ROYAUME DU MAROC
Université Mohammed V - Rabat
Faculté de Médecine et de Pharmacie
RABAT

Année 2022 **N°: MM182022**

MEMOIRE DE MASTER

MASTER DE « **Biotechnologie**»
OPTION: « **Médicale**»

Intitulé

Genomic analysis of *Mycobacterium tuberculosis* in Africa revealed lineage – drug resistance association

Soutenu le 18/10/2022, par:
Yasmine EL FATHI LALAOUI

Devant le jury compose de :

Pr. AANNIZ Tarik, Faculté de Médecine et de Pharmacie de Rabat, Président
Pr. EL ALLALI Achraf, Mohammed VI Polytechnic University of Benguerir, Encadrant
Dr. LAAMARTI Mariam, Mohammed VI Polytechnic University of Benguerir, Co-encadrante
Pr. BENTAYEBI Kaoutar, Faculté de Médecine et de Pharmacie de Rabat, Examinatrice

بسم الله الرحمن الرحيم

"قالوا سبحانك لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم "

صدق الله العظيم (البقرة -32)

Dedication

This work is heartily dedicated to everyone who has supported me, from parents to colleagues and friends who have helped and guided me throughthe hard stages of my life.

Aknowledgments

First of all, I would like to praise and thank God, the almighty, for the many blessings that have been offered to me.

This work would have been impossible to make without the kind supervision of Pr. **Achraf EL ALLALI** who has offered me the opportunity to join his lab and enhancemy knowledge in different fields.

I am heartily grateful to my co-supervisor, Dr. **Mariam LAAMARTI** for her motiva- tion, patience and precious help as well as for the valuable time she has devoted to this work.

I'd like to thank the members of the jury, Pr. **Tarik AANNIZ** and Pr. **Kaoutar BENTAYEBI** for accepting to devote some of their valuable time to examine this work.

Many thanks to all the African Genome Center - Bioinformatics lab members for their constant help and support, more importantly, Rachid **ELFERMI** without whom thejob wouldn't be completed.

Lastly, I offer my deepest gratitude to all the people in my life who encouraged me, my parents, my brothers and my sister, for supporting me through every phase of my life.

Abstract

Tuberculosis is a serious infectious respiratory disease caused by *Mycobacterium tuberculosis*.

In this study, we used genomics, to determine the population structure of *Mycobacterium tuberculosis* in Africa. Using the MTBseq pipeline, we analyzed samples from 28 different African countries in order to determine the distribution of *Mycobacterium tuberculosis* lineages as well as drug resistance in Africa and to verify the association between the two.

As a result, we have found out that the most dominant lineage is lineage 4 with a prevalence of 54.33%, followed by lineage 3 (8.62%), 2 and 1 with a percentage of 7.41% and 7.31% respectively. West African lineages 6 and 5 have a prevalence of 3.61% and 1.25% respectively, while lineage 7 with the lowest percentage of 0.58%.

We have also studied the association between lineages and drug resistance and the p value for the drugs variable was 0.014 while it was $3.24*10^{-7}$ for lineages proving the presence of an association between them.

Moreover, the number of sensitive and resistant samples was calculated and the result showcased that the majority (12034) of samples were sensitive, 2512 samples were Multi-Drug Resistant, 1674 samples are Mono-Drug resistant, 1198 samples are Pre-Extensively Drug Resistant and 223 are resistant to other drugs.

Keywords: *Mycobacterium tuberculosis*, Lineages, Drug-resistance, Genomic Analysis, Africa.

Title: Genomic analysis of *Mycobacterium tuberculosis* in Africa reveals lineage – drug resistance association.

Author: EL FATHI LALAOUI Yasmine

Resumé

La tuberculose est une maladie respiratoire infectieuse grave causée par *Mycobacterium tuberculosis* .

Dans cette étude, nous avons utilisé la génomique pour déterminer la structure de la population de *Mycobacterium tuberculosis* en Afrique. À l'aide du pipeline MTBseq, nous avons analysé des échantillons de 28 pays africains différents afin de déterminer la distribution des lignées de Mycobacterium tuberculosis ainsi que la résistance aux médicaments en Afrique et de vérifier l'association entre les deux.De ce fait, nous avons découvert que la lignée les plus dominante est la lignée 4 avec une prévalence de 54,33%, suivie de la lignée 3 (3.68%), 2 et 1 avec un pourcentage de 7,41% et 7,31% respectivement, les lignées ouest-africaines 6 et 5 ont une prévalence de 3,61% et 1,25% respectivement, et enfin, la lignée la moins répandue est la lignée 7 avec un pourcentage de 0,58%. les lignées des échantillons restants sont inconnues.

Nous avons également étudié l'association entre les lignées et la résistance aux médicaments et la valeur de p pour la variable médicaments était de 0,014 alors qu'elle était de $3,24*10^{-7}$ pour les lignées prouvant la présence d'une association entre eux. De plus, le nombre d’échantillons sensibles et résistants a été calculé et le résultat a montré que la majorité (12034) des échantillons étaient sensibles, 2512 échantillons étaient multi-résistants, 1674 échantillons sont mono-résistants, 1198 échantillons sont pré-ultra-résistants aux médicaments. et 223 sont résistants à d’autres médicaments.

Mot-clés: *Mycobacterium tuberculosis*, Lignées, resistance aux antibiotiques, Analyse Génomique, Afrique.

Titre: L’analyse génomique de *Mycobacterium tuberculosis* en Afrique revèle l’association entre les lignées et la resistance aux antibiotiques.

Auteur: EL FATHI LALAOUI Yasmine

ملخص

يعتبر السل مرضا تنفسيا معديا خطيرا سببه المتفطرة السلية.

في هذه الدراسة، استخدمنا علم الجينوم لتحديد التركيب السكاني لبكتيريا السل المتفطرة في إفريقيا. باستخدام MTBseq قمنا بتحليل عينات من 28 دولة إفريقية مختلفة من أجل تحديد توزيع السلالات و كذلك مقاومة الأدوية في افريقيا و للتحقق من الإرتباط بين الإثنين.

نتيجة لذلك، وجدنا أن السلالات الأكثر انتشارا هي السلالة 4 بنسبة انتشار 54.33%، تليها السلالة 3 بنسبة 3.68%، ثم السلالتين 2 و1 بنسب 7.41% و7.31% على التوالي، سلالات غرب إفريقيا 6 و5 تبلغ نسبة انتشارها 3.61% و1.25% على التوالي، وأخيرا، فإن السلالة الأقل انتشارا هي السلالة 7 بنسبة 0.58%. سلالات العينات المتبقية غير معروفة .

درسنا أيضا الارتباط بين السلالات و مقاومة الادوية و كانت قيمة p بالنسبة لمقاومة الأدوية هي 0.014 بينما كانت $3.24*10^{-7}$ بالنسبة للسلالات مما يثبت وجود ارتباط بينهما.

علاوة على ذلك ، تم حساب عدد العينات الحساسة والمقاومة وأظهرت النتيجة أن غالبية العينات (12034) كانت حساسة ، و 2512 عينة مقاومة للأدوية المتعددة ، و 1674 عينة مقاومة للأدوية الأحادية ، و 1198 عينة مقاومة للأدوية بشكل مكثف. و 223 مقاومة للأدوية الأخرى.

الكلمات المفتاحية: المتفطرة السلية، السلالات ، مقاومة الأدوية، التحليل الجيني، إفريقيا.

العنوان: التحليل الجينومي لبكتيريا المتفطرة السلية يكشف ارتباط السلالات بمقاومة الأدوية.

المؤلف: الفتحي العلوي ياسممين.

Table of contents

Dedication 2

Aknowledgments 3

Abstract 4

Resumé 5

Acronyms 12

1 Background theory 15

1.1 History 16

1.2 Epidemiology 16

1.3 Taxonomy and microbiology 17

1.3.1 Taxonomy and classification 17

1.3.2 Microbiology 17

1.4 Pathology 20

1.4.1 Transmission 20

1.4.2 Immune response 21

1.5 Latent TB infection and TB disease 24

1.5.1 Latent TB infection 24

1.5.2 TB disease 25

1.6 Diagnosis 26

1.6.1 Medical History 26

1.6.2 Physical examination 27

1.6.3 Tuberculin Skin Test 27

1.6.4 Interferon Gamma Release Assays 27

1.6.5 Chest X-ray 27

1.6.6 Diagnostic microbiology 28

1.6.6.1 Microscopic examination 28

1.6.6.2 Culture 28

1.7 Drug resistance 28

1.8 Treatment 29

1.8.1 Anti-tuberculosis drugs 29

1.8.2 Treatment regimens 32

1.8.2.1 Treatment Regimens for New Cases of Pulmonary Tuberculosis: 32

1.8.2.2 Treatment Regimens for Previously Treated Patients: 33

1.8.3 Treatment monitoring 33

1.8.4 Treatment of latent TB infection (LTBI) 34

1.9 *Mycobacterium tuberculosis* lineages 34

1.10 Genomics use in epidemiology 35

2 Genomic analysis 37

2.1 Materials and methods 38

2.1.1 DATA collection 38

2.1.2 Generating Fastq files 38

2.1.3 MTBseq pipeline 38

2.1.3.1 TBbwa 38

2.1.3.2 TBrefine 38

2.1.3.3 TBpile 39

2.1.3.4 TBlist 39

2.1.3.5 TBvariants 39

2.1.3.6 TBstats 40

2.1.3.7 TBstrains 40

2.1.3.8 TBjoin 40

2.1.3.9 TBamend 41

2.1.3.10 TBgroups 41

2.1.4 Lineage distribution analysis 42

2.1.5 Drug resistance analysis 42

2.1.6 Resistance associated genes analysis 42

2.1.7 Lineage-Drug resistance association 43

2.2 Results 44

2.2.1 Distribution of samples by country 44

2.2.2 Lineage distribution in African countries 45

2.2.3 Drug resistance by country 49

2.2.4 Lineage-Drug resistance association 52

2.2.4.1 Two-way ANOVA result 53

2.2.4.2 Two-way ANOVA result interpretation 53

2.2.5 New mutations 54

2.3 Discussion 56

Conclusion & Perspectives **58**

Bibliography **59**

List of figures

1.1 **Estimated TB incidence rates in 2020**..17

1.2 **Structure representation of the *Mycobacterium tuberculosis* cell wall**..19

1.3 **Transmission of *Mycobacterium tuberculosis***......................................21

1.4 **Life cycle of *Mycobacterium tuberculosis***..26

1.5 **distribution of the 7 *Mycobacterium Tuberculosis* lineages in the world**...35

2.1 **Workflow of the study**...44

2.2 **Distribution of 17641 samples by country**... ... 45

2.3 **Lineage distribution of TB in 28 African countries**...........................49

2.4 **Drug resistance by country**...52

2.5 **Two-way ANOVA result**...54

List of tables

1.1 **Posology and contraindications of first line Anti-TB drugs................** 31

2.1 **Number of resistances by drugs in each lineage** 52

2.2 **Normalized values of the number of drug resistances in each lineage** 53

2.3 **Number of newly found mutations in each gene ....** 55

Acronyms

AFB	Acid Fast Bacilli
AM	Arabinomannan
ATB	Active TB
Bam	Binary Alignment Format
BCG	Bacille Calmette t Guérin
DAMPs	Danger-associated Molecular Patterns
GATK	Genome Analysis Toolkit
GlcNac	N-acetyl-glucosamine
HIV	Human Immonodeficiency Virus
IFN-g	Interferon Gamma
IGRA	Interferon Gamma Release Assay
LAM	Lipoarabinomannan
LM	Lipomannan
LTBI	Latent Tuberculosis Infection
ManLAM	Mannose-capped LAM
MDR-TB	Multi-Drug Resistant Tuberculosis
MHC	Major Histocompatibility Complex
MOM	Mycobacterial Outer Membrane
MTB	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* Complex
MurNAc	N-acetyl-muramic acid
NCBI	National Center for Biotechnology Information
PAMPs	Pathogen-associated Molecular Patterns
RNA	Ribonucleic acid

AFB	Acid Fast Bacilli
SNP	Single Nucleotide Polymorphisms
SRA	Sequence Read Archive
TB	Tuberculosis
Th cell	T helper cell
TLR	Toll-like Receptors
TNF	Tumor Necrosis Factor
TST	Tuberculin Skin Test
WHO	World Health Organization
XDR-TB	Extensively-Drug-Resistant Tuberculosis

Introduction

Tuberculosis is a serious infectious respiratory disease that often affects the lungs, but can also affect other organs, it is the 13th leading cause of death and the second leading infectious killer after COVID-19 since 2020, according to the World Health Organization (WHO), which means that it is even more dangerous than HIV.

Tuberculosis is caused by *Mycobacterium tuberculosis*; an acid-fast, strict aerobic bacillus of which the cell wall content is rich in high-molecular-weight lipids and complex sugars. This slow-growing bacteria's flexible metabolic repertoire allows it to adapt to different environments encountered in host infection[1] which makes early, rapid and accurate diagnosis a crucial step in taking care of the patient, and an effective means in limiting the transmission of the infection and therefore eradicating the *Mycobacterium tuberculosis* bacteria[2].

Despite many TB control strategies, Tuberculosis pandemic is still far from being controlled[3], due to many reasons, including co-infection with human immunodeficiency virus (HIV) as well as the emergence of Multi Drug-resistant Tuberculosis (MDR-TB) and Extensively-Drug-Resistant Tuberculosis (XDR-TB).

This study aims to analyze the distribution of *Mycobacterium tuberculosis* lineages in order to determine the dominant ones and the least abundant, as well as determining the distribution of drug resistance within African countries and then verify the possible association between lineages and drug resistance, through bioinformatic analysis of the collected data using the MTBseq pipeline in order to determine the population structure of *Mycobacterium tuberculosis* in Africa.

Chapter 1

Background theory

1.1 History

TB in humans can be traced back to 9000 years ago when it was found in the remains of a mother and child buried together in Atlit Yam, an ancient submerged city under the Mediterranean Sea. Sea. [4]

TB caused approximately 25% of all deaths throughout the 1600-1800s in Europe and the United States. However, it is thought that its incidence peaked between the end of the 18th century till the end of the 19th century.[5] Before Johann L. Schonlein suggested the word ”tuberculosis” in 1834, various cultures in the world used different names to describe the illness; it was called ”phthisis” in ancient Greece, ”tabes” in ancient Rome, ”schachepheth” in ancient Hebrew, ”consumptio” in Latin America, ”yaksma” in India, and ”chaky oncay” in the Inka Empire. It was also called ”the white plague” in the 1700s due to the paleness of the patients. [6] During the Middle Ages, tuberculosis of the neck and lymph nodes was called ”scrofula” and was believed to be different from tuberculosisin the lungs. [7]

1.2 Epidemiology

Tuberculosis has claimed 1.7 million lives in 2016. Of the deaths attributable to TB in 2016, 22% occurred in people co-infected with HIV, and close to 5% of the 10.4 million incident cases of this disease were Multi-Drug Resistant (MDR-TB); i.e., they are resistant to at least two of the first-line TB drugs (Rifampicin and Isoniazid) [1]

Approximately one-quarter of the population is infected with the *Mycobacterium Tuberculosis* [8]. According to the World Health Organization (WHO), 10 million people contracted the disease in 2020, including 1,1 million children and causing more than 1,5 million deaths.

All regions are not uniformly affected by tuberculosis, some are more affected than others; Africa, for example, has the highest rates of mortality related to TB infection[8]

In Morocco, despite the significant efforts made to prevent and control tuberculosis, its frequency remains high. The incidence of TB in Morocco in 2019 was 97 cases per 100,000 people[9]

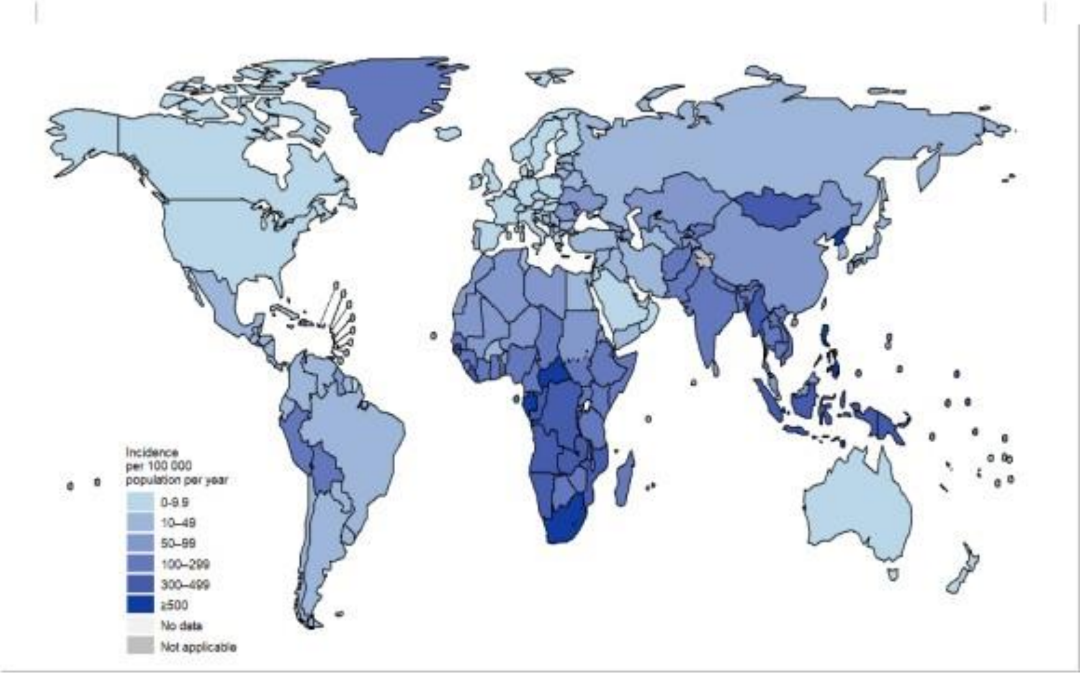


Figure 1.1: **Estimated TB incidence rates in 2020**

[10]

1.3 Taxonomy and microbiology

1.3.1 Taxonomy and classification

Mycoabcterium tuberculosis is a bacterium that belongs to the kingdom of Bacteria, its phylum is Actinobacteria, it belongs to the Actinomycetales order and the Corynebacterineae suborder, the family to which it belongs is called Mycobacteriaceae, the genus is Mycobacterium while the species is *Mycobacterium tuberculosis*.

1.3.2 Microbiological characteristics

Although the *Mycobacterium tuberculosis* cell wall is composed of an outer membrane,just like the cell walls of several gram-negative bacteria, it is considered to be a gram- positive bacterium[11]. The waxy cell wall of the bacteria can make the Gram staining method unreliable [12]. Because of the limitations of the Gram staining method acid- fast staining is an alternative method that can be used to determine the presence of mycobacteria[12]. The inclusion of phenol, which assists the dye in penetrating the cell, results in an improvement in the staining done with either fuchsin or the fluorescent dye auramine. In a subsequent decolorization step, only the mycobacterial cell wall is resistant to acidic solvents, resulting in the retention of the dye[13].

Tubercular bacteria usually appear as straight or slightly curved rods but may havevariable shapes and sizes ranging from short coccobacilli to long rods[14]. The size of the bacteria was reported to be 1-10 um long (usually 3-5 um), and 0.2-0.6 um wide.[14] Similar to other Mycobacteria, *Mycobacterium tuberculosis* has a high lipid content of approximately 40 to 60% of the dry weight of the cell envelope[15]. Among these lipids, mycolic acids confer a coloring particularity to mycobacteria called acid-fast resistance, which allows them to be distinguished from most other bacteria during Ziehl-Neelsen coloration.[13]

The cell wall is made of a layer of peptidoglycan surrounding an inner membrane that is believed to be similar to that of other bacteria, with a periplasmic space in between. Another layer of arabinogalactan is covalently attached to the peptidoglycan, which forms a hydrophobic barrier, while most arabinan residues are attached to long fatty acids. The mycolic acids that account for 60% of the mycobacterial cell wall generate the distinc- tive waxy coat of mycobacteria, also named the mycobacterial outer membrane or the mycomembrane, which allows the passage of hydrophilic substances thanks to transport proteins and porins in the mycolic acid layer.[16] [17]
Furthermore, the mycomembrane contains a number of free lipids such as glycolipids, phthioceroldimycocerosates, cord factor or dimicolyltrehalose, sulpholipids, and phosphatidylinositolmannosides, many of which are accounted as virulence factors with immunomodulatory properties, as well as lipoarabinomannan (LAM), which is found through- out the cell wall.[16] [17]

Biochemical analysis of the capsule of MTB suggested that it is mainly made of polysaccharides, with alpha-glucan being the most abundant. Additionally, arabinoman nan and mannan were also found. Some proteins and lipids have been found within the capsules, while others are mostly located in the inner part of the capsules. It is also sug- gested that the lipids may possibly interact with the mycolic acids to form a protective barrier.[16]

The mycobacterial cell wall helps protect the bacterium from dehydration, osmosis, and drugs, which helps contribute to the bacterium's inherent resistance to antibiotics.[16]

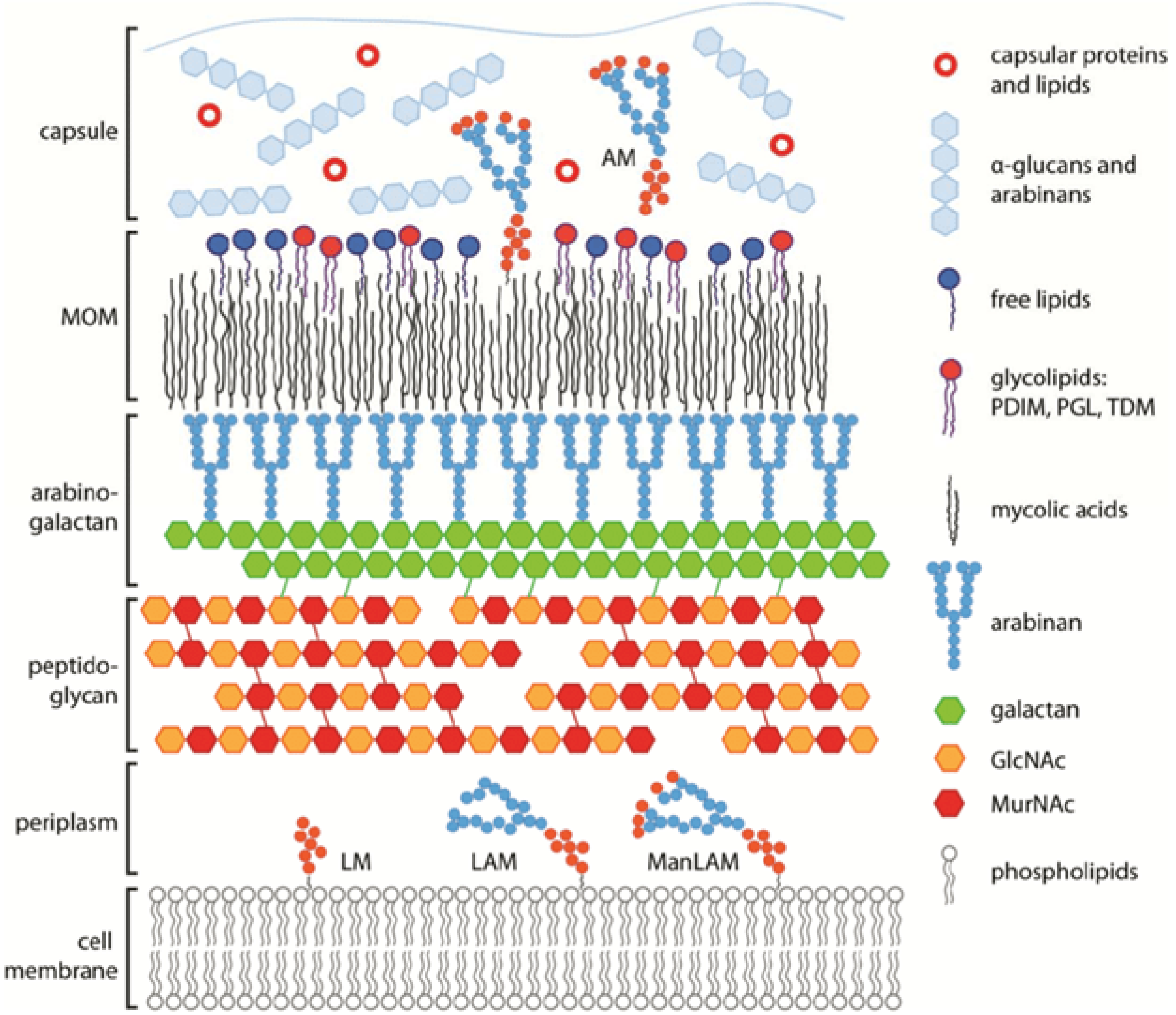


Figure 1.2: **Structure representation of the *Mycobacterium tuberculosis* cell wall** (MOM = mycobacterial outer membrane, LM = lipomannan, AM = arabinomannan, LAM = lipoarabinomannan, ManLAM = mannose-capped LAM, GlcNac = N-acetyl-glucosamine, MurNAc = N-acetyl-muramic acid.)

1.4 Pathology

1.4.1 Transmission

These droplet nuclei are generated and spread when people with active pulmonary TB disease cough, sneeze, shout, sing, or talk and can survive in the air for several hours,[18] [19] the infection then occurs when a surrounding person breathes in the bacteria. Therefore, the transmission of tuberculosis might be a serious danger to people who share the same space for long periods of time. The more bacteria are present in a person's sputum, the more likely they are to be contagious. Meanwhile, those that are positive for culturebut negative for microscopic examination are generally less contagious. Patients with negative microscopy and sputum culture are usually not contagious.[20] [18] [19]

The transmission of *Mycobacterium tuberculosis* depends on several factors including:

- **Immunity state and medical condition of the exposed person[21]:**
  - Co-infection with HIV or other immunosuppressive conditions.
  - Malnutrition: both micro-deficiency and macro-deficiency.
  - Age: people of young age are more susceptible to get infected than adults.
  - Diabetes: people with higher blood sugar levels are more at risk of being infected with TB.
- **Contagiousness of the source case[21]:**
  - Bacillary load.
  - Positive culture for *M. Tuberculosis*.
  - Positive AFB (Acid-Fast Bacilli) sputum smear.
  - Not covering the nose and mouth while coughing.
- **Environmental factors[22]:**
  - Indoor air pollution.
  - Overcrowded living conditions.
  - Tobacco smoke.

- **Exposure factors[23]:**
  - Proximity to infectious cases.
  - Duration of exposure.
  - Frequency of exposure.

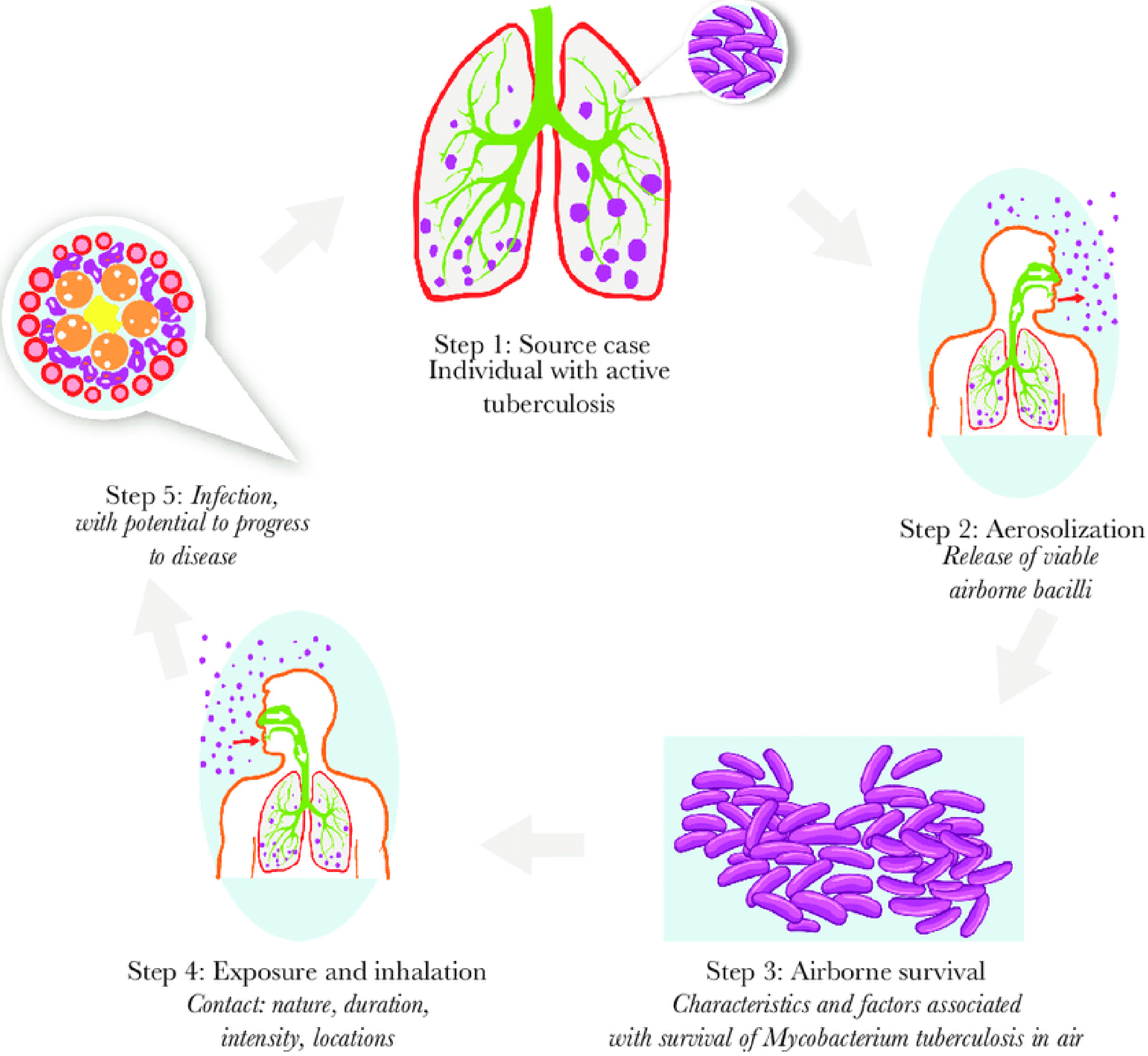

Step 1: Source case
Individual with active tuberculosis
Step 2: Aerosolization
Release of viable airborne bacilli
Step 3: Airborne survival
Characteristics and factors associated with survival of Mycobacterium tuberculosis in air
Step 4: Exposure and inhalation
Contact: nature, duration, intensity, locations
Step 5: *Infection, with potential to progress to disease*

Figure 1.3: **Transmission of *Mycobacterium tuberculosis***
[24]

1.4.2 Immune response

Once a pathogen manages to pass the physical barriers of the human body, the innate immunity instantly comes into play in order to eliminate it. The first cells to engage with the infection are macrophages, neutrophils, basophils, and eosinophils. However, in a tuberculosis infectoin the first cells to engage with the pathogen are alveolar macrophages. Pathogen-associated molecular patterns (PAMPs) are shared by most pathogens in re- sponse to pathogens and produce DAMPs, which allow the innate immune system to initiate a more tailored response [25].

Despite not being as specific as the adaptive immune system, the innate immune system still manages to detect pathogens thanks to the pattern recognition receptors present on the surface of the immunity cells, these receptors serve to recognize the PAMPs and the DAMPs. Toll-like receptors 4 (TLR4) exist on the surface of the cell and their role is to recognize lipo-proteins on gram - bacteria, meanwhile, Toll-like receptors 5 exist in the cytoplasm and they can recognize single-stranded RNA viruses. the pattern recognition receptors will be activated depending on the pathogen that attacks the human body and they will initiate the production of multiple pro-inflammatory proteins in order to eliminate the pathogen[25].

When it comes to the adaptive immune system, antigen-presenting cells such as dendritic cells or macrophages are necessary to alert the adaptive immune system of the presenceof a strange body, to do this, the antigen-presenting cells must engulf the pathogen before digesting it, and presenting portions of the pathogen also called antigens on a receptor known as MHC II to the cell receptors, then, to activate helper T cells, dendritic cells transmit different types of signals. Once the T cells are activated, they differentiate intoth1, th2 or th17 depending on the signal from the antigen-presenting cell[26].

The innate immune system also recruits cytokines to assist T cell differentiation into the suitable T helper cell, which can then activate specific B cells or cytotoxic cells to eliminate the infection[25].

Due to the fact that the adaptive immune system uses a precise antigen introduced to it by the innate immune system, it can mount a much more suitable response than the innate immune system[26].

Upon exposure to *mycobacterium tuberculosis*, alveolar macrophages detect PAMPs onthe bacterium using toll-like receptors 2 and 4 and then engulf the bacterium, normally,the phagosome is fused with the lysosome and this should digest the bacterium, which would clear the infection, however, to evade the immune system, TB usually blocks the fusion of lysosomes and phagosomes by interfering with MHC class II antigen presentation and thus delaying adaptive immune response and the bacteria may keep the phagosome altogether and live undetected in the cell's cytoplasm. Infected macrophages may clear the infection, be overrun by the bacterium and die or the bacteria may lie dormant undetected inside the macrophages. There are various factors that play an important role when it comes to that ability of the macrophages to eliminate the bacterium, but arguably, the most important factor is vitamin D as it is necessary for the production of cathylicitins and defensins, these proteins are important because they help disrupt the waxy cell membrane of *Mycobacterium tuberculosis*[27].

Sometimes, the primary innate response of macrophages is not sufficient to remove the bacteria and the bacteria continues to divide, in that case, the dendritic cells come into play; they are responsible to take the antigen to the lymph nodes where B and T cells are waiting to be useful, however, since the dendritic cells deliver the TB bacterium straight to the lymph nodes, it may set up a center of infection within the lymph node[28].

Despite mycobacterial attempts to escape the immune system, dendritic cells do pull through and present the antigens to the appropriate T cells, this interaction between Tcells and antigen presenting cells requires 3 types of signals; MHC class II T cell receptor interaction, a co-stimulated interaction and finally a cytokine signal, in this case, the cy- tokine signal is il-12 which is necessary to help T helper cells differentiate into th1 cells,so once the t helper cell receives all the signals, it becomes a th1 cell; th1 cells produce IFN-gamma: a cytokine that activates macrophages, those activated macrophages are able to destroy the bacteria and so the amount of bacteria begins to drop, however, atthis level, the immune system still struggles to fully eradicate the bacteria as they keep evading the immune system by reducing the replication rates in order to prevent detection. But because of the constant stimulation of interferon gamma, activated macrophages be- come epithelial macrophages, and they're called epithelial because their shape is similar to epithelial cells, their main job is to remove the infection and form a granuloma, so a gran- uloma is made of an aggregation of epithelial macrophages, fibroblasts also take place producing collagen to make this barrier even stronger, epithelial macrophages produceTNF-alpha that maintains the granuloma keeping the bacterium trapped. at this stage, the patient may have a latent TB infection as long as the bacteria are trapped, however,in case an immunosuppression occurs in the future, the bacteria will colonize the lungsby replicating rapidly once again, and since the body owns an immune memory against the TB bacteria, the adaptive response goes into override and causes a hypersensitivity reaction that results in cavitation within the lungs, so when the patient coughs, sneezes, or talks, the bacterium is exhaled and infects more people.[26] [29]

1.5 Latent TB infection and TB disease

1.5.1 Latent TB infection

According to the World Health Organization, latent tuberculosis infection, also known as LTBI is defined as a condition in which there is a persistent immune response to astimulus from the tuberculosis bacteria, with no clear signs of active tuberculosis. Most people who inhale and become infected with M. tuberculosis can fight the bacteria and stop their growth, without completely eliminating them. Therefore, M. tuberculosis remainsinactive until better conditions are found for it to grow again[30]. The length of the latency period varies from a person to another, and people in good health can carry LTBI throughout their lives. There is usually a 5 to 15% chance of re-infection in the first fewyears after infection, often occurring in the 2-5 years range. Reactivation refers to the process by which an inactive tuberculosis infection evolves into an active infection, which makes patients with LTBI a great reservoir of new cases of active tuberculosis[30].

People with latent tuberculosis infection:

- are asymptomatic.
- do not feel sick.
- do not spread tuberculosis to others.
- are usually positive for tuberculosis skin test reaction or tuberculosis blood test.
- may develop TB disease if they do not receive proper treatment for latent TB infection.

Despite not being fully understood yet, the underlying factors contributing to LTBI reactivation are thought to be of bacterial origin, environmental or other host-related factors. While the risk of lifetime reactivation varies between approximately 5% and 15% in healthy individuals with proven LTBI, various comorbidities and risk factors increase this risk and hence the rate of progression to active tuberculosis. The greatest risk factor is infection with the human immunodeficiency virus (HIV) as it has been proven that the risk of reactivation is 100 times higher in individuals co-infected with HIV, even after successful treatment[31]. other co-morbidities in relation with LTBI reactivation are divided in different categories depending on their significance:

The high-risk category includes, but is not limited to, patients with chronic kidney disease, transplanted immunosuppressants and patients with silicosis[32].

The risk is moderate in patients taking TNF-alpha inhibitors, which are prescribed for many autoimmune and inflammatory diseases, or glucocorticoids, as well as in patients with diabetes (all types) and children under four years of age who have recently been infected[33].

the risk is low amongst individuals with alcohol abuse, smokers, underweight patients or those suffering of malnutrition[21].

The incidence of tuberculosis is higher in these groups than in the general population.

1.5.2 TB disease

Unlike latent TB infection, TB disease or active TB (ATB) is characterized by a failure of the immune system to fight the infection which results in a spread of the bacteria, this makes tuberculosis a symptomatic and especially contagious disease.

Symptoms of active tuberculosis include[34]:

- a persistent cough that:
  - lasts longer than two weeks.
  - sometimes comes with hemoptysis (coughing up blood).
  - sometimes accompanied by mucus (thick fluid from lungs or airways).
- chest pain.
- weakness or fatigue.
- a loss of weight.
- loss of appetite.
- chills.

- fever.
- night sweats.

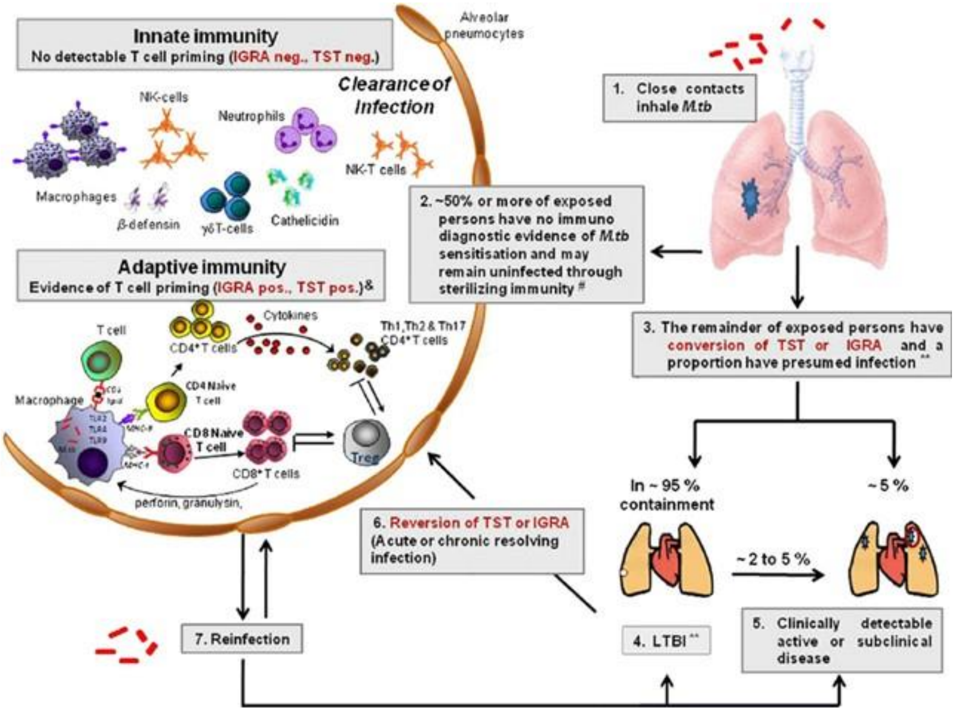

Innate immunity
No detectable T cell priming (IGRA neg., TST neg.)
Alveolar pneumocytes
Clearance of Infection
NK-cells
Neutrophils
NK-T cells
Macrophages
β-defensin
γδT-cells
Cathelicidin
Adaptive immunity
Evidence of T cell priming (IGRA pos., TST pos.)&
Cytokines
T cell
CD4⁺T cells
Th1,Th2 & Th17 CD4⁺T cells
Macrophage
CD4 Naive T cell
CD8 Naive T cell
CD8⁺T cells
Treg
perforin, granulysin,
1. Close contacts inhale M.tb
2. ~50% or more of exposed persons have no immuno diagnostic evidence of M.tb sensitisation and may remain uninfected through sterilizing immunity #
3. The remainder of exposed persons have conversion of TST or IGRA and a proportion have presumed infection **
In ~ 95 % containment
~ 5 %
~ 2 to 5 %
4. LTBI **
5. Clinically detectable active or subclinical disease
6. Reversion of TST or IGRA (Acute or chronic resolving infection)
7. Reinfection

Figure 1.4: **Life cycle of *Mycobacterium tuberculosis***

[35]

1.6 Diagnosis

Individuals suspected of being infected with TB should be referred for a complete medical evaluation, Various diagnostic tools are available to identify TB infection and TB disease:

1.6.1 Medical History

Doctors should enquire about the patient’s previous exposure to tuberculosis, infections, or illnesses while taking into consideration demographic factors that may increase the patient’s risk of exposure to tuberculosis or drug-resistant tuberculosis (such as country of origin, age, ethnic or racial group, and occupation). Furthermore, doctors should find out if the patient has any illnesses that increase the likelihood of latent TB infection progressing to active TB disease, such as HIV infection or diabetes. [36]

1.6.2 Physical examination

A physical exam can reveal essential aspects about a patient's general wellbeing as well as other conditions that might have an impact on how TB is treated, like HIV infection or other diseases. [36]

1.6.3 Tuberculin Skin Test

A person's Koch bacillus infection can be detected with the Tuberculin Skin Test (TST), commonly known as an intradermal reaction. This test involves injecting tuberculin into the forearm's dermis and monitoring the area for 3 to 5 days to see if a reaction develops. This points out possible tuberculous bacilli contamination. Additional tests are required to confirm TB disease. [36]

1.6.4 Interferon Gamma Release Assays

Consists of dosing the IFN-g or T cells that produce IFN when certain *Mycobacterium tuberculosis* antigens are present (ESAT-6 and CFP-10). It is considered an alternative to TST since there isn't a positive reaction if the patient is vaccinated with BCG.
The only methods for identifying latent TB infection are TST and IGRA testing. They do not, however, distinguish between the latent TB infection and active TB disease phases. A chest x-ray should be done when the TST or IGRA is positive. [36]

1.6.5 Chest X-ray

The initial test carried out when there is a suspicion of pulmonary TB is a chest X-ray. However, a bacteriological confirmation must be obtained in order to validate the suspicion raised by the abnormal radiological pictures. However, If a TST or TB blood test was positive but the patient showed no signs of the disease, a chest radiograph might be done to rule out the possibility of pulmonary TB.[37]

1.6.6 Diagnostic microbiology

1.6.6.1 Microscopic examination

Sputum smear analysis under a microscope is essential. When seen on a sputum smear or other specimen, Acid-Fast Bacilli (AFB) are frequently a sign of TB illness. Nevertheless, Acid-Fast microscopy does not confirm a TB diagnosis as not every Acid-Fast Bacilli is M. Tuberculosis. Therefore, a culture is essential to confirm the diagnosis.[38]

1.6.6.2 Culture

Culture on a suitable medium confirms the diagnosis of TB and makes it possible to perform an antibiogram. The culture is obtained in about 3 to 4 weeks in the solid medium of Löwenstein-Jensen and 10-15 days in a liquid medium. the more bacilli in the sample, the shorter the delay.[38]

1.7 Drug resistance

Regarding their sensitivity to antibiotics, mycobacteria are characterized by their natu- ral resistance to most common antibiotics such as Beta-lactamines, macrolides, cyclins, sulfonamides and glycopeptides. this resistance is due to the low permeability of the mycobacterial wall, which is 100 to 1000 times lower than that of Escherichia Coli for Beta-lactamines for example[39].

However, mycobacteria are naturally sensitive to few antibiotics used for treatment, such as rifampicin, isoniazid, pyrazinamide[40], streptomycin and ethambutol. However, these mycobacteria can acquire, by mutation, resistance traits antibiotics to which they are naturally sensitive[41].

acquired antibiotic resistance in mycobacteria is always linked to mutations in chromo-somic genes and is not transferable from one strain to another[41].

in strains resistant to several antibiotics, it is thought that each of the resistances is acquired independently of the others, usually in a successive way depending on the an- tibiotics used for the treatment[42]. These mutations take place either in the antibiotic target structure genes which results in a decrease in the target affinity for the antibioticor in the gene encoding an activating enzyme that serves to transform the antibiotic intoan active substance[43].

A bacterium is said to be Mono-resistant when it is resistant to a single antibiotic, when it is resistant to the two main drugs of the treatment which are Rifampicin and Isoniazid, the bacterium is considered Multi Drug-Resistant (MDR), further resistance to fluoroquinolones OR at least one second-line drug but not both leads to Pre-Extensively Drug-Resistant Tuberculosis (Pre-XDR-TB) while resistance to Rifampicin, Isoniazid, Fluoroquinolones AND at least one second-line drug leads to Extensively Drug-Resistant Tuberculosis (XDR-TB) [44].

the frequency of these mutations is assessed by the proportion of resistant mutants in a sensitive bacterial population, while the proportion of double mutant is equal to the product of the proportions of each mutant taken in isolation, this increases the likelihood of success of dual therapy.

1.8 Treatment

Once a diagnosis has been made, treatment for TB follows standardized patterns that depend on the category of patient. Tuberculosis is thus classified according to location, pulmonary or extrapulmonary, bacteriology, history of TB treatment, and serological status concerning the human immunodeficiency virus. From a public health perspective, adequate treatment of patients, especially contagious patients, is the most effective measure in the fight against tuberculosis.

A tuberculous lesion has 2 different bacillary populations[45]:

- a very rich population, whose rapid multiplication is responsible for the development of colonies that are resistant to each of the antibiotics, which prohibits monother-apy and justifies an initial phase of intensive treatment based on the simultaneous administration of several antibiotics.

- a population with slower multiplication, present in macrophages, less accessible to antibiotics and can be the cause of relapses. The eradication of these bacilli requiresa consolidation phase extended over several months.

1.8.1 Anti-tuberculosis drugs

Four 1st line drugs are usually used for TB treatment[40]:

- isoniazid and rifampicin, known as major anti-tuberculous drugs because they have the following properties: they are bactericidal; their good diffusion allows them to reach intra- and extracellular bacilli; the natural resistance of the TB bacillus to these drugs is relatively rare (1/10$\hat{8}$ for rifampicin and 1/10$\hat{5}$ for isoniazid).

- the simultaneous administration of these two antibiotics allows a rapid reduction in the number of extracellular bacilli and therefore rapid negativity of sputum smear.

- pyrazinamide, effective on intracellular bacilli, shortens treatment duration.

- ethambutol, bacteriostatic.

In addition to these 4 essential medicines, streptomycin can, in some cases, replace ethambutol.

Table 1.1: **Posology and contraindications of first line Anti-TB drugs** [40] [46]

Drug	Administration	Dosage	Contraindications
Rifampicin (R)	- Oral administration, - Intravenous administration in critically ill patients	- Daily: 10 mg/kg (600 mg max) - 3x/week: 10 mg/Kg (600 mg max) - 2x/week: 10 mg/Kg (600mg max)	- Known hypersensitivity to rifamycins, - Active, unstable hepatic disease
Isoniazid (H)	- Oral administration, - Intramuscular or intravenous administration in critically ill patients,	- Daily: 5 mg/Kg (300 mg max) - 3x/week: 15 mg/Kg (900 mg max) - 2x/week: 15 mg/Kg (900 mg max) - 1x/week: 15 mg/Kg (900 mg max	- Known hypersensitivity - Active, unstable hepatic disease
Pyrazinamide (Z)	Oral administration	*44 <weight<55Kg: - Daily: 1000 mg/Kg - 3x/week: 1500 mg/Kg - 2x/week: 2000 mg/Kg *56<weight<75Kg: - Daily: 1500 mg/Kg - 3x/week: 2500 mg/Kg - 2x/week: 3000 mg/Kg *76<weight<90Kg: - Daily: 2000 mg/Kg - 3x/week: 3000 mg/Kg - 2x/week: 4000 mg/Kg	- Known hypersensitivity - Active, unstable hepatic disease - Porphyria
Streptomycin (S)	- Deep intramuscular injection - Intravenous administration	- Daily: 15 mg/Kg (1000 mg max) - 3x/week: 15 mg/Kg - 2x/week: 15 mg/Kg * age >60 years: - Daily: 10 mg/Kg (500-750 mg max) * weight <50Kg: - Daily: 500-750 mg maximum	- Known hypersensitivity - Auditory nerve impairment - Myasthenia gravis - Pregnancy
Ethambutol (E)	Oral administration	*44 <weight<55Kg: - Daily: 800 mg/Kg - 3x/week: 1200 mg/Kg - 2x/week: 2000 mg/Kg *56<weight<75Kg: - Daily: 1200 mg/Kg - 3x/week: 2000 mg/Kg - 2x/week: 2800 mg/Kg *76<weight<90Kg: - Daily: 1600 mg/Kg - 3x/week: 2400 mg/Kg - 2x/week: 4000 mg/Kg	- Known hypersensitivity - Pre-existing optic neuritis from any cause

Fixed combinations of 2 or more anti-tuberculosis drugs exist and are recommended by the WHO, their interest is to limit prescription errors, simplify treatment by reducing the number of daily tablets to be taken, this promotes the patient’s adherence to treatment and avoids the risk of taking one or more antibiotics irregularly[47].

sometimes, in addition to first-line drugs, second-line drugs can also be used to avoid the survival of bacteria that are resistant to first-line drugs, the most used second-linedrugs are: fluoroquinolones and second-line injectables such as: Amikacin, Kanamycin and Capreomycin[48].

1.8.2 Treatment regimens

The treatment regimen varies depending on whether the patient has been previously treated or not, but in all cases, it consists of two distinct phases[49]:

- An initial or intensive two-month phase to rapidly destroy M. tuberculosis bacilli, prevent the development of resistant bacilli and eliminate contagiousness.
- A continuation phase: of varying duration depending on the clinical situation. This phase is used to prevent relapses.

In practice, there are three clinical situations requiring different treatment regimens:

- new cases;
- cases of reprocessing;
- Multi Drug-Resistant tuberculosis: MDR-TB.

1.8.2.1 Treatment Regimens for New Cases of Pulmonary Tuberculosis:

[40] Several therapeutic protocols have been validated. The standard regimen has 2 phases:

- an intensive initial phase combining first-line anti-tuberculosis drugs: Isoniazid (H), Rifampicin (R), Pyrazinamide (Z) and Ethambutol (E) for 2 months (2HRZE);
- followed by a consolidation phase involving the two major TB drugs for 4 months: 4HR.

The initial phase should be extended by one month if the direct examination is still positive to 2 months and if the result of the susceptibility test is not yet available.

1.8.2.2 Treatment Regimens for Previously Treated Patients:

[49] In all TB patients who have ever had treatment, the history of previous treatment should be well documented to assess the risk of resistance.
It is also recommended, whenever possible, to carry out a culture and an antibiogram at the beginning of the treatment in order to detect any possible resistance, including multi-resistance, this is 5 times more common in previously treated individuals than in new cases (15% vs. 3%).
The antibiotic treatment must be strictly adapted to the sensitivity of the bacilli.

- **In case of low risk of multi-resistance:** The risk is considered low when it is a relapse or re-treatment after interruption.
  In this situation, the recommended treatment consists of 5 drugs by adding streptomycin to HRZE for 2 months, then 4 drugs for 1 month and 3 drugs for 5 months (2SHRZE/1HRZE/5HRE).
  The treatment will be secondarily adapted to the antibiogram data.
  For some, streptomycin is replaced by Amikacin or Fluoroquinolone.

- **In case of high resistance risk:** These are cases of previous treatment failures and cases of exposure to a patient with Multi Drug-Resistant TB.
  It is recommended in these situations, to prescribe a regimen of Multi Drug-Resistant tuberculosis, while waiting for the antibiogram results.

1.8.3 Treatment monitoring

Regular follow-up of the patient is necessary in order to evaluate the effectiveness of the treatment and to detect any possible adverse effects[45].
The monitoring of treatment effectiveness is based on:

- clinical examination: monitoring temperature, weight and functional symptomatology.

- Chest X-ray: it is recommended to perform a chest X-ray at the end of treatment to ensure its favorable evolution.

- Bacteriological examinations should be performed after 2 months, 5 months and 6 months of TB treatment:

  the bacteriological negativity of expectoration is usually obtained during the first 2 months of treatment, if not, the patient's adherence to the therapy should bechecked as well as drug resistance, based on the antibiogram results;

  - if adherence to therapy is confirmed and in the absence of resistance, the initial treatment will be maintained and the search for the bacilli in the sputum will be repeated at the end of the 3rd month.

    The consolidation phase will be started once this exam is negative;
  - in the case of negative microscopic pulmonary tuberculosis:

    the effectiveness of the treatment is assessed on the clinical and radiological evolution;
  - in the case of extra-pulmonary tuberculosis:

    the effectiveness of the treatment is assessed on the clinical evolution and on various additional examinations deemed necessary by a specialist[45].

1.8.4 Treatment of latent TB infection (LTBI)

LTBI treatment is a preventive approach that aims to avoid progression to active TB. The most validated preventative treatment is Isoniazid monotherapy at 5 mg/kg daily for 6-9 months. But studies have shown that dual Isoniazid and Rifampicin therapy for three months would allow a better improvement in patient adherence.[50]

1.9 *Mycobacterium tuberculosis* lineages

Mycobacterium tuberculosis complex (MTBC), the cause of tuberculosis, comprises seven human-adapted, phylogenetically diverse lineages that are related to different geographical locations.

Some of these lineages have a greater worldwide reach than others.

Particularly, Lineages 2 and 4 are the most prevalent lineages. Although it is also found in Central Asia, Russia, and South Africa, Lineage 2 (also known as the East-Asian lineage, which contains the Beijing family) predominates in East Asia, while lineage 4 (also known as the Euro-American lineage) is commonly found in populations from Asia, Europe, Africa and America[51].

On the other hand, Lineages 1 (East-African-American lineage) and 3 (Central Asian Strain lineage) are more confined to East Africa, Central, South, and South-East Asia. The most geographically constrained lineages are Lineages 5-7, all of which are connectedto certain parts of Africa. lineages 5 and 6 named The West Africa 1 and West Africa2 lineages respectively, are found nearly exclusively in West Africa. While Lineage 5 predominates further east in areas bordering the Gulf of Guinea, Lineage 6 is mostly found in the western part of West Africa[51].

Similarly, for unknown reasons, Lineage 7 is restricted to Ethiopia and recent immigrants from that region of the world.

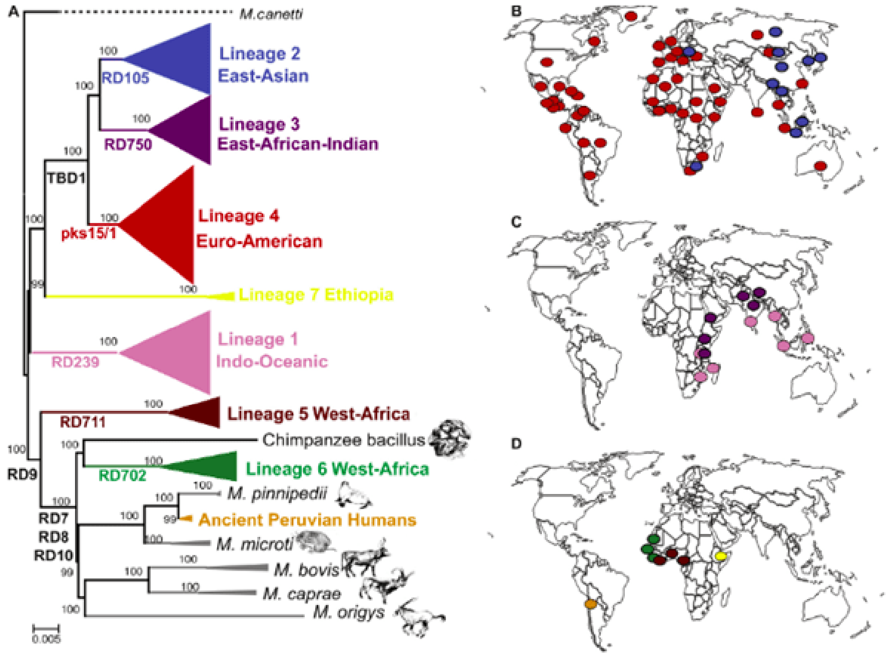


Figure 1.5: **distribution of the 7 *Mycobacterium tuberculosis* lineages in the world**

[36]

1.10 Genomics use in epidemiology

Genomics is a discipline based on the study of whole genomes, as opposed to genetics that deals with individual variants or individual genes. Genomics is a science that allows the study of the entire genetic information of an organism in order to understand the functioning of the human organism. It is a discipline that involves sequencing large amounts of DNA, a technique used in disease diagnostics and produces a huge amount of data. Genomics uses bioinformatic methods to analyze the structure and function of genomes. [52] [53]

Genomics has a significant role in epidemiology as it is sometimes necessary to analyze the samples of bacteria or viruses that may cause diseases in order to better understand the reasons of their transmission and eradicate them or at least minimize their effects.

Chapter 2

Genomic analysis

2.1 Materials and methods

2.1.1 DATA collection

Using the search words ”*Mycobacterium tuberculosis*” and the name of each country individually as keywords (*Mycobacterium tuberculosis* AND ”country name”), the metadata of 19286 paired-end *Mycobacterium tuberculosis* samples sequenced using Illumina was downloaded from the NCBI Sequence Read Archive (SRA) (http://ncbi.nlm.nih.gov/sra).

2.1.2 Generating Fastq files

The SRA accession numbers of the samples from Homo-sapien hosts were then taken from the metadata tables and used to generate the fastq files for each sample using the fastq-dump command line.

2.1.3 MTBseq pipeline

2.1.3.1 TBbwa

The raw fastq formatted sequences are mapped against the *Mycobacterium tuberculosis* H37Rv (GenBank accession number NC 000962.3) reference genome using the BWA- MEM [54] algorithm, this step generates SAM files that are then converted into binary alignment format (BAM) using SAMtools [55].

- Input files: FASTQ
- Output files: Bam
  Bai
  Bamlog

2.1.3.2 TBrefine

The scoring methods used by genome aligners to align reads relative to the reference might limit their ability to align reads well in the presence of insertions or deletions. Genome aligners treat each read independently. The aligner may favor alignments with mismatches rather than creating a gap in either the read or the reference sequence, depending on the variation occurrence and its position within a read.

Therefore, local realignment is used to correct the mapping errors that were made byaligners as it takes into consideration all reads that span a particular position, thus, mak-ing the read alignment more consistent in regions that have insertions or deletions.

This step, alongside base call recalibration, is made with GATK (Genome Analysis Toolkit) [56].

For base call recalibration, a set of known *Mycobacterium tuberculosis* resistance-associated variants is employed by the Mtbseq pipeline.

- Input files: Bam
- Output files: Gatk.bam
  Gatk.bai
  Gatk.bamlog
  Gatk.grp
  Gatk.intervals

2.1.3.3 TBpile

Pileup files (.mpileup) are created from the obtained GATK.bam files to make the display of SNP/indel calling and alignment easier using the SAMtools program [55].

- Input files: GATK.bam
- Output files: Gatk.mpileup
  Gatk.mpileuplog

2.1.3.4 TBlist

TBlist is used to create position lists from previously generated pileup files (.gatk.mpileup). Position list files display the essential data from mapping in a table format.

- Input files: Gatk.mpileup
- Output files: Gatk.position_table.tab

2.1.3.5 TBvariants

TBvariants is a step for variant detection from position tablesm, a step in which single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) are identified.

- Input files: Gatk.position_table.tab
- Output files: Gatk position_uncovered.tab
  Gatk_position_variants.tab

2.1.3.6 TBstats

Using the SAMtools flagstat tool, the TBstats step computes an overview of mapping quality and detected variations for a dataset. This step creates or modifies the ”Mapping and VariantStatistics.tab” file. All sample statistics for the studied datasets existing in the working environment are stored in this file.

- Input files: Bam
  Gatk_position_table.tab
- Output files: Mapping_And_Variant_Statistics.tab

2.1.3.7 TBstrains

Lineage classification is made based on a set of phylogenetic SNPs (Homolka et al., 2012; Coll et al., 2014; Merker et al., 2015). The output is a classification file with reported linkages for each dataset.
The file also gives an indication of the quality of the data for the positions used to infer the phylogenetic classification.

- Input files: Gatk_position_table.tab
- Output files: Strain_Classification.tab

2.1.3.8 TBjoin

This step is the first step in the comparative analysis, it aims to create a comparative SNP analysis of a set of samples. For the joint analysis, first a scaffold of all variant positions is built from the individual variant files. Second, for all positions with a detected variant, the allele information is recalculated from the original mappings to produce a comprehensive table.

- Input files:
  Gatk_position_variants.tab
  Gatk_position_table.tab

- Output files: joint samples.tab
  joint_samples.log

2.1.3.9 TBamend

In this step, post-processing of joint variant tables occurs. This step will produce a comprehensive variant table including calculated summary values for each position. In addition, the set of positions will be processed in consecutive filtering steps. Each sample needs to have either a SNP or wild-type base at the position, and positions within repetitive regions of the reference genome or within resistance-associated genes are excluded. This filtering step results in output files carrying the "amended[unambig] phylo" ending; a full table (ending in .tab), a FASTA file containing the aligned alleles of all samples for the given position (.fasta), and a corresponding FASTA file with the headers consisting solely of the respective sample ID (plainIDs.fasta).

- Input files: joint_samples.tab

- Output files:
  joint_samples_amended.tab
  joint_samples_amended_phylo.tab
  joint_samples_amended_phylo.fasta
  joint_samples_amended_phylo_plainIDs.fasta
  joint_samples_amended_phylo_removed.tab

2.1.3.10 TBgroups

TBgroups is a step for inferring likely related isolates based on the pairwise distance of distinct SNP positions. The output files consist of a text file listing the detected groups and ungrouped isolates, and the calculated pairwise distance matrix.

- Input files:
  joint_samples_amended_phylo.tab

- Output files:
  joint_samples amended phylo.matrix
  joint_samples_amended_phylo.groups

2.1.4 Lineage distribution analysis

The output files of the TBstrains step of the MTBseq pipeline are the classification tables that assign a lineage to each sample, the lineages are then associated to the country of the samples in order to calculate the percentage of each lineage in each of the studied African countries as well as in the continent.

2.1.5 Drug resistance analysis

The VCF files were used to determine the mutations that have occurred in each sample alongside the gene in which they occurred and the drug resistance they are responsible for. After comparing the mutations to the ones in the World Health Organization's 2021 *Mycobacterium tuberculosis* mutations database and selecting the ones that are associated with resistance, the samples are divided into 6 groups:

- Sensitive: samples that are not resistant to any drugs.
- Mono-Drug Resistant: samples that are resistant to either Rifampicin or Isoniazid but not both.
- Multi-Drug-Resistant (MDR): samples resistant to Rifampicin and Isoniazid.
- Pre-Extensively-Drug Resistant (Pre-XDR): samples that are resistant to Rifampicin, Isoniazid, and either a fluoroquinolone or a second line injectable but not both.
- Extensively-Drug Resistant (XDR): samples that are resistant to Rifampicin, Isoniazid, and at least one fluoroquinolone and one second-line injectable.
- Other: samples that are resistant to drugs other than the ones previously mentioned.

2.1.6 Resistance associated genes analysis

The mutations in resistance associated genes collected in the previous step were filtered to remove the ones not associated with resistance, In order to determine the unknown mutations, the following steps were followed:

- The resistance mutations were collected from the output files of the TBvariants step of the pipeline.
- A first filtration that consists of removing mutations not associated with resistance.
- The remaining mutations were compared to the World Health Organization's Database of *Mycobacterium tuberculosis* resistance mutations of 2021 and the matching mutations were filtered out.
- The mutations that were not in the Database were then individually looked for in the literature and all existing mutations were removed.
- A list of unknown mutations was made.

2.1.7 Lineage-Drug resistance association

The number of each drug resistance in each lineage was calculated, the resulting table contained 7 columns for 7 lineages and 10 rows for drugs.
considering the big variance of values in that table, those values underwent a min-max normalization also known as 0 1 normalization in order to get all the data in the range (0,1) while preserving the relationships among the original values of the data using the following statistical formula.

$$x_{\mathrm{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Since the 2 variables (Lineages and Drugs) have more than 2 levels, the association between them was studied using the **two-way Analysis of variance (two-way ANOVA)**

biostatistical test which serves to determine if there is an association between two variables by comparing the mean differences between the groups.
The value of alpha (significance level) was set to 0.05 or 5%, with the null hypothesis H0: the lineages and drug resistance are not associated.

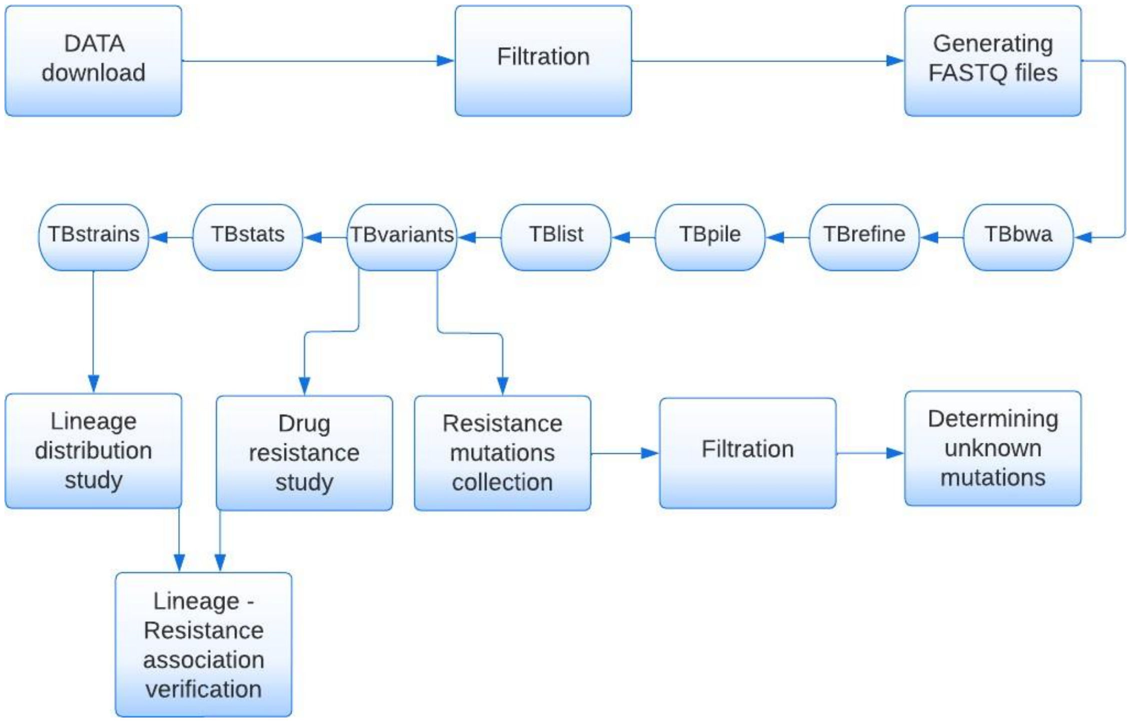

DATA download
Filtration
Generating FASTQ files
TBstrains
TBstats
TBvariants
TBlist
TBpile
TBrefine
TBbwa
Lineage distribution study
Drug resistance study
Resistance mutations collection
Filtration
Determining unknown mutations
Lineage - Resistance association verification

Figure 2.1: **Workflow of the study**

2.2 Results

2.2.1 Distribution of samples by country

After removing all the samples with non-Homo-Sapien hosts, 19198 samples from 28 countries remained of which the fastq files of 17641 samples were downloaded, thesesamples were distributed as shown in (figure 2.2).

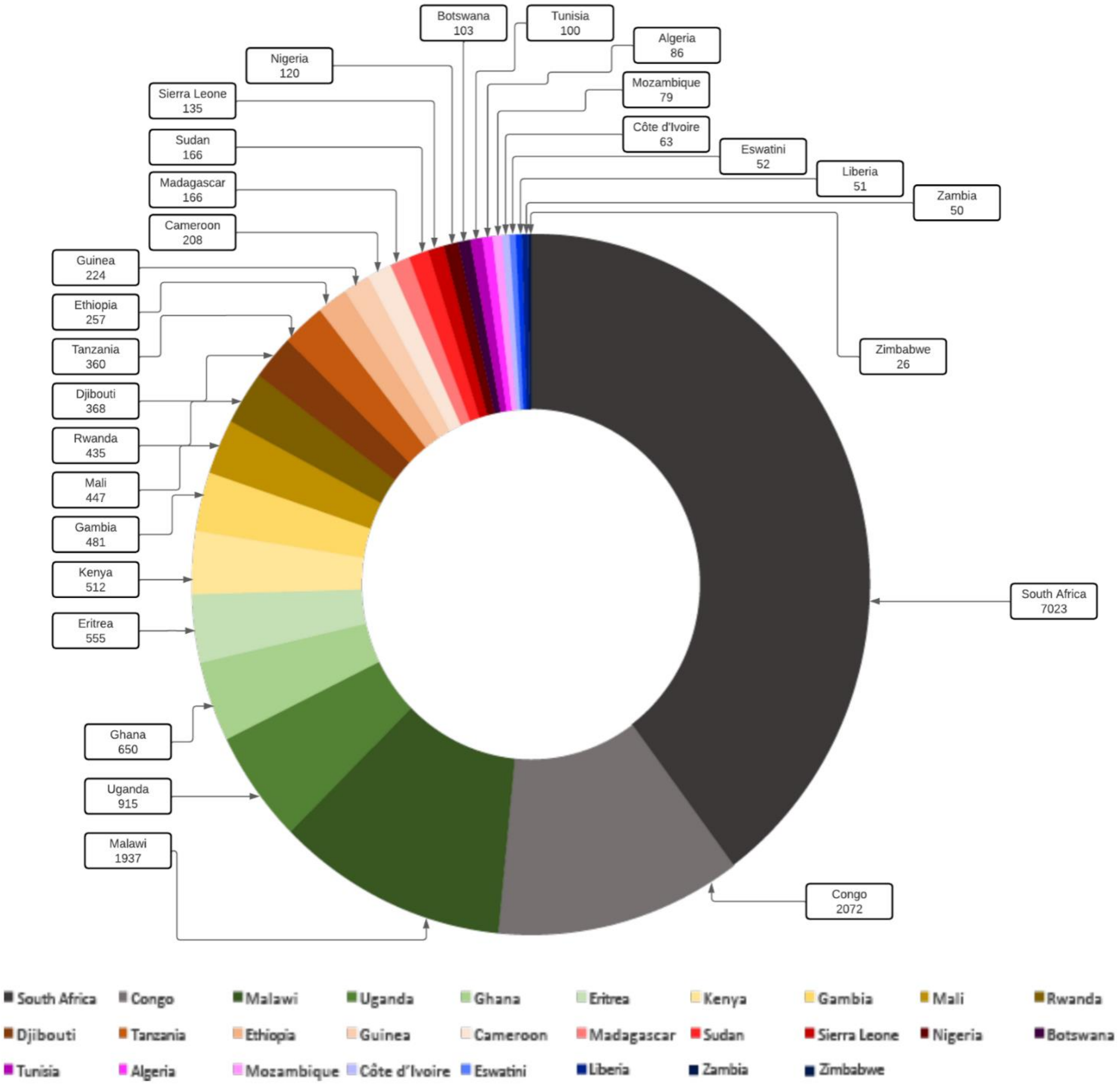


Figure 2.2: **Distribution of 17641 samples by country**

2.2.2 Lineage distribution in African countries

The classification tables obtained from the MTBseq pipeline have revealed the lineage to which each sample belongs, the most abundant lineage in the continent is lineage 4, it has a prevalence of 54.33%, followed by lineage 3 which has a prevalence of 8.62%, next is lineage 2 with a percentage of 7.41%, then lineage 1 with 7.37%, lineages 5 and 6 have a prevalence of 3.61% and 1.25% respectively, and the least abundant lineage is lineage 7 which has a prevalence of 0.58%, however, 16.83% of the samples were of unknown lineages. The percentages were also counted for each country which revealedthe following distribution:

In **Algeria**, the most prevalent lineage is lineage 4 with a percentage of 96.34%, followed by lineage 3 (2.44%) and finally lineage 1 with a percentage of 1.22%. The remaining lineages are not present in the studied samples from Algeria.

In **Botswana**, 50.49% of samples belong to lineage 4, 29.12% to lineage 1, the prevalence of lineage 2 is equal to 11.65% while lineage 3 shows a prevalence of 8.74%, meanwhile, lineages 5,6 and 7 do not exist amongst to samples used in this study.

Cameroon presents 90.82% prevalence of lineage 4, 0.97% of lineage 2, 0.48% of lineage 1. The remaining 0.48% of samples are of unknown lineages.

88.07% of samples originating from **Congo** belong to lineage 4, West African lineages 5 and 6 have a prevalence of 5.50% and 2.02% respectively, lineage 3 has a percentage of 1.83%, the least two prevalent lineages in Congo are lineages 1 and 2 with a percentageof 1.47% and 1.1% respectively, while lineage 7 is not present in the country.

Côte d'Ivoire has a 95.24% prevalence of lineage 4, 3.17% of lineage 2, 1.59% of unknown lineages.

24.37% of samples originating from **Djibouti** belong to lineage 4, 11.41% to lineage 1, lineage 3 has a percentage of 11.14% and lineage 2, a percentage of 2.45%. The lineages of the remaining 50.27% are unknown.

97.66% of the samples collected from **Eritrea** are of unknown lineage, and 100% of the remaining samples (2.34% of the total number of samples) belong to lineage 4.

In **Eswatini**, 98.08% of the samples are from lineage 4, 1.92% of lineage 2, while other lineages are not present in the country.

Ethiopia has 34.73% of samples from lineage 4, 17.57% of lineage 3, 16.32% belong to lineage 7, and a percentage of 2.09% and 0.42% for lineages 1 and 2 respectively. The remaining samples with a percentage of 28.87% are of unknown lineages.

Unlike most countries, the dominant lineage in **Gambia** is the West African lineage 6 (57.38%), followed by lineage 4 (28.07%), lineage 2 with 7.28% prevalence, lineage 1 hasa prevalence of 4.78%, the percentage of lineage 3 is 1.87% and finally, 0.62% for lineage

5.

The prevalence of lineage 4 in **Ghana** is equal to 26.15%, that of lineage 5 is 9.32%, lineage 6 has a percentage of 6.92% and 0.15% belongs to lineage 2. 57.54% of samplesare of unknown lineages.

The dominant lineage in **Guinea** is lineage 2 with 75% prevalence, followed by lineage 4 with a percentage of 22.77% and finally, lineage 1 with a percentage of 2.23%.

97.07% of samples from **Kenya** are of unknown lineages, the remaining samples are distributed as follows: 1.95% belong to lineage 4, 0.59% belong to lineage 3 and 0.39% to lineage 2.

Liberia has 66.67% samples from lineage 4, 21.57% belong to lineage 1, 9.80% belong to lineage 2, while the remaining 1.96% belong to lineage 5.

Madagascar has 63.25% of samples from lineage 1, 25.30% from lineage 4, 6.02% belong to lineage 2 and 5.42% to lineage 3.

68.27% of the samples collected from **Malawi** belong to lineage 4, 15.58% to lineage 1, lineage 3 follows with a percentage of 11.39%, then lineage 2 with 4.35% prevalence, and finally, 0.41% of the samples are of unknown lineages.

In **Mali**, 70.72% of samples are from lineage 4, 21.40% and 2.03% of them are from West African lineages 6 and 5 respectively, lineage 2 also has a percentage of 2.03%, while the percentage of lineage 1 is of only 1.80%. The lineages of 2.03% of the samples are unknown.

Mozambique has a 60.76% prevalence of lineage 4, 29.11% prevalence of lineage 1, lineage 2 has a prevalence of 8.86%, while lineage 3 has a low prevalence of 1.27%.

Nigeria has a high prevalence of lineage for (94.12%), a prevalence of 2.52% for each of the two lineages 2 and 5, and 0.84% of the samples belong to lineage 3.

The most dominant lineage in **Rwanda** is lineage 4 with a percentage of 96.32%, 0.92% is the percentage of prevalence of the 3rd lineage, 0.23% is that of lineage 2, and 2.53% is the percentage of samples of unknown lineages.

Sierra Leone also has a high prevalence of lineage 4 with a percentage of 74.07%, the second most abundant lineage is lineage 6 with a percentage of 13.33%, 5.93% is thepercentage of lineage 5, and lastly, are lineages 1 and 2 with a percentage of 4.44% and 2.22% respectively.

South Africa, the country with the most samples has a prevalence of 54.73% for lineage

4, 38.50% for lineage 2, the percentage of lineage 3 is equal to 2.43% and that of lineage 1 is 1.04%, meanwhile, the percentage of samples of unknown lineages is 3.30%.

Sudan shows a prevalence of 73.49% for lineage 3, followed by lineage 4 with a percentage of 23.49%, then 2.41% for lineage 1 and finally 0.60% of samples belonging to lineage 2.

70.19% is the percentage of prevalence of lineage 3 in **Tanzania**, followed by 13.65% for lineage 1, lineage 4 strangely only has a percentage of 8.91% and lastly, lineage 2 has a prevalence of 3.34% while 3.90% are of unknown lineages.

100% of the samples collected from **Tunisia** belong to lineage 4.

Uganda has a prevalence of 52.68% of lineage 4, 15.88% of lineage 3, the prevalence of lineage 2 is of 5.26% and that of lineage 1 is equal to 0.66%. the percentage of samplesfrom unknown lineages is 25.52%.

None of the samples collected from **Zambia** are of known lineages.

Zimbabwe has a prevalence of 65.38% of lineage 4, followed by 19.23% for lineage 2 and 15.38% for lineage 3.

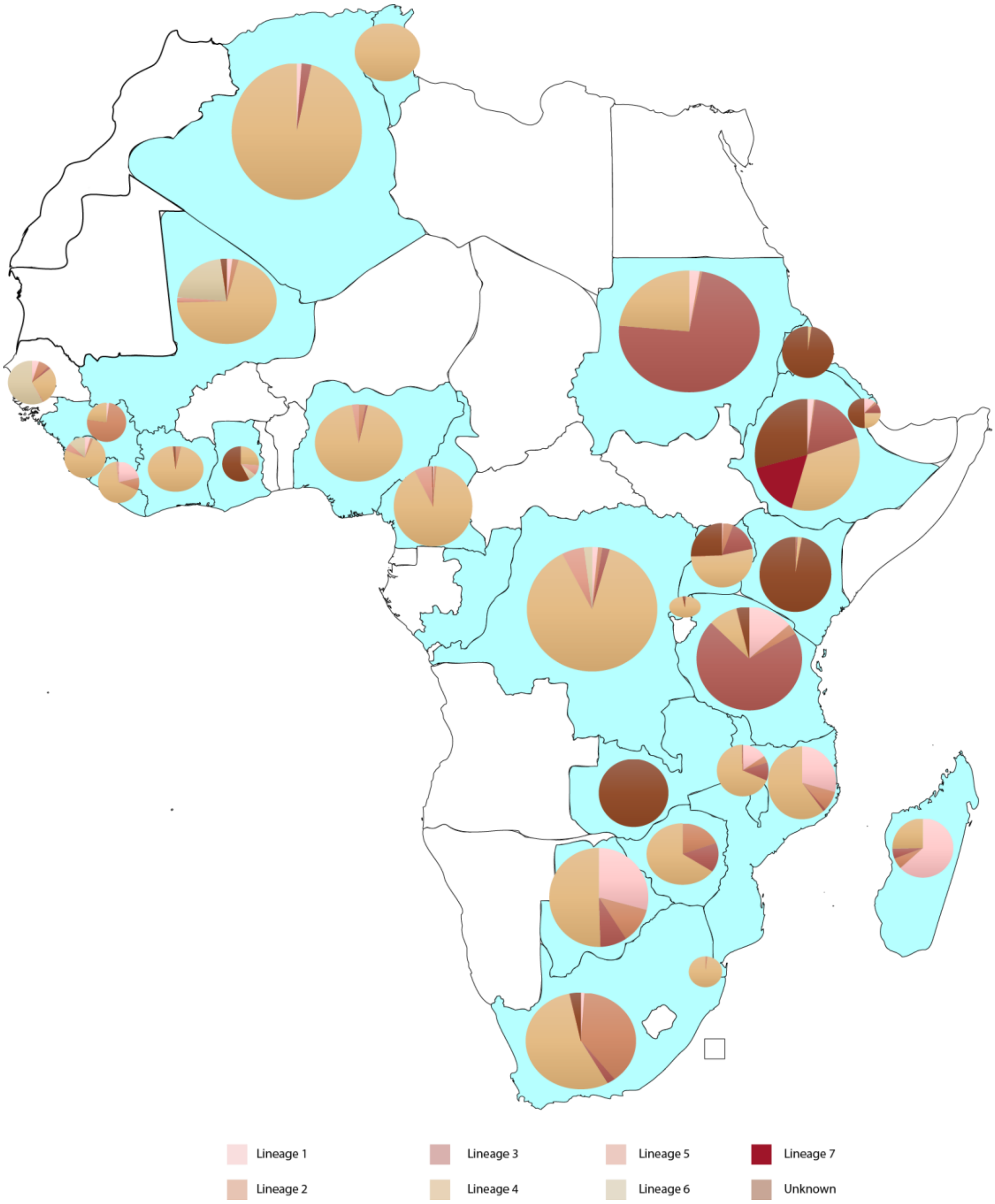


Figure 2.3: **Lineage distribution of TB in 28 African countries**

2.2.3 Drug resistance by country

The drug resistance analysis revealed that 68% are sensitive to all drugs, 9% are Mono-Drug Resistant, 14% are Multi-Drug Resistant, 7% are Pre-Extensively-Drug Resistant and 1% are resistant to other drugs.

In **Algeria**, 64 samples are sensitive to all drugs, 4 are Mono-Drug Resistant, 10 are

Multi-Drug Resistant (MDR), while 3 samples are Pre-Extensively-Drug Resistant (Pre-XDR) and 5 samples are resistant to other drugs.

Botswana has 55 sensitive samples, 19 Mono-Drug Resistant, 28 Multi-Drug Resistant (MDR) and a single sample is resistant to other drugs.

9 samples from originating from **Cameroon** are sensitive, 54 are Mono-Drug Resistant, 138 are Multi-Drug Resistant (MDR), 3 are Pre-Extensively-Drug Resistant, while 4 sam- ples are resistant to other drugs.

1740 of the samples from **Congo** are sensitive, 78 are Mono-Drug Resistant, 223 are Multi-Drug Resistant, with 11 Pre-Extensively-Drug Resistant samples and 20 that are resistant to other drugs.

Côte d'Ivoire has 30 sensitive samples, 11 Mono-Drug Resistant ones, 20 Multi-Drug Resistant samples, a single Pre-XDR sample as well as another single one that's resistant to other drugs.

294 of the samples from **Djibouti** are sensitive, 16 of them are Mono-Drug Resistant, 51 are Multi-Drug Resistant and 7 are resistant to other drugs.

All 555 samples from **Eritrea** are sensitive.

All 52 samples from **Eswatini** are resistant, of which, 3 are Mono-Drug Resistant, 48 are Multi-Drug Resistant and 1 sample is Pre-Extensively-Drug Resistant.

out of **Ethiopia**n samples, 158 are sensitive, 8 are Mono-Drug Resistant, 44 are Multi- Drug Resistant, 28 are Pre-Extensively-Drug Resistant and 19 samples are resistant to other drugs.

Gambia has 451 samples that are sensitive, 13 others are Mono-Drug Resistant, 4 Multi-Drug Resistant, a single Pre-Extensively-Drug Resistant sample and finally, 12 samples that are resistant to other drugs.

600 samples originating from **Ghana** are sensitive, 29 are Mono-Drug Resistant, 7 are multi-Drug resistant, while 14 are resistant to other drugs.

Guinea has 71 sensitive samples, 117 Mono-Drug Resistant samples, 18 are multi-Drug Resistant, while 15 are Pre-Extensively-Drug Resistant and 3 are resistant to other drugs.

out of all samples from **Kenya**, 499 are sensitive, with only 7 Mono-Drug Resistant and 6 Multi-Drug Resistant samples.

Liberia has 37 sensitive samples as well as 8 Mono-Drug Resistant, 5 Multi-Drug Resistant and a single sample that's resistant to other drugs.

alongside 115 sensitive samples from **Madagascar**, there are 21 Mono-Drug Resistant, 28 Multi-Drug Resistant and 2 samples that are resistant to other drugs.

Most samples from **Malawi** are not resistant, with 1839 samples being sensitive, 68 Mono-Drug Resistant, 13 Multi-Drug Resistant and 17 samples being resistant to other drugs.**Mali** has 261 sensitive samples, 36 Mono-Drug Resistants, 122 Multi-Drug Resistant, 18 Pre-Extensively-Drug Resistant and 10 resistant to other drugs.

26 samples from **Mozambique** are sensitive, 16 are Mono-Drug Resistant while 25 areMulti-Drug Resistant, 10 are Pre-Extensively Drug Resistant with only 2 samples resis- tant to other drugs.

Out of the samples from **Nigeria**, 31 are sensitive, 40 are Mono-Drug Resistant, 34 are Multi-Drug Resistant, 8 are Pre-Extensively-Drug Resistant while the number of samplesthat are resistant to other drugs is 7.

In **Rwanda**, 96 samples are sensitive, while 57 are Mono-Drug Resistant, 280 are Multi-Drug Resistant, a single sample is Pre-Extensively Drug Resistant as well as a single sample that is resistant to other drugs.

Sierra Leone has 94 sensitive samples, 19 Mono-Drug Resistant samples, 14 Multi-Drug resistant samples and 8 samples resistant to other drugs.

The country with the most samples, **South Africa**, has the highest numbers in all of the categories with 3867 sensitive samples, 848 Mono-Drug resistant samples, 1173 Multi- Drug Resistant samples, 1072 Pre-Extensively-Drug Resistants and 63 samples resistant to other drugs.

Sudan has 144 sensitive samples, as well as 2 Mono-Drug Resistant Samples, 12 Multi-Drug Resistant samples and 8 samples that are resistant to other drugs.

254 samples from **Tanzania** are sensitive, 39 are Mono-Drug Resistant, 58 samples are Multi-Drug resistant and 9 are resistant to other drugs.

Tunisia has 44 sensitive samples, it also has 8 Mono-Drug Resistant, 39 Multi-Drug resistant, 8 Pre-Extensively-Drug resistant and 1 sample with other resistances.

Uganda has 625 sensitive samples, as well as 153 Mono-Drug Resistants, 111 Multi-Drug Resistant, 18 Pre-Extensively-Drug Resistant and 8 samples with other resistances.

All 50 samples originating from **Zambia** are sensitive.

And finally, 25 out of the 26 samples from **Zimbabwe** are sensitive while the remaining sample is Multi-Drug Resistant.

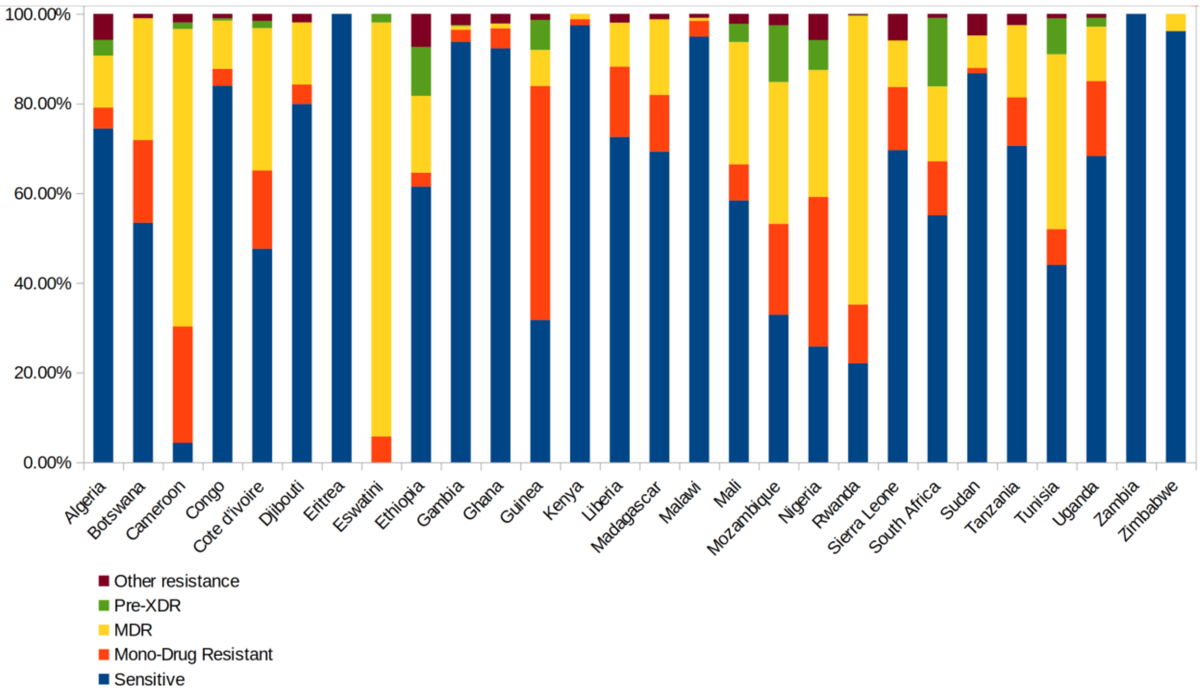


Figure 2.4: **Drug resistance by country**

2.2.4 Lineage-Drug resistance association

After calculating the number of samples resistant to each drug in each lineage, the result is displayed in table 2.2.

After min-max normalization of the values, they got in the (0 1) range, the result is displayed in table 2.3.

Table 2.1: **Number of resistances by drugs in each lineage** (CAP: Capreomycin, EMB: Ethambutol, ETH: Ethionamide, INH: Isoniazid, LEV: Levofloxacin, LZD: Linezolid, MXF: Moxifloxacin, PZA: Pyrazinamide, RIF: Rifampicin, STM: Streptomycin)

	Lineage 1	Lineage 2	Lineage 3	Lineage 4	Lineage 5	Lineage 6	Lineage7
CAP	0	0	0	29	0	0	0
EMB	74	1354	169	1952	29	17	0
ETH	14	1	1	97	2	0	0
INH	68	1089	204	2515	31	15	0
LEV	10	757	16	404	1	9	0
LZD	0	3	0	2	0	0	0

MXF	10	757	16	405	1	9	0
PZA	43	569	62	955	18	3	0
RIF	104	1774	274	2679	40	13	0
STM	26	395	43	810	28	12	0

Table 2.2: **Normalized values of the number of drug resistances in each lineage**

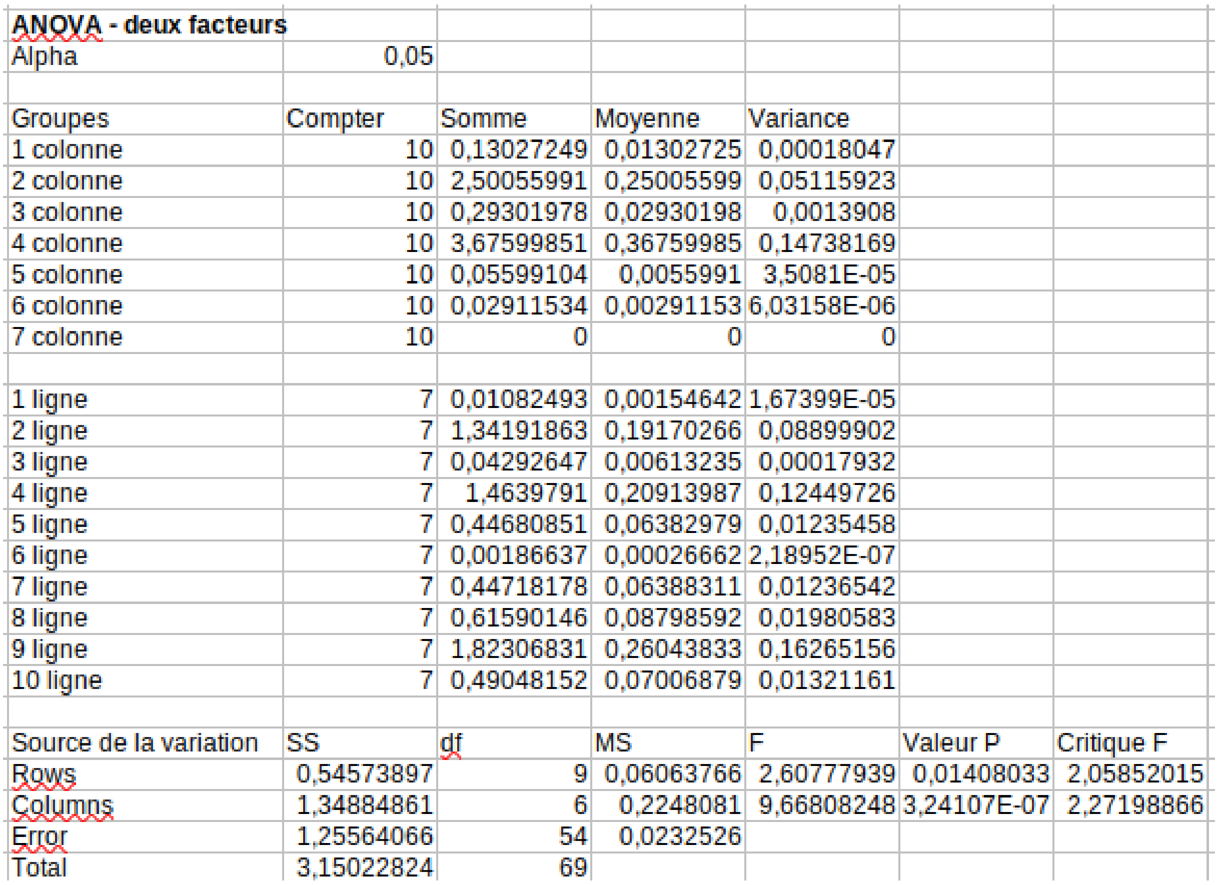
	Lineage 1	Lineage 2	Lineage 3	Lineage 4	Lineage 5	Lineage 6	Lineage7
CAP	0	0	0	0,0108249	0	0	0
EMB	0,0276222	0,5054125	0,0630832	0,7286301	0,0108249	0,0063457	0
ETH	0,0052258	0,0003733	0,0003733	0,0362075	0,0007465	0,0000000	0
INH	0,0253826	0,4064950	0,0761478	0,9387831	0,0115715	0,0055991	0
LEV	0,0037327	0,2825681	0,0059724	0,1508025	0,0003733	0,0033595	0
LZD	0	0,0011198	0	0,0007465	0	0	0
MXF	0,0037327	0,2825681	0,0059724	0,1511758	0,0003733	0,0033595	0
PZA	0,0160508	0,2123927	0,0231430	0,3564763	0,0067189	0,0011198	0
RIF	0,0388205	0,6621874	0,1022770	1	0,0149309	0,0048526	0
STM	0,0097051	0,1474431	0,0160508	0,3023516	0,0104517	0,0044793	0

2.2.4.1 Two-way ANOVA result

The two-way ANOVA was run and generated the following table as a result:

2.2.4.2 Two-way ANOVA result interpretation

- The P value for the rows (Drugs) = 0.01408033 is smaller than alpha = 0.05
- the statistical value of F for the rows (Drugs) F = 2.6063766 is bigger than F critical = 2.05852015
- The P value for the columns (Lineages) = 0.000000324107 is smaller than alpha = 0.05
- The statistical value of F for the columns (Lineages) F = 9.66808248 is bigger than F critical = 2.27198866



ANOVA - deux facteurs						
Alpha	0,05					
Groupes	Compter	Somme	Moyenne	Variance		
1 colonne	10	0,13027249	0,01302725	0,00018047		
2 colonne	10	2,50055991	0,25005599	0,05115923		
3 colonne	10	0,29301978	0,02930198	0,0013908		
4 colonne	10	3,67599851	0,36759985	0,14738169		
5 colonne	10	0,05599104	0,0055991	3,5081E-05		
6 colonne	10	0,02911534	0,00291153	6,03158E-06		
7 colonne	10	0	0	0		
1 ligne	7	0,01082493	0,00154642	1,67399E-05		
2 ligne	7	1,34191863	0,19170266	0,08899902		
3 ligne	7	0,04292647	0,00613235	0,00017932		
4 ligne	7	1,4639791	0,20913987	0,12449726		
5 ligne	7	0,44680851	0,06382979	0,01235458		
6 ligne	7	0,00186637	0,00026662	2,18952E-07		
7 ligne	7	0,44718178	0,06388311	0,01236542		
8 ligne	7	0,61590146	0,08798592	0,01980583		
9 ligne	7	1,82306831	0,26043833	0,16265156		
10 ligne	7	0,49048152	0,07006879	0,01321161		
Source de la variation	SS	df	MS	F	Valeur P	Critique F
Rows	0,54573897	9	0,06063766	2,60777939	0,01408033	2,05852015
Columns	1,34884861	6	0,2248081	9,66808248	3,24107E-07	2,27198866
Error	1,25564066	54	0,0232526			
Total	3,15022824	69				

Figure 2.5: **Two-way ANOVA result**

- Conclusion: From the previous results, we notice that both P values for Lineages and drugs are significant, we conclude that the drug resistance and the lineages are statistically associated, which means that some lineages are more responsible for certain drug resistances than others.
  We reject the null hypothesis H0.

2.2.5 New mutations

The collection and filtration of mutations associated with drug resistance in the samples on which the study was conducted resulted in the identification of 1026 new mutations affecting 28 different genes.

The numbers of mutations found in each gene are presented in table 2.4.

Table 2.3: **Number of newly found mutations in each gene**

Gene	Number of mutations
ahpC	3
atpE	2
clpC	109
DDN	4
eis	24
embA	52
embB	45
embC	43
embR	17
ethA	51
ethR	19
fbiB	18
FGD1	13
fprA	34
gyrA	55
inhA	5
katG	46
mmpL5	78
mmpS5	7
panD	4
pepQ	16
pncA	13
rplC	7
rpoA	45
rpoB	147
rpoC	149
tlyA	13
whiB7	7

2.3 Discussion

Tuberculosis is one of the most ancient diseases, it is caused by *Mycobacterium tuber- culosis* (MTB) that often affects the lungs, it causes high mortality rates[57] [58], and is considered the second most infectious killer after COVID-19 according to the World Health Organization. therefore, many strategies were considered in order to eradicate the disease[58], and to do so, we need to understand the underlying reasons for its sur- vival, such as the mechanism of tuberculous drug resistance and its association to the 7 *Mycobacterium tuberculosis* lineages as well as determining the resistance genes and new mutations that occur within them and that are associated with resistance.

The work carried out during this study helped showcase the uneven distribution of lineages and resistance throughout African countries.

Our results suggest that the African countries with the highest prevalence of tubercu- losis are South Africa, Congo and Malawi which is similar to what's suggested by an- other study[59], however, according to the same study, countries like Eswatini, Botswana, Mozambique, Zambia and Zimbabwe also have high rates of tuberculosis[59] which contradicts with our results where Zimbabwe and Zambia have the lowest rates of incidence.

In this study, we identified that the most dominant lineage in Africa is the Euro-American lineage also known as lineage 4, with a prevalence of 54.33%.

M. Senghore et al showed that the second most abundant lineage in the world is lineage 2[60], while M.B. Reed et al. suggest that lineage 1 is the second most prevalent lineage in Africa[61].

However, in our study, we identified that the second most abundant lineage in Africa is lineage 3 (also known as the Central Asian Strain lineage) and it has a percentage of prevalence that is equal to 8.62%.

Lineage 2, is known to be the most virulent lineage[62] [63] as well as being widespread worldwide, unlike lineages 5, 6 and 7 that are confined to West Africa and Ethiopia.

This diversity in lineages originating from other continents in Africa is supposed to be mainly because of immigration from and to different countries on different continents.

When it comes to drug resistance, it can be defined as the ability of the bacteria to resist to an antibiotic which causes the ineffectiveness of this antibiotic.

Many studies have shown that most resistant samples are in fact Multi-Drug resistant[64] [65], while a different study shows a higher number of MDR, followed by XDR and sensitives[65], in our case, we noticed that the number of sensitive samples in Africa is much high than the number of resistant samples, where, 12034 out of 17641 samples are sensitive to all the studied drugs.

In an effort to analyze a potential relationship between lineages and drug resistance in *Mycobacterium tuberculosis* samples, the association between them has turned out to be significant, determining the existence of a relationship between the two variables, this result is similar to that suggested in the literature where lineages 3 and 4 have been as- sociated with drug resistance including Multi-Drug Resistance in North and East Africa, and lineage 2 and 4 were proven to be highly associated to drug resistance.[66] As for re- sistance mutations, a lot of resistance-associated mutations identified in human-adapted *Mycobacterium tuberculosis* strains were mainly found in the following genes: embB, katG pncA, gidB, gyrA, fgd1,thyA and mshA.[67] [68]

While some of these genes were found to have a high number of resistance-associated mutations, more genes were also found to have high numbers of new identified mutations, such as: clpC (109 mutations), rpoC (149 mutations), rpoB (147 mutations), mmpL5 (78 mutations) and other genes with fewer new identified mutations.

Conclusion & Perspectives

This work presents genomic analysis of *Mycobacterium tuberculosis* samples originating from 28 African countries in an effort to showcase the distribution of the 7 lineages in the African continent as a first step.

The lineage analysis has revealed that lineage 4 is the more prevalent lineage in the African continent, while lineage 7 is the least abundant.

This work also aims to determine the drugs of which the ability to fight the infection might be reduced by certain mutations leading to drug resistance, sometimes to more than a single drug resulting in Multi-Drug Resistance, Pre-Extensively-Drug Resistanceor even Extensively-Drug Resistance, this analysis has revealed that most of the samples originating from Africa are susceptible to all drugs, while the most abundant type of resistance id Multi-Drug Resistance.

The project has also proved the association between the lineages and drug resistance which proves that some lineages are more prone to resistance against certain drugs than other lineages as we have found the p value for drugs is equal to 0.014 while it was equal to $3.24*10^{-7}$ for lineages which is smaller than the set value of alpha that's equal to 0.05.

New mutations that are associated with resistance have also been discovered during this project in 28 different genes, this is very much the key component in future attempts to determine the effect of the mutations on the affinity of the drug towards its target.

It is also a question of future research to investigate which drug is more associated with which lineage.

Bibliography

[1] Anastasia Koch and Valerie Mizrahi. Mycobacterium tuberculosis. *Trends in micro- biology*, 26(6):555–556, 2018.

[2] Arundhati Maitra, Tengku Karmila Kamil, Monisha Shaik, Cynthia Amaning Dan- quah, Alina Chrzastek, and Sanjib Bhakta. Early diagnosis and effective treatment regimens are the keys to tackle antimicrobial resistance in tuberculosis (tb): A report from euroscicon’s international tb summit 2016, 2017.

[3] Thomas R Frieden. Can tuberculosis be controlled? *International Journal of Epidemiology*, 31(5):894–899, 2002.

[4] Mark Spigelman, Helen D Donoghue, Ziad Abdeen, Suheir Ereqat, Issa Sarie, Charles L Greenblatt, Ildikó Pap, Ildikó Szikossy, Israel Hershkovitz, Gila Kahila Bar- Gal, et al. Evolutionary changes in the genome of mycobacterium tuberculosis and thehuman genome from 9000 years bp until modern times. *Tuberculosis*, 95:S145–S149,2015.

[5] Orsolya Anna Váradi, Dávid Rakk, Olga Spekker, Gabriella Terhes, Edit Urbán, William Berthon, Ildikó Pap, Ildikó Szikossy, Frank Maixner, Albert Zink, et al. Verification of tuberculosis infection among vác mummies (18th century ce, hungary) based on lipid biomarker profiling with a new hplc-hesi-ms approach. *Tuberculosis*, 126:102037, 2021.

[6] Giovanni Battista Migliori, Xhevat Kurhasani, Martin van den Boom, Dina Visca, Lia D’Ambrosio, Rosella Centis, Simon Tiberi, et al. History of prevention, diagno- sis, treatment and rehabilitation of pulmonary sequelae of tuberculosis. *La Presse Médicale*, page 104112, 2022.

[7] Rawle F Philbert, Andrew K Kim, and David P Chung. Cervical tuberculosis (scrofula): a case report1. *Journal of oral and maxillofacial surgery*, 62(1):94–97, 2004.

[8] J Cherif, N Ben Salah, S Toujani, Y Ouahchi, H Zakhama, B Louzir, and Rhouma Mehiri-Ben. N. and beji, m.(2014) epidemiology of tuberculosis. *Revue de Pneumologie Clinique*, 71:67–72.

[9] Oumnia Bouaddi, Mohammad Mehedi Hasan, Abdul Moiz Sahito, Pritik A Shah, Abdelrahman Zaki Ali Mohammed, and Mohammad Yasir Essar. Tuberculosis in the middle of covid-19 in morocco: efforts, challenges and recommendations. *Tropical Medicine and Health*, 49(1):1–4, 2021.

[10] Tb incidence. *World Health Organization*, 2020.

[11] LM Fu and CS Fu-Liu. Is mycobacterium tuberculosis a closer relative to gram- positive or gram–negative bacterial pathogens? *Tuberculosis*, 82(2-3):85–90, 2002.

[12] Dani S Zander and Carol F Farver. *Pulmonary Pathology E-Book: A Volume in Foundations in Diagnostic Pathology Series*. Elsevier health sciences, 2016.

[13] Gary Kaiser. *The Acid-fast cell wall*. Community College of Baltimore Country (Cantonsville), 2022.

[14] Juan Carlos Palomino, Sylvia Cardoso Leão, and Viviana Ritacco. Tuberculosis 2007; from basic science to patient care. 2007.

[15] Ciamak Ghazaei. Mycobacterium tuberculosis and lipids: Insights into molecular mechanisms from persistence to virulence. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, 23, 2018.

[16] Johanna Raffetseder. *Interplay of human macrophages and Mycobacterium tubercu- losis phenotypes*. PhD thesis, Linköping University Electronic Press, 2016.

[17] Y Timouyas. *Etude moléculaire de la résistance à la rifampicine des bacilles du complexe Mycobacterium tuberculosis*. PhD thesis, these de medecine. casablanca, 2017.

[18] Ashwin S Dharmadhikari, Matsie Mphahlele, Kobus Venter, Anton Stoltz, Rirhandzu Mathebula, Thabiso Masotla, Martie van der Walt, Marcello Pagano, Paul Jensen, and

Edward Nardell. Rapid impact of effective treatment on transmission of multidrug-resistant tuberculosis. *The International journal of tuberculosis and lung disease*, 18(9):1019–1025, 2014.

[19] Maxine Caws, Ben Marais, Dorothee Heemskerk, and Jeremy Farrar. *Tuberculosisin adults and children*. Springer Nature, 2015.

[20] M Singh, ML Mynak, L Kumar, JL Mathew, and SK Jindal. Prevalence and risk factors for transmission of infection among children in household contact with adultshaving pulmonary tuberculosis. *Archives of disease in childhood*, 90(6):624–628, 2005.

[21] Padmanesan Narasimhan et al. Risk factor for tuberculosis. the university of new south wales, kensington, sydney, nsw 2052, australia, 2013.

[22] Charles W Schmidt. Linking tb and the environment: an overlooked mitigation strategy, 2008.

[23] Michael U Shiloh. Mechanisms of mycobacterial transmission: how does mycobacterium tuberculosis enter and escape from the human host, 2016.

[24] Gavin Churchyard, Peter Kim, N Sarita Shah, Roxana Rustomjee, Neel Gandhi, Barun Mathema, David Dowdy, Anne Kasmar, and Vicky Cardenas. What we know about tuberculosis transmission: an overview. *The Journal of infectious diseases*, 216(suppl 6):S629–S635, 2017.

[25] Maximilian F Konig, Loreto Abusleme, Jesper Reinholdt, Robert J Palmer, Ri- cardo P Teles, Kevon Sampson, Antony Rosen, Peter A Nigrovic, Jeremy Sokolove, Jon T Giles, et al. Aggregatibacter actinomycetemcomitans–induced hypercitrulli- nation links periodontal infection to autoimmunity in rheumatoid arthritis. *Science translational medicine*, 8(369):369ra176–369ra176, 2016.

[26] Nigel Chaffey. Alberts, b., johnson, a., lewis, j., raff, m., roberts, k. and walter, p.molecular biology of the cell. 4th edn., 2003.

[27] Susanta Pahari, Gurpreet Kaur, Mohammad Aqdas, Shikha Negi, Deepyan Chatter- jee, Hilal Bashir, Sanpreet Singh, and Javed N Agrewala. Bolstering immunity through pattern recognition receptors: a unique approach to control tuberculosis. *Frontiers in Immunology*, 8:906, 2017.

[28] Adane Mihret. The role of dendritic cells in mycobacterium tuberculosis infection. *Virulence*, 3(7):654–659, 2012.

[29] Gail D Sckisel, Myriam N Bouchlaka, Arta M Monjazeb, Marka Crittenden, Brendan D Curti, Danice EC Wilkins, Kory A Alderson, Can M Sungur, Erik Ames, Annie Mirsoian, et al. Out-of-sequence signal 3 paralyzes primary cd4+ t-cell-dependent immunity. *Immunity*, 43(2):240–250, 2015.

[30] S Kiazyk and TB Ball. Tuberculosis (tb): Latent tuberculosis infection: An overview. *Canada Communicable Disease Report*, 43(3-4):62, 2017.

[31] AIM Olsen, HE Andersen, J Aßmus, JA Djupvik, G Gran, K Skaug, and O Mørkve. Management of latent tuberculous infection in norway in 2009: a descriptive cross-sectional study. *Public Health Action*, 3(2):166–171, 2013.

[32] David Rees and Jill Murray. Silica, silicosis and tuberculosis. *Occupational Health Southern Africa*, 26(5):266–276, 2020.

[33] Susan S Jick, Eric S Lieberman, Mahboob U Rahman, and Hyon K Choi. Gluco- corticoid use, other associated factors, and the risk of tuberculosis. *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, 55(1):19–26, 2006.

[34] Marriott Nliwasa, Peter MacPherson, Ankur Gupta-Wright, Mphatso Mwapasa, Katherine Horton, Jon Ø Odland, Clare Flach, and Elizabeth L Corbett. High hiv and active tuberculosis prevalence and increased mortality risk in adults with symp- toms of tb: a systematic review and meta-analyses. *Journal of the International AIDS Society*, 21(7):e25162, 2018.

[35] Stephan Schwander and Keertan Dheda. Human lung immunity against mycobac- terium tuberculosis: insights into pathogenesis and protection. *American journal of respiratory and critical care medicine*, 183(6):696–707, 2011.

[36] Christoph Lange and Toru Mori. Advances in the diagnosis of tuberculosis. *Respirol- ogy*, 15(2):220–240, 2010.

[37] MRA Van Cleeff, LE Kivihya-Ndugga, H Meme, JA Odhiambo, and PR Klatser. The role and performance of chest x-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in nairobi, kenya. *BMC infectious diseases*, 5(1):1–9, 2005.

[38] Chantal Truffot-Pernot, Nicolas Veziris, and Wladimir Sougakoff. Modern diagnosis of tuberculosis. *Presse Medicale (Paris, France: 1983)*, 35(11 Pt 2):1739–1746, 2006.

[39] Vincent Jarlier and Hiroshi Nikaido. Mycobacterial cell wall: structure and role in natural resistance to antibiotics. *FEMS microbiology letters*, 123(1-2):11–18, 1994.

[40] World Health Organization (WHO) et al. Essential first-line antituberculosis drugs. *Treatment of Tuberculosis. Guidelines. 4th ed. Geneva: WHO*, pages 103–114, 2010.

[41] Sebastian M Gygli, Sonia Borrell, Andrej Trauner, and Sebastien Gagneux. An- timicrobial resistance in mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS microbiology reviews*, 41(3):354–373, 2017.

[42] M Al-Saeedi and S Al-Hajoj. Diversity and evolution of drug resistance mechanisms in mycobacterium tuberculosis. infect drug resist 10, 333–342, 2017.

[43] Matthias Merker, Thomas A Kohl, Ivan Barilar, Sönke Andres, Philip W Fowler, Erja Chryssanthou, Kristian Ängeby, Pontus Jureen, Danesh Moradigaravand, Julian Parkhill, et al. Phylogenetically informative mutations in genes implicated in antibiotic resistance in mycobacterium tuberculosis complex. *Genome Medicine*, 12(1):1–8, 2020.

[44] Mandeep Jassal and William R Bishai. Extensively drug-resistant tuberculosis. *The Lancet infectious diseases*, 9(1):19–30, 2009.

[45] J Ben Amar, B Dhahri, H Aouina, S Azzabi, MA Baccar, L El Gharbi, andH Bouacha. Tuberculosis treatment. *Revue de pneumologie clinique*, 71(2-3):122–129, 2015.

[46] Edmund G Brown, Diana Dooley, and Howard Backer. *Drug-resistant tuberculosis*. National Collaborating Centre for Infectious Diseases, 2016.

[47] Christopher A Kerantzas and William R Jacobs Jr. Origins of combination therapyfor tuberculosis: lessons for future antimicrobial development and application. *MBio*, 8(2):e01586–16, 2017.

[48] Nicolas Veziris and Jerome Robert. Anti-tuberculosis drug resistance and therapeutic dead end. *Medecine Sciences: M/S*, 26(11):976–980, 2010.

[49] World Health Organization et al. *Implementing the WHO Stop TB Strategy:a handbook for national tuberculosis control programmes*. Number WHO/HT-M/TB/2008.401. World Health Organization, 2008.

[50] Prevention Committee of the Japanese Society for Tuberculosis, Treatment Committee of the Japanese Society for Tuberculosis, et al. Treatment guidelines for latent tuberculosis infection. *Kekkaku:[Tuberculosis]*, 89(1):21–37, 2014.

[51] Mireia Coscolla and Sebastien Gagneux. Consequences of genomic diversity in mycobacterium tuberculosis. In *Seminars in immunology*, volume 26, pages 431–444. Elsevier, 2014.

[52] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome biology*, 18(1):1–15, 2017.

[53] Lamiaa LAHLOU. Analyses bio-informatique de données génomiques et transcriptomiques dans deux pathologies: Tuberculose et maladie de parkinson. 2019.

[54] Heng Li and Richard Durbin. Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, 26(5):589–595, 2010.

[55] Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, et al. Twelve years of samtools and bcftools. *Gigascience*, 10(2):giab008, 2021.

[56] Geraldine A Van der Auwera and Brian D O'Connor. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media, 2020.

[57] Chou-Han Lin, Chou-Jui Lin, Yao-Wen Kuo, Jann-Yuan Wang, Chia-Lin Hsu, Jong- Min Chen, Wern-Cherng Cheng, and Li-Na Lee. Tuberculosis mortality: patient characteristics and causes. *BMC infectious diseases*, 14(1):1–8, 2014.

[58] Mario Raviglione and Giorgia Sulis. Tuberculosis 2015: burden, challenges and strategy for control and elimination. *Infectious disease reports*, 8(2):6570, 2016.

[59] Christopher Dye, Suzanne Scheele, Vikram Pathania, Mario C Raviglione, et al. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. *Jama*, 282(7):677–686, 1999.

[60] Madikay Senghore, Bassirou Diarra, Florian Gehre, Jacob Otu, Archibald Worwui, Abdul Khalie Muhammad, Brenda Kwambana-Adams, Gemma L Kay, Moumine Sanogo, Bocar Baya, et al. Evolution of mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in bamako, mali. *Scientific Reports*, 10(1):1–9, 2020.

[61] Michael B Reed, Victoria K Pichler, Fiona McIntosh, Alicia Mattia, Ashley Fallow, Speranza Masala, Pilar Domenech, Alice Zwerling, Louise Thibert, Dick Menzies, et al. Major mycobacterium tuberculosis lineages associate with patient country of origin. *Journal of clinical microbiology*, 47(4):1119–1128, 2009.

[62] Haixia Chen, Li He, Hairong Huang, Chengmin Shi, Xumin Ni, Guangming Dai, Liang Ma, and Weimin Li. Mycobacterium tuberculosis lineage distribution in xinjiangand gansu provinces, china. *Scientific reports*, 7(1):1–7, 2017.

[63] B Lopez, D Aguilar, H Orozco, M Burger, C Espitia, V Ritacco, L Barrera, K Kre- mer, R Hernandez-Pando, K Huygen, et al. A marked difference in pathogenesis and immune response induced by different mycobacterium tuberculosis genotypes. *Clinical & Experimental Immunology*, 133(1):30–37, 2003.

[64] Neel R Gandhi, Paul Nunn, Keertan Dheda, H Simon Schaaf, Matteo Zignol, Dick Van Soolingen, Paul Jensen, and Jaime Bayona. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet*, 375(9728):1830–1843, 2010.

[65] Luke T Daum, John D Rodriguez, Sue A Worthy, Nazir A Ismail, Shaheed V Omar, Andries W Dreyer, P Bernard Fourie, Anwar A Hoosen, James P Chambers, and Gerald W Fischer. Next-generation ion torrent sequencing of drug resistance mutations in mycobacterium tuberculosis strains. *Journal of clinical microbiology*, 50(12):3831–3837, 2012.

[66] Namaunga Kasumu Chisompola, Elizabeth Maria Streicher, Chishala Miriam Kapambwe Muchemwa, Robin Mark Warren, and Samantha Leigh Sampson. Molecular

epidemiology of drug resistant mycobacterium tuberculosis in africa: a systematic review. *BMC Infectious Diseases*, 20(1):1–16, 2020.

[67] Silke Feuerriegel, Claudio U Köser, and Stefan Niemann. Phylogenetic polymorphisms in antibiotic resistance genes of the mycobacterium tuberculosis complex. *Journal of Antimicrobial Chemotherapy*, 69(5):1205–1210, 2014.

[68] TC Victor, A Mi Jordaan, A Van Rie, GD Van Der Spuy, M Richardson,PD Van Helden, and R Warren. Detection of mutations in drug resistance genes of mycobacterium tuberculosis by a dot-blot hybridization strategy. *Tubercle and lung disease*, 79(6):343–348, 1999.