbreak of isoniazid resistant tuberculosis in north London. *Thorax* **59**, 279–285 (2004).

161. Barbier, M. & Wirth, T. The Evolutionary History, Demography, and Spread of the Mycobacterium tuberculosis Complex . *Microbiol Spectr* **4**, (2016).

162. Taune, M. *et al.* Successful implementation of bedaquiline for multidrug-resistant TB treatment in remote Papua New Guinea. *Public Health Action* **9**, S73–S79 (2019).

163. López, M. G. *et al.* Tuberculosis in liberia: High multidrug-resistance burden, transmission and diversity modelled by multiple importation events. *Microb Genom* **6**, (2020).

164. López, M. G. *et al.* Tuberculosis in Liberia: high multidrug-resistance burden, transmission and diversity modelled by multiple importation events. *Microb Genom* **6**, (2020).

165. Gehre, F. *et al.* The emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis research and drug resistance surveillance. (2016) doi:10.1186/s12916-016-0704-5.

166. Carter, B. B. *et al.* Survival analysis of patients with tuberculosis and risk factors for multidrug-resistant tuberculosis in Monrovia, Liberia. *PLoS One* **16**, (2021).

167. Dutschke, A. *et al.* Xpert MTB/RIF on urine samples to increase diagnosis of TB in people living with HIV in Guinea-Bissau. *International Journal of Infectious Diseases* **124**, S63–S68 (2022).

168. Desikan, P. *et al.* Line probe assay for detection of Mycobacterium tuberculosis complex: An experience from Central India. doi:10.4103/ijmr.IJMR_831_14.

169. WHO consolidated guidelines on tuberculosis. Module 3. Diagnosis : rapid diagnostics for tuberculosis detection. 160 (2024).

170. Mathys, V., Van De Vyvere, M., De Droogh, E., Soetaert, K. & Groenen, G. False-positive rifampicin resistance on Xpert®MTB/RIF caused by a silent mutation in the rpoB gene. *International Journal of Tuberculosis and Lung Disease* **18**, 1255–1257 (2014).

171. Heyckendorf, J. *et al.* What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi-and Extensively Drug-Resistant Tuberculosis MECHANISMS OF RESISTANCE crossm Downloaded from. **62**, 1550–1567 (2018).

172. Beckert, P. *et al.* MDR M. tuberculosis outbreak clone in Eswatini missed by Xpert has elevated bedaquiline resistance dated to the pre-treatment era. *Genome Med* **12**, 1–11 (2020).

173. Forson, A. *et al.* A cross-sectional study of tuberculosis drug resistance among previously treated patients in a tertiary hospital in Accra, Ghana: Public health implications of standardized regimens. *BMC Infect Dis* **18**, 1–6 (2018).

174. Muhammad Dayyab, F. *et al.* Emerging threat of drug-resistant tuberculosis and trends in the era of COVID-19: A descriptive study from northwestern Nigeria. *J Clin Tuberc Other Mycobact Dis* **28**, (2022).

175. Dreyer, V. *et al.* High fluoroquinolone resistance proportions among multidrug-resistant tuberculosis driven by dominant L2 Mycobacterium tuberculosis clones in the Mumbai Metropolitan Region. *Genome Med* **14**, 95 (2022).

176. World Health Organization. *Rapid Communication: Key Changes to the Treatment of Drug-Resistant Tuberculosis*. http://apps.who.int/bookorders. (2019).

177. Van Rie, A. *et al.* Balancing access to BPaLM regimens and risk of resistance. *Lancet Infect Dis* **22**, 1411–1412 (2022).

178. Omar, S. V, Ismail, F., Ndjeka, N., Kaniga, K. & Ismail, N. A. Bedaquiline-Resistant Tuberculosis Associated with Rv0678 Mutations. *N Engl J Med* **386**, (2022).

179. Guo, Q. *et al.* Whole Genome Sequencing Identifies Novel Mutations Associated With Bedaquiline Resistance in Mycobacterium tuberculosis. *Front Cell Infect Microbiol* **12**, 640 (2022).

180. Ismail, N. A. *et al.* Assessment of epidemiological and genetic characteristics and clinical outcomes of resistance to bedaquiline in patients treated for rifampicin-resistant tuberculosis: a cross-sectional and longitudinal study. *Lancet Infect Dis* **22**, 496–506 (2022).

181. Chesov, E. *et al.* Emergence of bedaquiline resistance in a high tuberculosis burden country. *European Respiratory Journal* **59**, (2022).

182. Sonnenkalb, L. *et al.* Bedaquiline and clofazimine resistance in Mycobacterium tuberculosis: an in-vitro and in-silico data analysis. *Lancet Microbe* **4**, e358–e368 (2023).

183. Sanchez-Padilla, E. *et al.* Detection of drug-resistant tuberculosis by Xpert MTB/RIF in Swaziland. *N Engl J Med* **372**, 1181–1182 (2015).

184. Makhado, N. A. *et al.* Outbreak of multidrug-resistant tuberculosis in South Africa undetected by WHO-endorsed commercial tests: an observational study. *Lancet Infect Dis* **18**, 1350–1359 (2018).

185. Nimmo, C. *et al.* Bedaquiline resistance in drug-resistant tuberculosis HIV co-infected patients. *Eur Respir J* **55**, (2020).

186. Chesov, E. *et al.* Emergence of bedaquiline resistance in a high tuberculosis burden country. *Eur Respir J* **59**, (2022).

187. Kumar, S. & Jena, L. Understanding Rifampicin Resistance in Tuberculosis through a Computational Approach. *Genomics Inform* **12**, 276 (2014).

188. Munir, A. *et al.* Identification and Characterization of Genetic Determinants of Isoniazid and Rifampicin Resistance in Mycobacterium tuberculosis in Southern India. *Sci Rep* **9**, (2019).

189. Sethi, S. *et al.* Elucidation of drug resistance mutations in Mycobacterium tuberculosis isolates from North India by whole-genome sequencing. *J Glob Antimicrob Resist* **20**, 11–15 (2020).

190. Ajawatanawong, P. *et al.* A novel Ancestral Beijing sublineage of Mycobacterium tuberculosis suggests the transition site to Modern Beijing sublineages. *Sci Rep* **9**, (2019).

191. Merker, M. *et al.* Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nat Genet* **47**, 242–249 (2015).

192. Merker, M. *et al.* Multidrug- And extensively drug-resistant mycobacterium tuberculosis Beijing Clades, Ukraine, 2015. *Emerg Infect Dis* **26**, 481–490 (2020).

193. Merker, M. *et al.* Transcontinental spread and evolution of Mycobacterium tuberculosis W148 European/Russian clade toward extensively drug resistant tuberculosis. *Nat Commun* **13**, (2022).

194. Klopper, M. *et al.* A landscape of genomic alterations at the root of a near-untreatable tuberculosis epidemic. *BMC Med* **18**, (2020).

195. Hang, N. T. Le *et al.* Whole genome sequencing, analyses of drug resistance-conferring mutations, and correlation with transmission of Mycobacterium tuberculosis carrying katG-S315T in Hanoi, Vietnam. *Sci Rep* **9**, (2019).

196. Casiraghi, E., Arriaga, M. B. & Anteneh, Z. Prognostication of treatment non-compliance among patients with multidrug-resistant tuberculosis in the course of their follow-up: a logistic regression–based machine learning algorithm. (2023) doi:10.3389/fdgth.2023.1165222.

197. Van Der Werf, M. J., Langendam, M. W., Huitric, E. & Manissero, D. Multidrug resistance after inappropriate tuberculosis treatment: a meta-analysis. *European Respiratory Journal* **39**, 1511–1519 (2012).

198. Mase, S. R. & Chorba, T. Treatment of Drug-Resistant Tuberculosis. *Clin Chest Med* **40**, 775 (2019).

199. Merker, M. *et al.* Phylogenetically informative mutations in genes implicated in antibiotic resistance in Mycobacterium tuberculosis complex. doi:10.1186/s13073-020-00726-5.

200. Desjardins, C. A. *et al.* Genomic and functional analyses of Mycobacterium tuberculosis strains implicate ald in D-cycloserine resistance. *Nat Genet* **48**, 544–551 (2016).

201. Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis, 27-29 October 2020. https://apps.who.int/iris/handle/10665/338776.

202. Otchere, I. D. *et al.* Detection and characterization of drug-resistant conferring genes in Mycobacterium tuberculosis complex strains: A prospective study in two distant regions of Ghana. *Tuberculosis* **99**, 147–154 (2016).

203. Grö Schelid, M. I. *et al.* Pathogen-based precision medicine for drug-resistant tuberculosis. (2018) doi:10.1371/journal.ppat.1007297.

204. Kontsevaya, I. *et al.* Perspectives for systems biology in the management of tuberculosis. *European Respiratory Review* **30**, (2021).

205. Weiner, M. *et al.* Effects of tuberculosis, race, and human gene SLCO1B1 polymorphisms on rifampin concentrations. *Antimicrob Agents Chemother* **54**, 4192–4200 (2010).

206. Naidoo, A. *et al.* Effects of genetic variability on rifampicin and isoniazid pharmacokinetics in South African patients with recurrent tuberculosis. *Pharmacogenomics* **20**, 225 (2019).

207. Chigutsa, E. *et al.* The SLCO1B1 rs4149032 Polymorphism Is Highly Prevalent in South Africans and Is Associated with Reduced Rifampin Concentrations: Dosing Implications. *Antimicrob Agents Chemother* **55**, 4122 (2011).

208. Beckert, P. *et al.* MDR M. tuberculosis outbreak clone in Eswatini missed by Xpert has elevated bedaquiline resistance dated to the pre-treatment era. doi:10.1186/s13073-020-00793-8.

209. Sahal, M. R. *et al.* Mycobacterium tuberculosis complex drug-resistance, phylogenetics, and evolution in Nigeria: Comparison with Ghana and Cameroon. *PLoS Negl Trop Dis* **17**, (2023).

210. Asare, P. *et al.* Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. *International Journal of Infectious Diseases* **73**, 30–42 (2018).

211. Niobe-Eyangoh, S. N. *et al.* Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon. *J Clin Microbiol* **41**, 2547 (2003).

212. Niobe-Eyangoh, S. N. *et al.* Molecular Characteristics of Strains of the Cameroon Family, the Major Group of Mycobacterium tuberculosis in a Country with a High Prevalence of Tuberculosis. *J Clin Microbiol* **42**, 5029 (2004).

213. Van Soolingen, D. *et al.* Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. *J Clin Microbiol* **33**, 3234–3238 (1995).

214. Connor, P. *At Least a Million Sub-Saharan Africans Moved to Europe since 2010*. https://www.travelmath.com/flying-time/from/Africa/to/Europe (2018).

215. Holiday Destinations by Flight Duration. https://www.streetdirectory.com/travel_guide/218039/holidays/holiday_destinations_by_flight_duration.html.

216. Revolutionizing Ghana's Economy: Makola Market take their busine - NEWSnet Northern Michigan. https://michigan.newsnetmedia.com/story/48620223/revolutionizing-ghanas-economy-makola-market-take-their-business-online.

217. MAKOLA MARKET LONDON LIMITED overview - Find and update company information - GOV.UK. https://find-and-update.company-information.service.gov.uk/company/13329554.

218. Kruijshaar, M. E., Abubakar, I., Stagg, H. R., Pedrazzoli, D. & Lipman, M. Migration and tuberculosis in the UK: targeting screening for latent infection to those at greatest risk of disease. *Thorax* **68**, 1172–1174 (2013).

219. Pareek, M., Greenaway, C., Noori, T., Munoz, J. & Zenner, D. The impact of migration on tuberculosis epidemiology and control in high-income countries: A review. *BMC Med* **14**, 48 (2016).

220. Pescarini, J. M. *et al.* Migration and tuberculosis transmission in a middle-income country: A cross-sectional study in a central area of São Paulo, Brazil. *BMC Med* **16**, 1–10 (2018).

221. Boudville, D. A., Joshi, R. & Rijkers, G. T. Migration and tuberculosis in Europe. *J Clin Tuberc Other Mycobact Dis* **18**, (2020).

222. Walker, T. M. *et al.* Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet Infect Dis* **13**, 137 (2013).

223. Karmakar, M., Trauer, J. M., Ascher, D. B. & Denholm, J. T. Hyper transmission of Beijing lineage Mycobacterium tuberculosis: Systematic review and meta-analysis. *Journal of Infection* **79**, 572–581 (2019).

224. Peil, M. Ghanaians Abroad. *Afr Aff (Lond)* **94**, 345–367 (1995).

225. Mensah, J., Teye, J. K. & Setrana, M. B. The Janus-Face of Contemporary Migration: Perspectives on West African Return Migration and Transnationalism with a Focus on Ghana and Senegal. *IMISCOE Research Series* 237–259 (2022) doi:10.1007/978-3-030-97322-3_12/TABLES/2.

226. Ribeiro, S. C. M. *et al.* Mycobacterium tuberculosis Strains of the Modern Sublineage of the Beijing Family Are More Likely To Display Increased Virulence than Strains of the Ancient Sublineage. *J Clin Microbiol* **52**, 2615 (2014).

227. Ramazanzadeh, R. & Sayhemiri, K. Prevalence of Beijing family in Mycobacterium tuberculosis in world population: Systematic Review and Meta-Analysis. *Int J Mycobacteriol* **3**, 41–45 (2014).

228. Liu, Y. *et al.* The study on the association between Beijing genotype family and drug susceptibility phenotypes of Mycobacterium tuberculosis in Beijing. *Scientific Reports 2017 7:1* **7**, 1–7 (2017).

229. Rivi Ere, E., Whitfield, M. G., Nelen, J., Heupink, T. H. & Rie, A. Van. Systematic review Identifying isoniazid resistance markers to guide inclusion of high-dose isoniazid in tuberculosis treatment regimens. *Clinical Microbiology and Infection* **26**, 1332–1337 (2020).

230. Zeng, M. C., Jia, Q. J. & Tang, L. M. rpoB gene mutations in rifampin-resistant Mycobacterium tuberculosis isolates from rural areas of Zhejiang, China. *J Int Med Res* **49**, (2021).

231. Shahi, F., Khosravi, A. D., Tabandeh, M. R. & Salmanzadeh, S. Investigation of the Rv3065, Rv2942, Rv1258c, Rv1410c, and Rv2459 efflux pump genes expression among multidrug-resistant Mycobacterium tuberculosis clinical isolates. *Heliyon* **7**, e07566 (2021).

232. Jiang, X. *et al.* Assessment of efflux pump gene expression in a clinical isolate Mycobacterium tuberculosis by real-time reverse transcription PCR. *Microb Drug Resist* **14**, 7–11 (2008).

233. Machado, D. *et al.* Ion Channel Blockers as Antimicrobial Agents, Efflux Inhibitors, and Enhancers of Macrophage Killing Activity against Drug Resistant Mycobacterium tuberculosis. *PLoS One* **11**, (2016).

234. Machado, D. *et al.* Contribution of efflux to the emergence of isoniazid and multidrug resistance in Mycobacterium tuberculosis. *PLoS One* **7**, (2012).

235. Oh, T. S. *et al.* RNA expression analysis of efflux pump genes in clinical isolates of multidrug-resistant and extensively drug-resistant Mycobacterium tuberculosis in South Korea. *Infect Genet Evol* **49**, 111–115 (2017).

236. Fonseca, J. D., Knight, G. M. & McHugh, T. D. The complex evolution of antibiotic resistance in Mycobacterium tuberculosis. *Int J Infect Dis* **32**, 94–100 (2015).

6. Appendices

6.1. Appendix A

Materials

Product	***Catalog # number***	***Company***
Reagents		
1x Phosphate buffered solution	MFCD0013855	Sigma- Aldrich
CTAB (N-Cetyl-N,N,N-trimethylammonium-bromide)	1023421000	Merck
CyQuant TM LDH Cytotoxicity assay	C20301	Thermofisher Scientific
RPMI 1640(without phenol)	11835030	Gibco
DEPC water	T143.3	Roth
Difco^{TM} Centralbrook 7H9 broth base 500g	271310	Becton Dickinson
Difco^{TM} Centralbrook 7H10 agar base	262710	Becton Dickinson
Elution Buffer AE (5mM Tris/HCL, pH 8,5)	12716563	Fisher Scientific
Ethanol 100%	2005786	Roth
Glycerol (Glycerin)	A2926,1000	PanReac AppliChem
HCl 0.1 mol/L	2315957	Roth

Isoamyl alcohol	T870.1	Roth
Oleic acid dextrose catalase (OADC) enrichment supplement	6037	Condalab
Tween 80	P8074	Sigma-Aldrich
Histopaque (6x100ml)	11191	Sigma
Sytox green Nucleic Acid stain	S7020	Thermofisher Scientific (Invitrogen)
Triton X-100	22-332-046	Apacor
Trypan blue solution	15250061	Gibco
EasySep TM Direct human neutrophil isolation kit	50300	Stemcell
Fecal bovine serum	SH3039602	Thermofisher Scientific
L-glutamate	SH3003401	Thermofisher Scientific
DNF-473 Standard Sensitivity NGS Fragment Analysis Kit, (1bp – 6000 bp)	DNF-473-1000	Advanced Analytical
NextSeq 500/550 High Output Kit v2.5 (300 cycles)	20024908	Illumina
NextSeq 500/550 Mid Output Kit v2.5 (300 cycles)	20024905	Illumina
Nextera XT DNA Sample Preparation Kit (96 samples)	FC-131-1096	Illumina
Nextera XT Index Kit v2 Set A (96 indices, 384 samples)	FC-131-2001	Illumina
Nextera XT Index Kit v2 Set B (96 indices, 384 samples)	FC-131-2002	Illumina
Nextera XT Index Kit v2 Set C (96 indices, 384 samples)	FC-131-2003	Illumina
Nextera XT Index Kit v2 Set D (96 indices, 384 samples)	FC-131-2004	Illumina

Qubit dsDNA HS assay kit (500)	Q32854	Life Technologies
Consumables		
Biosphere Filter Tips 10µL	70.1114.210	Sarstedt
Biosphere Filter Tips 100µL	70.760.212	Sarstedt
Biosphere Filter Tips 200µL	70.760.211	Sarstedt
Biosphere Filter Tips 1250µL	70.1186.210	Sarstedt
Cannula 0.4 x 23mm (26G)	59507	M+W Select
Cannula 1.2 x 40mm (18G)	4038088-01	Brau - Sterican®
Cuvettes 10x4x45 mm	67.742	Sarstedt
Cuvette single-use lids	XK25.1	Roth
Dispenser tips (for multi-stepper pipette)	613-1012	VWR
Flat bottom 96-well plate	833.924	Sarstedt
Inoculation loops (sterile)	612-2495	VWR
Latex gloves (Medium & Small)	MAG 08029	Dahlhausen
Microplate 96 well, F bottom (Chimney well) black	655077	Greiner bio-one
Microscope slides	03-0060	R. Langenbrinck
Nitrile gloves (Small)	MAG 08047	VWR
Nonstick, RNase-free Tubes, 1.5 mL	AM12450	Thermo Fisher
Parafilm®	P7793-1EA	Pechiney
Petri dishes 92 x 16 mm	82.1472	Sarstedt
Reagent reservoir (55mL, sterile)	E831.2	Roth
Roller bottle (490cm2)	430195	Corning
Screw Cap Tube 1.5mL with lid	72.692	Sarstedt
SealMat for 96 well plates	710889	Biozym
Square media bottle, sterile, PETG, 30 mL	2019-0030	NalgeneTM
Stripette® serological pipette 5mL	4487	Corning Inc.

Stripette® serological pipette 10mL	4488	Corning Inc.
Stripette® serological pipette 25mL	4489	Corning Inc.
Syringe 1mL	303172	Becton Dickinson
Syringe filtration unit Filtropur 0.2µm	83.1826.001	Sarstedt
Syringe filtration unit PTFE membrane 0.2µm	SLLG013SL	Millex
Tube 1.5mL safety cap	S 72690001	Sarstedt
Tube 15mL, polypropylene screw caps	62.554.502	Sarstedt
Tube 50mL, polypropylene screw caps	62.547.254	Sarstedt
Tubes 0.5 ml (Qubit)	732-0675	Axygen
UKMYC5 plate	*	Thermo Fisher
UKMYC6 plate	*	Thermo Fisher
Round bottom 96-well plate		Sarstedt
Equipment		
Cell culture incubator		Thermo Fisher Scientific
Water bath Precision GP 10		Thermo Fisher Scientific
EasySep magnet (Easy 50)		StemCell
Multifuge X Pro Centrifuge Series		Thermo Fisher Scientific
Claire pro-safety cabinets		Berner
Synergy 2 Multi-Detection microplate reader		Agilent Technologies
Analytical Balance EMB 2000-2		KERN
Autoclave		Systec
BD BACTEC™ MGIT™ 320		Becton Dickinson
Synergy 2 plate reader		BioTek (Agilent)
Cell Density Meter model 40		Fisher Scientific
Centrifuge 5430		Eppendorf
Centrifuge 5430		Eppendorf
DynaMag™-PCR Magnet		Thermo Fisher Scientific
Fragment Analyzer™		Advanced Analytical
Heating Block		Biozym Scientific GmbH
Hera Safe - Safety bench		Thermo, Germany

Multi-channel pipettes (100µl, 200 µl, 300µl)		Eppendorf
NextSeq500		Illumina
Öko Refrigerator		Privileg
Pipet boy		Integra Biosciences
Pipettes (10µl, 100µl, 200µl, 1000µl)		Eppendorf
Premium No Frost Freezer		Liebherr
Qubit 3.0		Life technologies
Roller apparatus		Wheaton
Spectrophotometer DS-11 FX		DeNovix
Thermocycler		Analytik Jena
Thermocycler C1000 Touch		Biorad
Vortex Genie2		Scientific Industries
Vortex Genius 3		IKA Vortex
Water bath		Köttermann
Water bath 14 L		Gesellschaft für Labortechnik®

6.2. Appendix B

Table B.1. **Proportion of clusters based on =<12 SNP distances**

Sublineage	Number Clustered	Total	Percent
5 West Africa 1	0	1	0.0
1.1.1 EAI	3	10	30.0
2.2.1 Beijing Ancestral 3	19	22	86.4
3 Delhi-CAS		2	0.0
4.1 Euro-American	13	38	34.2
4.1.1.1 X-type		3	0.0
4.1.1.3 X-type		3	0.0
4.1.2.1 Haarlem	28	45	62.2
4.3.3 LAM	7	14	50.0
4.3.4.1 LAM	2	2	100.0
4.3.4.2 LAM	3	7	42.9
4.4.1.1 S-type	2	6	33.3
4.6.2.2 Cameroon	2	3	66.7
4.8 mainly T	13	23	56.5
4.9 H37Rv-like		3	0.0
5 West Africa 1		2	0.0
5.1.3 West Africa 1		1	0.0
5.1.5 West Africa 1		2	0.0
5.3 West Africa 1		1	0.0
6.1.3 West Africa 2		2	0.0
6.2.2 West Africa 2	2	23	8.7
6.2.3 West Africa 2		5	0.0
6.3.1 West Africa 2		3	0.0
6.3.3 West Africa 2	10	17	58.8

Table B.2. **Distribution of drug resistance in Sierra Leone**

Drug	Number Resistant	Percent
RR	238	100
INH	196	82
EMB	129	54
PZA	92	39
BDQ	6	3
CFZ	5	2
ETH/PTH	23	10
CPR	1	0
CS	57	24
PAS	5	2

Table B.3. **Distribution of strains based on treatment**

	Number	Percent
MDR	196	82.4
1st-line	61	25.6
MDR+BDQ +CFZ	5	2.1
1st-line + BDQ + CFZ	3	1.3
Ist-line + BDQ + SM	2	0.8

Table B.4. **Distribution of strains based on lineages and resistances**

	Number	L1 (%)	L2(%)	L3(%)	L4(%)	L5 (%)	L6 (%)
MDR	196	10(5.1)	22(11.2)	1(0.5)	116(59.1)	2(1)	45(23)
1st-line	61	5(8.2)	8(13.1)	0	31(50.8)	2(3.2)	15(24.6)
BDQ+CFZ	5	1(20)	1(20)	0	1(20)	0	2(40)
MDR+BDQ +CFZ	5	1(20)	1(20)	0	1(20)	0	2(40)
1st-line + BDQ + CFZ	3	1(33.3)	0.0	0	0	0	2(66.7)
Ist-line + BDQ + SM	2	0	0.0	0	0	0	2(100)

Table B.5. **Distribution of Isoniazid resistance mutations in Sierra Leone**

Mutations	**Number**	**Percent**
fabG1 -15c>t	12	6.12
fabG1 -15c>t; katG 63_del_c; katG 36_del_c	1	0.51
fabG1 -15c>t; **katG S315T**	4	2.04
fabG1 -15c>t; katG W191R	1	0.51
fabG1 -17g>t; fabG1 -8t>c; **katG S315T**	1	0.51
fabG1 -17g>t; **katG S315T**	14	7.14
fabG1 -17g>t; katG Y98C	1	0.51
fabG1 -8t>a; **katG S315T**	1	0.51
fabG1 -8t>c	2	1.02
fabG1 -8t>c; **katG S315T**	3	1.53
fabG1 L203L	2	1.02
fabG1 L203L; **katG S315T**	10	5.10
inhA S94A	1	0.51
katG 1003_ins_g	2	1.02
katG 1079_del_gg	1	0.51
katG 1253_del_gc	1	0.51
katG 1335_ins_t	1	0.51
katG 1432_del_g	1	0.51
katG 1745_ins_g	1	0.51
katG 1861_del_gtgactcgcatt	1	0.51
katG 2010_del_c	1	0.51
katG 2087_ins_c	1	0.51
katG 23_ins_a	1	0.51
katG 29_ins_g	1	0.51
katG 371_del_g	1	0.51
katG 402_del_cgacaattcgc; katG 177_del_tgggtagcccagtc	1	0.51
katG 521_ins_t	1	0.51
katG 540_del_gcagaacttc	1	0.51
katG 89_ins_g	1	0.51
katG 978_del_g	1	0.51
katG D142G	1	0.51
katG E334_	1	0.51
katG G14_	1	0.51
katG G279D	1	0.51
katG N138S	1	0.51
katG Q500_	1	0.51
katG S315I	4	2.04
katG S315N	12	6.12

Mutations	Number	Percent
katG S315T	99	50.51
katG S315T; katG 15_del_caca	1	0.51
katG V1A	1	0.51
katG W412C	1	0.51
katG W668_	1	0.51
Total	**196**	

Table B.6. **Distribution of Rifampicin resistance mutations found in Sierra Leone**

Mutations	Number	Percent
rpoB 1277_del_gcacca	1	0.42
rpoB 1277_del_gcacca; rpoB S450L	1	0.42
rpoB 1288_del_ctgagccaa	1	0.42
rpoB 1291_ins_gcc	1	0.42
rpoB 1295_del_aattcatgg	1	0.42
rpoB 1296_ins_ttc	1	0.42
rpoB 1304_del_accaga	1	0.42
rpoB 1307_del_aga	1	0.42
rpoB 1309_del_aac	2	0.84
rpoB 1309_del_aac; rpoB S450L	1	0.42
rpoB D435F	4	1.68
rpoB D435L	2	0.84
rpoB D435V	25	10.50
rpoB D435V; rpoB 1307_del_aga	1	0.42
rpoB D435V; rpoB H445Y	2	0.84
rpoB D435V; rpoB S450L	2	0.84
rpoB D435Y	13	5.46
rpoB D435Y; rpoB I491L	1	0.42
rpoB D435Y; rpoB L452M	1	0.42
rpoB D435Y; rpoB N437H	1	0.42
rpoB D435Y; rpoB S441L; rpoB H445Y	1	0.42
rpoB H445C	5	2.10
rpoB H445D	9	3.78
rpoB H445D; rpoB L452P	1	0.42
rpoB H445L	6	2.52
rpoB H445N	2	0.84
rpoB H445P	1	0.42
rpoB H445R	6	2.52
rpoB H445Y	15	6.30

Mutations	Number	Percent
rpoB L430P	3	1.26
rpoB L452P	8	3.36
rpoB Q429H; rpoB L430P; rpoB D435Y; rpoB H445Q	1	0.42
rpoB Q429L; rpoB D435Y; rpoB N437Y	1	0.42
rpoB Q432E	1	0.42
rpoB Q432K	2	0.84
rpoB Q432K; rpoB S450L	1	0.42
rpoB Q432L	1	0.42
rpoB S431R; rpoB S450W	1	0.42
rpoB S441L	1	0.42
rpoB S441L; rpoB H445R	1	0.42
rpoB S441L; rpoB S450L	1	0.42
rpoB S450F	1	0.42
rpoB S450L	94	39.50
rpoB S450L; rpoB I491V	1	0.42
rpoB S450Q	2	0.84
rpoB S450W	7	2.94
rpoB V170F	1	0.42
rpoB V170F; rpoB M434I; rpoB H445N; rpoB L452V	1	0.42
Total	**238**	

6.3. Appendix C

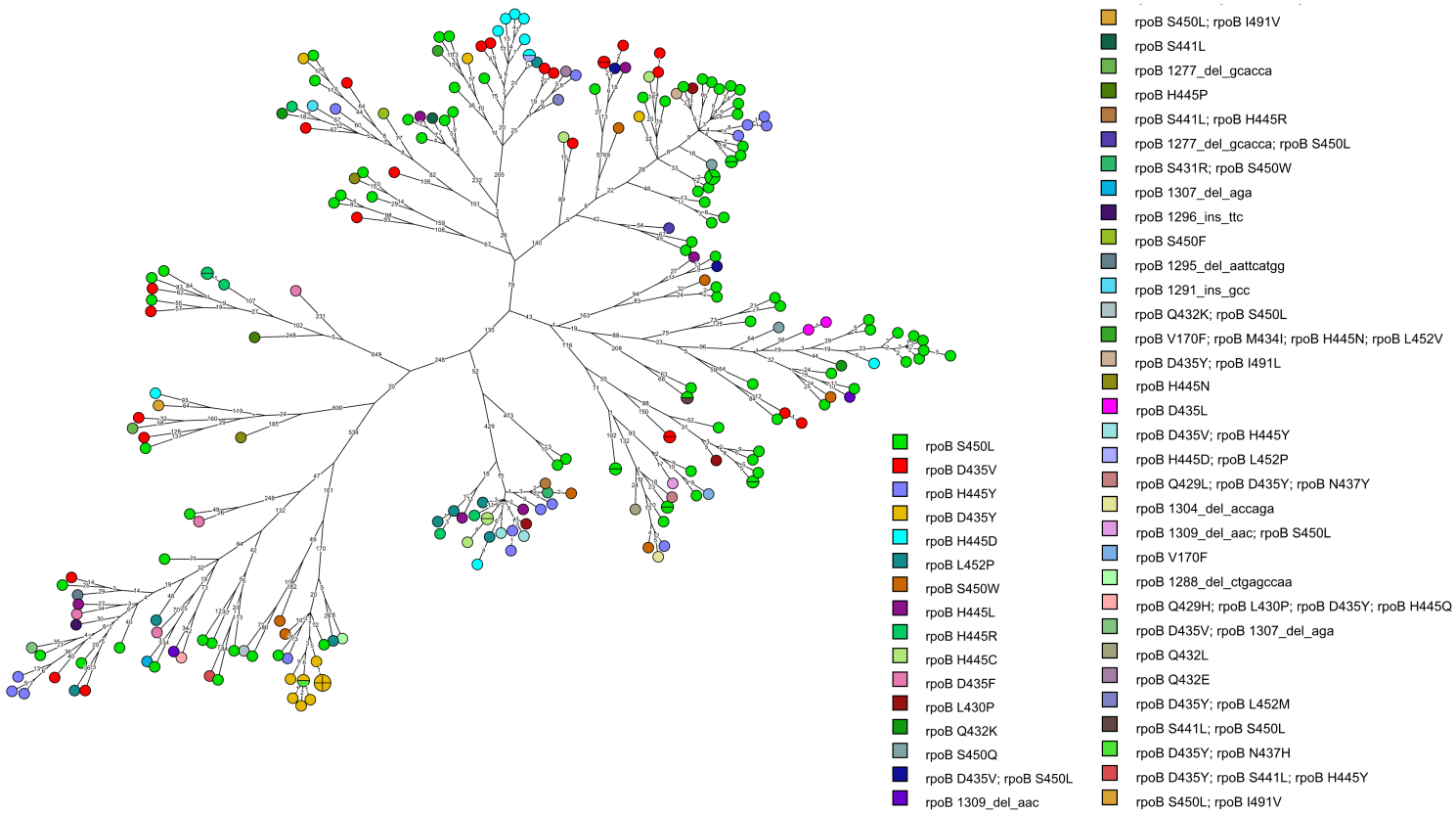


Figure C.1. Phylogeny of rifampicin resistance mutations of *Mycobacterium tuberculosis complex* strains from Sierra Leone. Maximum parsimony tree of 238 Sierra Leone *Mycobacterium tuberculosis* based on the alignment of 15,089 informative single nucleotide polymorphisms. The different mutations detected are color-coded.

6.4. Appendix D

Table D.1. **Selected group-specific single nucleotide polymorphisms for the designations of the strains of the Cameroon sublineage**

Clade Name	Reference	Position	Gene	Function	Group Allele	Group Substitution
C.1	G	2486741	*Rv2219*	Probably conserved transmembrane protein	C	A169A (gcg/gcC)
C.1	C	4345445	*Rv3869*	ESX conserved component EccB1 ESX-1 type VII secretion system protein Possible membrane protein	A	A136D (gcc/gAc)
C.2	C	1586471	*Rv1410c*	Aminoglycosides/tetracycline-transport integral membrane protein	T	G432G (ggg/ggA)
C.2	C	2823247	*Rv2507*	Possible conserved proline-rich membrane protein	T	D270D (gac/gaT)
C.3	G	1776435	*Rv1568*	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase BioA	C	L348L (ctg/ctC)
C.4	G	3470335	*Rv3101c*	Putative cell division protein FtsX (septation component-transport integral membrane protein ABC transporter)	A	D115D (gac/gaT)
C.5	C	2375883	*Rv2115c*	Mycobacterial proteasome ATPase Mpa	T	K136K (aag/aaA)
C.6	G	4354077	*Rv3876*	ESX-1 secretion-associated protein EspI Conserved proline and alanine-rich protein	A	G356G (ggg/ggA)
C.7	C	3584089	*Rv3207c*	hypothetical protein	T	M190I (atg/atA)
C.8	C	151535	*Rv0125*	Probable serine protease PepA (serine proteinase) (MTB32A)	T	L130L (ctg/Ttg)

Table D.2. **Proportion of rifampicin resistance mutations of strains of the Cameroon sublineage**

Rifapicin Mutation	Frequency	Percent
rpoB 1301_del_tgg; rpoB S450L	1	0.65
rpoB 1306_del_cag	1	0.65
rpoB D435F	8	5.23
rpoB D435V	6	3.92
rpoB D435Y	5	3.27
rpoB D435Y; rpoB H445D	2	1.31
rpoB H445D	17	11.11
rpoB H445L	6	3.92
rpoB H445N	1	0.65
rpoB H445R	3	1.96
rpoB H445Y	7	4.58
rpoB L430P	4	2.61
rpoB Q432K; rpoB H445D	1	0.65
rpoB S441L; rpoB S450L	1	0.65
rpoB S450L	76	49.67
rpoB S450W	1	0.65
rpoB V170F	13	8.50

Total =153

Table D.3. **Proportion of isoniazid resistance mutations of strains of the Cameroon sublineage**

Isoniazid mutations	Frequency	Percent
fabG1 -15c>t	435	75.39
fabG1 -15c>t; inhA I21T	1	0.17
fabG1 -15c>t; inhA S94A	1	0.17
fabG1 -15c>t; katG S315T	16	2.77
fabG1 -15c>t; katG W477	1	0.17
fabG1 -17g>t	1	0.17
fabG1 -8t>c	1	0.17
fabG1 L203L	1	0.17
inhA S94A	3	0.52
katG 1706_del_c; katG 388_del_gcttg	1	0.17
katG S315N	2	0.35
katG S315T	112	19.41
katG V1A	1	0.17
katG W191G	1	0.17

Total =577

Table D.4. **Proportion of ethambutol resistance mutations of strains of the Cameroon sublineage**

Ethambutol Mutations	Frequency	Percent
embA -12c>t; embB M306I	1	1.54
embA -16c>g	1	1.54
embA -16c>t	3	4.62
embA -11c>a	2	3.08
embA -16c>a; embB M306V	1	1.54
embA -8c>a	1	1.54
embB D354A	9	13.85
embB G406D	1	1.54
embB G406S	1	1.54
embB M306I	20	30.77
embB M306I; embB G406A; embB Q497R	1	1.54
embB M306I; embB G406D	1	1.54
embB M306V	18	27.69
embB Q497K	2	3.08
embB Q497R	3	4.62

Total =65

Table D.5. **Proportion of pyrazinamide resistance mutations of strains of the Cameroon sublineage**

Pyrazinamide Mutations	Frequency	Percent
pncA -11a>g	1	1.75
pncA 244_ins_a; pncA A46E; pncA -11a>g	1	1.75
pncA 298_del_a	1	1.75
pncA 392_ins_gg	2	3.51
pncA 449_ins_gt; pncA 392_ins_a	1	1.75
pncA 99_del_ca	1	1.75
pncA A171T	1	1.75
pncA A171T; pncA L85P; pncA -11a>g	1	1.75
pncA A28D	2	3.51
pncA C14_	1	1.75
pncA D63A	1	1.75
pncA E15G	1	1.75
pncA F13I	1	1.75
pncA G132S	1	1.75
pncA G132V	2	3.51
pncA G78A; pncA A46E; pncA I6L	1	1.75
pncA H82R	1	1.75
pncA I6T	1	1.75
pncA L120R	4	7.02
pncA M175I	1	1.75
pncA P69A	5	8.77
pncA P69S	11	19.30
pncA T47A	1	1.75
pncA T47I	7	12.28
pncA T76P	1	1.75
pncA V9A	1	1.75
pncA Y103_; pncA -11a>g	1	1.75
pncA Y103H	3	5.26
pncA Y95_; pncA 195_ins_ctt; pncA C14G	1	1.75

Total = 57

Table **D.6. Proportion of bedaquiline resistance mutations of strains of the Cameroon sublineage**

Bedaquiline Mutations	Frequency	Percent
Rv0678 E113	1	50
Rv0678 R132	1	50

Total = 2

6.5. Appendix E

R script for RhierBAPs

```
library(rhierbaps)

library(ggtree)

library(phytools)

library(ape)

library(ggplot2)

Camphylo_fasta <-
"Cam_phylo_v1_cf4_cr4_fr75_ph4_l0_x0_445_combined_amended_u95_phylo_w12.plainIDs.f
asta"

snp.matrix <- load_fasta(Camphylo_fasta)

#hb.results <- hierBAPS(snp.matrix, max.depth = 2, n.pops = 20, quiet = TRUE)

hb.results <- hierBAPS(snp.matrix, max.depth = 3, n.pops = 20, n.extra.rounds = 0,

                       quiet = FALSE)

nexus.camphylo <- "Camphylov1_rooted.nex"

Camphylo.iqtree <- read.nexus(nexus.camphylo)

###plot tree with groups

gg <- ggtree(Camphylo.iqtree, layout = "circular", branch.length = "none")

gg <- gg %<+% hb.results$partition.df

gg <- gg + geom_tippoint(aes(color = factor(`level 1`)))

gg <- gg + theme(legend.position = "right")
```

```
gg <- gg + geom_tiplab(aes(label = `level 1`), size = 1, offset = 1)

#gg <- gg + geom_tiplab(size = 1, offset = 10, geom = "text")

gg


gg <- ggtree(Camphylo.iqtree, layout = "circular")

gg <- gg %<+% hb.results$partition.df

gg <- gg + geom_tiplab2(aes(label = `label`, angle = angle), align = TRUE, linetype = "dotted",

                        linesize = 0.3, geom = "text", offset = 0.0001, size = 1.5)

gg <- gg + geom_tippoint(aes(color = factor(`level 2`)))

gg <- gg + theme(legend.position = "right")

gg <- gg + scale_colour_discrete(name = "LEVEL 2")

gg

write.table(hb.results[["partition.df"]], file="hierbabs_level", sep="\t", quote=FALSE
```

6.6. Appendix F

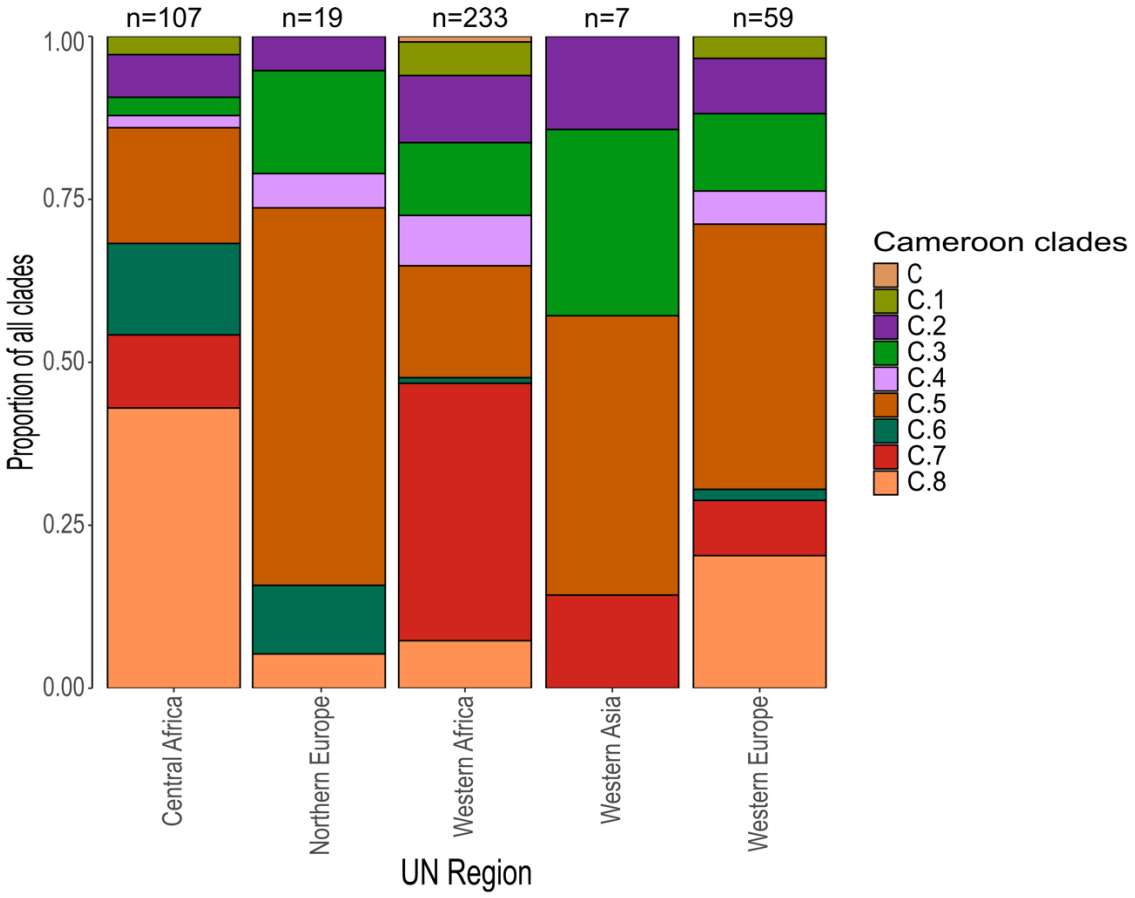


Figure F.1. Distribution of strains of the Cameroon clades based on United Nations regional designation. The bar chart was based on 444 strains of the Cameroon sublineage selected for clade designation. United Nations regions with less than five strains were not included. The Total number of strains in each region is written on top of the bar chart. Clades are color-coded, n = number of strains, UN = United Nations

7. Dedication, Acknowledgements, Disclosure

7.1. Dedication

In loving memory of my late mother, Mrs. Sarah Yetunde Akiwumi Abbey, I dedicate this thesis. Throughout my journey, she was my biggest cheerleader and her pride in my every little achievement was evident whenever she spoke about me to her friends. Though she is no longer with us today, I am certain that she watched over me from Heaven with a beaming smile. This work is a tribute to you, Mum – thank you for everything.

7.2. Acknowledgment

I am deeply grateful for the support and contributions of numerous individuals who made this dissertation possible. It honestly took a village, and I am indebted to every one of you.

First and foremost, I offer my heartfelt thanks to the Almighty God, whose grace and blessings made this journey possible. Without His guidance and support, I would not have reached this point.

I extend my sincere gratitude to Prof. Dr. Stefan Niemann, my primary supervisor, whose exceptional expertise and visionary approach enriched my research. Stefan's brilliance as a scientist is genuinely awe-inspiring. He has a remarkable ability to see the big picture, and when presented with a data, uncovers trends and generates ideas that no one could fathom. His touch seemingly elevates everything he works on, making it much better. As a supervisor, Stefan's openness to new ideas and unwavering support for my scientific development were invaluable. I feel incredibly fortunate to have had the opportunity to work under his guidance. Thank you, Stefan, for your mentorship, expertise, and inspiration. Your influence has profoundly impacted both my research and personal growth.

I am also thankful to Dr. Christian Utpatel, another vital figure who served as my supervisor and played a pivotal role in shaping this research. Christian's intelligence and patience are truly

remarkable. He took the time to guide me through designing and analyzing my work. I am grateful for Christian's meticulous review of my thesis, ensuring it was ready for submission. His support and constructive feedback have been instrumental in shaping the outcome of this work. Thank you, Christian, for your guidance; I am sincerely grateful to have learned from you.

I acknowledge the valuable contributions of Dr. Tobias Dallenga, my third supervisor, who played a crucial role in my infection-model experiments. His support and willingness to review my work were highly appreciated.

My heartfelt appreciation goes to Dr. Christiane Gerlach, who was always there for me as a go-to person, offering guidance and support whenever I needed it. Her openness and willingness to assist with any queries or concern made a significant difference during my time in Germany.

I am grateful to Dr. Susanne Homolka and Dr. Thorsten Goldberg. Even though they were not my direct supervisors, they generously shared their time in numerous meetings, discussing my experimental work and analysis. Their support was really valuable.

The technicians in my group were exceptional, and I want to express my thanks to Silvia Maass, Doreen Beyer, Menie Wiemer Winey, and Larissa Mohr for their instrumental assistance in my experimental work. Their support ensured smooth progress in my research.

I also acknowledge the expertise and assistance of our molecular lab technicians, including Tanja Niemann, Vanessa Mohr, Tanja Struve-Sonnenschein, Anja Lüdeman, Carina Hahn, Maya Mundzeck, and Jasmin Scharnberg, who provided crucial training and support for my molecular work.

Special thanks go to Jacqueline Eich from the Cellular Microbiology group for patiently training me in working with neutrophils, and to Fenja Boysen and Ilse Radzio from the NRZ for their technical support.

My fellow PhDs and office mates Nicole Ullrich, Emilie Rousseau, Teresa Waltz, Jana Schönfeld, Sabine Petersen, Annemarie Hintz-Rüter, Darsha Nadarajan, and Simon Pennuttis made the work environment enjoyable and supportive. You have been a source of invaluable support and camaraderie throughout this journey.

The postdoctoral fellows Dr. ir. Margo Diricks, Dr. Thomas Kohl, Patrick Beckert, Leonardo de Araujo, Meriem Belheouane, Dr. Viola Dreyer, and Ivan Barilar provided invaluable suggestions to enhance my projects. Special thanks also go to Dr. ir. Margo Diricks and Dr. Viola Dreyer, who generously dedicated their time to read through my thesis and offered valuable suggestions for improving my project. Their insights were instrumental in shaping the outcome. I am also thankful to Ivan Barilar for his guidance and training in data analysis. His expertise and willingness to share knowledge have been instrumental in enhancing the quality of my research.

I cannot express my gratitude enough to Dr. Lindsay Sonnenkalb. She was not only my first friend in the group and institute, but she has been a constant source of support, providing constructive criticism, discussing project ideas, and offering emotional support throughout. Her friendship has made this journey even more meaningful.

Gratitude goes to my collaborators in Sierra Leone, Ghana, and the Cameroon study consortium for their data and the opportunity to work together, notably Dr. Tim Walker, who offered kind and patient assistance.

I would like to thank the funding institutes and corporations, including the Government of Ghana, German Academic Exchange Service (DAAD), the Research Center Borstel Leibniz Lung Center, German Center for Infection Research (DZIF), and Cluster of Excellence "Precision Medicine in Chronic Inflammation" (PMI), for their financial support that made this journey possible.

Special thanks go to my friends Adeela Sana, Olaitan Hawanot Tijani, Irene Yeboah, Dr. Yassir Schwaib, and Dr. Ireti Ogunsulire for their unwavering support and friendship.

A heartfelt thank you to my dear friend and adopted sister, Mrs. Jessica Akoh-Array. Jessie has been an incredible support system throughout this entire journey, and I am profoundly grateful for her presence in my life. I am grateful for her unwavering support, providing a platform for bouncing off ideas and being there whenever I needed her.

My family has been an incredible source of support throughout this journey, and I am deeply grateful for their presence in my life. I thank my father, Mr. Silas E Abbey, for his unwavering support, encouragement, and confidence in my abilities. His belief in me has been a constant source of motivation. My sisters, Ms. Angela N Abbey, Mrs. Dilys Hobenu, Dr. Mrs. Judith Glover,

Mrs. Lilian Glover-Quartey, and Mrs. Belinda Ansah, have been my pillars of strength. They have always been by my side, ready to lend a listening ear, provide companionship, and offer their unwavering support. Sister Angie, in particular, ensured I had everything I needed and lacked nothing. Sis Maa D's words of encouragement and presence during challenging times were incredibly reassuring.

I am immensely grateful to my sister Ju, my brother-in-law Mr. Obed Glover and their whole family for their magnanimous hearts and support. They cared for me and my family, ensuring I had the time to focus on my work, rest, and not worry about my kids. Their selfless assistance was truly a blessing. Lilian's constant presence, encouragement, and emotional support throughout this journey have been invaluable. Her unwavering belief in me kept me going during difficult times. I sincerely thank Belinda and Eithan for their company, care, and love. The time they spent with me and my daughter Naana brought immense joy and support to our lives.

I extend my thanks to my mother and father-in-law, Madam Gifty Selby and Mr. Fiifi Blankson, as well as Uncle Kofi, for their enduring support.

I am grateful to my nephew, Vere Dziedzorm Glover, for selflessly babysitting my kids, allowing me the time to work on my thesis. I am also thankful to Elinam, Sena, Eyram, Kelvin and Selasi who always made sure my kids were happy and not lonely.

To my beloved children, Kaitlyn Nana-Esi Frimpomaa Blankson and Killian Ebo Blankson, your love and presence kept me strong and grounded throughout this journey.

Lastly, I owe a debt of gratitude to my husband, Dr. Paa Kwesi Blankson, for his unwavering support and strength in every aspect of my life. Paa's commitment to our family was evident as he adjusted his working schedule to accommodate mine. His emotional, financial, and intellectual support played a crucial role in my success. He willingly reviewed my work, engaged in discussions, and helped me find solutions to challenges along the way.

Paa-Kwesi has been a pillar of support for both me and our family throughout this entire journey. Despite the sacrifices, he had to endure being separated from his own family for several years so that I could pursue my Ph.D., and he did so without complaint.

To my dear husband, I want to say thank you from the bottom of my heart. Your love, encouragement, and unwavering belief in me have been the driving force behind my achievements. I am grateful for your presence by my side through all the ups and downs. I love you more than words can express, and I am forever thankful for your constant support.

Thank you all for being a part of this journey with me. Your support and encouragement have made all the difference.

7.3. Full disclosure

A manuscript based on the sections discussing MDR-TB in Sierra Leone has been authored and submitted for publication. The primary author of the article is the doctoral candidate who conducted the experiments, analyzed the outcomes, and composed the manuscripts.