UNIVERSITÄT ZU LÜBECK

From the Research Center Borstel

Leibniz Lung Center

Priority Research Area Infections

Molecular and Experimental Mycobacteriology

Director: Prof. Dr. Ulrich E. Schaible

Impact of *Mycobacterium tuberculosis* complex Strain Diversity on Pathobiology

Dissertation for the fulfillment of the requirements for
the Doctoral Degree
of the University of Lübeck

From the Department of Natural Sciences

Submitted by

Francy Johanna Pérez Llanos M. Sc.

From Bogotá, Colombia

Lübeck 2023

First referee: Prof. Dr. Stefan Niemann

Second referee: Prof. Dr. med Jan Rupp

Date of oral examination: 15.03.2024

Approved for printing. Lübeck, 02.04.2024

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I am overwhelmed with gratitude and inspiration after completing my doctoral thesis. Working alongside Prof. Dr. Stefan Niemann has been an exceptional experience, enriching my life in countless ways and allowing me to make a meaningful contribution to society through my scientific research. My work, which employs a genomic epidemiology approach and provides new insights into host-pathogen interactions, will help the public health authorities and this field of science better control tuberculosis, the deadliest ancestral infectious disease in human history. Prof. Dr. Stefan Niemann's unwavering commitment and expertise in this field of science have been a constant source of motivation for me to strive for genuine fulfillment, no matter the circumstances.

I want to express my sincere gratitude to Dr. Susanne Homolka and Prof. Dr. Matthias Merker for their support, guidance, and encouragement during my research journey. Their mentorship has been invaluable, and I am grateful for the countless hours dedicated to helping me develop my ideas and refine my work. I would also like to extend my appreciation to the members of the Molecular and Experimental Mycobacteriology group, especially Dr. Yassir Adam Shuaib, Dr. Christiane Gerlach, Dr. Viola Dreyer, Dr. Patrick Beckert, Silvia Maaß, Doreen Beyer, Vanessa Mohr, Franziska Daduna and all anonymous blood donors whose contribution and constructive support have been instrumental in shaping the direction and quality of my research. My work also benefited from the interdisciplinary input from Dr. Bianca Schneider from the coinfection division, Dr. Thomas Scholzen, Dr. Jochen Behrends, Martina Hein from the flow cytometry core, Dr. Julius Brandenburg from the microbial interface biology division at the Research Center Borstel, Dr. Maha R Farhat from the department of biomedical informatics at Harvard Medical School and Dr. Matthias Gröschel from the infectious diseases and respiratory medicine division at the Charité Universitätsmedizin, the international collaborators from Colombia, United States of America, The Netherlands and Germany.

I would also like to express my sincere gratitude to my fellow doctoral colleagues and the “Borstel Ausländerbande” at the time for their kindness throughout my doctoral journey. I would also like to thank my friends in Hamburg, Lübeck, Berlin, Düsseldorf, and Köln for supporting me during the challenging times. Furthermore, I extend my heartfelt appreciation to Germany and its funding parties, including Deutsche Forschungsgemeinschaft, the Leibniz Gemeinschaft, the landeregiwrung Schleswig-Holstein, the Research Center Borstel and the Universität zu Lübeck, for their support. Their contributions, particularly through the Leibniz Science Campus EvoLUNG, have been critical to the success of my research. I appreciate their resources, warm welcome, and support provided to me as an international scientist to advance knowledge and innovation in this field of science.

My heart is filled with gratitude for my family's unwavering love and support, my lovely parents, Nivio Jesus Perez Lozano and Bertha Mercedes Llanos Niño, and the something out there who has inspired me to follow my passions with unwavering determination. Finally, I owe it to myself to continue pushing forward, traveling far and wide with the courage to explore the unknown.

DEDICATION

"*We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about science and technology. Therefore, popularizing science trying to make its methods and findings accessible to non scientists then follows naturally and immediately.*"

C A R L S A G A N

a beautiful mind (9.11.1934 in Brooklyn, New York City; †20.12.1996 in Seattle, Washington)
The Demond haunted world: Science as a candle in the dark

"We are people who value science: scientists, educators, journalists, students, neighbors, friends, and family. We come from all races, all religions, all gender identities, all sexual orientations, all abilities, all socioeconomic backgrounds, all political perspectives, and all nationalities. Our diversity is our greatest strength: a wealth of opinions, perspectives and ideas are critical for a scientific process. What unites us is the love of science and an insatiable curiosity. We all recognize that science is everywhere and affects everyone.

Science is often an arduous process, but it is also thrilling. A universal human curiosity and dogged persistence are the greatest hopes for the future. This movement cannot and will not end with a march. Our plans for policy change and community outreach will start with marches worldwide, but it is imperative that we continue to celebrate and defend science at all levels from local schools to federal agencies throughout the world."

M A R C H F O R S C I E N C E

a meaningful purpose (04.22.2017 in Washington-Present)

LIST OF ABBREVIATIONS

AA	Alternative Allele
AIDS	Acquired Immunodeficiency Syndrome
ANOVA	Analysis of Variance
AR	Amazonas River
ATCC	American Type Culture Collection
A4	Animal Clade 4
BCG	Bacillus Calmette-Guérin
bMDMs	Bovine Monocyte Derived Macrophages
BSL3	Biosafety Level 3
°C	Degrees Celsius
CCE	Counterflow Centrifugal Elutriation
CFU	Colony Forming Unit
COL	Colombia
COVID-19	Coronavirus Disease 2019
CRAN	Comprehensive R Archive Network
CSS	Closed Source Software
diH_2O	Deionized Water
dH_2O	Distilled Water
DNA	Deoxyribonucleic acid
DS-TB	Drug Susceptible Tuberculosis
ENA	European Nucleotide Archive
EvoLUNG	Evolutionary Medicine of the Lung
FCM	Flow Cytometry
FSC	Forward Scatter
HBSS	Hanks Balanced Salt Solution
HIV	Human Immunodeficiency Virus
hMDMs	Human Monocyte Derived Macrophages
hpi	Hours post infection
IL	Interleukin
LRS	Long Read Sequencing
LAM	Latin American Mediterranean
L4	Lineage 4 (Euro American)
L6	Lineage 6 (West African II)
Maf	*Mycobacterium africanum*
Mbv	*Mycobacterium bovis*
MCM	Macrophage Culture Media
MCP-1	Monocyte Chemoattractant Protein - 1
MDR-TB	Multi Drug Resistance Tuberculosis
MDR/RR-TB	Multidrug resistant /Rifampicin Resistant TB
MGIT	Mycobacteria Growth Indicator Tube

min	Minute
MIM	Macrophage Infection Media
MIRU-VNTR	Mycobacterial Interspersed Repetitive Unit- Variable Number of Tandem Repeats
mL	Milliliter
MLM	Mycobacteria liquid Media
MOI	Multiplicity of infection
Mtbc	*Mycobacterium tuberculosis* complex
Mϕ	Macrophage
ng	nanogram
NGS	Next Generation Sequencing
NRC	National Reference Center for Mycobacteria
OADC	Oleic Acid Albumin Dextrose Catalase
OD	Optical Density
OSS	Open Source Software
PAHO	Pan American Health Organization
PBMC	Peripheral Blood Mononuclear Cell
pg	Picogram
R	Reference Allele
RCB	Research Center Borstel
RD	Region of Difference
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
rpm	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic Acid
RT	Room Temperature
SEM	Standard Error of the Mean
SNP	Single Nucleotide Polymorphism
SMRT	Single Molecule Real Time
SRS	Short Read Sequencing
TB	Tuberculosis
TbD1	Tuberculosis deletion 1
µL	Microliter
WA	West Africa
WGS	Whole Genome Sequencing
WHO	Word Health Organization

1. ABSTRACT

Tuberculosis (TB) has been a significant public health issue for thousands of years, affecting every region of the world. Until the COVID-19 pandemic, it was the top infectious disease resulting in human fatalities, now ranking second. TB is caused by bacterial pathogens that belong to the *Mycobacterium tuberculosis* complex (*Mtbc*), which are 99.9% similar at the nucleotide level. However, these pathogens differ in phenotype, pathogenicity, and host tropism. The *Mtbc* strains are classified into nine lineages, adapted to humans or animals, and grouped into modern and ancient clades based on genomic diversity. Their genomic diversity has exposed a complex evolutionary history of the *Mtbc* lineages, with distinct geographical distributions as a result of forming speciation allopatric and sympatric relationships while coevolving with humans since their migration out of Africa. Modern Lineage 4 (L4) strains are the most widespread and responsible for TB endemic in Europe, the Americas, and certain parts of Africa. On the other hand, the ancient Lineage 6 (L6) strains remain restricted to West Africa.

Understanding the interaction between different strains of *Mtbc* with varying virulence and hosts with different disease susceptibility is crucially important. However, this relationship is not yet fully comprehended. This research aimed to help bridge this knowledge gap. Firstly, by conducting a genomic investigation of a TB outbreak that affected an indigenous population in a remote region of the Amazon in Colombia, Latin America. Secondly, by carrying out *in vitro* research to understand how *Mtbc* strain diversity impacts pathobiology in human host cells, particularly the causative lineage of the TB outbreak.

The genomic investigation of the TB outbreak was based on whole genome sequencing (WGS), which has become the main public health surveillance tool for studying the transmission dynamics of *Mtbc* strains. However, the clonal expansion of one strain often limits its application in local outbreaks of *Mtbc* straing Using an alternative reference genome and the includin repetitive regions in the analysis could potentially increase the resolution, but the added value has not yet been defined. First, this work leveraged short and long read WGS data of a previously reported *Mtbc* outbreak in the Colombian Amazon Region to analyze possible transmission chains among 74 patients in the indigenous setting of Puerto Nariño (March to October 2016). In total, 90.5% (67/74) of the patients were infected with one distinct *Mtbc* strain belonging to lineage 4.3.3. Employing a reference genome from an outbreak strain and highly confident single nucleotide polymorphisms (SNPs) in repetitive genomic regions, e.g., the proline-glutamic acid/proline-proline-glutamic-acid (PE/PPE) gene family, increased the phylogenetic resolution compared to a classical H37Rv reference mapping approach. Specifically, the number of differentiating SNPs increased from 890 to 1,094, which resulted in a more granular transmission network as judged by an increasing number of individual nodes in a maximum parsimony tree, i.e., 5

versus 9 nodes. It was also found 29.9% (20/67) of the outbreak isolates, heterogenous alleles at phylogenetically informative sites, suggesting that these patients are infected with more than one clone.

Furthermore, this study sought biological evidence of the success of the Lineage 4 strains that caused the outbreak. To achieve this, a diverse panel of *Mtbc* clinical and reference strains were tested in an *in vitro* infection cell model, using blood human monocyte derived macrophages (hMDMs) from individuals of European and African ancestries. The panel of *Mtbc* strains included the human-adapted strains of L4 and L6, which are widespread-modern and restricted-ancient respectively, as well as the animal-adapted strains of A4 (*M. bovis*).

The study showed that these lineages had specific growth patterns, transcriptomic profiles, phagocytosis rates, and intracellular growth performances, indicating that they had a varied impact on the pathobiology of the disease while inducing a consistent inflammatory immune response in hMDMs. The results revealed that *Mtbc* strains of L4 are more virulent than L6 and A4 strains. L4 strains grew faster and reached a higher number of *Mtbc* cells, colony forming units compared to the other strains. L4 strains also displayed transcriptomic signatures associated with virulence, adaptation, cell wall and cell processes/PPE, intermediary metabolism, and conserved hypothetical proteins. L4 strains were able to adapt to hMDMs and cause pathobiology irrespective of the human ancestries considered in the study. L6 strains showed moderate virulent behavior and limited pathobiology in allopatric hMDMs cells. However, when infecting hMDMs cells from sympatric ancestral hosts, L6 strains displayed higher intracellular growth rates and IL-8 inflammatory response. A4 strains showed superior virulence in bovine host cells.

In conclusion, this study has demonstrated that customized SNP calling thresholds and a mapping approach based on local reference genome can significantly improve the phylogenetic resolution in highly clonal *Mtbc* populations and helps to elucidate the within-host *Mtbc* diversity. Moreover, the research has yielded valuable insights into the diversity of *Mtbc* strains and their relationships with sympatric and allopatric hosts, crucial factors in the pathobiology of TB disease. These findings will enable public health authorities to better investigate clonal TB outbreaks. Additionally, the *in vitro* characteristics of *Mtbc* strain diversity in these host-pathogen relationships may reflect different strategies of *Mtbc* strains to exploit ecological niches. These strategies may influence the disease outcome in different human populations. These findings can help improve treatments, vaccines, and control strategies to combat TB.

2. ZUSAMMENFASSUNG

Die Tuberkulose (TB) stellt das öffentliche Gesundheitswesen in aller Welt bereits seit Jahrtausenden vor große Herausforderungen. Die TB ist derzeit die zweittödlichste Infektionskrankheit beim Menschen nach COVID-19. Die Erkrankung wird durch bakterielle Erreger verursacht, die zum *Mycobacterium tuberculosis* complex (*Mtbc*) gehören. Auf Nukleotidebene ähneln sich diese Erreger zu 99,9 %, unterscheiden sich jedoch im Phänotyp, in der Pathogenität und im Wirtstropismus. Es werden neun *Mtbc*-Stämme unterschieden, die sich entweder an Menschen oder Tiere angepasst haben. Zudem werden die *Mtbc*-Stämme in moderne und ancestrale Kladen eingeteilt. Die genomische Vielfalt der *Mtbc*-Stämme ist Spiegel ihrer komplexen Evolutionsgeschichte. Als Menschen den afrikanischen Kontinent verließen, kam es zu einer Koevolution der *Mtbc*-Stämme mit dem Menschen. Die *Mtbc*-Stämme bildeten allopatrische und sympatrische Speziesbeziehungen aus, die zu einer unterschiedlichen geographischen Verbreitung der *Mtbc*-Stämme führten. Derzeit sind die modernen Stämme der Linie 4 (L4) am weitesten verbreitet und für die in Europa, Amerika und bestimmten Teilen Afrikas endemische TB verantwortlich. Die Verbreitung der alten Stämme der Linie 6 (L6) beschränkt sich hingegen nach wie vor auf Westafrika.

Bislang kaum erforscht ist die Interaktion zwischen unterschiedlich virulenten *Mtbc*-Stämmen einerseits und Wirten mit unterschiedlicher Krankheitsanfälligkeit andererseits. Ziel dieser Arbeit war es, diese Wissenslücke zu schließen. Dazu wurde ein Tuberkuloseausbruch, der sich innerhalb einer indigenen Bevölkerung in einer abgelegenen Region des Amazonasgebiets in Kolumbien ereignete, zunächst genomisch untersucht. Anschließend wurde *in-vitro* untersucht, wie sich die Vielfalt der *Mtbc*-Stämme auf die Pathobiologie in menschlichen Wirtszellen auswirkt, wobei jene Linie besonders in den Blick genommen wurde, die für den Tuberkuloseausbruch ursächlich war.

Die genomische Untersuchung des Tuberkuloseausbruchs erfolgte mittels Ganzgenomsequenzierung (WGS, Whole Genome Sequencing). Diese hat sich zum wichtigsten Instrument zur Untersuchung der Übertragungsdynamik von *Mtbc*-Stämmen entwickelt, über die das öffentliche Gesundheitswesen verfügt. Die klonale Ausbreitung eines *Mtbc*-Stammes schränkt jedoch bei lokalen Ausbrüchen die Anwendung der Ganzgenomsequenzierung ein. Die Verwendung eines alternativen Referenzgenoms und die Einbeziehung repetitiver Regionen in die Analyse könnten die Auflösung potenziell erhöhen. In dieser Arbeit wurden zunächst Short- und Long-Read-WGS-Daten eines zuvor gemeldeten *Mtbc*-Ausbruchs im kolumbianischen Amazonasgebiet genutzt, um mögliche Übertragungsketten bei 74 Patienten in der indigenen Umgebung von Puerto Nariño im Zeitraum von März bis Oktober 2016 zu analysieren. Insgesamt waren 90,5 % (67/74) der Patienten mit einem bestimmten *Mtbc*-Stamm infiziert, der zur Linie

4.3.3 gehört. Die Verwendung eines Referenzgenoms aus einem Ausbruchsstamm und hochsichere Einzelnukleotid-Polymorphismen (SNPs, Single Nucleotide Polymorphisms) in repetitiven Genomregionen, z. B. der Prolin-Glutaminsäure/Prolin-Prolin-Glutaminsäure (PE/PPE)-Genfamilie, erhöhte die phylogenetische Auflösung im Vergleich zu einem klassischen H37Rv-Referenzkartierungsansatz. Insbesondere stieg die Zahl der differenzierenden SNPs von 890 auf 1.094, was zu einem granulareren Übertragungsnetzwerk führte. Dies zeigte sich durch eine steigende Zahl einzelner Knoten in einem Maximum-Parsimony-Baum, nämlich 5 gegenüber 9 Knoten. Außerdem wurden bei 29,9 % (20/67) der Ausbruchsisolate heterogene Allele an phylogenetisch informativen Stellen gefunden, was darauf hindeutet, dass diese Patienten mit mehr als einem Klon infiziert waren. Darüber hinaus wurde in dieser Studie nach biologischen Beweisen für den Erfolg der Stämme der Linie 4, die den Ausbruch verursacht hatten, gesucht. Zu diesem Zweck wurden eine Reihe von klinischen *Mtbc*-Stämmen und Referenzstämmen in einem *In-vitro*-Infektionszellmodell getestet. Dabei wurden aus menschlichen Monozyten gewonnene Blutmakrophagen (hMDMs, Human Monocyte Derived Macrophages) von Personen europäischer und afrikanischer Abstammung verwendet. Die Gruppe von *Mtbc*-Stämmen umfasste die an den Menschen angepassten Stämme L4 und L6 sowie den an Tiere angepassten Stamm A4 (*M. bovis*).

Die Untersuchung zeigte, dass diese Linien spezifische Wachstumsmuster, transkriptomische Profile, Phagozytoseraten und intrazelluläre Wachstumsleistungen aufwiesen. Dies deutete darauf hin, dass diese Linien einen unterschiedlichen Einfluss auf die Pathobiologie der Krankheit haben und gleichzeitig eine einheitliche entzündliche Immunreaktion in hMDMs hervorrufen. Die Ergebnisse zeigten, dass *Mtbc*-Stämme von L4 virulenter sind als L6- und A4-Stämme. L4-Stämme wuchsen schneller und erreichten im Vergleich zu den anderen Stämmen eine höhere Anzahl von *Mtbc*-Zellen und koloniebildenden Einheiten. L4-Stämme wiesen auch transkriptomische Signaturen auf, die mit Virulenz, Anpassung, Zellwand- und Zellprozessen/PPE, intermediärem Stoffwechsel und konservierten hypothetischen Proteinen in Zusammenhang stehen. L4-Stämme waren in der Lage, sich an hMDMs-Zellen anzupassen und Pathobiologie zu verursachen, unabhängig von der Abstammung den in der Studie berücksichtigten Personen. L6-Stämme zeigten mäßig virulentes Verhalten und begrenzte Pathobiologie in allopatrischen hMDMs-Zellen. Bei der Infektion von hMDMs-Zellen aus sympatrischen Vorfahren zeigten die L6-Stämme jedoch höhere intrazelluläre Wachstumsraten und eine IL-8-Entzündungsreaktion. A4-Stämme zeigten eine höhere Virulenz in bovinen Wirtszellen.

Zusammenfassend zeigt diese Arbeit, dass maßgeschneiderte SNP-Calling-Schwellenwerte und ein auf dem lokalen Referenzgenom basierender Kartierungsansatz die phylogenetische Auflösung in stark klonalen *Mtbc*-Populationen erheblich verbessern können und dazu beitragen, die *Mtbc*-Diversität

innerhalb des Wirts zu untersuchen. Darüber hinaus hat die Arbeit wertvolle Erkenntnisse zur Vielfalt der *Mtbc*-Stämme und ihrer Beziehungen zu sympatrischen und allopatrischen Wirten erbracht, die für die Pathobiologie der TB von entscheidender Bedeutung sind. Diese Erkenntnisse ermöglichen es den öffentlichen Gesundheitsbehörden, klonale Tuberkuloseausbrüche besser zu untersuchen. Darüber hinaus dürften die *In-vitro*-Merkmale der veschiedenen *Mtbc*-Stämme in den Wirt-Pathogen-Beziehungen Ausdruck unterschiedlicher Strategien der *Mtbc*-Stämme zur Nutzung ökologischer Nischen sein. Dies dürfte den Krankheitsverlauf in verschiedenen menschlichen Populationen beeinflussen. Diese Ergebnisse können dazu beitragen, Behandlungen, Impfstoffe und Kontrollstrategien zur Bekämpfung der TB-Krankheit zu verbessern.

3. INTRODUCTION

3.1 An Ancient Disease: Tuberculosis and its Global Epidemiology

Tuberculosis (TB) is an ancient infectious disease that has affected humanity for over 40000-75000 years [1]. Throughout history, it has been referred to by various names, including “phthisis or consumption” by the Hippocratic Medicine during 410-400 before Christ, “the great white plague,” and “The Captain of all these Men of Death” during the 18th and 19th centuries [2]. TB has been a medical challenge over the millennia [3] and has killed one billion people over the last bicentennial [4]. Until the pandemic of coronavirus disease 2019 (COVID-19), TB was the first leading cause of mortality caused by a single contagious agent globally and the second in 2022[5], [6]. The contagious agent of TB is *Mycobacterium tuberculosis* (*Mtb*), which Robert Koch discovered in 1882 [7]. Today, the causative agents of TB are defined as members of the so-called *Mycobacterium tuberculosis* complex (*Mtbc*).

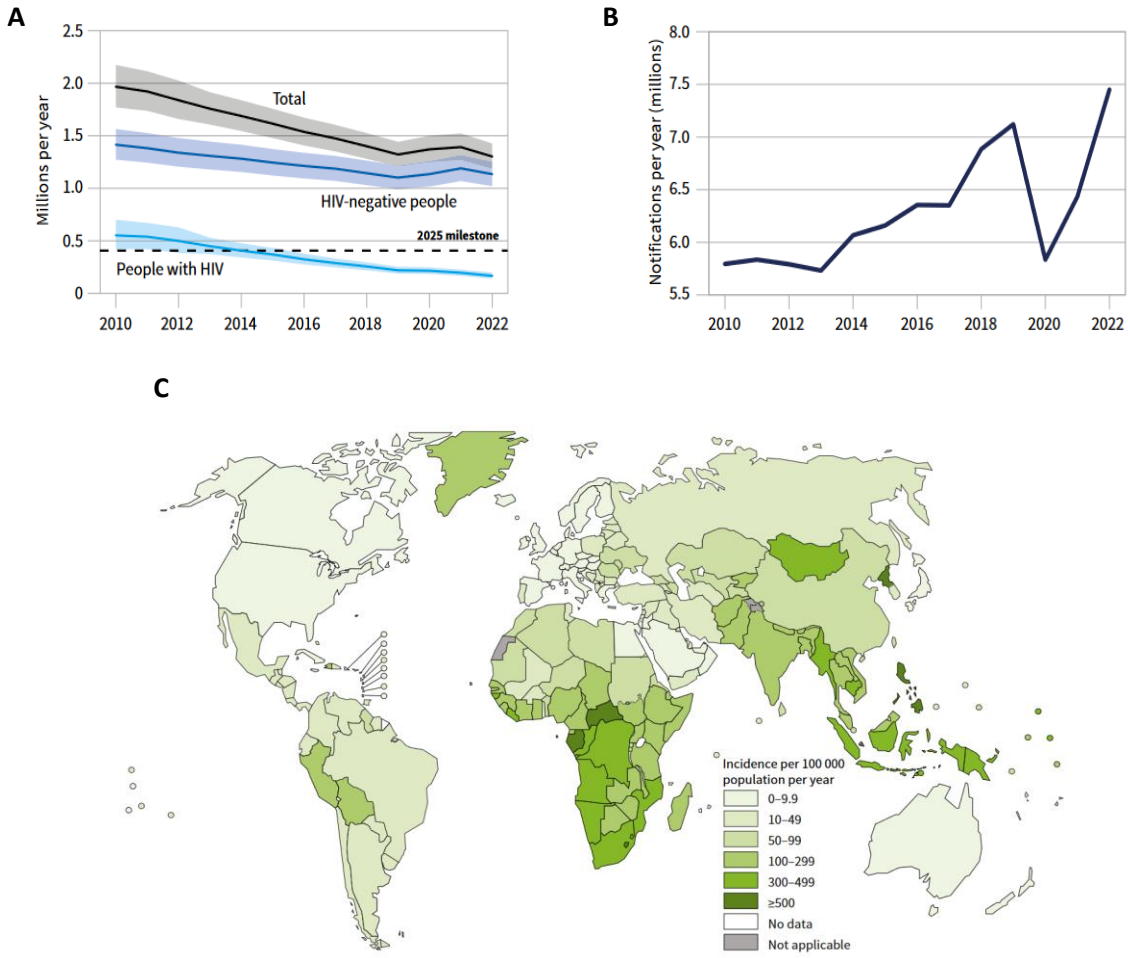
TB disease spreads through the air when an infected person coughs, releasing *Mtbc* bacteria into small droplets that can be inhaled by others who become infected [8]. The disease primarily affects the lungs of both humans and animals, but it can also spread to organs through the bloodstream [9]. It is estimated that only 5-10% of infected individuals will develop active TB disease, with most cases remaining in an incubation period of 3-24 months until symptoms arise [10], [11]. These symptoms include intense coughing with sputum and occasionally blood, extreme tiredness, fever, and night sweats. Those individuals with underlying medical conditions, such as Human Immunodeficiency Virus (HIV) infection, diabetes, cancer, and undernutrition, are at a higher risk of mortality [12]. In 90-95% of the infected individuals, the TB bacillus can persist in a state known as Latent Tuberculosis Infection (LTBI), in which the pathogen replication is contained and the inflammatory repose subsided, which complicates the determination of those who have cleared the bacillus, those who are still latently infected or those who progress to active disease[13]. LTBI poses a critical challenge to TB elimination efforts, as a quarter of the world´s population is estimated to be infected with LTBI [10], creating a bomb for a potential pandemic.

According to the World Health Organization (WHO)'s 2022 report, the number of TB cases has leveled off at 10.6 million, with 1.3 million deaths. This is a return to the 2019 report, after a period of increasing mortality rates that saw 1.4 million deaths between 2010 and 2021 (as depicted in Figure 1A)[6], [10], [14]. These striking indicators accounted for the COVID-19 pandemic and might rise significantly over the coming years [15]. Access to TB diagnosis and treatment was reduced in 2020, leading to a significant decline in TB case notifications globally; however, it has been partially increasing in 2022 (Figure 1B).

In terms of TB burden, thirty countries with high incidence rates accounted for a staggering 87% of global cases in 2022[14]. These countries are located in various WHO regions, including South-East Asia (with India, Indonesia, and Bangladesh contributing 27%, 10%, and 3.6% respectively), West Pacific (with China and the Philippines contributing 7.1% and 3.6% respectively), Western Pacific (with Pakistan contributing 5.7%), and Africa (with Nigeria and the Democratic Republic of the Congo contributing 4.5% and 3% respectively). Notably, there has been a concerning increase in TB cases in the Americas since 2016 (as shown in Figure 1C). Regrettably, the COVID-19 pandemic has had an adverse effect on the advancements made in TB control in previous years. As a result, the WHO has declared that the benchmarks and objectives established to eradicate the TB endemic by 2035 are improbable to be achieved [14], [16]. Nonetheless, the leaders of nations and governments recently committed at the 2023 United Nations high-level meeting to progress science, finance, and innovation to terminate the TB global endemic[17].

Figure 1: Global tuberculosis incidence in 2022.
A Number of deaths were caused by tuberculosis (TB) between 2010-2022. Shaded areas along the lines indicate 95% confidence intervals. **B** Notifications of TB cases between 2010-2022. **C** Estimated incidence of TB cases by country in 2023. Incidence thresholds are depicted per country on the world map based on a code-colored green spectrum, the brightest being 0-9.9 and the darkest being ≥ 500 per 100,000 population/ year. Gray and white color indicates not applicable and no data, respectively. HIV, Human Immunodeficiency Virus. Figures were obtained and adapted from the most recent year's global tuberculosis report[14].

Figure 1 (Continued)



3.1.1 Tuberculosis Burden in the Indigenous Peoples: A Public Health Threat in Latin America and Colombia.

Latin America (LATAM) accounted for 2.9% of the TB cases reported globally in 2021 [6], [18]. At the same time, Colombia, with 37.3 cases per 100000 population, ranks fifth on the list of high TB burden countries in the Americas in 2020 [18]. Although drug resistance in LATAM affects mainly Peru (38%), Brazil (23%), Mexico (7%), Ecuador (7%), Dominican Republic, and Colombia (4%) [18], spillover effects to other countries are likely within the current antibiotic crisis [19], [20]. These estimations represent an increasing TB burden in LATAM due to recent socio-political and economic crises, which have influenced the rise of TB disease in vulnerable populations, particularly in the indigenous populations [18], [21], [22], [23]. TB detection in indigenous people remains underrepresented and understudied globally. Compelling evidence suggests that the TB burden is unknown in the estimated 370 million indigenous people worldwide, of which 41.81 million are settled in LATAM [24]. Indigenous people tend to be found

in the most remote settings in every continent of the world, for example, in deserts, the most extended mountain chains, and the Amazon region, among others; however, in the last years, the indigenous people are increasingly migrating or being displaced to the largest cities [25].

In recent years, TB incidence in the WHO/The Pan American Health Organization (PAHO) Region of the Americas has been on the rise [10], especially in the indigenous population of LATAM, with nine times more burden than in the general population with 11,608 cases reported per 100,000 indigenous people in 2018 of which 5.4% were from Colombia, particularly in communities of the Amazon [26]. These communities are still living at the margin of society and are at the most risk of developing TB disease given their isolated location, health, sociological, and economic inequalities (i.e., overcrowding, immunodeficiency, high burden of infectious diseases, malnutrition, tobacco, substance, and alcohol abuse, poor access to resources and their cultural believes) that have hindered timely access to medical services [25], [27]. In addition, the ongoing COVID-19 pandemic has worsened their situation, threatening public health efforts to end TB by 2035 [10], [26], [28], [29].

3.2 Tuberculosis Diagnostics: Gold Standards and New Technologies

For more than a century, the determination of TB has been based on clinical examination, medical records, chest X-rays, and classical bacteriological testing, known as the gold standard[30]. Unfortunately, these remain the primary diagnostic tools in many countries and remote settings with limited resources. Just 63% of the new TB cases reported globally in 2021 were diagnosed with bacteriological testing, which comprises the Ziehl-Neelsen (ZN) smear microscopy and bacteria culturing in liquid media [10], [30].

Bacteriological testing falls short in sensitivity and faster turnaround times; for example, ZN smear microscopy cannot distinguish species, dead and live bacteria and has a limited detection in paucibacillary samples (under 5000 bacilli per mL), which is critical for individuals with extrapulmonary TB, e.g., coinfected with HIV and those infected with nontuberculous mycobacteria (NTM) [31], [32]. Even though mycobacterial culture increases the sensitivity significantly by detecting a minimum of 10 bacilli per mL, the time consumption of 6-8 weeks is a significant constraint for rapid and successful treatment [33].

From the last decade forward, the WHO has been emphasizing the usage of rapid microbiology (i.e., Liquid culture BACTEC ™ MGIT ™ system from Becton, Dickinson and Company) and molecular biology (i.e., Genotype ® MTBDRplus from Hain Life Science, Xpert ® MTB/RIF from Cepheid, and Targeted Next-generation sequencing [NGS]) testing in correlation with the clinical examination to accelerate TB

diagnostics [34]. The BACTEC ™ MGIT ™ system is based on the detection of O_2 broken down in the liquid culture media of the mycobacteria growth indicator tube (MGIT) and reduces the culture time up to 10 days, serving as a reference for faster drug susceptibility testing (DST)[35].

Standard molecular tests are based on the amplification of genetic targets specific for *Mtbc* strains, while targeted-NGS further performs deep sequencing at the nucleotide level on these amplified targets and, therefore, resolve major and minor variants with high precision (i.e., those present in mixed populations) Targeted[36], [37]. Thus, molecular testing enables in a matter of hours not only the detection of the TB bacillus and species but also the screening of resistances against essential drugs, e.g., Rifampicin [RIF], Isoniazid [INH] and Fluoroquinolones [FQs] directly in clinical specimens and cultures[30], [34]. Furthermore, extensive research has proved that molecular testing is more accurate and specific and, therefore, imperative to shorter diagnosis turnaround times and the selection of proper treatments[33]. Although these tools are, for the most part, affordable and with high confidence, they require special infrastructures, equipment, and technical expertise, which can still be a limitation in low-income settings[30], [38], [39], [40], [41]. The diagnosis journey does not stop merely in the identification of the TB bacillus in single patients/ samples but also escalates towards the determination of circulating genotypes and drug resistance prediction of TB outbreaks affecting an entire population[42], [43].

3.3 Genomic Epidemiology of Tuberculosis: Solving the Transmission Puzzle.

In order to improve the effectiveness of public health operations aimed at controlling and eliminating tuberculosis (TB), it is crucial to gain a comprehensive understanding of the complex dynamics of TB transmission both locally and globally. Genomic epidemiology provides a powerful tool for achieving this by combining patient clinical data, contact tracing, and analysis of the DNA fingerprints of the TB bacillus[42], [44], [45]. This approach enables molecular surveillance to be carried out, which can help to identify outbreaks, transmission chains, and geographical and host population links, as well as facilitate investigation into pathogen evolution and its interaction with the host[42], [44].

Genomic epidemiology is, therefore, an essential public health tool in controlling and preventing the spread of TB, especially in vulnerable populations[44]. This is particularly important for dangerous strains like multidrug resistance (MDR-TB) and extensively drug resistance strains (XDR-TB), which pose a significant threat. MDR-TB is caused by *Mtbc* strains that are resistant to at least isoniazid (INH) and rifampicin (RMP), while XDR-TB is caused by *Mtbc* strains that are resistant to not only INH and RMP but also any fluoroquinolone (FQ) and at least one additional Group A drug such as levofloxacin (LFX), moxifloxacin (MFX), bedaquiline (BDQ), and linezolid (LZD)[46].

Genomic epidemiology falls into two main categories: Classical and whole genome sequencing (WGS) based methods. These methods can either target a specific marker or the entire bacterial genome[44]. Classic methods have been used for several decades and include IS*6110* Restriction Fragment Length Polymorphism (RFLP), Spacer Oligonucleotide Typing (Spoligotyping), and Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR) typing[47]. MIRU-VNTR Typing targets the number of DNA tandem repeats in the 15 or 24 loci of the multiple *Mtbc* genome locus. By amplifying up to 24 loci with primers and comparing the lengths of these amplicons to a standard allele table, the size of the repetitive regions can be translated to a 15 or 24 number code for lineage discrimination and clonal stability assessment (source: www.miru-vntrplus.org) [48].

Although classical methods have been used extensively and given enormous insights into the international landscape evolution and transmission of *Mtbc* strains, these suffer from significant limitations, such as low discriminatory power, labor-intensive processes, errors associated with sample inversion, gel misreading and allele misinterpretation that can reduce the reproducibility up to 55% [39]. However, In the past decade, these limitations have been significantly addressed and overcome with the development of WGS approaches based on NGS technology[49]. WGS analyses virtually the entire pathogen genome with an astonishing resolution and speed, which, by comparison with known sequences controlled by global *Mtbc* strains databases, enables in short-term turnarounds times reliable strain genotyping, drug resistance prediction and identification of transmission chains[42], [50], [51], [52]. Moreover, it allows the study of microevolution dynamics in host-pathogen interactions, which play a key role in precise genome-based personalized medicine in healthcare (i.e., disease severity, treatment success, and vaccine development) [42], [53], [54].

The WHO has endorsed WGS as a tool for the Public Health investigation of TB [12], [13]. While TB genomic surveillance is already implemented successfully in developed countries such as the United Kingdom and the United States [55], [56], it is poorly integrated into the public health practices of low-income countries due to high initial investments, equipment restrictions, and lack of technical and bioinformatics expertise [14]. However, with the massively dropping sequencing costs, the release of more accurate, affordable, and portable NGS platforms, and its demonstrated control of local and global pandemics (e.g., COVID-19), WGS is now potentially affordable to everyone.

The most widely used NGS platforms can be classified into two categories: 1) Second Generation, which is based on massively parallel sequencing of short-reads (up to 300bp) on a solid phase using reversible terminators and “washing and scanning “cycles (e.g., Illumina). 2) Third Generation, which is based on single molecule sequencing of large reads of up to 1Mb and the DNA elongation by DNA polymerase [57].

The third generation of sequencing technology has opened up new possibilities for exploring complex genomic sites, producing the novo genomes, discovering large structural variations, and analyzing epigenetic modifications[58]. Currently, there are three platforms for third-generation sequencing: synthesis-based, nanopore technology-based, and optical whole genome mapping. The synthesis-based platform allows for real-time base detection with highly accurate long-reads (e.g., PacBio), while the nanopore technology platform utilizes DNA strands that pass through a protein nanopore and create disruptions in current to allow for base detection (e.g., Oxford Nanopore Technology). Lastly, the optical whole genome mapping platform analyzes extremely long strands of genomic DNA, such as entire chromosomes (e.g., BioNano Genomics)[59].

3.4 Tuberculosis Treatment and Drug Resistance: The Antibiotic Crisis

TB is a curable disease; the standard treatment for Drug-susceptible TB (DS-TB), where the *Mtbc* strains are susceptible to INH or RMP, lasts for four months for patients who are 12 years or older, as well as for patients who are between 3 months and 16 years old and do not have severe TB[60]. This four-month treatment includes INH, rifapentine (RPT), MFX, and Pyrazinamide (PZA). However, the curation success rate decreases drastically to less than 50% or impossible to cure in resistant TB patients [10], [61], [62]. Patients infected with *Mtbc* strains resistant to TB medicines require specialized treatment customized to their condition's severity and any additional resistance. The duration of treatment can range from 6 to 18 months and may involve various medication combinations with bedaquiline (BDQ), pretomanid (PA), linezolid (LZD), and MFX; all oral regimen or all three Group A agents and at least one Group B agent (Table 1)[63].

However, these treatments pose a significant challenge as they require a long time under burden medications, which might have severe side effects[64]. This could lead to the patient stopping treatment early, which increases the risk of death and the spread of antibiotic resistance[63]. To address these concerns, the WHO recommends a 9-month all-oral treatment regimen, which includes 6 months of BDQ, in combination with LFX and MFX, ethionamide (ETO), ethambutol (EMB), INH (high dose), PZA and clofazimine (CFZ) for 4 months. This program has shown an 80% cure rate without relapse under ideal conditions. If a patient still tests positive for sputum smear, the treatment can be extended[65], [66] [63].

Drug-resistant bacteria pose a significant challenge as they require long and difficult treatments that may have severe side effects. This could lead to the patient stopping treatment early, which increases the risk of death and the spread of antibiotic resistance. To address this issue, the WHO recommends a 9-month treatment plan that involves multiple medications.

Table 1: Overview of drug regimen to treat tuberculosis.

WHO Group [67]	**Medication Name (abbreviation)**	**Target Genes**	**e.g., key non-synonymous mutations** ¥[68]	**Type**[69], [70]	**Mode of Action** [69]
First Line	Rifampicin (RIF)	*rpoB*	rpoB_S450L rpoB_D435 rpoB_H445Y rpoB_H445D	Bactericidal	Bind to *rpoB* and block protein synthesis by inhibiting RNA synthesis
	Isoniazid (INH)	*katG* *inhA* *fabG1*	katG_S315T inhA_c-777t (fabG1_c-15t) inhA_g-154a (fa L203L) katG_S315N	Bactericidal	Inhibits mycolic acid synthesis
	Ethambutol (EMB)	*embC* *embA* *embB*	embB_M306V embB_M306I embB_Q497R embA_c-12t	Bacteriostatic	Inhibits synthesis and polymerization of cell wall component arabinan. Affect lipid cell wall synthesis.
	Pyrazinamide (PZA)	*pncA* *rpsA*	pncA_H57D pncA_a-11g pncA_AQ10P pncA_H51D	Bacteriostatic	Formation of pyrazinoic acid that lowers pH and inactivates fatty active synthase. Inhibition of trans-translation by disrupting tmRNA interaction.
A	Levofloxacin (LFX)/ Moxifloxacin (MFX) †	*gyrA* *gyrB*	gyrA_D94G gyrA_A90V gyrA_D94A gyrA_D94N	Bactericidal	Inhibits DNA gyrase by binding of Topoisomerase II
	Bedaquiline (BDQ)	atpE *mmpR*	mmpL5_D767N mmpL5_T194I	Bactericidal	Inhibits the synthesis of ATP

		(Rv0678) pepQ	mmpL5_I948V Rv1979c_a-129g		
	Linezolid (LZD)	*rrL rpIC*	rpIC_C154R rrs_c-187t rrl_c344t	Bactericidal	Inhibits protein synthesis
B	Clofazimine (CFZ)	*mmpR (Rv0678)*	mmpL5_I948V Rv1979c_a-129g mmpL5_T794I mmpL5_D767N	Bacteriostatic	Disrupt redox balance, inhibits cell growth and ROS prediction
	D-cycloserine (DCS)	*alr ddl*	_	Bacteriostatic /Bactericidal	Inhibits peptidoglycan synthesis
C	Delamanid (DLM)	*fgd1 fbiA/B/C Ddn*	ddn_L49P	Bactericidal	Inhibits protein and mycolic acid synthesis
	Imipenem-Cilastatin (IMP-CLN) / Meropenem (MPM)	*ftsI*	_	Bactericidal	Inhibits cell wall synthesis by binding to penicillin to penicillin-binding proteins Cilastatin
	Amikacin (AMK)	*rrs eis whiB7*	rrs_a1401g eis_c-14t rrs_g1484t rrs_c1402t	Bactericidal	Disrupt protein synthesis by binding to the 70s ribosomal subunit.
	Streptomycin (STR)	*rpsL rrs gidB*	rpsL_K43R rpsL_K88R rrs_c517t rrs_a514c	Bactericidal	Inhibits protein synthesis by targeting
	Ethionamide	*ethA*	inhA_c-	Bactericidal	Inhibits mycolic acid synthesis

	(ETO)/ Protheonamide (PTO)	*inhA* *Ndh* *mshA* *ethR* *kasA*	777t(fabG1_c-15t) ethA_111_del_1_ct_c ethA_R207G ethA_M1R	
	Aminosalicylic Acid (PAS)	*ribD* promotor *thyA* *folC*	-	Bacteriostatic	Inhibits folic acid synthesis

† FQ, Fluoroquinolones; ¥ High confidence mutations involved in drug resistance of *Mycobacterium tuberculosis* strains. WHO, World Health Organization; RNA, Ribonucleic Acid

Drug resistance in TB patients occurs due to the infection with drug-resistant *Mtbc* strains harboring mutations to drug associate resistance genes (primary resistance) or can emerge during treatment failure (e.g., relapse, patient metabolism, inadequate serum, and tissue concentration), incomplete treatment or the combination of altogether along with host genetic factors and pathogen virulence traits (secondary resistance) [71], [72]. Patients infected with drug-resistant *Mtbc* strains can be classified as rifampicin resistance (RR-TB), MDR-TB, - Pre- extensively drug-resistant (Pre-XDR-TB, MDR strains with additional resistances to one of the WHO group A), XDR -TB and totally drug-resistant (TDR-TB, resistance to all first- and second-line drugs), (Table 1) [73] [67]. According to the WHO, in 2022, there were an estimated 0.41 million incident cases of MDR/RR-TB (95% UI: 370 000–450 000). The majority of drug resistance (MDR/RR-TB) cases accounted for 42 % in three countries: India (27%), The Philippines (7.5%) and the Russian Federation (7.5%) (Figure 2)[14].

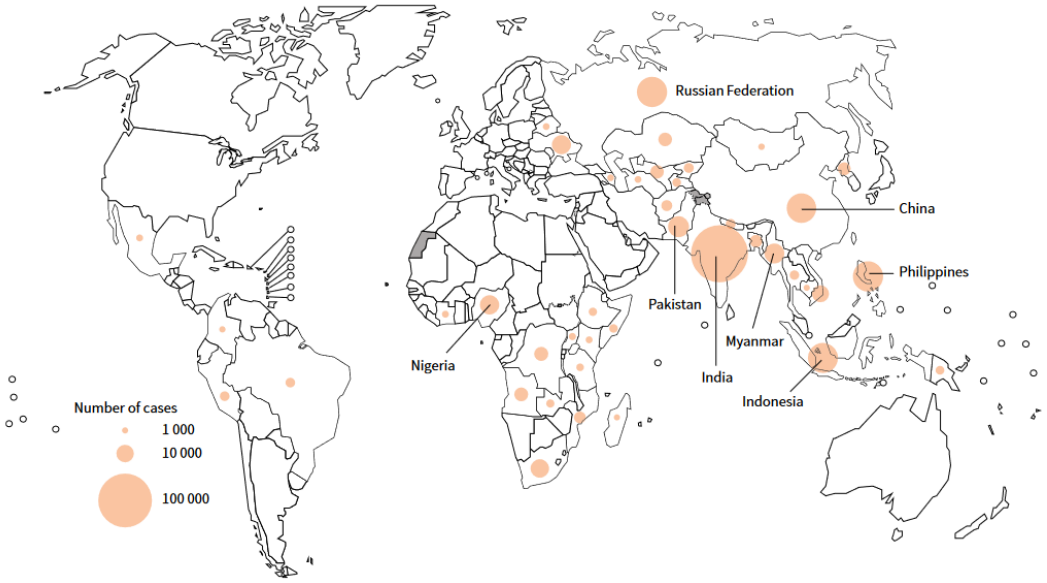


Figure 2: Global multidrug resistance/Rifampicin resistance tuberculosis cases in 2022.
Estimated incidence of resistant tuberculosis cases by country with at least 1000 cases. The map uses a color code that ranges from bright (lower incidence, 0-2.9%) to dark (higher incidence, ≥20%). Gray indicates that data is not applicable, while white indicates no data. The figures were obtained and adapted from the global tuberculosis report of 2023[14].

Additionally, around 20% of MDR/RR-TB cases were resistant to fluoroquinolone antibiotics. It's important to note that a quarter of deaths caused by antimicrobial resistance are due to RR-TB[74]. The COVID-19 pandemic has had a negative impact on achieving the global goal of universal therapy access.

In 2021, treatment coverage was below 50% in 30 countries with the highest TB burden. However, there has been a recovery to pre-pandemic levels with 70% coverage in 2022[14].

The increasing number of *Mtbc strains* that are resistant to RMP, the most effective medication against TB, poses a significant challenge to global health and economy and threatens TB control[6], [75], [76], [77]. The lack of funds for the development of new drugs since the discovery of quinolones in 1962[78], [79], coupled with the progressive increase in transmission of drug-resistant *Mtbc* strains, makes research in TB resistance a priority[80] [76] [81], [82], [83] [6]. Key areas of research include the time to sputum conversion, treatment success, relapse rates, mortality, adverse effects of medication, molecular markers of certain medications, cross-resistance with other drugs, and biomarkers of treatment failure[84] [63].

3.5 The Host-Pathogen Interaction: Transmission and Pathobiology of Tuberculosis

TB in humans and animals is caused mainly by pathogens of the *Mtbc.* These are airborne pathogens, gram-positive and acid-fast bacteria, slow growers, non-motile, no capsulated, no spore-forming, no producer of poisonous toxins, organotrophic, particularly enriched with lipidic cell wall [85], and of intracellular lifestyle [86]. In contrast to the recombinant nature of *M. canettii*, which has been exceptionally isolated from humans living in Horn Africa [12], the members of *Mtbc* comprise a clonal group of mycobacteria with 99.9% of nucleotide sequence similarity, identical 16s RNA sequences, and with distinct mammalian host preferences [87].

Mtbc cells are transmitted by aerosol when a person with active TB coughs up infected droplets nuclei into the air, which are inhaled by a susceptible person (Figure 3)[88]. The droplets are estimated to have a size of about 1-5µm diameter and can last several hours in the air [86]. Once *Mtbc* cells have entered and settled in the lower lungs of the new host, they are recognized and engulfed by alveolar macrophages. Alveolar macrophages become infected and transport *Mtbc* cells through the epithelium into deeper tissue layers of the lung [89]. Hence, in the infection site, *Mtbc* cells elicit an inflammatory response by producing a storm of cytokines and chemokines such as tumor necrosis factor (TNF) that recruit new macrophages and other cells from the phagocyte forefront (i.e., macrophages, neutrophils, monocyte, and T lymphocytes) [90].

The new phagocytes engulf the debris from the first infected macrophages, which had an apoptotic fate. However, T cells fail microbial eradication and become part of the permissive cells containing *Mtbc* cells in a cell structure called a granuloma, in which it spreads and grows restricted[91], [92]. At the same time, the adaptative immune response has evolved. From here, depending on host immunity and

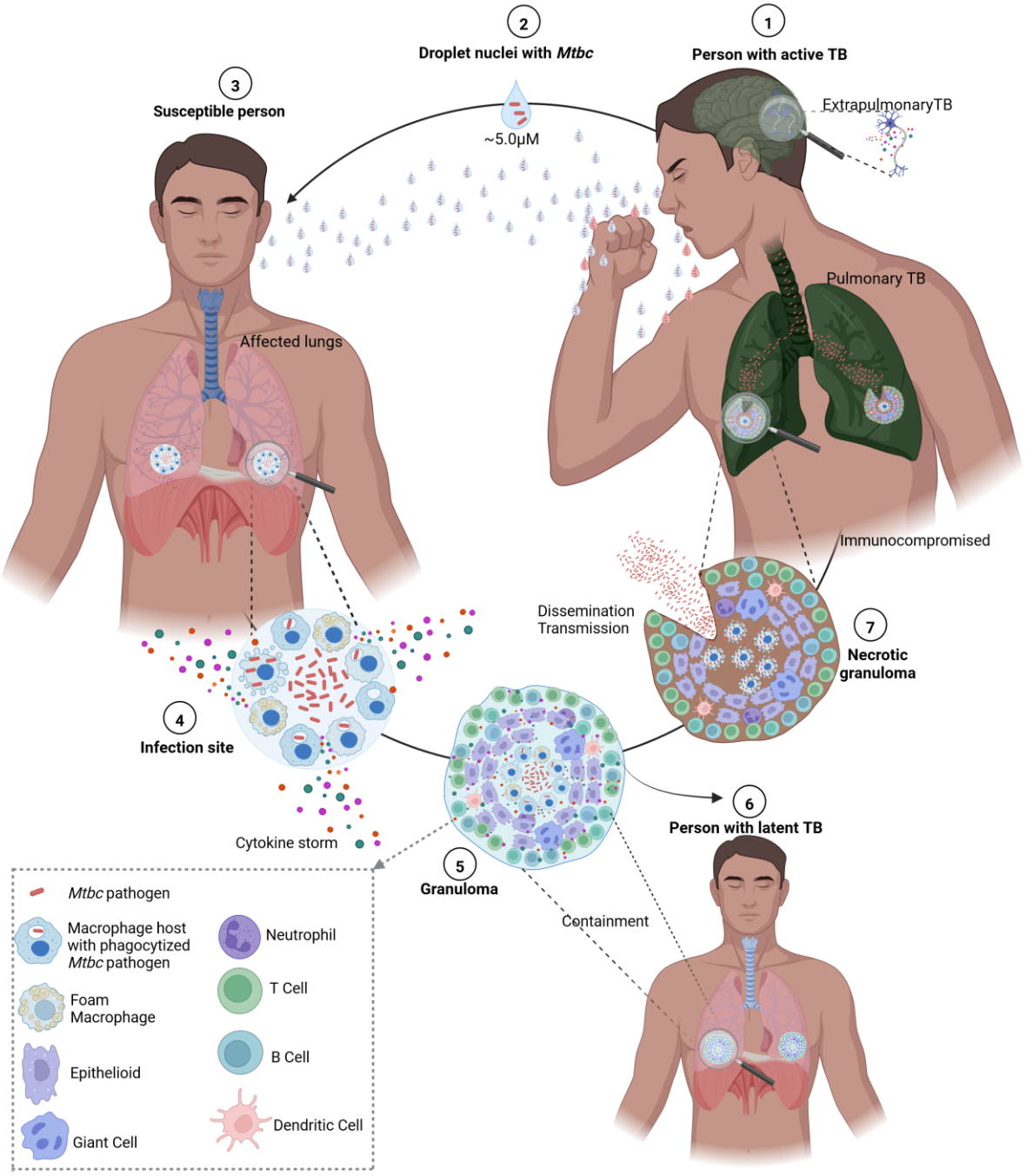
potential comorbidities, the outcome of the infection can take three possible routes: 1) Abortion of the infection manifested as a spontaneous healing. 2) *Mtbc* cells can be contained for an indefinite time causing latent disease. 3) *Mtbc* cells can escape innate or T cell immunity elimination and, therefore, cause active disease [93].

In active TB, *Mtbc* cells escape macrophage killing by avoiding phagosome-lysosome fusion, tolerating the acidic nature of the phagosomes, inhibiting apoptosis and autophagia, and manipulating host defenses through the induction of high levels of TNF, T-cell responses, and Matrix Metalloproteinase 1(MMP1). The combination of these events inflicts disintegration and destruction of the granuloma, which becomes a caseous niche full of cell debris and lipid material, which promotes growth and accelerated dissemination of *Mtbc* cells [90]. As a result, several cavitate lessons are developed in the lung, which eventually reaches the alveoli where *Mtbc* cells exit the host through the airways and infect up to 15 close contact individuals [12] (Figure 3).

Figure 3: Tuberculosis transmission.

1-3 Cells of *Mycobacterium tuberculosis* complex (*Mtbc*) strains are transmitted from an infected to a susceptible individual through the expelling and further inhalation of floating infected particles. These particles pass the upper lungs till they encounter the alveolar macrophages in the lower lungs. **4-7** Alveolar macrophages are infected and produce a storm of inflammatory cytokines that recruit other cells such as macrophages, dendritic cells, multinucleated giant cells, epithelioid and foamy cells, granulocytes, and lymphocytes. These cells form an organized aggregate called granuloma, which restricts the growth of *Mtbc* cells and, depending on the impact of the immune response, can lead to a latent (**6**) or active disease (**7**). The active disease occurs when the granuloma breaks down, which enhances the replication and release of *Mtbc* cells to the lung airways. Next, *Mtbc* cells are expelled as infectious droplets nuclei and infect new hosts. In some cases, *Mtbc* cells can cause life-threatening disease by infecting vital organs, for example, the brain (extrapulmonary TB). TB, Tuberculosis; Graphical representation elaborated with BioRender [94]

Figure 3 (Continued)



Macrophages are the hallmark host cells for the survival of *Mtbc* cells; as such, they play a crucial role in TB control and elimination by eliciting innate and acquired protective immune responses [90]. Macrophages are derived from promonocytic cells in the bone marrow, mature into monocytes in the bloodstream, then differentiate into macrophages, and migrate into distinct body tissues [95]. Macrophages become specialized in a given tissue depending on the site environment (i.e., localization, cell structure, and the composition of growth factors and cytokines) and contribute to cell homeostasis and immunological responses [96].

Macrophages possess an extensive array of receptors and compartments that enable them to identify, bind, absorb, and break down foreign entities[97]. Macrophages detect the pathogen associated molecular patterns (PAMPs) located on the surface of *Mtbc* cells and initiate phagocytosis by means of their surface and cytoplasmic pattern recognition receptors (PRRs)[98]. These PRRs consist of scavenger receptors (SR) like SR-A and MARCO, [99] C-type lectins[91], mannose receptor (MR), Dectin, Micle, and Mannose-capped Lipoarabinomannan, among others[92]. Furthermore, TLR (Toll-like Receptor) 1/2, 2/6, 4, 5, and 9[93], Fc-gamma receptors I-III [100], molecules for pathogen opsonization such as mannose-binding lectin [101], [102], surfactant proteins[103], and complement factors (C1q, C3, and C4) are also involved in this process[104].

However, *Mtbc* pathogens have developed defense mechanisms to protect themselves from the host's immune system. They have a cell wall made up of complex lipids such as phospholipids, peptidoglycan, mycolic acids, phthiocerol dimycocerosates, and sulpholipids[105], [106]. They also have specialized protein secretion systems like the Sec-dependent general secretory pathway, the Twin Arginine Translocation (TAT) system, and the ESX/Type VII Secretion Systems (T7SS) [107] [108] [109], [110]. Once phagocytosis begins, *Mtbc* pathogens use their own arsenal to manipulate the host's defense system for survival and to proliferate in other hosts[111].

When *Mtbc* cells are internalized, they can block phagosome maturation [112] and survive the harsh acidic environment, unlike other pathogens, where normally phagosomes maturate by fusing them with lysosomes, thus becoming extremely harmful structures called phagolysosomes[113]. By blocking phagosome/lysosome fusion, *Mtbc* cells make the phagosome an environment less acidic, absent of lysosomal markers and vesicular H^+ -ATPase[114]. Furthermore, LAM inhibits interferon gamma (IFN-γ) secretion and the infected macrophages neutralize the oxide nitric acid and produce higher levels of interleukin (IL)-10 that inhibits IL-12 and consequently IFN-γ production (Figure 4) [115] [116].

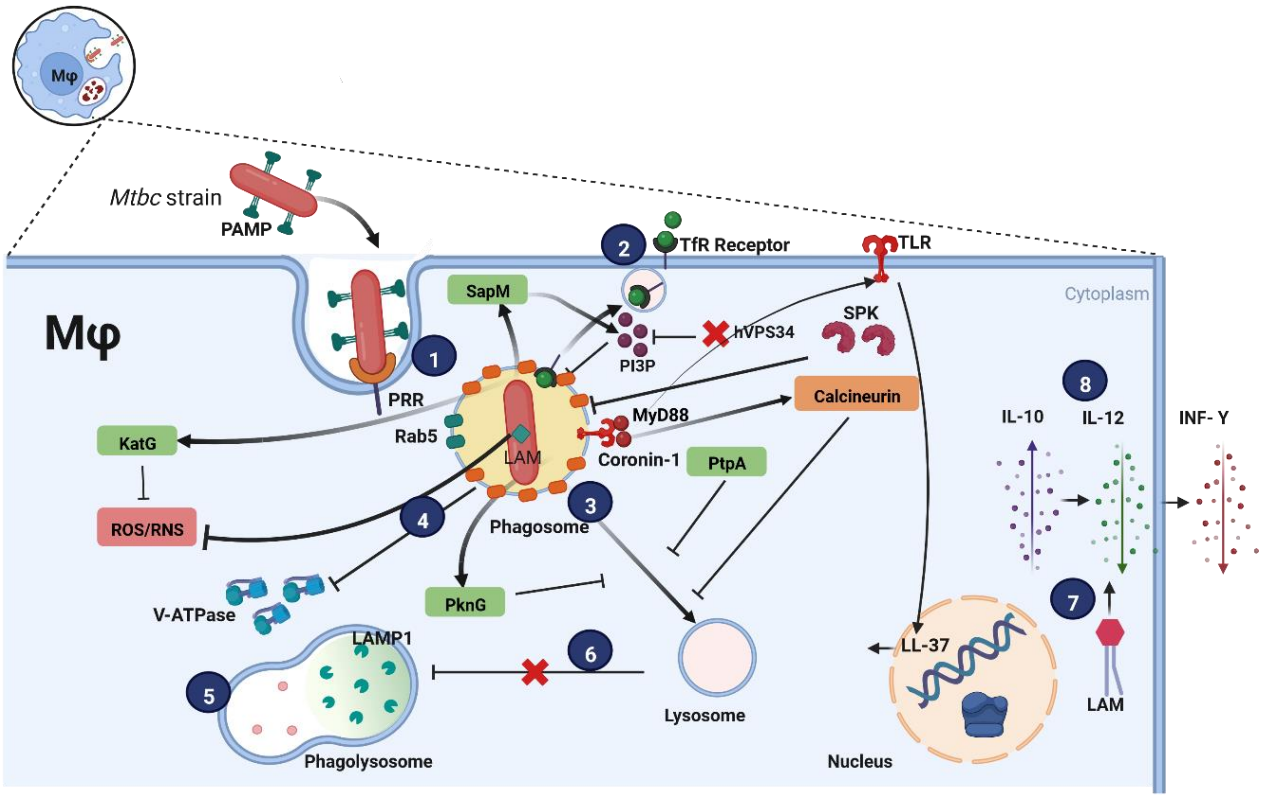


Figure 4: Molecular and immunological mechanism of the interaction of the host (macrophage)-pathogen (*Mycobacterium tuberculosis*).

1-2 Cells of *Mycobacterium tuberculosis* complex (*Mtbc*) strains are engulfed by the macrophage while obtaining iron from the extracellular environment with the transferrin receptor (TfR). **3-4** A phagosome containing *Mtbc* cells is formed from which the pathogen not only prevents the fusion of the phagosome with the lysosome but also with lysosomes by continuously associating with Rab5, preventing the Lysosomal-Associated Membrane Protein 1 (LAMP1), lowing V-ATPase levels, decreasing Ca^{2+}, discharging Secreted Acid Phosphatase (SapM), Protein Tyrosine Phosphatase (PtpA), Protein Kinase G (PknG), and Catalase-Peroxidase enzyme (KatG) and eliciting the host protein coronin 1/TACO. In addition, the action of the nitric oxide produced within macrophages is turned off by the proteasome of the *Mtbc* cells. **5-6** Toll-like Receptor (TLR) signaling moderate the phagosome-lysosomal fusion via the activation of p38MAPK and trigger the production of the antimicrobial peptide LL37 or cathelicidin by activating the vitamin D receptor via 1,25-dihydroxyvitamin D3 (1,25D3). **7-8** Lipoarabinomannan (LAM) inhibits IFN-γ secretion, and infected macrophages produce higher levels of IL-10 that inhibit IL-12 secretion, which subsequently suppresses IFN-γ production. PAMP, Pathogen-Associated Molecular Pattern Molecules; PRR, Pattern Recognition Receptor; IL, Interleukin. Representation adapted from Pieters [111] and Bekale et al. [114] and elaborated with BioRender [94].

3.5.1 The Pathogen: *Mycobacterium tuberculosis* complex

Strains of *Mtbc* share a single common ancestor, are 99% similar at the genomic level, differ in a maximum of ~2500 single nucleotide polymorphisms (SNPs) between two strains, and are classified into 9 distinct phylogenetic human and animal phylogenetic lineages [3]. *M. canettii* strains that are rarely isolated from patients from the Horn of Africa have been proposed to resemble the putative ancestor of *Mtbc* strains due to the much larger genomic diversity [98]. The human-adapted *Mtbc* lineages are grouped in ancestral or modern clades based on the presence or absence of a 2153 bp genomic region termed *Mtb* specific deletion 1 (TbD1). TbD1 codifies for the mycobacterial membrane protein families (MmpL), and it is absent in the modern clade (Lineage 2 [L2]/East Asian, Lineage 3 [L3]/ "Delhi/CAS" and Lineage 4 [L4]/ Euro American) in which the mmpS6 gene is deleted and the mmpL6 gene truncated[117]. On the other hand, TbD1 is present in strains of the ancestral clade comprising strains of Lineage 1 (L1)/East African Indian, Lineage 5 (L5)/West African I, and Lineage 6 (L6)/West African II, and partially present in strains of Lineage 7 (L7)/Ethiopia. Recently, strains of new lineage L8 and L9 have been reported in Africa, L8 being an ancestral sister clade to the known *Mtbc* lineages strains, and L9 being closely related to L6 strains (Figure 5A) [7],[118], [119].

In recent years, research has focused on understanding the success of strains of the different lineages in different geographical regions of the world, e.g., based on their transmission dynamics among human populations. WGS data have shown that *Mtbc* strains have a phylogeographical population structure with strains of particular lineages either occurring worldwide (generalist) or just in a few regions (specialist) [120], [87]. For instance, strains of several sub-lineages of L4 are spreading worldwide [121], L3 strains are mainly found in North-East Africa and Central- South Asia[122], L2 strains are dominating East Asia [123], L1 strains are found mainly in the surrounding regions of the Indian Ocean [124]. *M. africanum* (*Maf*) strains of L5 and L6 are restricted to West Africa [125], L7 strains solely of Ethiopia [126], L8 strains are found in the African Great Lakes, and L9 happening in East Africa [30],[119], (Figure 5B). These remarkable findings suggest that a complex interplay between the pathogen and the host might play a role in disease pathogenesis and the success of *Mtbc* strains in different host populations via sympatric and allopatric host-pathogen interactions [127], [128], [129].

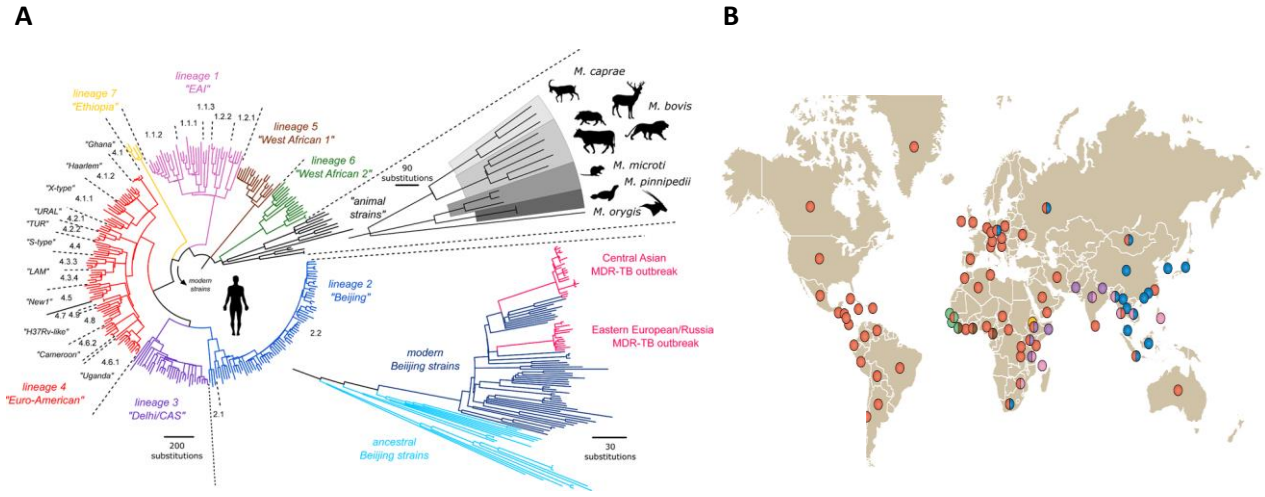


Figure 5: *Mycobacterium tuberculosis* lineage strains diversity and its global expansion.
A Host distribution of *Mycobacterium tuberculosis* complex (*Mtbc*) strains. The whole genome sequencing-based phylogeny comprises the main seven lineages and preferred hosts. Branches are code colored per lineage strains: red (Lineage 4), violet (Lineage 3), blue (Lineage 2), pink (Lineage 1), brown (Lineage 5), and green (Lineage 6). Image obtained from Niemann et al. [130] **B** Geographical distribution of *Mtbc* lineage strains. Localization of six major lineages of *Mtbc* strains represented as circles code colored per lineage: red (Lineage 4), violet (Lineage 3), blue (Lineage 2), pink (Lineage 1), brown (Lineage 5) and green (Lineage 6). MDR-TB, Multidrug resistance-Tuberculosis. Image obtained from Gagneux [118].

3.5.1.1 Modern Human-Adapted Lineages: *Mycobacterium tuberculosis* L4 Strains

Strains of L4, also known as Euro American strains, are among the most successful strains, spreading worldwide and causing a large number of human TB [131]. Several studies have suggested that Europe was the initial source of its global expansion due to human migrations, trading, and land explorations. Indeed, European colonization has been strongly associated with the dominance of L4 strains in Africa and America [3], [132].

L4 strains can be subdivided into ten sub-lineages, out of which some are generalist and some are specialist, which are restricted geographically[118], [121], [122], [127], [128], [129], [130]. Sub-lineages termed by the Coll scheme are summarized in Table 2 [133]. Strains of generalist sub-lineages are spreading globally, mainly in Europe and the American regions such as Haarlem, Latin American Mediterranean (LAM)[121], while strains of specialist sub-lineages are found in particular geographic regions, e.g., in in Sub-Saharan Africa, and Southeast Asia (e.g., Ghana, Uganda, and Cameroon) [90][122].

Table 2: Sub-lineages of Lineage 4 strains.

Coll Nomenclature Sub-lineages	Name
L4.1.1	X-Type
L4.1.2	Haarlem
L4.1.3	Ghana
L4.2	Ural
L4.4.1.1	S-type
L4.3	LAM
L4.5	-
L4.6.1	Uganda
L4.6.2	Cameroon
L4.10	Principal Genetic Group 3

L4, Lineage 4; LAM, Latin American Mediterranean

Strains of L4.3/LAM, are the most dominant sub-lineage found in LATAM followed by strains of L4.1.2/Haarlem, particularly in the northern South America Region (e.g., Mexico, Brazil, Peru, Colombia, etc..)[134]. They were presumably introduced into the indigenous populations after Christopher Columbus and Portuguese invasions during the 16th -19th centuries[132]. Despite the massive trade of African slaves at that time, *Maf* strains never succeeded in spreading in the region, potentially due to lower virulence in the host population compared to L4 strains that outcompete them in the region of the Americas [134], [135].

In vitro experiments in cells, animal models, as well as studies in human patient cohorts indicate that L4 strains are potentially more virulent than strains from other lineages (i.e., bacterial uptake, intracellular growth, transcriptomic signatures, mycolic acid biosynthesis, oxygen starvation) [136], [137], [138]. Furthermore, it has been suggested that L4 strains induce a delayed immune inflammatory response (e.g., IL-6, IL-12p40/p70, TNF-α, Macrophage Inflammatory Protein [MIP-1a] and reduction of the adaptative response and therefore causing a more severe disease [139], [140], [141]. The delayed response favors bacterial proliferation at the early stages of the infection, which leads to an increased virulence later, which benefits from a displayed conservative human T cell epitope repertory; however, the diversity of the T cell epitopes can vary among sub-lineages strains[142], [143].

3.5.1.2 Ancient Human-Adapted Lineages: *Mycobacterium africanum* L6 Strains

The *Maf* strains are predominantly found in West Africa and are grouped into L5 and L6 strains, also known as strains of West African I and II lineages [125]. Strains of these lineages are considered ancient as they emerged early in the evolution of the *Mtbc* from the modern lineages occurring in the regions of America, Europe, and Asia [3]. *Maf* strains differ from their *M. tuberculosis* (*Mtb*) sensu stricto contra parts at the biological level, in terms of transmission dynamics and the ability to cause disease in humans[125]. *Maf* strains display a phenotypical and biochemical heterogeneity that is comparable to *M. bovis* (*Mbv*) animal-adapted strains, for instance: dysgonic morphology, slow growth rate, microaerophilic using pyruvate as preferred carbon source, catalase activity, and negative nitrate reduction. Moreover, *Maf* strains differ from *Mbv* strains by producing niacin weakly and being susceptible to Pyrazinamide (PZA); however resistance to PZA has been found in MDR *Maf* strains[144], [145], [146], [147], [148]. In addition, depending on their susceptibility or resistance to Thiophene -2-carboxylic acid hydrazide (TCH), *Maf* strains are classified into East African or West African I and II variants [149].

Maf strains have been mostly found in countries of west and central Africa such as Ghana, Sierra Leone, Nigeria, and Benin[150], [151]. However, there have also been a few cases reported of African immigrants residing in other regions of the world, Europe and America [152], [153] [154], [155], [156]. Despite this, the transmission of *Maf* strains has not occurred on a large scale outside of Africa, indicating that host-pathogen adaptations may play a role, which fits the generalist and specialist concept described for L4 strains[121], [125]. Although *Maf* strains share similarities in their phylogenetic, phenotypical, and geographical distributions with animal-adapted *Mtbc* lineage strains (such as *Mbv*, *M. mungi*, and the Dassie bacillus), no animal reservoir has been identified based on overall *Mtbc* population structure analysis using SNPs-based phylogenies, absence or presence of Regions of Difference (RD)7, RD8, RD9, and RD1, and microbiology properties[138], [157]. However, *Maf* strains have been isolated from animals like cows, chimpanzees, and wild animals, leading to speculation of a possible animal reservoir that could cause zoonotic infection in humans [125], [151], [158], [159], [160], [161].

Moreover, *Maf* strains have been described as more opportunistic pathogens in HIV patients compared to *Mtb* strains[162]; however, this statement has been contradicted by Meyer and colleagues, who reported no differences in pathology caused by *Maf* and *Mtb* strains in HIV patients[163]. Additionally, *Maf* strains of the L6 have shown a less virulent response in animal and human cell models, a delayed

progression to active disease, and an attenuated response of ESAT-6, eliciting an early strong response of proinflammatory cytokines, as compared to other lineages strains [139], [164]

3.5.1.3 Animal-Adapted Lineages: *Mycobacterium bovis* Animal Clade 4 Strains

In 1898, it was revealed for the first time that the pathogen causing TB in cattle was different from the one in humans. This pathogen was identified as *Mbv*, which led to the discovery of other strains infecting different animal species [165], [166]. These strains are distinguished from the other *Mtbc* members based on their animal host preference, the presence or absence of certain genetic targets, WGS-based classifications, and physiological features. For instance, RD12 and RD13 are present in *M. microti* (small animals and rodents)[167], the Dassie Bacillus (Cape hydrax) [168], *M. mungi (*banded mongoose) [157], and *M. suricattae* (meerkats) [169]. Strains of the last three species also have the RD1 deleted [170]. RD12 is absent in *M. orygis* (*Bovidae* family mainly, but also recovered from waterbucks, species of deer, and rhesus monkeys among others)[166], PiD1 is absent, and PiD2/RD2 are present in *M. pinnipedii* (seals, sea lion, tapirs, and camels)[171]. The deletions of several RD regions (e.g., 4, 5, 7, 9, 6, 7, 8, 9, 10, 11, 12, 13, 17) in *Mbv* (cattle, deer, buffalo, goats bison, badger among others)[172] [173], [174], [175].

A recent study has classified animal-adapted strains also known as ecotypes based on WGS data into three phylogenetic clades termed A1 (*M. mungi*, *M. suricattae*, the dassie bacillus, and the chimpanzee bacillus), A2 (*M. microti and M. pinnipedii)*, A3 (*M. orygis*) and Animal Clade 4 (A4) such as *Mbv*, *M. caprae* and the BCG vaccine strains of *Mbv* [170]. Furthermore, there has been evidence of the animal-adapted strains isolated from humans, for instance, *M. pinnipedi* in seal trainers and Peruvian human skeletons[176] [177], M. *orygis* in African and South Asian patients, *Mbv* in abattoir workers and cattle herders and immunocompromised individuals [178], [179], [180], [181].

Mbv has been extensively studied in comparison to other animal-adapted strains due to its economical relevance [182]. Cattle are controlled for TB before going to the global livestock market as meat and milk production. When cattle have a tuberculin skin test (TST) positive, they are slaughtered, which represents an enormous financial loss. Although *Mbv* is generally attenuated in humans in comparison with *Mtb*, it's a public health zoonotic threat [183].

Further differences among strains from different lineages of the *Mtbc* have also been observed at the immunological, transcriptomic, and pathological levels in macrophages and bovine animal models [184] [185], for instance in the production of antimicrobial activity agents (e.g., nitric oxide and TNF [186], differential expression of virulence-associated genes (e.g., CCL3, CCL3, CCL4, CCL5, CXCL8, and CCL23) and increased gross pathology and histopathology[184] [187]. These findings have intrigued the scientific community to better understand the landscape of differences in host-pathogen interaction in relation to

the *Mtbc* phylogeny. While the concept of generalist and specialist lineages, allopatric and sympatric host-pathogen interactions have been established, and differences in host tropism have been observed, the underlying pathobiological mechanisms are mainly undefined.

3.5.2 The Host: Disease Susceptibility

Over the past years, numerous studies have sought to understand the mechanisms that make individuals susceptible to TB and why certain genes that cause sickness have not been eliminated through the Darwinian process [188]. Researchers have been looking into why some strains of TB, such as L4 strains, have spread globally while others, like L3, L1, L5, and especially L6 strains, remain confined to specific geographical regions[76], [120], [121], [122], [127], [189] despite the mass transatlantic migrations waves. They have shown evidence that strains of these TB lineages might adapt to certain human populations as a result of complex co-evolutionary dynamics at play between different strains of TB and their hosts[127], [128].

Throughout human evolution, pathogens have been a significant selective pressure [190]. Hosts have had to evolve ways to combat these pathogens, while pathogens have attempted to devise tactics such as antigen variation to evade host immune systems and successfully infect. When the resistance of both the host and the pathogen meet, an adaptation status arises, and the interactions between hosts and pathogens impact the genetic diversity of both organisms, which can determine the susceptibility and severity of the disease; this process is known as an arms race[191] [192]. Interestingly, the *Mtbc* pathogens do not seem to resist being recognized by the human immune system. In fact, 95 percent of its T cell antigens remain the same [142], [143], which suggests that it wants to be detected. This may actually be beneficial for the pathogen, as it could promote the formation of granulomas and ensure its survival and successful transmission in the long term[193].

Host- Pathogen relationships can be classified as *allopatric*, in which a geographical barrier forms in nature from the host; thus pathogens originate from separated or *sympatric*, when the host and pathogen share a common ancestral geographic location [194], [195]. For example, West Africans infected with West African/L6 strains, East Asians infected with East Asian/L2 strains, and Europeans/Caucasians infected with Euro American/L4 strains are examples of sympatric relationships[127]. In contrast, Asians infected with Euro-American/L4 strains or Europeans and Americans infected with West African/L6 strains are examples of allopatric relationships. Moreover, there is evidence to suggest that the type of relationship between hosts and pathogens may affect the outcome of TB disease. A recent study discovered that individuals with pulmonary TB caused by

sympatric genotypes are less likely to experience pulmonary impairment compared to those with allopatric genotypes[196].

When it comes to *Mtbc* strains, it's important to note that their ability to cause damage to the lungs, survive outside of the host, overcome the host response, and successfully transmit can vary [136], [137], [197], [198], [199]. The host's genetic makeup also might play a role in their resistance or susceptibility to a particular strain, which can ultimately impact the course of infection, disease severity, and overall resistance to the pathogen, regardless of any sociodemographic risk factors[200]. Specific genetic regions, such as Human Leukocyte Antigen (HLA) alleles, as well as polymorphisms and epigenetic contexts in certain human populations, may also affect the success of a given *Mtbc* lineage in a certain population [201], [202].

R52C [rs5030737] and TYK*2* P1104A. R52C [rs5030737] are among the most interesting human polymorphisms. R52C is found predominantly in West African populations and affects the Mannose-binding Lectin (MBL) 2 family, which plays a crucial role interacting with the TB pathogen. This polymorphism is thought to provide protection against TB caused by *Maf* strains[102], [203]. In contrast, TYK2 P1104A is linked to an increased likelihood of developing clinical TB in people of European descent. Recent findings indicate that this genetic variant has played a significant role in TB prevalence among European populations for the past 2000 years[204], [205]. Environmental factors may also influence host susceptibility; for example, it is believed that the West African populations consume a high amount of vitamin B12, which is advantageous for *Maf* strains because they cannot synthesize the vitamin [206]. Vitamin B12 is essential for maintaining the intracellular growth of *Maf* in the host, and studies have shown that its addition to growth cultures greatly increases *Maf* growth[34], [207].

Overall, the interactions between the host and pathogen in TB seem to have formed a co-adaptation pattern, where the success of certain strains of TB in specific populations could be explained by sympatric and allopatric relationships[120], [127], [129], [208], [209]. However, in cases where individuals are co-infected with HIV, the sympatric relationship between the host and pathogen can be disrupted, allowing different *Mtbc* strains to cause disease regardless of the host's genetic makeup [210]. The success of *Mtbc* strains from different lineages ultimately seems to depend on their genetic makeup concerning the host, which can affect disease progression, pathobiology, and transmission. Despite this understanding, the impact of the *Mtbc* pathogen's diversity on pathobiology is not fully comprehended. Further research is necessary to address this knowledge gap.

3.6 Main Objective

This dissertation aims to investigate the impact of the diversity of *Mycobacterium tuberculosis* complex strains on transmission and the pathobiological aspects of tuberculosis disease.

3.6.1 Specific Objectives

I. Investigate the causative *Mycobacterium tuberculosis* complex lineage strains and the transmission dynamics of a clonal tuberculosis outbreak affecting a remote indigenous population in the Colombian Amazon Region.
II. Explore an alternative whole genome sequencing approach analysis to increase the phylogenetic resolution of the tuberculosis clonal outbreak.
III. Identify growth and transcriptomic signatures of ancient and modern *Mycobacterium tuberculosis* complex strains *in vitro*.
IV. Characterize host-pathogen relationships between distinct *Mycobacterium tuberculosis* complex lineage strains and host cells using an *in vitro* macrophage infection model.
V. Analyze the effects of different *Mycobacterium tuberculosis* complex lineage strains on the inflammatory response of human macrophages.
VI. Link of pathogen and host responses and its impact on the susceptibility and outcome of tuberculosis disease.

4. MATERIAL

Consumables, reagents, kits, antibodies, primers, laboratory instruments and facilities, software, and coding tools are listed in the supplementary material (see

10. SUPPLEMENTARY MATERIAL)

4.1 Buffers and Solutions

Buffer or Solution Name	Preparation
Bovine Macrophage Culture Media	Dissolve 250 µL penicillin/streptomycin 1%, 250 µL L-glutamine 1%, 1 mL bovine serum 1% and 2.5 µL of Bovine Macrophage Colony-Stimulating Factor (M-CSF) (10 ng/mL) in 25 mL Very Low Endotoxin-Roswell Park Memorial Institute Medium (VLE-RPMI) 1640.
CFU solution Tween^{80} 0.05%	Dilute 2.5 mL Tween^{80} 20% with 997,5 mL dH_2O, sterilize by filtration 0.2 µm and store at 4°C.
EDTA-free protease inhibitor solution	Stock solution 7 X: Add one cOmplete Mini EDTA (Ethylenediaminetetraacetic acid**)**-free tablet in 1.5 mL of Dulbecco´s Phosphate-Buffered Saline (D-PBS) 1X.
Erythrocyte lysis Buffer	Dilute nine parts of solution 1 (0,16 M NH_4 CL) and one part of solution 2 (0,1M Tris pH 7.5).
Ethanol 80%	Dilute 40 mL Ethanol 100% with 10 mL of Milli-Q water and store at room temperature (RT).
Flow Cytometry Staining Buffer	Bovine serum albumin (BSA) 1%, fetal bovine serum (FBS) 1% in PBS 1X.
Fixation Solution for Ziehl-Neelsen staining	Dissolve 38 mL Ethanol 70% and 2 mL Phenol 5%.
Heat-inactivated Human Serum AB+	Inactivate serum AB+ by heating at 56°C for 30 min.
Human Macrophage Culture	Dissolve 250 µL penicillin/streptomycin 1%, 250 µL L-glutamine 1%, 1

Media	mL heat-inactivated human AB + serum 1% and 2.5 µL of Human M-CSF (10 ng/mL) in 25 mL VLE-RPMI 1640.
Macrophage Infection Media	Add 5 mL L-glutamine 1% and 50 mL Heat-inactivated Fetal Calf Serum (FSC) 10% in 45 mL VLE-RPMI 1640. Prepare on the day of the infection.
Mycobacteria Solid Media	Dissolve 19 g of the Middlebrook 7H10 dehydrated base powder and 5 mL glycerol in 900 mL dH_2O. Autoclave and add 100 mL of inactivated and sterile bovine serum and 10 mL of Cycloheximid (1 mg/ mL). Serve 25 mL in a sterile plastic plate and store at 4°C.
Mycobacteria Liquid Media	Dissolve 4.7 g of Middlebrook 7H9 dehydrated base powder and 2 mL glycerol in 900 mL dH_2O and autoclave. Next, add 2.5 mL $Tween^{80}$ 20%, OADC 10% shortly before use, and store at 4 °C.
Macrophage lysis solution SDS 0.05%	Dissolve 50 mg of Sodium Dodecyl Sulfate (SDS) in 100 mL of dH_2O and sterilize by filtration 0.2 µm.
PBS 1X	Dilute 100 mL PBS 10X with 900 mL sterile water.
Permeabilization Buffer for Intracellular Staining	Saponin 0.5%, Bovine Serum Albumin (BSA) 1% in PBS 1X.
Fixation Solution PFA 4%	Add 2.5 mL of aqueous PFA (Paraformaldehyde) solution 16% to 7.5 mL of PBS.

5. METHODS

5.1 Study Material

5.1.1 Outbreak Study in the Colombian Amazon Region

Next-generation sequencing analysis of the outbreak included 74 patients with pulmonary TB. *Mtbc* isolates were recovered by primary isolation from sputum samples on Löwenstein Jensen (LJ) media and previously characterized with molecular and microbiological standard methods [211]. The TB patients were identified through house-to-house medical consultations in 16 indigenous settlements among a population of 6310 individuals based on the shore of the Amazon (AR) and Loretoyaco rivers (LR) in Puerto Nariño, Colombia, in 2016 [212]. The localization of the settlements is shown in Figure 7. Geographical maps are an adaptation of Google and Colombian National Maps: Comunidades indígenas de Puerto Nariño, Asociación de Indígenas, Tikuna, Cocama and Yagua (ATICOYA). Open data by Gobernación del Amazonas at https://www.datos.gov.co was included [213], [214]. Sociodemographic and clinical characteristics of the TB patients, such as age, sex, health system affiliation, household conditions, history of tuberculosis, Bacillus Calmette-Guérin (BCG) vaccination status, recreational drugs usage, alcohol, smoke, self-reported HIV status, and other medical diagnoses were recorded. Genomic DNA of the isolated *Mtbc* strains and epidemiological data of the infected patients were provided by Dr. Martha Murcia from the School of Medicine, Microbiology Department of the National University of Colombia, Colombia.

5.1.2 Host-Pathogen Interaction Study

5.1.2.1 Host Cells

Human Donors: Blood donation of healthy donors was approved by the ethics committee of the Research Center Borstel (RCB) (Borstel, Germany) and the Universität of Lübeck (Lübeck, Germany) under the license AZ-19-071. A total of ~ 300 mL of whole blood heparin- anti-coagulated was obtained via venipuncture by the blood service unit at the RCB. Human healthy donors were recruited from the European Caucasian and African ancestries with a medical checkup no longer than two years ago.

Cattle Donors: ~ 300 mL of whole blood (heparin- anti-coagulated) was obtained from German Holstein cattle at slaughter from a local abattoir (Landschlachterei Rögner, Todesfelde, Germany). The competent German ethics committee approved using bovine cells (file number: II-39.20-05-04 Date: 19.10.2017).

5.1.2.2 *Mycobacterium tuberculosis* complex Strains

A panel of nine *Mtbc* clinical susceptible isolates representing the animal-adapted and human-adapted clades were selected from a reference collection at the RCB (Table 3). Isolates were recovered from clinical samples or provided by the American Type Culture Collection (ATCC)and previously characterized by drug susceptibility testing and sequencing [211]. Four main animal-adapted clades have been reported in a new phylogenetic framework, including *Mbv* within the Clade A4 [170]. Tb d1 is absent in the ancient lineages but present in the modern lineages. The regions of difference (RD)7, RD8, and RD10 are not present in the ancient *Maf* and animal-adapted *Mbv* strains. RD 900 is present in *Maf* and absent in the modern lineages and animal-adapted *Mbv*. In *Maf*, RD900 contains a single gene called maf1, which encodes a putative ATP-binding cassette transport protein [215].

Table 3: *Mycobacterium tuberculosis* complex strains investigated in host-pathogen interaction experiments.

***Mtbc* Member**	**Clade**	**Lineage**	**ID NRC**	**Type**	**Genotype**	**Tb d1**	**RD**[117], [215]				
							7	**8**	**9**	**10**	**900**
Mtb	Modern Human-Adapted	4 (Euro-American)	9679/00	Reference strain	ATCC 27294 H37Rv	-	+	+	+	+	-
			7968/03	clinical isolate	LAM	-	+	+	+	+	-
			8885/03	clinical isolate	LAM	-	+	+	+	+	-
Ecotype Mbv	Animal-Adapted Clade 4	Animal-Adapted	9564/00	Reference strain	ATCC 19210	+	-	-	-	-	-
			751/01	clinical isolate		+	-	-	-	-	-
			7540/01	clinical isolate		+	-	-	-	-	-
Maf	Ancient Human-Adapted	6 (West-African II)	9550/00	Reference strain	ATCC 25420	+	-	-	-	-	+
			10517/01	clinical isolate	II	+	-	-	-	-	+

	10514/01	clinical isolate	II	+	-	-	-	-	+

Mtbc, *Mycobacterium tuberculosis* complex; ID, Identification; NRC, National Reference Center; Tb d1, Tuberculosis deletion 1; RD, Region of Difference; LAM, Latin American Mediterranean; *Mtb*, *Mycobacterium tuberculosis*; *Mbv*, *Mycobacterium bovis*; *Maf*, *Mycobacterium africanum*; ATCC, American Type Culture Collection

5.2 MICROBIOLOGY

5.2.1 Growth Conditions of *Mycobacterium tuberculosis* complex strains

Mtbc strains were grown on Löwenstein-Jensen (LJ) medium and further passaged in 10 mL of Mycobacteria liquid media (MLM), Middlebrook 7H9 broth supplemented with oleic acid albumin dextrose catalase (OADC) 10%, glycerol 0.2%, Tween 80 0.05%. *Mtbc* bacilli were grown by shaking 100 revolutions per minute (rpm) until mid-log phase at an Optical Density $(OD)_{600nm}$ of 0.5 - 0.6 at 37 °C, harvested, and stored as stocks in MLM at -80 °C. Precultures were performed by inoculating 1000 µL of unfrozen *Mtbc* bacilli stock in 9 mL of MLM and were grown to shake (100 rpm) until mid-log phase (OD_{600nm} of 0.5 - 0.6) at 37 °C, harvested, and stored in MLM with glycerol 20% at -80 °C. After three weeks of storage, six aliquots were selected randomly per isolate and titrated per triplicate by Colony-forming units (CFU) assay onto Middlebrook 7H10 agar plates supplemented with heat-inactivated bovine serum 10% and cycloheximide (10 mg/mL), followed by incubation at 37 °C for 3-4 weeks.

5.2.2 Growth Curves of *Mycobacterium tuberculosis* complex Strains

Precultures were performed by inoculating 1000 µL of unfrozen stock in 9 mL of MLM and were grown by shaking (100 rpm) until mid-log phase (OD_{600nm} of 0.5 - 0.6) at 37 °C; the calculated volumes of these precultures were transferred into roller bottles (490cm^2) containing 100 mL of MLM for an inoculum of *Mtbc* bacilli at $OD_{600nm} \sim 0.02$. *Mtbc* growth was monitored by OD and CFU measurements (see 5.2.4 and 5.2.5) every day and two days, respectively, for 14 days. Sterility controls such as inoculation in Luria broth (LB), brain heart infusion broth (BHI), blood agar, and Ziehl-Neelsen staining were performed upon growth (5.2 MICROBIOLOGY 5.2.3.1 Ziehl-Neelsen Staining in 5. METHODS).

5.2.3 Sterility Controls

Mtbc liquid and solid cultures were inspected for contamination with fungi or other bacterial species with the following methods:

5.2.3.1 Ziehl-Neelsen Staining

Fresh fixation solution consisting of Ethanol 70% plus Phenol 5% was prepared every time**.** A sample of 10 µL of *Mtbc* culture was dropped onto a slide, spread with a pipet tip, and air-dried for 20 min. Next, slides were introduced in a coplin staining jar filled with fixation solution and incubated for 10 min at RT (*Mtbc* killing time). Afterward, the slides were left in a jar, disinfected with Optiset 7%, and taken out of the Biosafety Level 3 laboratory (BSL3). Slides were overlayed with filtered carbolic fuchsine solution, heated over a flame until vapor raised, cooled down for 1-2 min (repeated two times more). The staining solution was rinsed off and the slide was dipped in HCL-ethanol 3% and washed again with deionized water (diH_2O). The slide was submerged in methylene blue solution (1:10 diluted) for 1 min, washed by dipping in diH_2O, and air-dried. Finally, the smear was inspected for acid fast bacilli (AFB), visualized as red bacilli using light microscopy (Olympus BX41 MODEL BX41TF) with the 100x oil immersion objective.

5.2.3.2 Blood Agar

A sample of 10 µL was taken from *Mtbc* culture, dropped and distributed on blood agar, incubated at 37 °C for one week, and checked for growth every day.

5.2.3.3 Brain-Heart Infusion and Luria Broth Culturing

A sample of 200 µL of *Mtbc* culture were inoculated in 5 mL of these two media in polystyrene sterile tubes of 15 mL, incubated standing at 37 °C for two weeks and checked for contamination every week.

5.2.3.4 Inactivation of *Mycobacterium tuberculosis* complex Grown in Mycobacteria Solid Media

Mycobacterium tuberculosis complex (*Mtbc*) bacilli were inactivated to extract DNA to track cross-contamination among the strains with MIRU-VNRT 24 Loci. Inactivation was performed: Grown *Mtbc* bacilli were scraped from plates with sterile loops and transferred into a 1.5 mL screw-cap tube containing 400 µL distilled water. Screwcap tubes were closed and incubated in a water bath at 80 °C for 20 min for deactivation (*Mtbc* killing time).

5.2.4 Optical Density Readout

Light scattering upon the growth of *Mtbc* bacilli was measured by OD_{600nm} with a spectrophotometer (Cell density meter 40, Fisher Scientific). The instrument was calibrated by measuring OD_{600nm} of a no-light scattered sample (blank) of 1 mL of MLM added into a single-use cuvette covered with a lid. After that, *Mtbc* cultures were measured in 1 mL aliquots in single-use cuvettes covered with caps. Cultures were diluted when the OD_{600nm} measurements exceeded the value desired.

5.2.5 Colony Forming-Unit Assay

To assess *Mtbc* strains in MLM, calculate an infectious dose, or monitor intracellular growth, 25 µL of unclamped *Mtbc* strains suspension (passed 20 times through a 25 G cannula) or 25 µL of infectious macrophage lysate was taken and added into 225 µL of CFU solution Tween ^{80} 0.05% in a 96 well plate, mixed 20 times. 25 µL of the dilution transferred to the next well, extending with the subsequent dilutions (10^{-8}x). Then, 100 µL from each dilution was plated on half of a plate of mycobacterial solid media (MSM), meaning two dilutions were platted on one agar plate until dry with a sterile loop and incubated at 37 °C for three or four weeks. MSM comprises 7H10 agar containing 10% (w/v) heat-inactivated bovine serum (Biochrom, Germany) and cycloheximide (10 mg/mL). Three technical CFU measurements were conducted to access *Mtbc* strain concentration in MLM or to calculate infection dose. Three technical CFU measurements were conducted per macrophage infection well, and three infection wells were assayed per strain. Therefore, nine CFU measurements were performed per strain per infection time point. After the incubation period, CFU readouts were visually conducted and calculated using the following formula:

$$\text{CFU/ml} = \frac{\text{Colonies count} * \text{dilution factor}}{\text{Volume plated (µl)}}$$

5.3 CELL BIOLOGY

5.3.1 Isolation of Peripheral Blood Mononuclear Cells

Whole blood was diluted 1:1 (human) and 1:2 (bovine) with warm 1x PBS. 35 mL of this dilution was overlayed slowly on a 15 mL polysucrose solution (Pancoll, density 1.077 g) contained in sterile polystyrene tubes of 50 mL. These tubes were centrifugated at 1500 rpm (TX-1000 max. 4200 rpm Thermo Scientific) at RT for 45 min, acceleration: 3, deceleration: 0, with no brake. The yellow upper

phase was removed, and the buffy coats containing the peripheral blood mononuclear cells (PBMCs) were kept.

PBMCs were separated by density gradient centrifugation method as follows: PBMCs of two buffy coats were placed into a new 50 mL sterile polypropylene tube and filled with ice-cold Dulbecco´s phosphate-buffered saline (D-PBS) 1x up to 50 mL, homogenized and centrifugated at 1500 rpm (TX-1000 max. 4200 rpm Thermo Scientific) at 4 °C for 15 min with full brake (washing step). PBMCs pellets were polled together and washed by adding ice D-PBS 1x up to 50 mL and centrifugated as described previously.

Next, supernatants were discarded, and pellets were incubated with erythrocyte buffer lysis for 5 min to remove contaminating red blood cells and resuspended in Macrophage Culture Media (MCM) for Counterflow Centrifugal Elutriation (CCE) or MACS buffer for $CD14^+$ monocyte separation. PBMCs were counted by electronic current exclusion technology with CASY (Innovatis) or automatic trypan blue dye exclusion with the Vi-CELL XR instrument (Beckman) according to the manufacturer's instructions.

5.3.2 *In vitro* Generation of Macrophages

5.3.2.1 Counterflow Centrifugal Elutriation

Monocyte fraction was separated from PBMCs using a liquid clarification technique based on distinct sedimentation properties and cell size. Centrifugation was carried out as described elsewhere [216]. Monocytes isolated by this method were used for trial experiments dedicated to establishing the macrophage cell culture.

5.3.2.2 Isolation of $CD14^+$ Monocytes

Human and bovine PBMCs underwent immunomagnetic separation of monocyte cells by using beads coated with anti-human CD14 antibodies (Miltenyi Biotec, Germany), which has shown cross-reaction with bovine monocytes; isolation was conducted according to manufacturer´s instructions [217]. Monocytes with purity and viability greater than 85-90% were used for cell culture. Purity was evaluated with flow cytometry (see 5.3.5.1 Cell Immunophenotyping in 5. METHODS). Viability and counting were measured using electronic current exclusion technology with CASY or automatic trypan blue dye exclusion with the Vi-CELL XR instrument (Beckman). Monocytes isolated by this method were used for the official cell infection experiments.

5.3.3 Cell Culture

Monocytes were seeded in Macrophage Culture Media (MCM): RPMI 1640, penicillin/streptomycin 1%, L-glutamine with 1% heat-inactivated human serum, or bovine fetal calf serum. Human monocyte-derived macrophages (hMDMs) were seeded in 24 well plates at a density of 0.5 x10^6 cells per well, and bovine monocyte-derived macrophages (bMDMs) were seeded in 48 well plates at a density of 0.2 x10^6 cells per well due to limited cell yield (Table 4). Monocytes were further differentiated into monocyte-derived macrophages (MDMs) after six to seven days of culture in the presence of 10 ng/mL of human or bovine recombinant M-CSF. Cells were incubated at 37°C in CO_2 5%, MCM being changed every three days until cells achieved ~ 80% confluence before infection. Monocytes isolated by CCE were used only in the cell culture optimization experiments, and CD14+ monocytes separated with CD14 MicroBeads were used for all experiments based on the best yield, purity, and viability.

Table 4: Cell culture conditions.

Experiment Type	Mϕ * Host	Seeding Density	Volume (µL) of MCM	Plate	Plate Type
***Mtbc* strains survival and Mϕ Immunological response**	Human	0.5 x10 ^{6} cells	500	24 well	Flat bottom transparent
***Mtbc* strains survival and Mϕ Immunological response**	Bovine	0.2 x10 ^{6} cells	500	48 well	Flat bottom transparent

Mtbc, *Mycobacterium tuberculosis* complex; MCM, Macrophage Culture Media; Mϕ, Macrophage

5.3.4 Cell Infections

Human monocyte-derived macrophages from donors of European Caucasian and African ancestries and bMDMs from cattle donors were infected with a panel of genetically distinct *Mtbc* strains (Table 3). Three or four independent infection experiments per host were conducted at a multiplicity of infection (MOI) of ~1:1 (*Mtbc* bacilli: macrophage [Mϕ] cells) (see Table 5). Higher MOI ratios were not considered due to its pronounced interference with host viability *in vitro* experiments lasting up to 7 days [218]. Each experimental setup included three macrophage infection well replicates per *Mtbc* strain and three negative controls (macrophages not infected) wells.

Table 5: Macrophage infection experiments.

Experiment Type	Host Specie	Ancestry	Independent Experiments	Blood Donor Born Country	Blood Donor Nationality and ID §	MOI ~1 No. *Mtbc* bacilli: No. Mϕ cells
***Mtbc strains* survival and Mϕ Immunological response**	Human	African	No.1†	Nigeria	BS-14505†	0.5 x10^6: 0.5 x10^6
			No.2 ϕ	Cameroon	BS-14445 ϕ	
			No.3 ¥	Ghana	BS-14451¥	
***Mtbc strains* survival and Mϕ Immunological response**	Human	European Caucasian	No.1†	Germany	BS-12097†	0.5 x10^6: 0.5 x10^6
			No.1†	Germany	BS-13049†	
			No.2 ϕ	Germany	BS-13461 ϕ	
			No.3 ¥	Germany	BS-14089¥	
					BS- 14249¥	
***Mtbc* strains survival**	Bovine	German Schleswig Holstein	No.1*	Germany	DE 01 224 53551†	0.2 x10^6: 0.2 x10^6
			No.1†	Germany		
			No.2 ϕ	Germany	DE 01 229 66130†	
			No.2 ϕ	Germany		
			No.3 ¥	Germany	DE 01 224 28 299 ϕ	
			No.4 *	Germany		
					DE 01 229 13 543 ϕ	
					DE 01 230 16494¥	
					DE 01 228 08 396*	

Donors are pointed with a symbol to indicate the experiment they belong to. † = Experiment No.1, ϕ = Experiment No.2**,** ¥ = Experiment No.3 and * = Experiment No.4. *Mtbc*, *Mycobacterium tuberculosis* complex; MCM, Macrophage Culture Media; Mϕ, Macrophage; MOI, Multiplicity of Infection; BS, Blutspender in German; DE 01, Deutschland Schleswig-Holstein

Macrophage infections were conducted as follows: *Mtbc* strain stocks were previously grown at mid-log phase (OD_{600} 0.5-0.6), and CFU titrated. These *Mtbc* strain stocks were unfrozen for 10 min in a heat block at 37 °C and centrifugated at 4500 rpm (Eppendorf 5427 R FA-45-24-11 max. 16220 rpm) for 10 min at 4 °C to remove medium traces. Next, supernatants were removed, and *Mtbc* bacilli pellets were resuspended in 1 mL of MIM by inserting a syringe with a 25-gauge needle into the tube and passing the bacterial suspension in and out of the syringe ten times (20 times in total passing through the needle). No more than four samples were processed simultaneously to avoid bacterial clumping.

After that, an infection bacterial suspension (*Mtbc* strains infection dose) was prepared per each strain in MIM at a MOI ~1:1 based on the CFU titration of the frozen stocks and the number of assay wells per plate. Each experimental setup comprised 4 plates for 4, 24, 96, and 168 hours post-infection (hpi), 0.5 × 10^6 *Mtbc* bacilli and 0.2 × 10^6 *Mtbc* bacilli per 500 µL were used to infect 0.5 × 10^6 hMDMs and 0.2 × 10^6 bMDMs per well respectively. Next, MCM was aspirated and replaced with infection bacterial suspension in the assay wells. MIM was only added in the negative control (macrophage not infected) wells. Plates were incubated at 37 °C in CO_2 5% for 4 h. In the meantime, a CFU assay was performed to titrate the infection dose (5.3.4 Cell Infections in 5. METHODS). Before bacteria harvesting at different time points (see above), the morphological characteristics of live macrophages were visually inspected with an inverted light microscope after completing 4 hpi and over the infection period of 7 days.

Only supernatants from the 24 hpi and 96 hpi plates were collected, filtered in a spin- X ® centrifuge tube filter 0.22 µm Cellulose Acetate by centrifugation at 800 g for 2 min and stored in EDTA-free protease inhibitor solution 7X at -80 °C for further cytokine analysis. After that, all plates of the whole experimental setup were washed three times with 500 µL of warm Hanks Salt Solution (HSS) per well to remove extracellular bacteria, as verified elsewhere[216].

To calculate *Mtbc* strains uptake or phagocytosis rate 4 hpi, 200 µL of macrophage lysis buffer (MLB) was added to the infected wells of the 4 hpi plate, incubated for 5 min at RT, and the infectious cell lysates assessed by CFU assay (see 5.3.4 Cell Infections in 5. METHODS). MLB comprises SDS 0.05% (w/v) in double-distilled water (ddH_2O). 500 µL of MIM was added to all wells of 24 hpi, 96 hpi, and 168 hpi plates and incubated at 37 °C with CO_2 5%. After three days of incubation, the infected macrophages were fed with fresh MIM.

Supernatant and bacteria harvesting was conducted as described previously in each time post-infection. The number of intracellular bacteria (CFU) was determined immediately after 4 hpi and at different time points.

$$\textit{Mtbc}\ \text{uptake}\ (\%) = \frac{\text{Colonies count at 4 hpi} * 100\%}{\text{Colonies count of infection dose.}}$$

5.3.4.1 Macrophage Infection Experiments Setup

5.3.4.1.1 European Human Monocyte-Derived Macrophages

Five separate experiments (five donors) were conducted to infect three European donors with two strains of L4, three strains of A4, one strain of H37Rv ATCC strain, and three distinct European donors with three strains of L6 and one strain of H37Rv. Each lineage strain had three distinct European donors.

Experiment 1 for Germany 1: 1 Donor = L4 (2 strains), A4 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 2 for Germany 1: 2 Donor = L6 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 3 for Germany 2: 3 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 4 for Germany 3: 4 Donor = L4 (2 strains), A4 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 5 for Germany 3: 5 Donor = L6 (3 strains), and the reference H37Rv ATCC (1 strain).

Due to limited cell yield, it was only possible to assay both lineage strains simultaneously, only in one experiment (one donor). Strains of L4, A4, and L4 were assayed together in one experiment (Germany 2), while strains of L4-A4 and L6 were each assayed in two separate experiments using hMDMs from two different donors with the H37Rv strain as a control in each experiment (Germany 1 and donor Germany 3). The separated experiments of the strains of L4-A4 and L6 were represented in one plot (Germany 1 and Germany 3), the CFU values of H37Rv ATCC strains were averaged between donors as no significant differences were found among H37Rv ATCC strains (Figure S7 and Figure S8).

5.3.4.1.2 African Human Monocyte-Derived Macrophages

Three separate experiments (three donors) were conducted to infect three distinct European donors with two strains of L4, three strains of A4, three strains of L6, and one strain of H37Rv ATCC. Each lineage strain had three distinct African donors.

Experiment 1 for Nigeria: 1 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain)
Experiment 2 for Cameroon: 2 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain)
Experiment 3 for Ghana: 3 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain)

5.3.4.1.3 Bovine Monocyte-Derived Macrophages

Six separate experiments (six donors) were conducted to infect four distinct bovine donors with two strains of L4, three strains of A4, three strains of L6, and one strain of H37Rv ATCC. Each lineage strain had three distinct bovine donors.

Experiment 1 for Bovine 1: 1 Donor = L4 (2 strains), A4 (3 strains) and the reference H37Rv ATCC (1 strain).
Experiment 2 for Bovine 1: 2 Donor = L6 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 3 for Bovine 2: 3 Donor = L4 (2 strains), A4 (3 strains) and the reference H37Rv ATCC (1 strain).
Experiment 4 for Bovine 2: 4 Donor = L6 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 5 for Bovine3: 5 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain).
Experiment 6 for Bovine 4: 6 Donor = L4 (2 strains), A4 (3 strains), L6 (3 strains), and the reference H37Rv ATCC (1 strain).

The separated experiments of the strains of L4-A4 and L6 were represented in one plot (Bovine 1 and Bovine 2), the CFU values of the H37Rv ATCC strains were averaged between donors as no significant differences were found among H37Rv ATCC strains.

5.3.5 Flow Cytometry

5.3.5.1 Cell Immunophenotyping

An antibody panel was designed to determine the purity of the isolated monocytes and differentiated macrophages (Table 6). Live/dead cells were assessed by Propidium Iodide Solution (Miltenyi Biotec, Germany). Experimental cells were stained, including unstained and stained controls. Next, cells were fixed with Paraformaldehyde (PFA) 4 % and acquired with LSR II (Becton Dickson Biosciences) and MACS Quant 10 (Miltenyi Biotec, Germany) flow cytometers in a BSL2 environment. The gating strategy included Side Scatter (SSC) and Forward Side Scatter (FSC) patterns and doublet exclusion.

Table 6: Monocyte and macrophage phenotyping targets.

Antibody	CD	Clone
Lipopolysaccharide-binding protein	14	REA599
Signal-regulatory protein	172a/b	SE5A5
FcγRIIIa Transmembrane receptor	16	KD1
ITGAM	11b	ICRF44
Macrophage mannose receptor	206	19.2
Scavenger receptor cysteine-rich type 1 protein M130	163	GHI/61
Macrosialin	68	Y1/82A

CD, Cluster of Differentiation; ITGAM, Integrin alpha M

5.3.5.2 Multiplex Cytokine-Chemokine Profiling

Cell culture supernatants from infected and uninfected macrophages were recovered at indicated time points (see 5.3.4 Cell Infections in 5. METHODS). Frozen supernatants were unfrozen while being kept on ice and processed for the determination of 13 cytokines/chemokines with the LEGENDplex™ Human inflammatory panel with V- bottom Plate[219] (Table 7). Manufacturer protocol instructions (Biolegend, San Diego, CA) were followed with the subsequent modifications: based on previous work performed by the FCM core facility, half of the reagents and samples were used. Each LEGENDplex™ experiment comprised eight controls of known concentration, experimental samples as a pool of three Mϕ infection replicates, and negative controls such as PBS and MIM. All of these were measured per duplicate and, beads were acquired on a MACSQuant 10 (Milteny Biotec, Germany) capable of detecting phycoerythrin (PE) and Allophycocyanin (APC) fluorescent signals in a BSL3 environment.

Preliminary experiments were conducted to establish acceptable cytometer settings and an appropriate dilution factor; dilutions were performed with MIM to ensure accurate measurements. IL-8 and MCP1

were determined individually when their concentrations were higher; these cytokines were assayed with match subpanels in which the individual captured bead had a 13-x fold higher concentration. The calculation to obtain 1x capture bead concentration was performed as follows: For the customized 13 cytokines/ chemokines panel, 12.5 µL of capture beads is used for 13 cytokines/chemokines for a single captured bead, 0.96 µL was used per well (12.5 µl divided by 13 µL).

Table 7: Proinflammatory human panel 1.

Analyte	Bead ID	Analyte	Bead ID
IL-1β	A4	**IL-8**	B2
IFN-α2	A5	**IL-10**	B3
IFN-γ	A6	**IL-12p70**	B4
TNF-α	A7	**IL-17A**	B5
MCP-1	A8	**IL-18**	B6
IL-6	A10	**IL-23**	B7
IL-8	B2	**IL-33**	B9

IL, Interleukin; INF, Interferon; TNF, Tumor Necrosis Factor; MCP, Monocyte Chemoattractant Protein; ID, Identification

5.4 MOLECULAR BIOLOGY

5.4.1 Isolation of Nucleic Acids

Harvesting and isolation of nucleic acids were conducted in Biosafety Level 2 and 3 (BSL3 and BSL2) environments.

5.4.1.1 Prokaryotic DNA

Outbreak *Mtbc* bacteria were killed in an 80°C water bath for 20 minutes. Genomic DNA was extracted using the PureLinkGenomic DNA Mini Kit (Thermo Fisher Scientific Inc) according to the manufacturer´s instructions. Quality metrics were assessed with NanoDrop ™ 2000 (Thermo Fisher Scientific Inc) and Qubit ® dsDNA HS assay kit (Life Technologies). DNA extraction was performed at the School of Medicine, Department of Microbiology, Universidad Nacional de Colombia, Bogota, Colombia. Genomic DNAs were shipped directly to the Molecular and Experimental Mycobacteriology Laboratory, Research Center Borstel, Borstel, Germany, where end-to-end genotyping and next-generation sequencing were conducted.

5.4.1.2 Prokaryotic RNA

Mtbc bacteria used for the macrophage infection experiments were harvested from bacterial cultures grown in roller bottles containing 100 mL of mycobacterial media and incubated in a roller bottle system at 37°C with a rolling speed of 100 rpm. Until OD_{600} of 0.5-0.6. 30 mL were removed, and 10 mL was aliquoted into three tubes of 15 mL, followed by 20 min of centrifugation at 3500 rpm at RT. Cell harvesting and RNA extraction were carried out as described elsewhere [211] with the following modifications:

I. Cell Harvesting: *Mtbc* cell pellets were washed with PBS 1x and centrifugated 10 min at 2.200 x g at 4°C, then supernatants were discarded and incubated with 1 mL of RNAprotect Bacteria Reagent (Qiagen) for 5 min at RT and centrifugated again. After removing the supernatant, *Mtbc* cell pellets were preserved at -80°C for some days until further processing.
II. Cell Disruption: On the day of the RNA isolation, each bacterial pellet was resuspended in 1 mL TRIzol® reagent (Thermo Scientific), subjected to mechanical bead at 50s 6.5 m/s in the FastPrep™-24 instrument (MP Biomedicals), and incubated on ice for 5 min in between (repeated four times). After that, *Mtbc* cell lysates were centrifugated for 15 min at 10000 x g at 10°C; the supernatant was transferred into a new 2 mL RNase-free Tube.

Total RNA recovery was performed with Direct-zol™ RNA Miniprep and DNase treatment with DNase Max kit following manufacturer instructions[220], [221]. Purity and no DNA contamination were assessed by absorbance with nanodrop and negative amplification of the 16s gene by PCR; samples were purified with RNA clean & concentrator ™ kit (ZYMO RESEARCH) and preserved at -80 °C.

5.4.2 Mycobacterial Interspersed Repetitive Unit-Variable Number of Tandem Repeats 24 Loci Genotyping

Mycobacterial Interspersed Repetitive Unit (MIRU)- Variable Number of Tandem Repeats (VNTR) was performed as described previously [48]. Fragment analysis was performed using genemapper software version 5.0.1.3 (PE Applied Biosystems). Genotypes were identified on the web-based databank (MIRU-VNTR plus https://www.miru-vntrplus.org/MIRU/index.faces). Phylogenetic analysis was performed with BioNumerics software version 7.6 (Applied Maths). MIRU clusters comprise nodes with more than one isolate with identical genotypes. 24-Locus MIRU - VNTR dendrogram was calculated using the categorical distance and unweighted pair group method with arithmetic mean algorithm (UPGMA).

5.4.3 Next-Generation Sequencing

Short (Illumina)and long (Pacific Biosciences [PacBio]) read sequencing technologies were used.

5.4.3.1 Whole Genome Sequencing

DNA libraries were prepared based on the Baym protocol with the Nextera XT kits and sequenced with the NextSeq 500 sequencing platform from Illumina (151bp, paired-end) according to the manufacturer's instructions[222].

5.4.3.2 Single Molecule Real-Time Sequencing

A new reference genome of one outbreak local strain named COL-2 was generated using the Single Molecule Real Time (SMRT), long-read technology of the Sequel II system (PacBio] to increase the resolution of the outbreak study. The local isolate COL-2 from the largest clonal cluster was selected as the outbreak reference genome. The DNA library was prepared with the SMRTbell Express Template Prep Kit 2.0 with barcoded adapters from IDT (Integrated DNA Technologies) and sequenced on the Sequel II system following manufacturer instructions[223].

5.4.3.3 RNA Sequencing

The quantity and quality of total purified RNA were analyzed using Qubit® RNA BR Assay (Life Technologies) and DNF-472 High Sensitivity RNA Analysis 15nt kits (Advanced Analytical Technologies, Inc.) with the Qubit and Fragment Analyzer instruments. Quality metrics such as ribosomal 23s/16s picks and RNA quality numbers were determined by using the prokaryotic mode analysis of ProSize Data Analysis Software (Agilent). Samples with good RNA quality were used for library preparation: eight samples with RNA Quality Number (RQN) > 8 and one with RQN of 6.7. RNA minus stranded libraries were prepared with 500 ng of total RNA for each sample using the Zymo-Seq RiboFree ™ Total RNA Library kit (ZYMO RESEARCH) following manufacturer indications. Minus stranded libraries for the read one sequence will be antisense to the RNA transcript from which it originates. RNA libraries were sequenced using Illumina technology, eight (2 x 75 bp paired-end) and one sample in the NextSeq 500 (2x 150 bp) according to manufacturer instructions[224].

5.5 DATA ANALYSIS

5.5.1 Statistics

5.5.1.1 Outbreak Study in the Colombian Amazon Region

Descriptive statistics were calculated for demographic and clinical data of the TB cases of the outbreak. Continuous data (age) was evaluated for Gaussian distribution with the D'Agostino and Pearson test, and IQRs and media were calculated. Mann- Whitney U post hoc test was used for median age comparison and Fisher´s exact test for categorical comparison. Primary statistics were computed using the rstatix v0.6.0 package with R version 4.0.2 (https://www.r-project.org/). All p-values were two-sided and considered significant if less than 0.05.

5.5.1.2 Host-pathogen Interaction Study

All data were analyzed using GraphPad Prism version 9.0.0 (GraphPad Software Version 9, Inc). Outliers were identified by Grubb's outlier with n α-value ≤ 0.05. Statistical analysis was performed with unpaired Students' t-tests and multiple comparison tests, one-way- two-way ANOVA with Dunnet or Bonferroni post-hoc test corrections, accordingly, as described in the figure legends. Values of $*p \leq 0.05$, $**p \leq 0.01$ and $***p \leq 0.001$ were considered significant.

5.5.1.3 Flow Cytometry

Flow Cytometry Standard (FSC) data files were inspected with FACs express™ 7 and analyzed with LEGENDplex™ software and an in-house R script to quantify the molecule concentrations in each sample. The in-house R script was developed and tuned by Dr. Thomas Scholzen from the flow cytometry core facility, and it comprises the following R packages for flow cytometry data of the Bioconductor project (Table 8 and Table 9).

Table 8: Main flow cytometry packages used from the Bioconductor project.

Package	Function	Source
Flow Core	Basic structures for flow cytometry data	[225]
FlowViz	based on principles of Trellis for quality assessment and visualization	https://git.bioconductor.org/packages/flowViz
Flow Density	Sequential Flow Cytometry Data Gating. It provides tools for automated sequential gating analogous to the manual gating strategy based on the density of the data.	https://www.bioconductor.org/packages/release/bioc/html/flowDensity.html
FlowStats	Statistical methods for the analysis of flow cytometry data, such as data normalization and automated gating.	http://www.github.com/RGLab/flowStats

Table 9: Main flow cytometry R packages used from cran.r project.

Package	Function	Source
nplr	N-Parameter Logistic Regression. descriptive statistics	https://github.com/fredcommo/nplr
ggplot2	Data visualization based on grammar graphics	https://ggplot2.tidyverse.org https://github.com/tidyverse/ggplot2
qantreg	Quantile Regression. Estimation and inference methods for models for conditional quantile functions.	https://www.r-project.org/
car	Companion to Applied Regression	https://r-forge.r-project.org/projects/car
Scales	Scale function for visualization	https://scales.r-lib.org, https://github.com/r-lib/scales
gridExtra	Miscellaneous functions for Grid Graphics	https://cran.r-project.org/web/packages/gridExtra/index.html

5.5.2 Bioinformatics

5.5.2.1 Whole Genome Sequencing

WGS data was analyzed with the MTBseq pipeline using a reference mapping approach and inferring the phylogenetic lineage and resistance-associated variants[76], [133], [211] [50]. Briefly, FASTQ files were mapped to the reference genome *Mycobacterium tuberculosis* H37Rv (Gen Bank ID: NC_000962.3) using

the Burrows-Wheeler Aligner Alignment tool [226] and mapping refined with the Genome Analysis Toolkit software package [227]. For inference of the phylogeny, a concatenated SNPs alignment was built from SNPs positions with the MTBseq default thresholds: 4 reads mapped in forward and reverse orientations, a minimum of 4 reads supporting the allele with a Phred score not less than 20 and 75% allele frequency.

SNPs positions within repetitive regions and drug resistance-associated genes were excluded, and combined positions that matched these thresholds in > 95% of the isolates. The Illumina short-read sequences of the 74 genomes of the outbreak and the PacBio long-read sequence corresponding to the local reference genome COL-2 are publicly accessible in the European Nucleotide Archive under the projects PRJEB57971 and PRJEB57950. In addition, a concatenated SNPs alignment with the thresholds mentioned above but also including repetitive regions was generated. Monomorphic positions were retained from both multi-FASTA alignments with SNPs-sites [228]. Cluster analysis was performed by the MTBseq pipeline with a threshold of 5 SNPs[51]

De-novo genome assembly was performed using the PacBio SMRTlink software v9.0 and its "Microbial Assembly" application, with the genome length setting to 4.4 Mb and a seed coverage of 30. The assembly of 54,267 subreads (N50 length = 6696 bp, coverage = 73.5x) resulted in two polished contiguous sequences (contigs) 3,947,682 and 454,994 bp long, respectively. Maximum parsimony trees (MPT) were generated with BioNumerics either based on a distance matrix derived from the standard 24-loci MIRU-VNTR data or derived from the three reference mapping approaches based on a concatenated SNPs alignment with classical SNPs calling thresholds: (1) classical comparative *Mtbc* genomic pipeline, (2) including repetitive regions and drug resistance-associated regions, and (3) reference mapping to a new outbreak genome and including repetitive regions. All approaches employed the default MTBseq conservative SNPs calling parameters described above.

Low-frequency variants associated with drug resistance genes were searched in the Binary Alignment and Map (BAM) files of the whole outbreak sequenced genomes and investigate further the frequency diversity of 16 SNPs in 67 patients comprising the clonal events by using the smart tool binoSNP (a p-value of 0.05 and 0.01 were considered accordingly)[229]. Low-frequency resistance variants detected were represented with scatter plots considering the frequencies of the reference versus alternative alleles of all SNPs positions of the core-genome of the *Mtbc*. The plots were checked for clusters with the "density-based spatial clustering of applications with noise" (DBSCAN) algorithm. The frequency diversity analysis was represented with a heatmap in which the frequencies of any 16 SNPs are depicted per strain analyzed.

To identify the corresponding positions of variants of interest in the PacBio assembly, a multiple genome alignment was performed between H37Rv (L4) (GenBank ID: NC_000962.3)[230], [231], [232] and PacBio COL-2 (L4), Latin American Mediterranean [LAM]) references genomes using the mauve-plugin [233] with the Geneious Prime software. In addition, the reference-mapping-based genome assemblies calculated with the standard H37Rv and COL-2 genome for representative isolates from distinct nodes were visualized using the integrative genome viewer (IGV) [234].

Mtbc genotypes were identified on the web-based databank MIRU-VNTR*plus* (Source: https://www.miru-vntrplus.org/MIRU/index.faces). Phylogenetic analyses were performed with BioNumerics software version 7.6 (Applied Maths). MIRU clusters comprise nodes with more than one isolate with identical genotypes. A 24-loci MIRU-VNTR dendrogram was calculated using the categorical distance and the unweighted pair group method with the arithmetic mean algorithm (UPGMA).

5.5.2.2 Bulk RNA Sequencing

Mtbc RNA Seq data was analyzed with an in-house pipeline based on the reference genome approach. Briefly, FASTQ files were inspected for quality and the presence of contaminators, such as sequence adaptors and sequences that are not part of the target with RSEQC. Next, reads were aligned to *Mycobacterium tuberculosis* H37Rv (L4) (NC_000962.3) reference genome, and transcript abundance was estimated by counting the reads were aligned to the most probable gen. Read counts per condition were grouped in a feature count matrix and analyzed for differential expression analysis based on the negative binomial distribution with the R package DESeq2 of the Bioconductor project to identify genes with significant changes in the mean expression levels between three distinct conditions (L4, L6 and A4 strains). Three biological replicates were assessed per condition. P-adjusted values of ≥0.05 and a cutoff of 1.5 for $\log_2$ (fold change) were considered for the differential expression analysis. Plots for data visualization were generated with R.

6. RESULTS

6.1 Genomic Investigation of the Transmission Dynamics of a *Mycobacterium tuberculosis* complex Outbreak in an Indigenous Population in the Colombian Amazon Region.

In 2016, an outbreak of *Mycobacterium tuberculosis* complex (*Mtbc*) bacteria ravaged an indigenous population in Puerto Nariño, a region in the Colombian Amazon known for its high TB prevalence rate of 1,267/100,000 people[212]. This devastating outbreak impacted 23 settlements of the Ticuna, Cocama, and Yagua ethnicities and raised concerns among national and international public health authorities due to the resulting TB prevalence of 1,267/100,000 inhabitants[235]. However, the existing molecular epidemiology methods in the country were insufficient to fully understand the outbreak because of the low genetic diversity of the circulating strain[212], [236]. To better comprehend the outbreak's spread, 74 *Mtbc* isolates obtained from March to October 2016 were subjected to 24 MIRU-VNTR genotyping and WGS. A standard WGS approach was used to investigate the outbreak isolates, with an extended analytical range to repetitive regions, low-frequency variant detection, and mapping on a newly established "outbreak genome" to improve the understanding of the transmission dynamics in this indigenous and remote patient population[237].

6.1.1 Demographic Characteristics of the Populations Studied

Between March and October 2016, 6310 individuals were screened for TB, resulting in 80 positive cases[212]. Only 74 isolates were recovered from single patients, with 70 identifying as indigenous and residing in 16 settlements on the Loretoyaco and Amazon rivers (Figure 6 and Figure 7). The demographic and clinical characteristics of the patients are listed in Table 10 and Table S1. The mean age of the patients was 30, with a range of 2-82 years. Of the 74 patients, 45 were male, and 29 were female, with 41.9% being children. Half of the patients were from the Ticuna ethnicity, and 86.5% of them were covered by the subsidized program. Additionally, 31/74 (41.9%) had no education. Risk factors for TB included malnutrition in 26 patients and overcrowding conditions (8-14 people in a household) in 34 patients. 98.6% of the patient population had at least one symptom of TB.

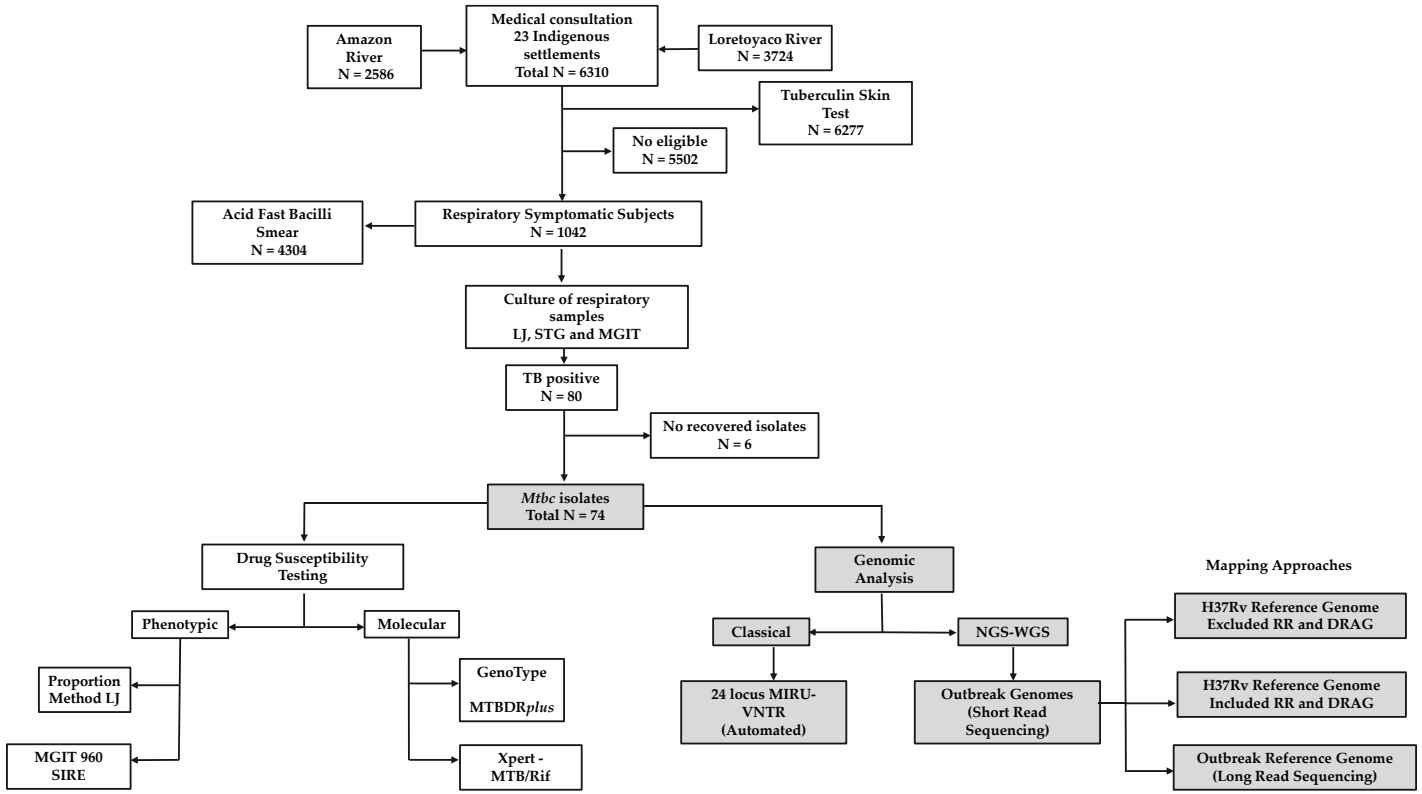


Figure 6: General workflow of the tuberculosis investigation in the remote setting of Puerto Nariño, Amazonas, Colombia.

Gray highlighted squares point to the genomic investigation of this study. TB, Tuberculosis; LJ, Löwenstein-Jensen; MGIT, mycobacterial growth indicator tube; Mtbc, *Mycobacterium tuberculosis* complex; SIRE, streptomycin isoniazid rifampicin ethambutol; MTBDR, *Mycobacterium tuberculosis* drug resistance; MIRU-VNTR, mycobacterial interspersed repetitive-unit–variable-number tandem-repeats; NGS, next-generation sequencing; WGS, whole-genome sequencing; RR, repetitive regions; DRAG, drug resistance-associated genes.

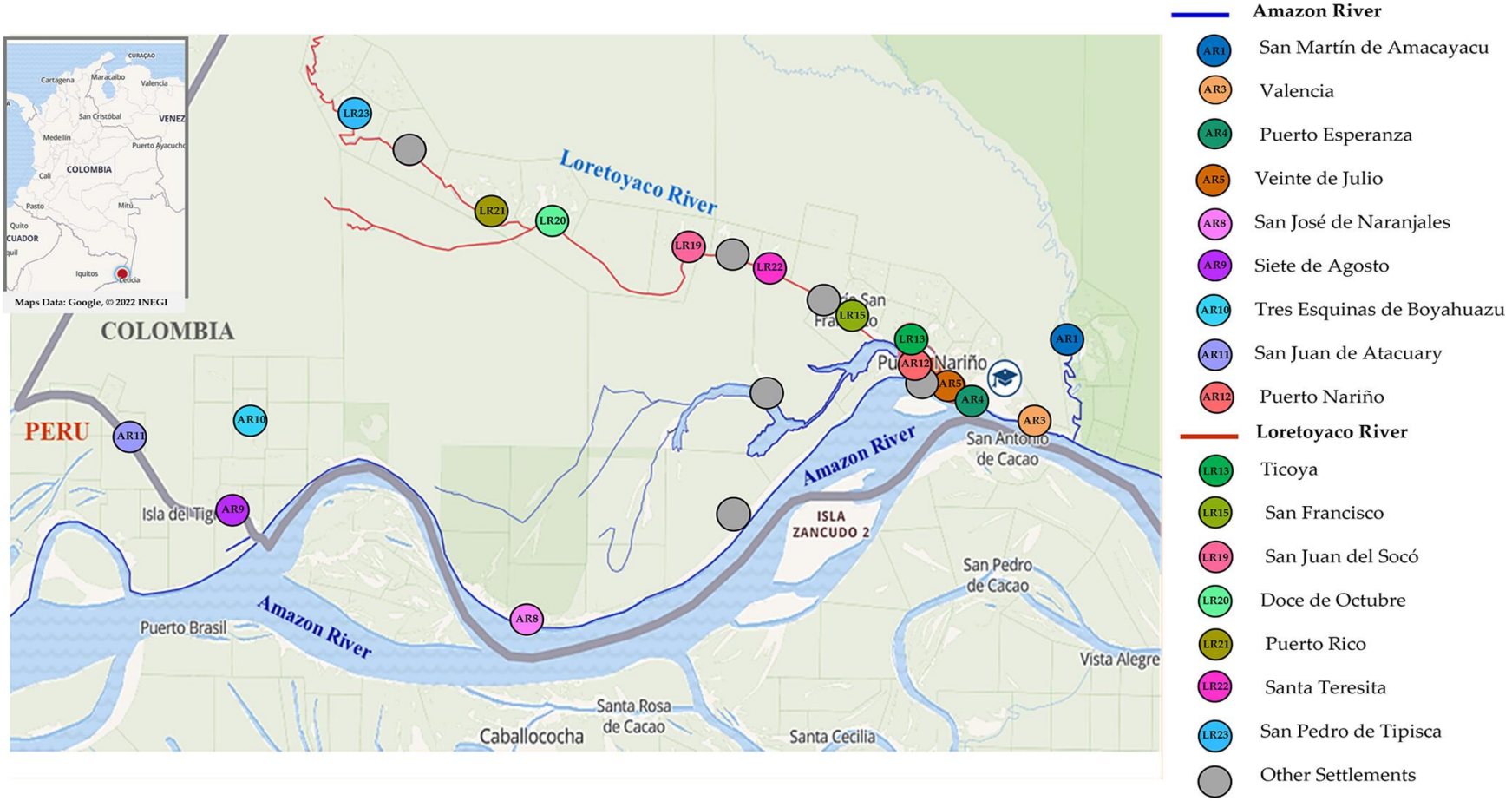


Figure 7: Geographical representation of the 16 indigenous settlements in Puerto Nariño, Amazonas, Colombia.

Colored circles indicate individual settlements on the shore of the Loretoyaco (red line) and the Amazon River (blue line). AR, Amazon River; LT, Loretoyaco River.

Table 10: Demographic and clinical characteristics of 74 patients from 16 indigenous settlements in Puerto Nariño[a].

Characteristic	No. of cases (%)[b]	Characteristic	No. of cases (%)[b]
Georeferencing		**Bacillus Calmette-Guérin Vaccination Scar**	
Amazon River	31 (41.89)	Presence	47 (63.5)
Loretoyaco River	43 (58.1)	Absence	27 (36.5)
Detection Case		**Tuberculin skin test**	
Primary Consultation	62 (84)	Positive	28 (37.8)
Contact Tracing	12 (16,2)	Negative	34 (45.9)
		Unknown	12 (16.2)
Gender			
Male	45 (60.8)	**TB risk factors and comorbidities**	
Female	29 (39.2)	HIV unknown	64 (86.5)
		Drug use	1 (1.4)
Age categories		Alcohol	13 (17.6)
Child:1-15 years	31 (41.9)	Tobacco Smoking	9 (12.2)
Youth: 16-26 years	7 (9.5)	Malnutrition[c]	26 (35.1)
Adult: 27-59 years	27 (36,4)	Other medical diagnoses	5 (6.8)
Old Adult: ≥ 60 years	9 (12,16)		
		Previous TB diagnosis and treatment	
Insurance Type		Reported	6 (8.1)
No insured	6 (8.1)	Not reported	68 (91.9)
Subsidized	64 (86.5)		
Employer-based	3 (4.1)	**Recent TB contact**	
Other	1 (1.4)	Reported	26 (35.1)
		Not reported	48 (64.9)
Schooling Level			
No education	31 (41.89)	**Symptoms of TB[e]**	
Primary	43(58,10%)	Presence	73 (98.6)
Secondary incomplete	14 (18.9)	Absence	1 (1.4) ‡
Indigenous Ethnicity		**Pulmonary aggregates**	
Ticuna	37 (50)	Presence	13 (17.6)
Cocama	14 (18.9)	Absence	61 (80)
Yagua	19 (25.7)		

No Indigenous	4 (5.4)	**Acid fast bacilli smear**	
		Positive	73 (98.8)
Overcrowding Conditions[d]		Negative	1 (1.4)
Household 2-7 people**	40 (54.1)		
Households of 8-14 people	34 (45.9)	**Drug sensitivity rifampicin-isoniazid**	
		Sensitive	73 (98.6)
		Resistance	1 (1.4)

[a]Data are depicted in the table as absolute numbers followed by the percentages (%), calculated based on the number of people for a given characteristic out of the total number of 74 TB cases with a single *Mtbc* isolate of the Puerto Nariño outbreak. Percentages were rounded and may not total 100.
[b]Percentages were calculated based on a total number of 74 patients (100%).
[c]Malnutrition was defined as a body mass index (BMI) of <18 kg/m^2.
[d]Household crowding was defined as two or more people living in the same house.
[e]TB symptoms such as coughing of any duration, hemoptysis, weight loss, thoracic pain, fever, asthenia, and others. Abnormal respiratory sounds: rhonchus, coarse crackles, bronchial breath, or rales over upper lobes or other regions from both lung fields.

6.1.2 Classical Genotyping of Latin American Mediterranean-*Mycobacterium tuberculosis* complex Outbreak

The genomic molecular epidemiological investigation began by analyzing the genotypes of 74 isolates using 24-loci MIRU-VNTR typing. The *Mtbc* population structure was found to be highly uniform (Table S2), with all 100% (74/74) isolates belonging to Euro American Lineage (also known as Lineage 4 or L4). Of these, 97.2% (72/74) were Latin American Mediterranean (LAM), and 2.7% (2/74) were Haarlem genotype. Among the LAM isolates, 43.05% (31/72) were from indigenous settlements in the Amazon, and 56.94% (41/72) were from settlements in the Loretoyaco rivers (Table S1). Two clusters were observed: One large cluster with 49 patients with identical 24-loci MIRU-VNTR genotypes (including 16 children) and a smaller cluster with 11 patients and two allele distances. These patients came from different settlements in the study region. Additionally, 10.8% (8/74) of isolates had mixed MIRU-VNTR alleles (as shown in Figure S1A-B).

6.1.3 Whole Genome Sequencing Analysis

WGS was successfully performed for all 74 *Mtbc* isolates. Phylogenetic classification based on Coll *et al.* revealed 72 LAM sublineage 4.3.3., and two Haarlem sublineage 4.1.2.1 isolates [133] (Table S1). Maximum parsimony (MPT) tree-based analysis on an SNP (single nucleotide polymorphism) distance

matrix of 890 SNPs confirmed the clonal population structure of the outbreak. Cluster analysis revealed two main molecular transmission chains of LAM 4.3.3. strains named cluster I and cluster II (Figure 8). Furthermore, two other clusters were noted, i.e., cluster III comprising 4 LAM 4.3.3. strains separated by more than 200 SNPs from clusters I/II and IV, comprised two Haarlem strains (Figure 8).

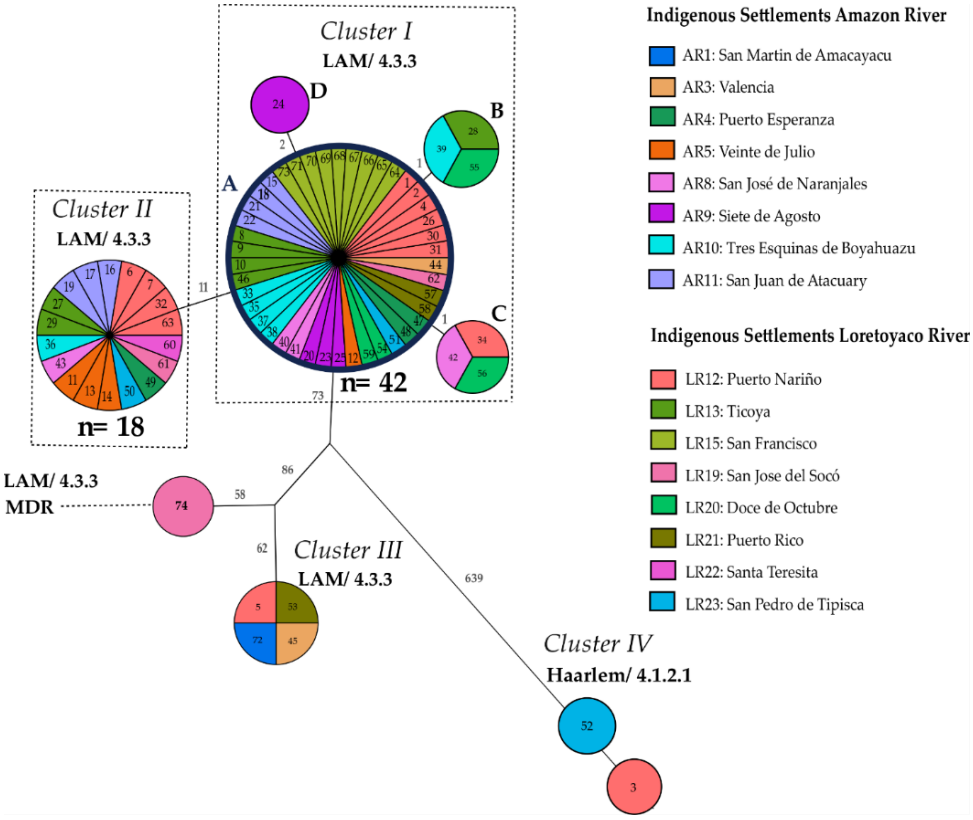


Figure 8: Genetic relationship of patient isolates based on an H37Rv reference mapping approach.
Maximum parsimony tree (MPT) is based on 890 concatenated single nucleotide polymorphisms (SNPs) of 74 *Mycobacterium tuberculosis* complex (*Mtbc*) *isolates*. The size of each node is proportional to the number of isolates. The genetic distance is indicated on branches as the number of SNPs that differ between nodes. Samples are color-coded based on the patient's settlement and identified with an Identification (ID) number. The sample ID number in the datasets is preceded by the label COL, which refers to Colombia. Outbreak clusters are named I, II, III, and IV. Clusters comprising the clonal events are shown with dashed boxes, and nodes of cluster I are termed A, B, C, and D. LAM, Latin American Mediterranean; MDR, Multidrug-resistant; AR, Amazon River; LR, Loretoyaco River.

Cluster I contained 42 identical (42/74, 56.8%) isolates and 7 isolates differing by a maximum of 2 SNPs (Figure 8). Of these 49 patients, 20 were children from distinct settlements and river locations (Table S1 and Figure S2). Cluster II comprised 18 identical (18/74, 24.3%) isolates, with 11 SNPs separating both

clusters, suggesting a relatively recent common ancestor; however, the clear separation by 11 SNPs indicates two independent short-term transmission chains in the study region.

Molecular drug resistance prediction detected multidrug resistance in one LAM/4.3.3 isolate from a child (COL-74). Although this case was identified by contact tracing, it was not in a molecular cluster (Figure 8). The resistance-mediating mutations were *rpoB* Ser450Leu (99.64%), *katG* Ser315Thr (100%), *pncA* His82Asp (99.68%), *embB* Met306lle (100%), and *rpsL* Lys88Thr (100%) conferring resistance to rifampicin, isoniazid, pyrazinamide, ethambutol, and streptomycin, respectively (Figure 8). In addition, resistance-associated SNPs with a low allele frequency were found in two isolates/patients (LAM 4.3.3 /COL-18 and Haarlem 4.1.2.1/ COL-3), i.e**.,** *rpoB* Ser431Thr (6.62%), *rpsA* Thr210Ala (8.21%) in COL-18, and *rpsA* Thr5Ala (10.57%) in COL-3.

6.1.4 Extended Whole Genome Sequencing Analysis of Latin American Mediterranean-*Mycobacterium tuberculosis* complex Outbreak

Single nucleotide polymorphisms in repetitive and resistance-associated genes were included to increase the genotyping resolution of the outbreak. This approach led to a subdivision of the largest cluster I (Figure 9). Overall, the number of SNPs included in the concatenated SNP alignment rose from 890 to 1062 SNPs, increasing in the number of nodes in the MPT from 9 to 12 (Table 11, Figure 9). Cluster I was split further into three new nodes: one isolate at 2 SNPs distance (Node E), two isolates at 4 SNPs distance (Node F), and 12 isolates (Node D) at 5 SNPs distance towards parental node A (Figure 9). The 12 isolates within node G were characterized by 5 novel SNPs (Table 12). *Rv1194c* Trp244Arg, *Rv3621c* (*PPE65*) Gln253Arg, *Rv3478* (*PPE60*), *Rv3478* (*PPE60*) Ala326Ala, and *Rv1945* Ala327Ala.

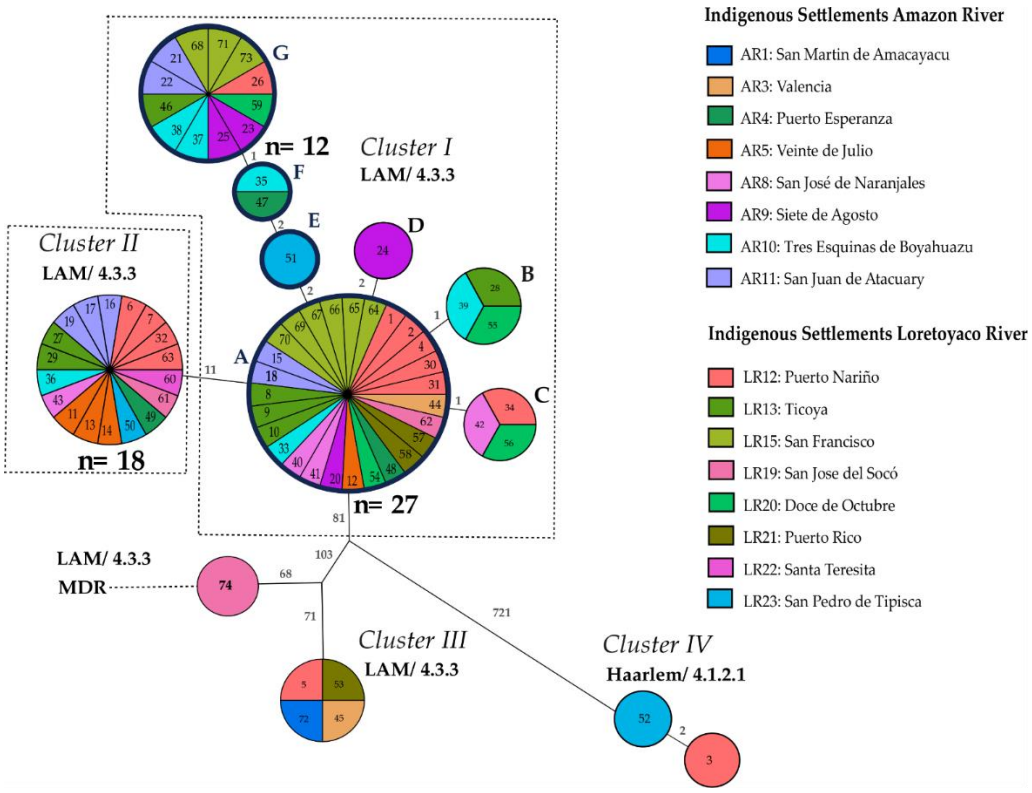


Figure 9: Genetic relationship of patient isolates based on an H37Rv reference mapping approach, including repetitive regions and drug resistance-associated genes.

Maximum parsimony tree (MPT) is based on 1062 single nucleotide polymorphisms (SNPs) of 74 *Mycobacterium tuberculosis* complex (*Mtbc*) isolates. The genetic distance is indicated on branches as the number of SNPs that differ between nodes. Samples are color-coded based on the patient's settlement with an identification (ID) number. The sample ID number in the datasets is preceded by the label COL, which refers to Colombia. Outbreak clusters are named I, II, III, and IV. Clusters comprising the clonal events are shown with dashed boxes, and nodes of cluster I are termed A, B, C, and D. New nodes derived from Node A are termed E, F, and G. LAM, Latin American Mediterranean; MDR, Multidrug-resistant; AR, Amazon River; LR, Loretoyaco River.

Table 11: Details of the reference mapping approach.

Mapping Approach	Standard	Including RR And DRAG	New Outbreak Genome [a]
Reference Genome	H37Rv	H37Rv	LAM/4.3.3 (COL-2)
GeneBank accession no.	NC_000962.3	NC_000962.3	ERR10558090
Reference Genome Size (bp)	4,411,532	4,411,532	4,402,676
GC Content (x)	65.61	65.61	65.61
No. of CDS	4,069	4,069	4,067
No. of differentiating SNPs	890	1062	1094
No. outbreak nodes in the MPT	5	8	9
No. of branches differentiating the outbreak	4	7	9

[a]Including repetitive regions and resistance genes.

LAM, Latin American Mediterranean; RR, Repetitive Regions; DRAG, Drug Resistance Associated Genes; bp, bases pairs; GC, Guanine Cytosine; CDS, Coding Sequences based on Prokka annotation; SNPs, Single Nucleotide Polymorphisms; MPT, Maximum Parsimony Tree

Table 12: Description of 16 mutations differentiating the clonal clusters of Latin American Mediterranean /4.3.3 outbreak.

SNP No.	Genomic Position	R	A	*Locus*	Gene Name	Annotation	Amino acid Change (codon change)	LAM Cluster
1	1337784	A	G	*Rv1194c*	-	hypothetical protein	W244R (tgg/Cgg)	I
2	4061132	T	C	*Rv3621c*	*PPE65*	PPE family protein PPE65	Q253R (cag/cGg)	I
3	3895400	A	G	*Rv3478*	*PPE60*	PE family protein PPE60	P325P (cca/ccG)	I
4	3895403	A	C	*Rv3478*	*PPE60*	PE family protein PPE60	A326A (gca/gcC)	I
5	2196969	G	C	*Rv1945*	-	hypothetical protein	A327A (gcg/gcC)	I
6	78383	T	C	*Rv0070c*	*glyA2*	Serine hydroxymethyltransferase GlyA2 (serine methylase 2) (SHMT 2)	I172V (atc/Gtc)	II
7	413981	T	C	*Rv0343*	*iniC*	Isoniazid inductible gene protein IniC	S409P (tcg/Ccg)	II

8	637055	C	T	*Rv0545c*	*pitA*	Probable low-affinity inorganic phosphate transporter integral membrane protein PitA	V137M (gtg/Atg)	II
9	758066	G	C	*Rv0663*	*atsD*	Possible arylsulfatase AtsD (aryl-sulfate sulphohydrolase) (arylsulphatase)	G644R (ggc/Cgc)	II
10	960798	C	G	*Rv0862c*	-	hypothetical protein	A605A (gcg/gcC)	II
11	1062381	C	T	*Rv0951*	*sucC*	Probable succinyl-CoA synthetase (beta chain) SucC (SCS-beta)	P140S (ccg/Tcg)	II
12	1158574	C	T	*Rv1033c*	*trcR*	Two components of transcriptional regulator TrcR	G55S (ggc/Agc)	II
13	1290447	G	C	*Rv1161*	*narG*	Respiratory nitrate reductase (alpha chain) NarG	M1040I (atg/atC)	II
14	1403781	C	T	*Rv1256c*	*cyp130*	Probable cytochrome P450 130 Cyp130	G275S (ggc/Agc)	II
15	2229203	C	T	*Rv1985c*	-	Probable transcriptional regulatory protein (probably LysR-family)	G234R (ggg/Agg)	II
16	3461997	G	C	*Rv3093c*	-	Hypothetical oxidoreductase	S256S (tcc/tcG)	II

SNP, Single Nucleotide Polymorphism; R, Reference Allele; A, Alternative Allele; A, Adenine; G, Guanine; C, Cytosine; T, Thymine; LAM, Latin American Mediterranean; PE, Proline-Glutamic acid; PPE, Proline-Proline-Glutamic acid

Next, the outbreak-specific genome from isolate COL-2 *de-novo* was assembled and used as a reference genome comprising genomic regions absent in the H37Rv genome. The COL-2 genome comprised 4,402,976 nucleotides (two contigs, 3,947,682 and 454,994 bp) and was slightly smaller than the H37Rv reference genome with 4,411,532 nucleotides (Table 11). The gene content between COL-2 and H37Rv was highly similar; however, 6 disrupted and 10 newly annotated coding sequences were observed due to larger insertions and deletions (Table 11 and Table S3). Employing the COL-2 genome as a reference for the mapping approach resulted in a further increase in the number of differentiating SNPs in the concatenated sequence from 1,062 to 1,094, which also resulted in an increase in outbreak nodes in the

MPT from 8 to 9 (Table 11). The overall population structure in the MPT remained consistent, and node F was further differentiated into F1 and F2, resulting in two new branches in the MPT (Figure 10).

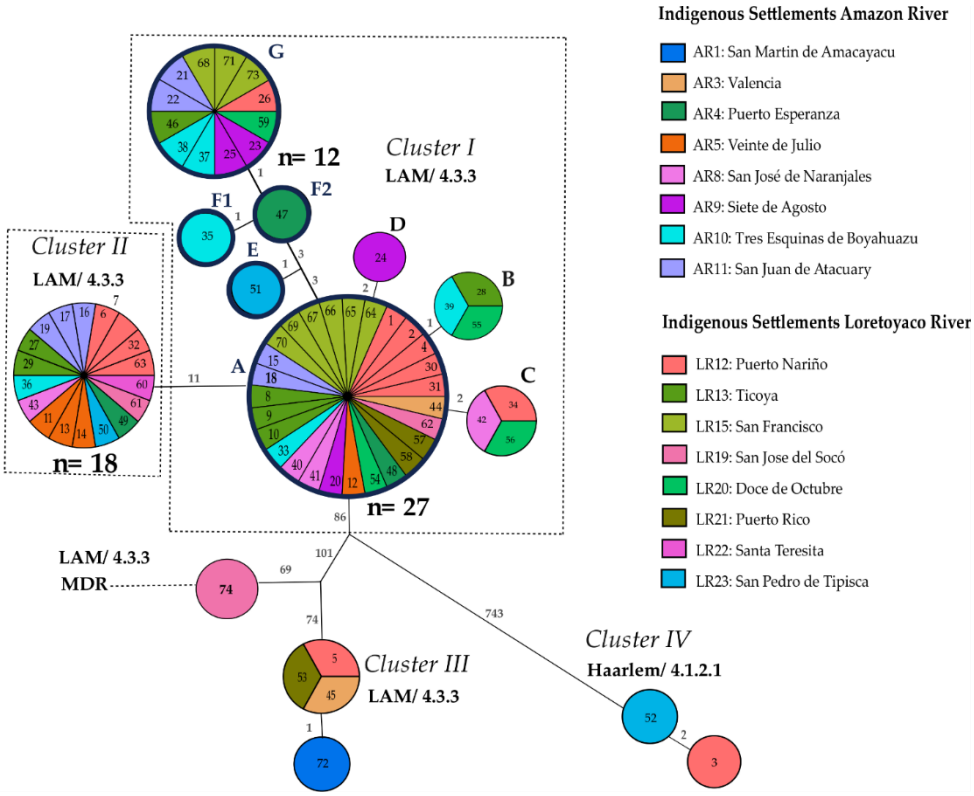


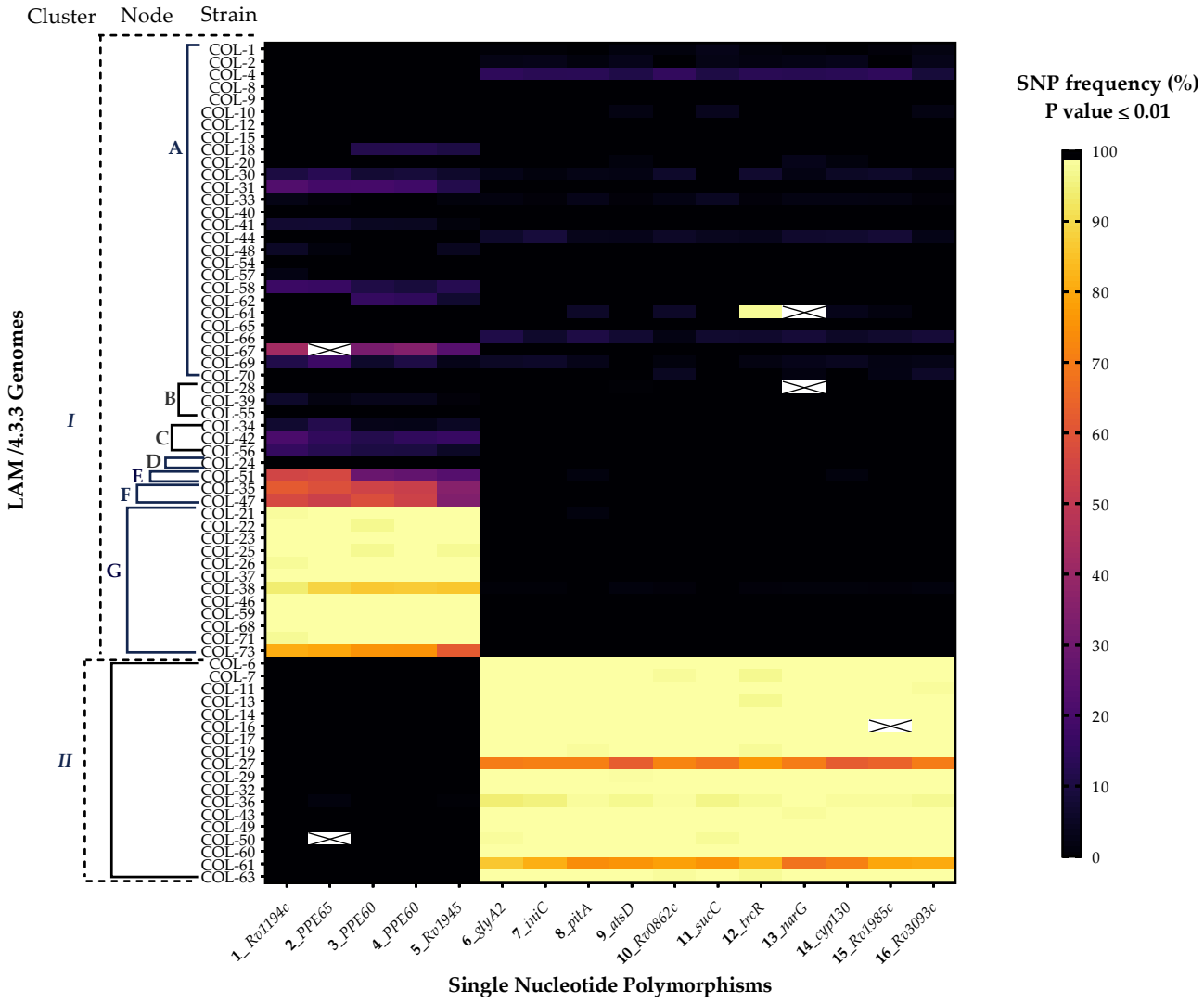
Figure 10: Genetic relationship of patient isolates based on a mapping approach including repetitive regions, drug resistance-associated genes, and a newly assembled genome from the outbreak strain COL-2 as reference.

Maximum parsimony tree (MPT) is based on 1094 single nucleotide polymorphisms (SNPs) of 74 *Mycobacterium tuberculosis* complex (*Mtbc*) isolates. The size of each node is proportional to the number of isolates. The genetic distance is indicated on branches as the number of SNPs that differ between nodes. Samples are color-coded based on the patient's settlement with an Identification (ID) number. The sample ID number in the dataset is preceded by the label COL, which refers to Colombia. Outbreak clusters are named I, II, III, and IV. Clusters comprising the clonal events are shown with dashed boxes, and their nodes are termed A, B, C, and D. New nodes derived from Node A are termed E, F, and G. The partition of node F into F1 and F2 is shown. LAM, Latin American Mediterranean; MDR, Multidrug-resistant; AR, Amazon River; LR, Loretoyaco River.

Lastly, potential mixed infections were investigated in all outbreak isolates (n=67) based on the allele frequencies of 16 SNPs, differentiating especially Cluster I, node G (5 SNPs), and Cluster II (11 SNPs) isolates (Figure 11). Overall, 22/67 isolates were considered clonal infections, i.e., heterogeneous alleles were not found in any of the 16 genome positions (Table S4). However, in 20/67 isolates, either all Cluster II specific SNPs or all Cluster I, node G specific SNPs were identified at a lower frequency (<100%), suggesting a mixed infection with at least two clones of the outbreak. In the remaining 25/67 isolates, heterogeneous alleles were present in some node-specific positions.

Figure 11: Heat map representing the interpatient diversity of 16 outbreak-differentiating variants.
The y-axis shows 67 Latin American Mediterranean (LAM) /4.3.3 genomes comprising the clonal outbreak; their corresponding nodes are indicated with solid and dashed brackets, respectively. The x-axis shows 16 genes comprising single nucleotide polymorphisms (SNPs) differentiating nodes E, D, and G in cluster I (5 SNP) and cluster II (11 SNP) (see Figure 9). SNP frequency is color-coded from a continuous dark violet to yellow. SNPs were called in confidence positions with a p-value of ≤0.01 considering an H37Rv reference mapping. Empty crossed cells represent individual SNPs that could not be identified in the Illumina short-read data due to a possible deletion (gaps in the alignment of the respective genome positions). MDR, multidrug resistant; AR, Amazon River; LR, Loretoyaco River.

Figure 11 (Continued)



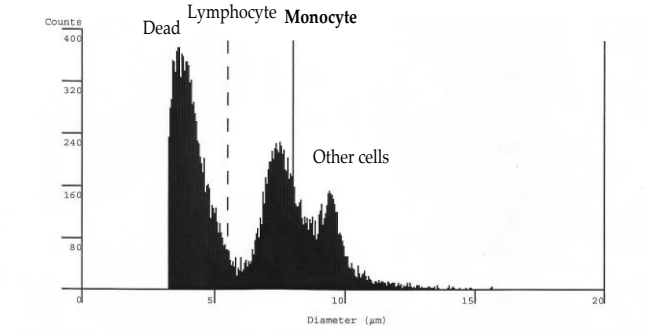
6.2 Establishment of an Infection Model to Study Disease Susceptibility

Although murine and cancer cell lines as infection models have provided valuable insights into understanding TB disease and offer experimental advantages, they do not fully capture the human response to the disease[238], [239]. That is especially important given the unique global distribution, host preference, and phylogeographical population structure of the *Mtbc* pathogen [127]. In the following, I will describe an optimized infection cell model using blood monocyte-derived macrophages (MDMs) from human donors of European and African ancestry and bovine donors.

6.2.1 Generation of *in vitro* Blood Monocyte-Derived Macrophages

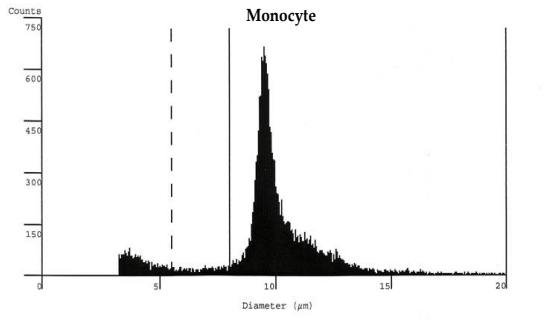
Human (hM) and bovine (bM) monocytes were isolated from ~ 300 mL of whole blood heparin-anti-coagulated by obtaining the PBMCs fraction with the density gradient centrifugation method. Monocytes had a cell yield between 50-120x10^6 cell counts with a viability of 85-98% for humans and 20-70 x10^6 cell counts with a viability of 84-92% for bovine; both cell types had a cell diameter of 8-9.99 µm and a purity of 97.42% and 85% CD14+ (Figure 12 and Figure S3) respectively. Next, monocytes were seeded, cultivated, and differentiated into MDMs for seven (bovine) and six (human) days, respectively (see 5.3.2 *In vitro* Generation of Macrophages in 5. METHODS). MDMs showed the typical morphological characteristics of macrophages under light microscopy visualization (e.g., spindle, and extended shape and cell confluency over 80%), and the expression of the pan-macrophage marker CD11b and the mannose receptor CD206 (human, 99.61% and bovine,92.83%) (Figure 13).

A



Capillary	150 µm	Conc. Range	Ok	Volume/ml	2.225e+09 fl
X-Axis Scale	20 µm	Counts	5573		
Eval Cur. left	8.00 µm	Counts> 20µm/ml	5.223e+03		
Eval Cur. right	20.00 µm	Aggr. Correct.	Off	Mean Diameter	9.40 µm
Norm Cur. left	5.50 µm	Aggr. Factor	Off	Mean Volume	4.586e+02 fl
Norm Cur. right	20.00 µm	Viable Cells/ml	4.851e+06	Peak Diameter	8.03 µm
Sample Volume	3 x 400 µl	Total Cells/ml	9.896e+06	Peak Volume	2.706e+02 fl
Dilution	1.000e+03	% Viability	49.02 %	Interface	Parallel
%Calculation	Viability	Cell Debris/ml	8.863e+06	Print Mode	Manual

B



Capillary	150 µm	Conc. Range	Ok	Volume/ml	1.042e+10 fl
X-Axis Scale	20 µm	Counts	17941		
Eval Cur. left	8.00 µm	Counts> 20µm/ml	1.349e+05		
Eval Cur. right	20.00 µm	Aggr. Correct.	Off	Mean Diameter	10.51 µm
Norm Cur. left	5.50 µm	Aggr. Factor	Off	Mean Volume	6.671e+02 fl
Norm Cur. right	20.00 µm	Viable Cells/ml	1.562e+07	Peak Diameter	9.54 µm
Sample Volume	3 x 400 µl	Total Cells/ml	1.646e+07	Peak Volume	4.546e+02 fl
Dilution	1.000e+03	% Viability	94.89 %	Interface	Parallel
%Calculation	Viability	Cell Debris/ml	1.714e+06	Print Mode	Manual

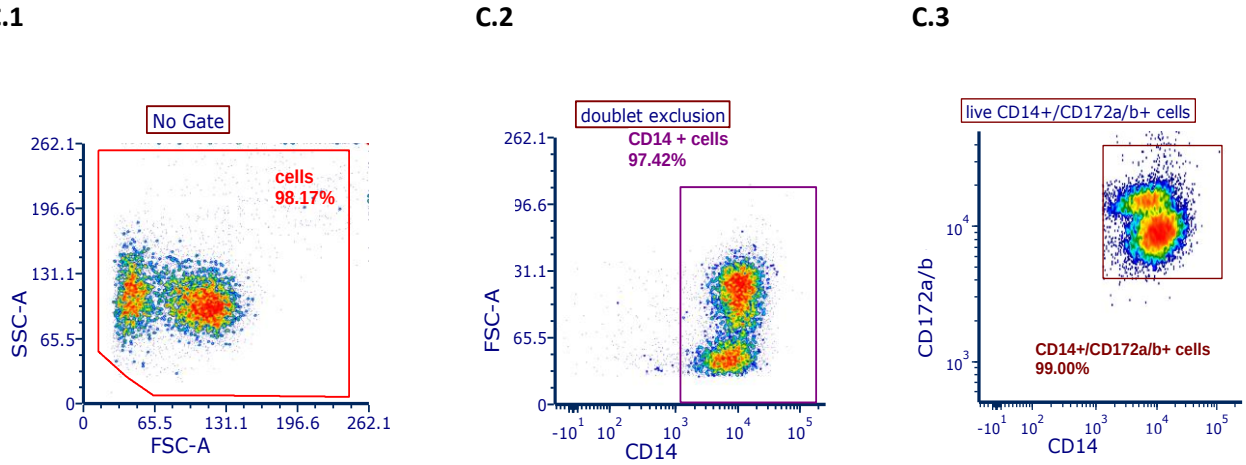


Figure 12: Cell parameters of the human cells isolated.

A Human blood peripheral blood mononuclear cells (PBMCs) before magnetic isolation. **B** CD14 positive selected cells after magnetic isolation from a representative experiment. Cell counts (y-axis) are plotted against cell diameter (x-axis). Cell parameters results are depicted below the graphs and were calculated with electronic current exclusion technology with the CASY instrument (capillary of 150µm). The results of cell parameters are depicted below the graphs. Cell counts are given per mL (1:2 dilution). Monocyte cells have a diameter of 8-9.99 µm. Viable and dying/dead cells are detected based on their volume. Viable cells with a polarized membrane generate a high resistance signal, which correlates with the whole volume of the cell (> 5 µm), while dying/ dead cells generate a much lower resistance due to depolarized or completely disrupted membrane (5 µm). **C.1**-**C.3** Immunophenotyping profiling of CD14+ isolated Human Monocytes. **C.1** Side Scatter (SSC) and FSC (Forward Scatter) patterns of the monocyte population. **C.2** Gating strategy doublet exclusion showing the percentage of cells expressing the monocyte differentiation antigen CD14+. **C.3** Monocytes CD14+ were further gated on the expression of CD172a/b. Unstained and stained controls were included (not shown). CD, Cluster of Differentiation.

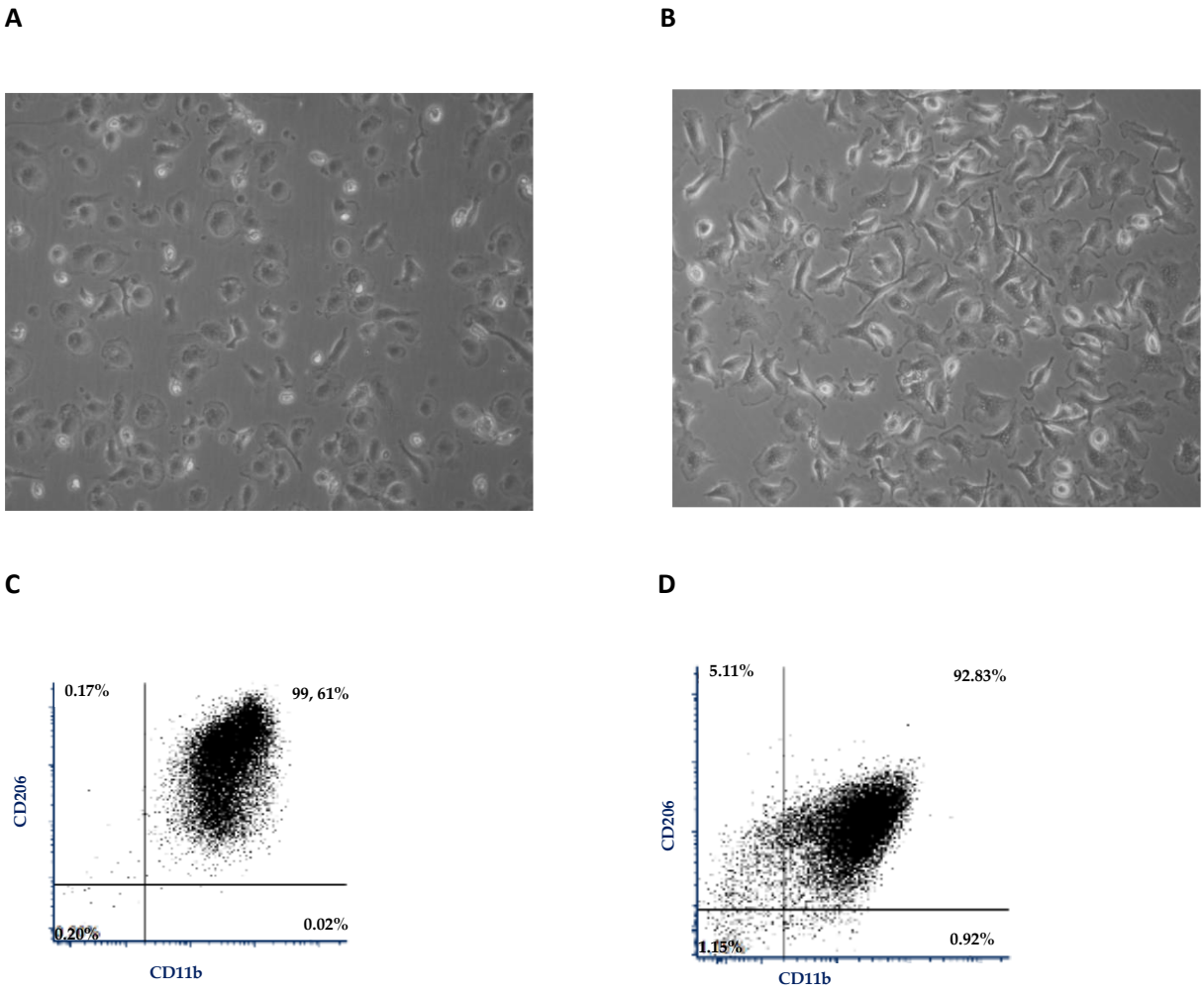


Figure 13: Characteristics of blood monocyte-derived macrophages cultured.
Light microscopy images magnified 20x of human (**A**) and bovine (**B**) monocyte-derived macrophages (MDMs). Large cells with extensive and diffuse cytoplasm, eccentric nucleus, and islands of aggregated and proliferating cells are observed. Immunophenotyping analyzed data of Human (**C**) and Bovine (**D**) MDMs with CD11b and CD206 represented in contour plots. Cells double positive for CD11b, and CD206 was classed macrophages. One representative experiment out of three is shown, fluorescence minus one, unstained and stained controls were included (no shown). CD, Cluster of Differentiation.

6.3 Virulence Profiling of *Mycobacterium tuberculosis* Complex Lineages in Human and Bovine Infection Cell Models

The common ancestral geographical origin between a host and a pathogen is known as sympatric host-pathogen relationships[127], [240]. These relationships seem to play a critical role in the outcome of TB infection, given the global success of the modern human generalist L4 strains among human populations and the restricted success of the specialist ancient L6 strains in populations of West Africa despite the

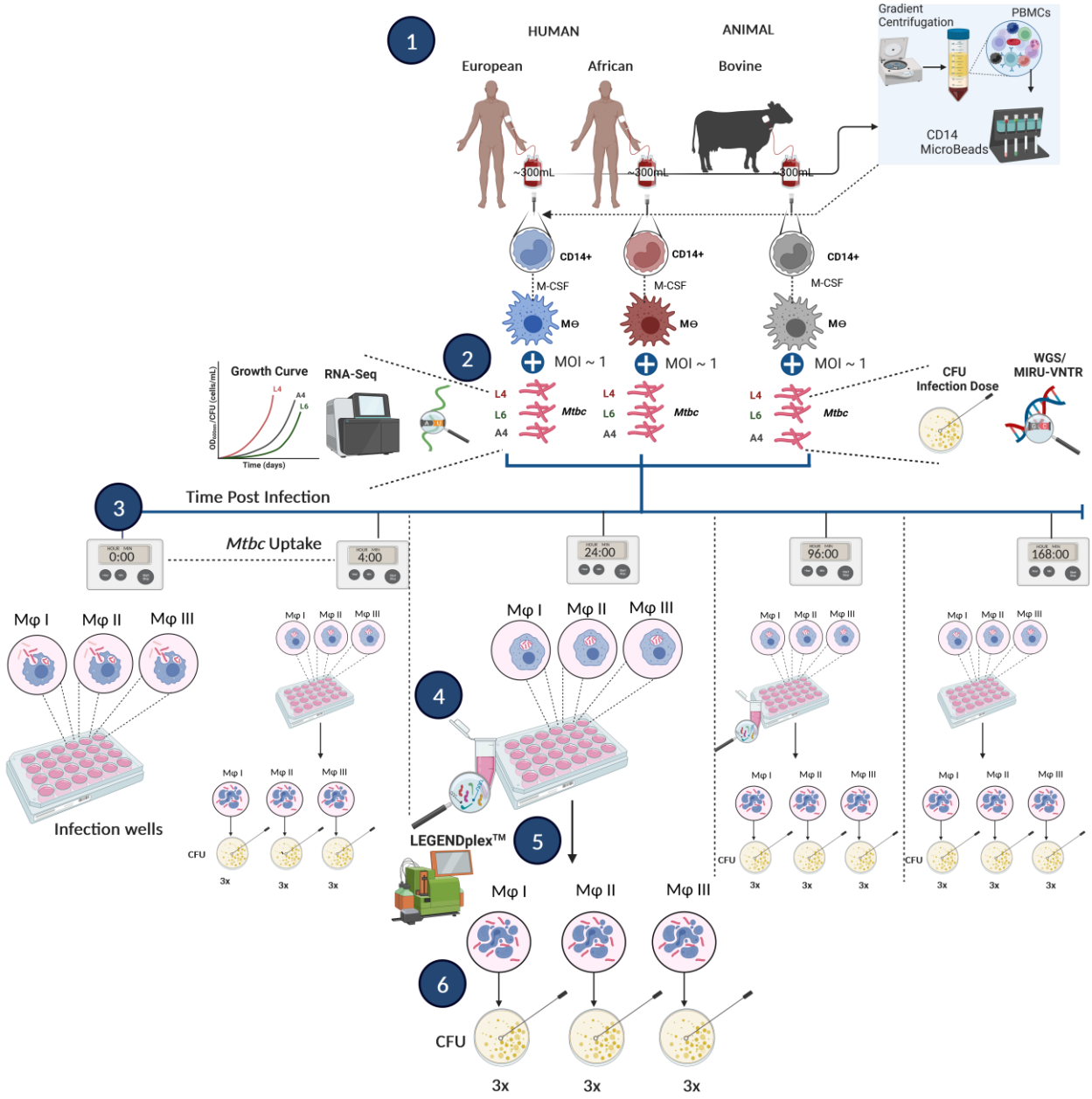
mass transatlantic migration[3], [121], [127], [128], [130], [241], [242], [243]. Furthermore, *Mbv* strains from A4 seem to play a similar dynamic by causing a more successful infection in the bovine than the human hosts[166]. To investigate this co-adaptation of the pathogen to the host, three distinct *Mtbc* strains from the human-adapted generalist L4 and specialists L6 and A4 (*Mbv)* strains were investigated in cells derived from three different host types. Each lineage included clinical and ATCC reference strains (

Figure 14).

Figure 14: Experimental design of host (macrophage)-pathogen (*Mycobacterium tuberculosis* complex strains) interaction.

General workflow: **1** Total blood (~300 mL) was obtained from human and bovine healthy donors. Human donors comprised African (Nigerian, Cameroonian, and Ghanaian) and European (German) ancestries. $CD14^+$ monocytes were obtained by magnetic separation with beads conjugated to monoclonal antihuman CD14 antibodies and further differentiated into macrophages in the presence of Macrophage Colony-Stimulating Factor for 6-7 days. **2** Monocyte-derived macrophages (MDMs) were infected with a panel of *Mycobacterium tuberculosis* complex (*Mtbc*) strains of L4, A4, and L6 previously genotyped and colony-forming units (CFU) titrated. These strains were assessed for growth kinetics and transcriptomics profiling. **3** Macrophages were incubated with *Mtbc* strains for 4 h, 24 h, 96 h, and 168 h. **4** Cell culture supernatants were harvested from human MDMs infections at 24 hpi and 96 hpi and stored for cytokine analysis. **5** Infected MDMs were lysed, and intracellular growth was assessed by CFU assay (**6**). One independent experiment comprised four plates of 24 wells for the four infection time points evaluated per donor; each plate comprised control wells with non-infected macrophages and experimental wells with macrophages infected with six or four strains of a given lineage, depending on the cell yield obtained. The reference strain H37Rv ATCC (9679/00) was used as a control of infection. Three infection macrophage well replicates were assessed per strain, and CFU measurements were performed per well for a total of 9 CFU replicates per strain and time points were assessed. Figures were elaborated with BioRender. MΦ, Macrophage; L4, Lineage 4; L6, Lineage 6; A4, Animal clade 4; MOI, Multiplicity of Infection, WGS, Whole Genome Sequencing; MIRU-VNTR, Mycobacterial Interspersed Repetitive Units-Variable-Number Tandem Repeats.

Figure 14 (Continued)



6.3.1 Growth Profile of *Mycobacterium tuberculosis* Complex Strains in Liquid Media

Strains of *Mycobacterium tuberculosis* belonging to L4, L6, and A4 lineages were grown in liquid 7H9 Middlebrook to assess their growth rate over 10 days (Figure 15A-B and Figure 16A-B). Two independent growth experiments were performed. L4 strains grew higher in comparison with L6 strains and A4 strains towards the end of 5 days and 7 days ($p = <0.001$), respectively (Figure 15A-B and Figure 16A-B). L6 and A4 strains showed similar kinetics upon growth (ns, $p = >0.05$). Based on classical Ziehl-Neelsen staining, bacterial density evaluation revealed lineage-specific differences (Figure 15C). The maximum number of colonies forming units (C_{max}) and the time required to reach C_{max} were calculated (T_{max}). All strains reached T_{max} at 10 days, from which the LAM clinical 7968/03 strain showed the highest C_{max} with $1.57x10^9 \pm 0.28x10^9$ at 10 days (Table 13). In general, L4 strains reached the maximum number of CFUs compared to L6 and A4 strains at 10 ($p = <0.05$, Figure 15D and Figure 16C). No significant differences were found among the C_{max} values of L6 and A4 strains at 10 days (Figure 16C $p= >0.05$).

Table 13: Growth profile of *Mycobacterium tuberculosis* complex strains in liquid culture.
Comparison of the maximum number of colony-forming units reached at the maximum time.

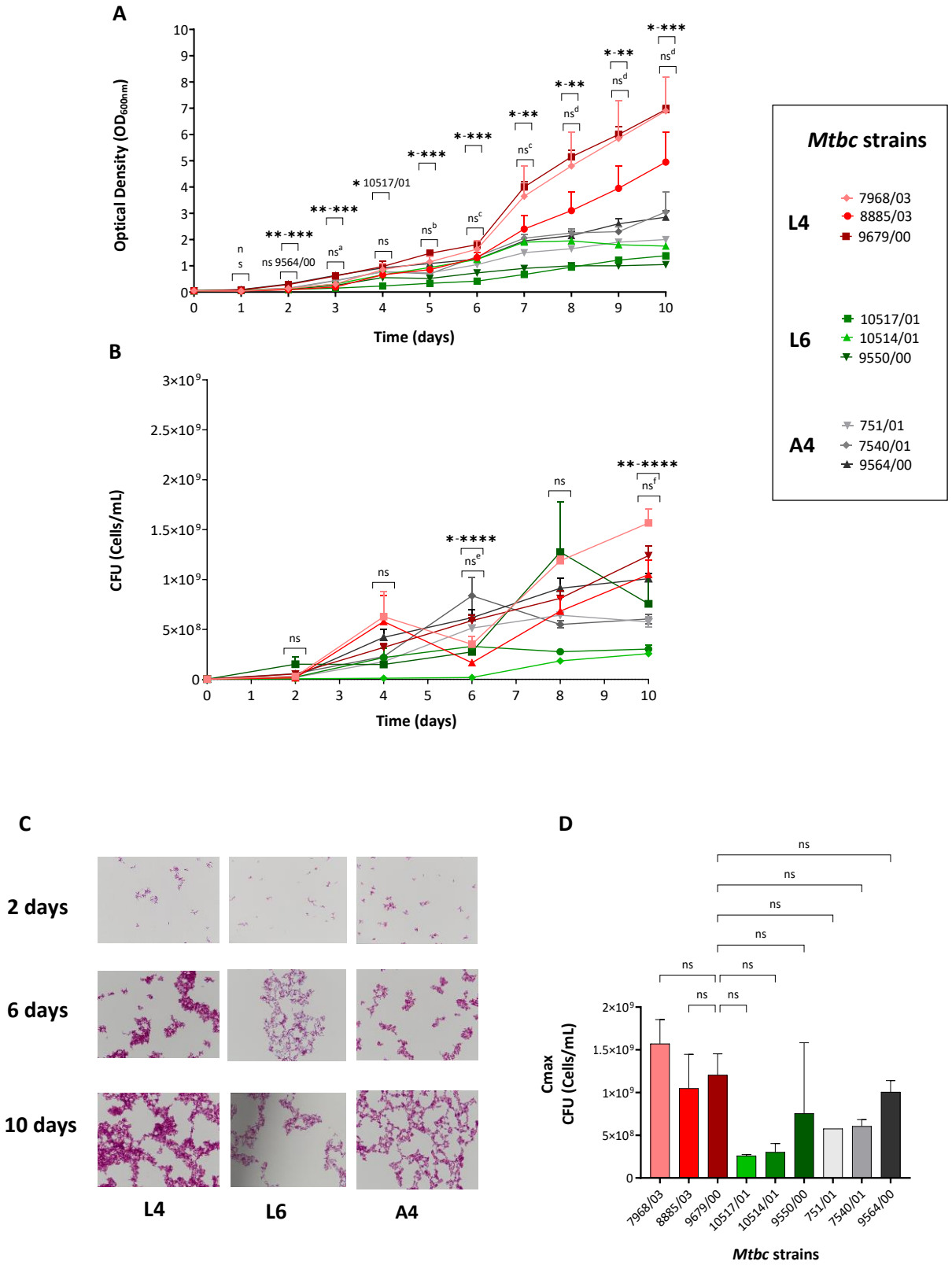
Specie	Lineage Strain	Strain ID	C_{max} (CFU/mL)[a]	T_{max} (day)
M. tuberculosis	Lineage 4 LAM	7968/03	$1.57 \times 10^9 \pm 0.28x10^9$	10
M. tuberculosis	Lineage 4 LAM	8885/03	$1.04 \times 10^9 \pm 0.39x10^9$	10
M. tuberculosis ATCC	Lineage 4 H37Rv	9679/00	$1.20 x10^9 \pm 0.24x10^9$	10
M. africanum	Lineage 6 WAII	10517/01	$2.60 x10^8 \pm 0.14x10^8$	10
M. africanum	Lineage 6 WAII	10514/01	$3.03x10^8 \pm 0.98x10^8$	10
***M. africanum* ATCC**	Lineage 6 WAII	9550/00	$7.55x10^8 \pm 8.27x10^8$	10
M. bovis	Animal Clade 4	751/01	$5.77x10^8 \pm 0.00$	10
M. bovis	Animal Clade 4	7540/01	$6.05x10^8 \pm 0.77x10^8$	10
***M. bovis* ATCC**	Animal Clade 4	9564/00	$1.00x10^9 \pm 0.13x10^9$	10

[a]Data represents mean ± SD obtained from two independent growth curves.

C_{max}, Maximum number of colony-forming units; T_{max}, Time required to reach C_{max}; ATCC, American Typing Culture Collection; LAM, Latin American Mediterranean; WAII, West African Two; SD, Standard Deviation.

Figure 15: Growth profiles of *Mycobacterium tuberculosis* complex strains in liquid media.
A Optical density (OD_{600nm}) **B** Colony-forming units(cells/mL). Mean, standard deviation and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are shown. Statistical analysis for OD_{600nm} and CFU are based on One-way ANOVA with Bonferroni post hoc test correction (n= 2 biological replicates, n=6 technical replicates). **C** Representative Ziehl-Neelsen staining of three reference strains at 2, 6, and 10 days of growth. **D** Maximum Colony-forming units number reached during the growth period. No significant comparison results are depicted in Figure **A**: ns^a = 9679/00 Vs. 9564/00, 751/01, 7540/01; ns^b = 9679/00 Vs. 7968/03, 10514/01, 9564/00; ns^c = 9679/00 Vs. 7968/03, 8885/03, 10514/01,7540/01, 9564/00; ns^d = 9679/00 Vs. 7968/03, 8885/03; Not significant results of comparisons depicted in figure **B**: ns^e = 9679/00 Vs. 7968/03, 10514/01, 9564/00, 751/01, 7540/01; ns^f = 9679/00 Vs. 7968/03, 8885/03;9550/00; 9564/00. ns, not significant results (p > 0.05); OD, Optical Density; CFU, Colony-forming units; C_{max}, the maximum number of colonies forming units reached at the maximum point during growth; L4, Lineage 4; L6, Lineage 6 and A4, Animal Clade 4.

Figure 15 (Continued)



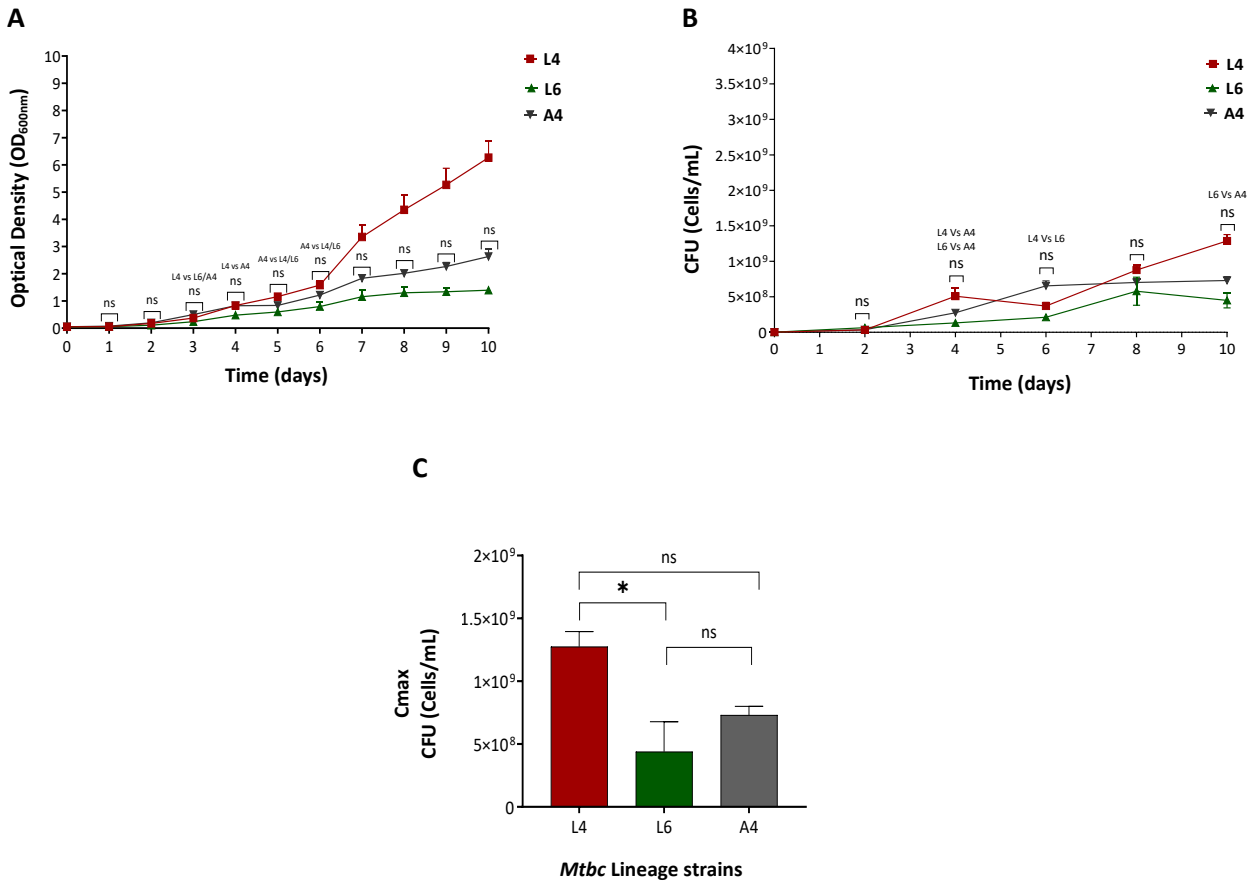


Figure 16: Growth kinetics of the *Mycobacterium tuberculosis* complex strains per lineage.

A Optical density (OD_{600nm}). **B** Colony-forming units (cells/mL). **C** Comparison of C_{max} values among lineages strains. OD and CFU values comprise the mean of three strains for L4, L6, and A4. Growth curves represent the mean of two independent experiments. *Mycobacterium tuberculosis* complex (*Mtbc*) strains were cultured at 0.05 OD_{600nm} in roller bottles containing 7H9 broth for 10 days. OD measurements were taken every day until 10 days, and Three CFU technical measurement replicates were performed at 0, 2, 4, 6, and 10 days. Statistical results between CFU/OD values among lineages were calculated with One-way ANOVA with Bonferroni post hoc test correction. Mean, standard deviation and only not significant statistical results (ns >0.05) are depicted in Figures **A** and **B**. Significant and Not significant results are depicted in Figure **C** results (ns>0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001). OD, Optical Density; CFU, Colony-forming units; C_{max}, the maximum number of colonies forming units reached at the maximum point during growth; L4, Lineage 4; L6, Lineage 6 and A4, Animal Clade 4.

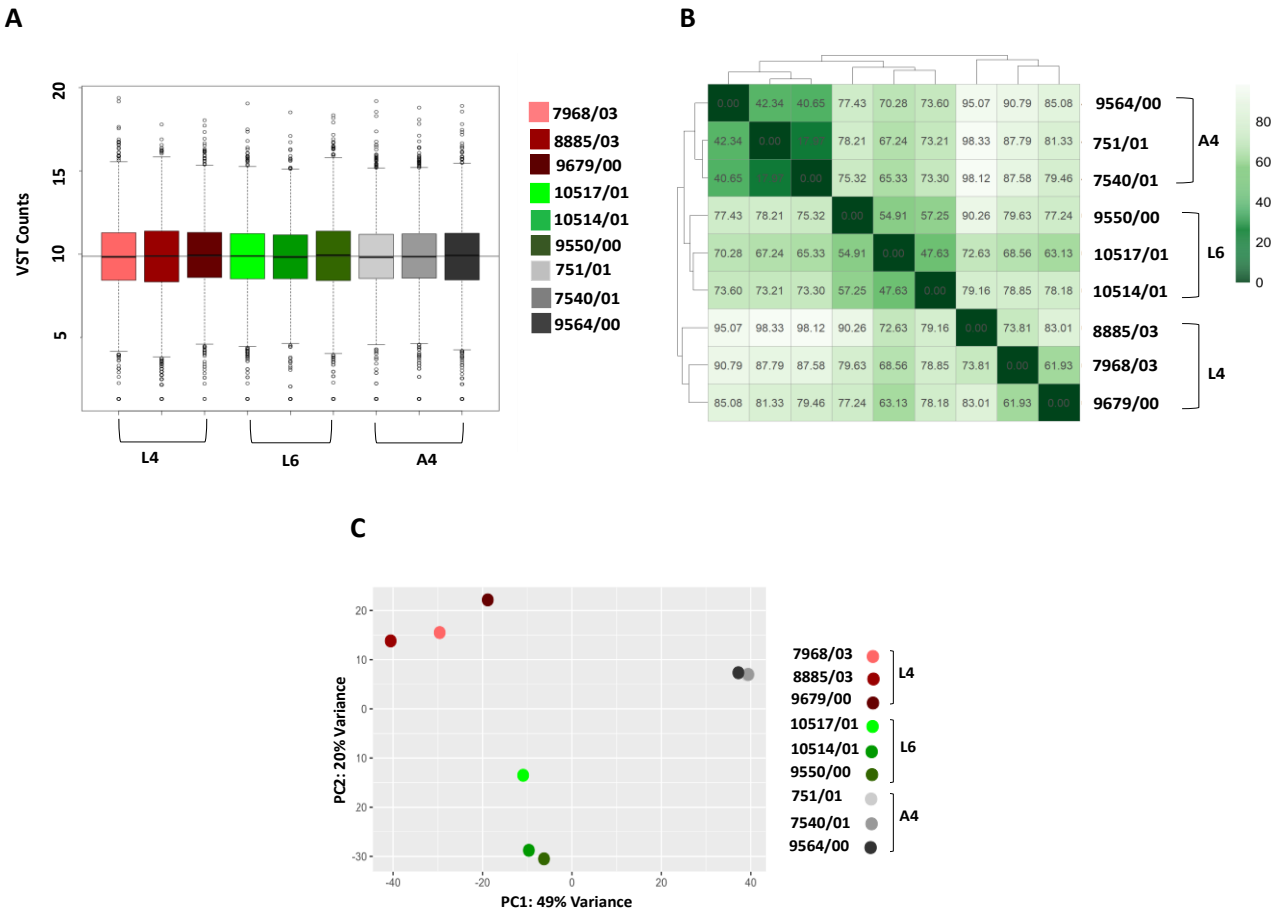
6.3.2 Transcriptomics Profiling of *Mycobacterium tuberculosis* complex Strains

6.3.2.1 RNA-Seq Data Structure

Three clinical *Mtbc* isolates representing the clades of the human widespread-modern (L4) and restricted-ancient (L6), as well as the animal (A4), were characterized at the transcriptomics level. *Mtbc* strains were cultivated in Middlebrook 7H9 supplemented with 10% oleic acid albumin dextrose catalase (OADC), 0.2% glycerol, 0.05% Tween 80, and harvested at an OD_{600nm} of 0.5-0.6. The RNA was isolated and sequenced with Illumina technology and analyzed with the DESeq2 method (see 5.4.3 Next-Generation Sequencing and 5.5.2 B). Figure 17A shows the transformation of the RNA-Seq counts per sample. Next, an unsupervised hierarchical clustering with the transformed data based on the Euclidean distance clustering calculation was performed. The sample-to-sample calculated distances showed that the strains grouped according to their phylogenetic lineage (Figure 17B), further confirmed by the Principal Component Analysis (PCA). The PCA nicely shows the strains belonging to a given lineage clustered closely (Figure 17C), indicating a clear difference between the transcriptome profile among lineages. For example, the *Maf* strains comprising the two clinical strains (10517/01 and 10514/01) and the reference strain ATCC (9550/00), and the *Mtb* strains comprising the two clinical LAM strains (7968/03 and 8885/03) and the reference strain H37Rv ATCC (9679/00) separated along the first component (49% variance) compared to the A4 (*Mbv*) strains.

Figure 17: RNA-Seq profiling of the *Mycobacterium tuberculosis* complex strains.
A Comparison of the means of variance stabilizing transformation (VST) of RNA-Seq counts among samples. The horizontal line highlights the mean of the normalized counts across samples. **B** Transcriptome heatmap was calculated from the variance stabilizing transformation of count data for the overall gene expression showing transcriptional relationships between the investigated strains. Rlog values of the dispersion estimates were clustered to calculate the Euclidean distance between samples. **C** Principal component analysis (PCA) scatter of the gene expression among *Mycobacterium tuberculosis* complex (*Mtbc*) strains of lineages (L4: red, L6: green, and A4: gray). PCA plot showing the variance of three biological replicates of each lineage. The percentages on each axis represent the percentage of variation explained by principal components. *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6 and A4, Animal Clade 4.

Figure 17 (Continued)



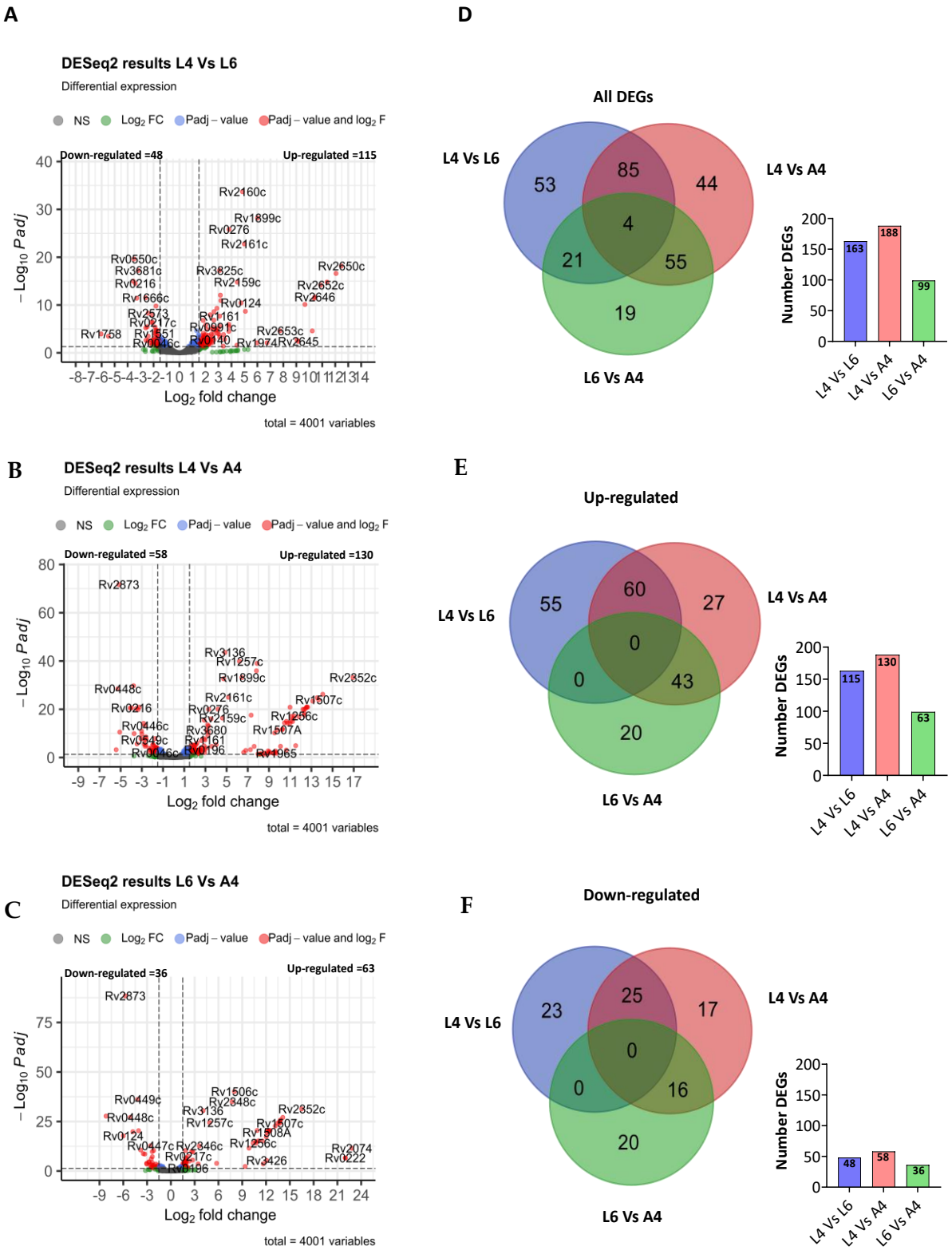
6.3.2.2 Differential Expression Analysis

The differentially expressed genes (DEGs) of lineage strains of L4, L6, and A4 were determined using the DESeq2 method based on negative binomial generalized models. Genes with adjusted p-values <0.05 and absolute Log_2FC >1.5 were considered significant DEGs. Comparative transcriptomic analysis revealed that L4 Vs. A4 strains had the higher number of DEGs (188), followed by L4 Vs. L6 (163) and L6 Vs. A4 (99) strains, these results are represented in Figure S4, Figure S5, and Figure S6, respectively, briefly, out of 188 genes differentially expressed in L4 Vs A4 strains. In A4 strains, 130 genes were up-regulated and 58 down-regulated. A total of 163 DEGs in L4 Vs. L6 strains comparison, including 115 genes up-regulated and 48 down-regulated, in L6 Vs. A4 strains comparison, there were 99 DEGs, of which 63 were up-regulated and 36 down-regulated (Figure 18A-C).

Venn diagrams were drawn to investigate the distribution of DEGs among pairwise comparisons. Figure 18D shows that 4 genes were commonly expressed between comparisons; these genes are involved in intermediary metabolism and respiration (Rv0771, Rv0770, and Rv0769) and cell wall and cell processes (Rv2875). The number of genes that were exclusively up-regulated was 55, 27, and 20, and down-regulated genes were 23, 17, and 36 in L4 Vs. L6, L4 Vs. A4 and L6 Vs. A4 strains, respectively. In addition, there were not commonly expressed up and down genes between all pairwise groups (Figure 18D-F). However, between L4 Vs. A4 and L4 Vs. L6 strains pairwise comparisons, 60 DEGs were commonly up-regulated, of which 25% (15/60) are involved in the cell wall and cell processes, and 18% (11/60) are involved in insertion sequences and phages (Figure 18E). 25 DEGs were commonly down-regulated, of which 32% (8/60) are involved in intermediary metabolism and respiration and 24% (6/25) in the cell wall and cell process (Figure 18F). Between L4 Vs. A4 and L6 Vs. L4 strains, 43 DEGs were commonly up-regulated, of which 30% (13/43) are involved in the cell wall and cell processes and 19% in proline-glutamic-acid/proline-prolin-glutamic-acid (PE/PPE) (Figure 18E). Out of 16 DEGs commonly down-regulated, 44% (7/16) are involved in the cell wall and cell processes, and 31% (5/16) are involved in information pathways (Figure 18F).

Figure 18: Transcriptional profiling of *Mycobacterium tuberculosis* complex strains.
Volcano plots of differentially expressed genes (DEGs) for the pairwise comparisons representing the logarithm of fold change (log_2) on the x-axis and log_{10} of the p-adjusted value of each transcript on the y-axis. **A** L4 Vs. L6 strains (48 Down-regulated and 115 Up-regulated). **B** L4 Vs. A4 strains (58 Down-regulated, 130 Up-regulated). **C** L6 Vs. A4 strains (36 Down-regulated, 63 Up-regulated). The Venn diagram represents the distribution of all DEGs (**D**), up-regulated (**E**), and down-regulated (**F**) among the pairwise comparisons of L6 Vs. L4, A4 Vs. L4, and L6 Vs. A4 strains. Venn diagrams were generated using the web tool: http://bioinformatics.psb.ugent.be/webtools/Venn/. L4, Linage 4; L6, Lineage 6; A4, Animal Clade 4; Padj-value, P adjusted value.

Figure 18 (Continued)



6.3.2.3 Functional Annotation of the Differential Expressed Genes

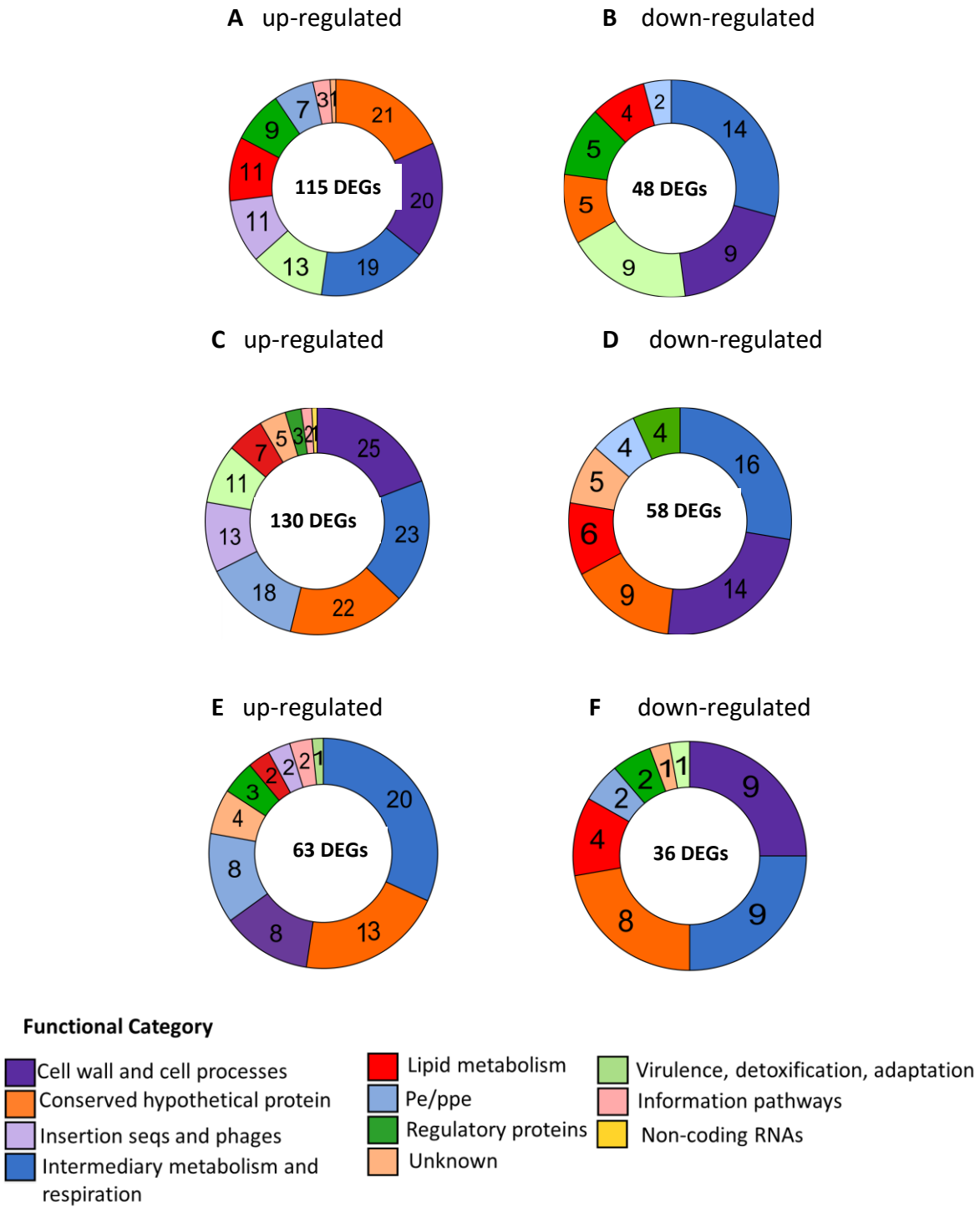
Mycobrowser data repository with genomic and proteomic for pathogenic mycobacteria was interrogated to understand the biological functions of the DEGs identified (https://mycobrowser.epfl.ch/), out of the 163 DEGs identified in L4 Vs. L6 strains, 115 genes up-regulated were found to be involved in eleven distinct functional categories (Figure 19A), including conserved hypothetical proteins (n=21), cell wall and cell processes (n=20), Intermediary metabolism and respiration (n=19), virulence, detoxification, and adaptation (n=13), insertion sequences ad phages (n=13). On the other hand, 48 down-regulated genes were involved in eight categories, such as Intermediary metabolism and respiration (n=14), cell wall and cell processes (n=9), virulence, detoxification and adaptation (n=9), and conserved hypothetical proteins (n=5) (Figure 19B).

In the pairwise comparison between L4 and A4 strains 188 DEGs were identified. Of these, 130 genes were found in L4 strains distributed across 12 functional categories. These categories included, for example, cell wall and cell processes (n=25), intermediary metabolism and respiration (n=23), conserved hypothetical proteins (n=22), and PE/PPE (n=18) (Figure 19C). On the other hand, Figure 19D shows a total of 58 DEGs down-regulated were involved in eight categories, including Intermediary metabolism and respiration (n=16), cell wall and cell processes (n=14), conserved hypothetical proteins (n=9), and lipid metabolism (n=6).

In the pairwise comparison between L6 and A4 strains 99 DEGs were identified. Out of these, 63 up-regulated genes were involved in 12 functional categories, including intermediary metabolism and respiration (n=20), conserved hypothetical protein (n=13), cell wall and cell processes (n=8), and PE/PPE (n=8) (Figure 19E). Meanwhile, 36 down-regulated DEGs were involved in the cell wall and cell processes (n=9), Intermediary metabolism and respiration (n=9), conserved hypothetical protein (n=8), and lipid metabolism (n=4), as shown in Figure 19F.

Figure 19: Donut plots of differentially expressed genes according to their functional category.
Differentially expressed genes (DEGs) up (**A**) and down (**B**) regulated among L4 Vs. L6 strains. DEGs up (**C**) and down (**D**) regulated among L4 Vs. A4 strains. DEGs *u*p (**E**) and down (**F**) regulated among L6 Vs. A4 strains. These categories were assigned according to Mycobrowser (www.mycobrowser.epfl.ch), a data repository that provides manually curated annotations based on genomic and proteomic studies of *Mycobacterium tuberculosis* H37Rv[244]. L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4.

Figure 19 (Continued)



6.3.2.4 Overview of the Top 50 Differential Expressed Genes among Pairwise Comparisons

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs. 24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean 0.22x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean 0.97x106 CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean 0.17x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However, a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

illustrates an overview of the top 50 DEGs for each pairwise comparison, showing that strains belonging to a given lineage show a similar expression and clustering pattern. Contrasting transcriptional profiles can be observed in the pairwise comparisons, e.g., the first 16 genes are significantly up-regulated in L6 strains but down-regulated in L4 strains and the other way around with the remaining 34 genes (

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs. 24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean 0.22x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean 0.97x106 CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean 0.17x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However,

a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

A). Heatmaps representing the total number of DEGs identified in each pairwise comparison can be seen in 11. SUPPLEMENTARY RESULTS (Figure S4, Figure S5 and Figure S6).

A functional categorical analysis has revealed that the first 16 up-regulated genes in L6 strains, as compared to L4 strains, are involved in cell wall and cell processes. For instance, Rv3806c/ubiA is responsible for producing Decaprenylphosphoryl-5-phosphoribose (DPPR) synthase, which is a decaprenyl-phosphate 5-phosphoribosyltransferase. Overexpression of this gene is associated with ethambutol resistance in *Mtb* [243].

Functional categorical analysis revealed that the first 16 up-regulated genes in L6 strains, compared to L4 strains, are involved in cell wall and cell processes. For instance, Rv3806c /ubiA, encodes a Decaprenylphosphoryl-5-phosphoribose (DPPR) synthase (decaprenyl-phosphate 5-phosphoribosyltransferase) whose overexpression has been associated with ethambutol resistance in *Mtb* [245]; Intermediary metabolism and respiration (Rv1092c/coaA, Rv0216, Rv1666c/cyp139, and Rv0217c/lipW); Virulence, detoxification, and adaptation, e.g., the Rv0549c/vapC3 gene encodes a possible toxin VapC3, and Rv0550c/vapB3 encodes a possible antitoxin VapB3, these proteins belong to the Toxin-Antitoxin (TA) systems that support the survival of the mycobacteria pathogen under stressful conditions [246]; Regulatory proteins (Rv0465c, Rv3681c/whiB4, and Rv2779c), e.g., whiB4 which regulates oxidative stress response to enhance survival and dissemination of *Mtb* inside the macrophages[247]; Conserved hypotheticals (Rv3531c, Rv2573, and Rv3123) and lipid metabolism (Rv1867) (see Table S5).

The 34 genes up-regulated genes in L4 strains are involved in several biological processes (

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs. 24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean 0.22x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean 0.97x106 CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean 0.17x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However, a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

A and Table S5): Cell wall and cell processes, e.g., Rv1899c/ lppD encodes a possible lipoprotein LppD; Conserved hypothetical, insertion sequences and phages (e.g., Rv2646, Rv2657c, Rv2653c, and Rv2655c) involved in the production of a possible PhiRv2 prophage protein; Intermediary metabolism and respiration, e.g., the Rv1161/narG, Rv1162/narH, and Rv1163/narJ genes involved in the production of nitrate reductases that are highly expressed during aerobic growth of *Mtb*; [248] Lipid metabolism, e.g., Rv3824c/papA1, Rv1182/papA3, Rv3825c/pks2, and Rv1925/fadD31, these genes produce essential acyltransferases for the biosynthesis of sulfolipid factors in virulent strains of *Mtb* [249], [250].

Furthermore, some of these genes are also involved in PE and polymorphic GC-rich (PE-PGRS) family proteins, regulatory proteins, and virulence, detoxification, and adaptation, e.g., the Rv2416/eis gene encodes the enhanced intracellular survival proteins Eis.

As can be seen in

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs. 24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean 0.22x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean 0.97x106 CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean 0.17x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains

infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However, a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

B, a similar contrasting transcriptomic profiling was observed in L4 Vs. A4 strains. For instance, the first 10 genes were up-regulated in A4 and down-regulated in L4 strains, and conversely, the residual 40 genes were up-regulated only in L4 strains. The up-regulated genes in L4 strains are implicated in: Cell wall and cell processes (e.g., Rv1899c, Rv3679, Rv1510, and, Rv1899c /lppD); Conserved hypothetical proteins (e.g., Rv2159c, Rv1977, Rv1514c and Rv1508A); Insertion sequences and phages (e.g., Rv2651c, Rv2650c, Rv2659c, and Rv2652c); Intermediary metabolism and respiration (e.g., Rv3119/moaE1, Rv2351c/plcA, Rv1512/epiA, Rv2349c/plcC, and Rv2350c/plcB); lipid metabolism (e.g., Rv3825c/pks2); PE/PPE (e.g., Rv3429/PPE59, Rv2162c/PE_PGRS38, Rv3136/PPE51, Rv3739c/PPE67, and Rv2352c/PPE38) and also genes grouped in the unknown category (Rv2348c, Rv1509, and Rv150c). On the other hand, the 10 genes up-regulated in A4 strains are involved in Cell wall and cell processes (Rv2873/mpt83, and Rv0446c); Conserved hypotheticals (Rv0448c and Rv0449c); Intermediary metabolism and respiration (Rv1666c/Cyp139, Rv2874/dipZ, and Rv0216); lipid metabolism (Rv0447c/ufaA1); Regulatory proteins (Rv3681c/whiB4) and virulence, detoxification, and adaptation (Rv0550c/vapB3) (see Table S6).

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs.

24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean 0.22x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean 0.97x106 CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean 0.17x106 CFU/mL Vs. 0.66x106 CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However, a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

C illustrates the top 50 DEGs from the L6 and A4 strains comparison; the heat map shows that the first 15 genes were strongly up-regulated in A4 strains and down-regulated in L6 strains and, conversely, the remaining 35 genes. The 15 up-regulated genes in A4 strains are implicated in cell wall and cell processes, e.g., the Rv2873/mpt83 and Rv2875/mpt70 genes which encode for the immunodominant antigensMPB70 and MPB83 that are involved in a delayed hypersensitive response [251], [252]; Conserved hypotheticals (e.g., Rv1519, Rv0449c, and Rv0448c); Intermediary metabolism and respiration (Rv0770, Rv2874/dipZ, Rv1170/mshB); Lipid metabolism e.g., Rv2386c/mbtI that encodes a protein of anthranilate synthase which is essential for the *in vitro* growth of *Mtb* [253]and in the functional

category of PE/PPE, the Rv0124/ PE_PGRS2 gene. The 35 up-regulated genes in L6 strains are implicated in Cell wall and cell processes (e.g., Rv2346c/esxO, Rv3806c/ubiA, and Rv2873/mpt83, the esxO gene is a member of the early secretory antigenic target 6 kDa (ESAT-6) involved in host cell conquest and intracellular persistence of *Mtb* [254]; Conserved hypotheticals (e.g., Rv1976c, Rv3120, Rv1515c, Rv1514c); Intermediary metabolism and respiration (e.g., Rv2074, Rv2349c/plcC, and Rv2351c/plcA); lipid metabolism (e.g., Rv0222/echA1, Rv2386c/mbtI, and Rv0447c/ufaA1); PE/PPE (e.g., Rv3739c/PPE67, Rv3136/PPE51, Rv3429/PPE59, Rv2352c/PPE38, and Rv0124/PE_PGRS2); and virulence, detoxification, adaptation with the Rv3501c/yrbE4A (see Table S7).

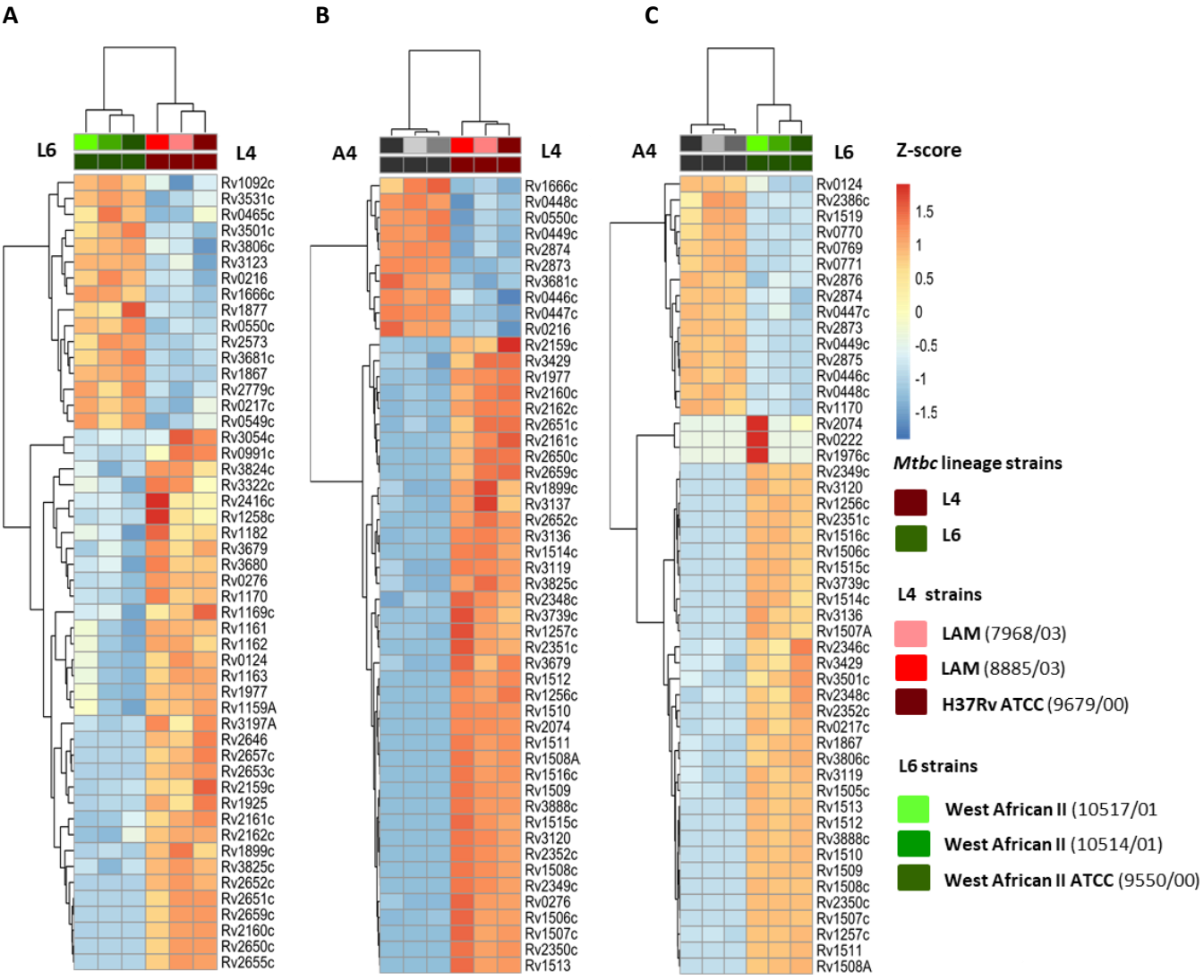


Figure 20: Overview of the top 50 differential expression genes among *Mycobacterium tuberculosis* complex strains based on RNA sequencing.

Heatmaps display the top 50 differential expression genes (DEGs) expression profiles for each strain. Pairwise comparisons: L6 Vs. L4 strains (**A**), A4 Vs. L4 strains (**B**) and A4 Vs. L6 strains (**C**). Hierarchical

clustering among strains is depicted in the figures. Normalized read counts underwent row scaling calculation with the function named scale of the pheatmap R package for visualization purposes for comparison across rows. Heatmap is based on the feature counts with an adjusted p-value of 0.05, Log_2 Fold Change (log_2FC) >1.5 for up-regulated (orange color spectrum) and <-1.5 for down-regulated (Blue color spectrum). *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4.

6.3.3 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Human Monocyte-Derived Macrophages from Donors with African Ancestry

To define potential pathobiological differences of strains of different phylogenetic lineages, I tested first the phagocytosis/uptake in hMDMs derived from donors with African ancestry. No significant differences in phagocytosis/uptake were found among *Mtbc* strains at 4 hpi compared to the reference *Mtb* H37Rv ATCC (9679/00) strain (Figure 21A-C), neither among L4 Vs. L6 strains (mean 29.47% Vs. 30.43%, p = 0.8919), and A4 Vs. L4 strains (mean 24.66% Vs. 29.47%, p = 0.4909), L6 Vs. A4 strains (mean 30.43% Vs. 24.66%, p = 0.2037) (Figure 21D-F). However, it should be noted that the animal-adapted A4 strains were phagocytized at similar rates compared to the human-adapted strains.

Furthermore, after the uptake of *Mtbc* cells into hMDMs, their intracellular growth kinetics were investigated over time (4 h, 24 h, 96 h, and 168 h post-infection). The progression of the infection caused by L6 strains was similar to L4 strains. However, significant differences were found among the L6 (10517/01 and 10514/01) clinical strains and *Mtb* H37Rv ATCC (9676/00) strain at 24 hpi (p = 0.0336 and p =0.0296) (Figure 22A) and observed among strains of L4 and L6 (mean $0.22x10^6$ CFU/mL Vs. $0.66x10^6$ CFU/mL, respectively, p = 0.0003) (Figure 22B). Significant differences were found among L4 and A4 strains at 96 hpi (mean $0.97x10^6$ CFU/mL Vs. 0.67x106 CFU/mL, respectively, p = 0.0493) (Figure 22D). Clinical strains of L6 (10517/01 and 10514/01) and the *Mbv* (751/01) significantly differed to *Mbv* ATCC (9564/00) strain at 24 hpi (p = 0.0387 and p =0.0341) and 96 hpi accordingly (p =0.0150) (Figure 22E), A4 and L6 lineages strains differed at 24 hpi (mean $0.17x10^6$ CFU/mL Vs. $0.66x10^6$ CFU/mL, respectively, p = <0.0001) and at 96 hpi (Figure 22F). No further significant differences were determined among other *Mtbc* strains and lineages at a given time post-infection. These data indicate that the *Mtbc* strains infected, survived, and grew in the African host similarly, irrespective of their genetic makeup. However, a distinct response was triggered after 24 hpi for L4 Vs. L6 and L6 Vs. A4 strains, which can be related to the onset of the infection.

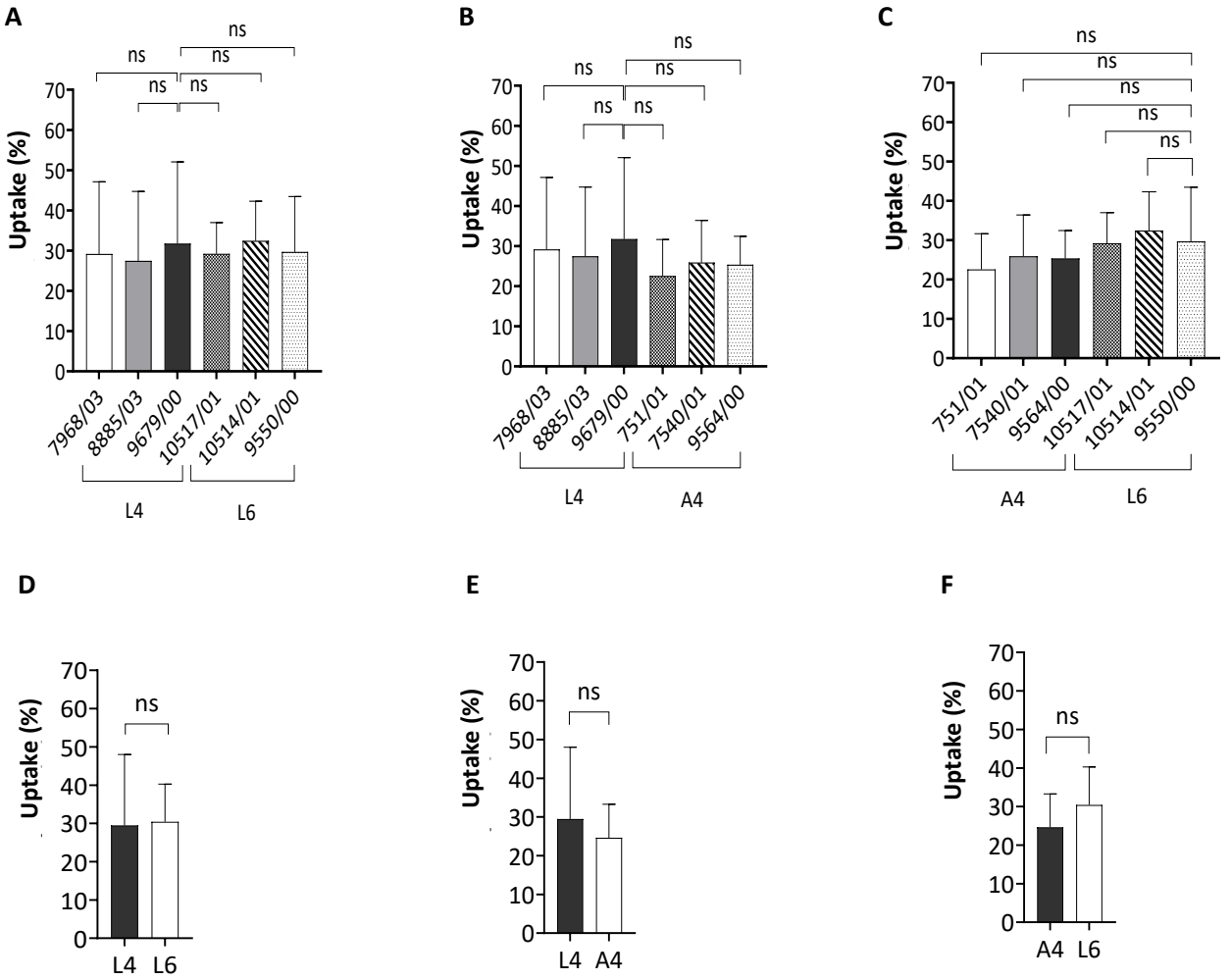
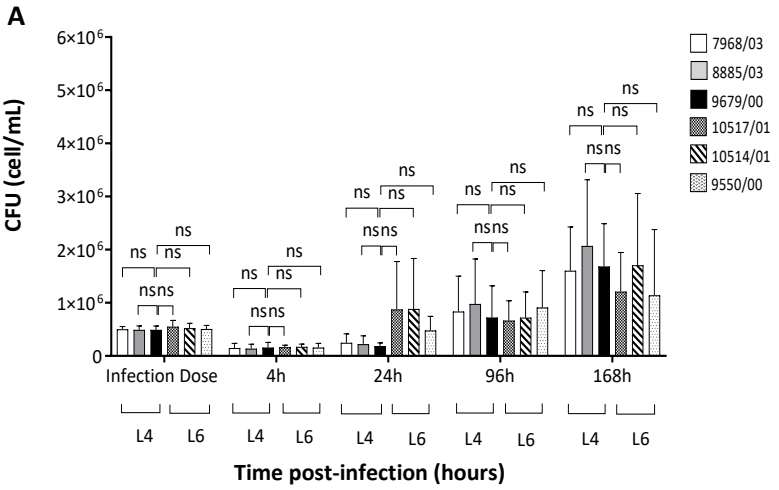
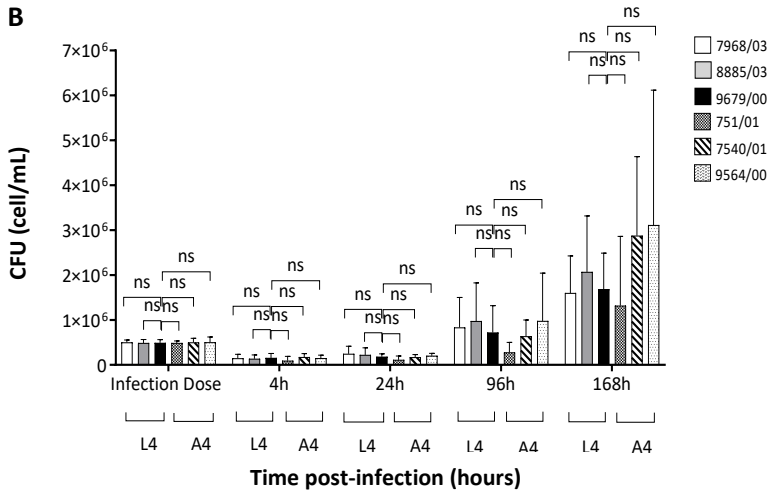


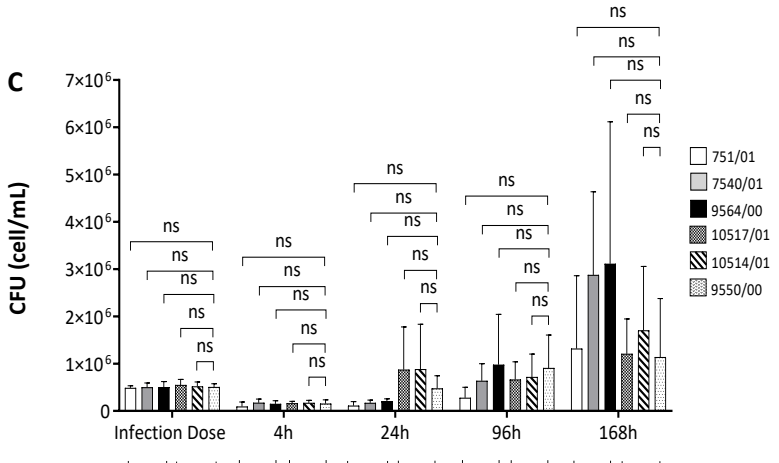
Figure 21: Phagocytosis of distinct *Mycobacterium tuberculosis* complex strains in African human monocyte-derived macrophages.

Comparative analysis per strain (**A**-**C**) and lineage (**D**-**F**) of the uptake of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (**B** and **E**), and A4 Vs. L6 strains (**C** and **F**), respectively. Human monocyte-derived macrophages (hMDMs) were infected at an MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately after uptake at 4 hours post-infection (hpi). Figures represent mean and standard deviations comprising a total of three independent experiments. Figures **A**-**C** show the mean and standard deviation per *Mtbc* strain (n=9 technical replicates of infection well,

n=27 CFU replicates). Figures **D**-**F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=27 technical replicates) and CFU replicates per lineage (n=81). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant; *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; h, hours.







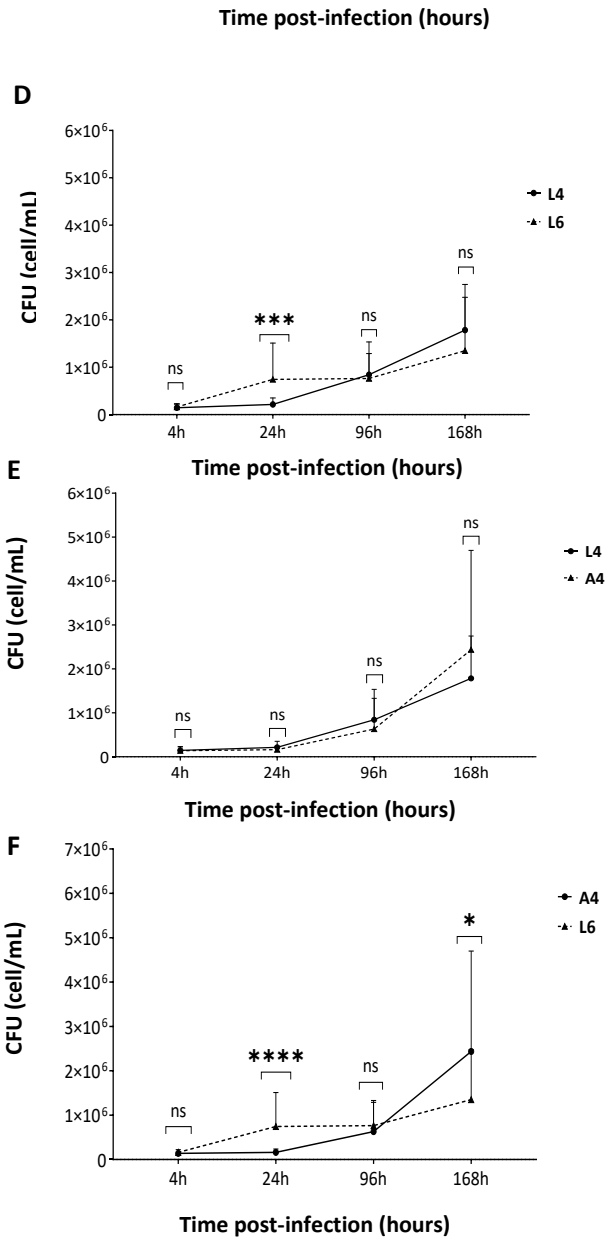


Figure 22: Intracellular growth of *Mycobacterium tuberculosis* complex strains in African human monocyte-derived macrophages.

Comparative analysis of intracellular survival and growth kinetics per strain (**A**-**C**) and lineage (**D**-**F**) of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (**B** and **E**), and A4 Vs. L6 (**C** and **F**) strains, respectively. Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli). Survival and intracellular growth of *Mtbc* strains were determined immediately at 4 h, 24 h, 96 h, and

168 h post-infection. Macrophage infections comprise three independent experiments. Figures **A**, **B**, and **C** show the mean and standard deviation per *Mtbc* strain (n=9 technical replicates of infection, n=27 CFU replicates). Figures **D**, **E**, and **F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=27 technical replicates of infection, n=81 CFU replicates). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; h, hours.

6.3.4 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex Strains in Human Monocyte-Derived Macrophages from Donors with European ancestry

It was observed that the L4 strains were significantly more phagocytized than H37Rv ATCC strain when compared to L6 strains in European hMDMs (Figure 23A) Interestingly, the LAM (8885/03 and 7968/03) clinical strains within L4 also showed significant differences compared to H37Rv ATCC (9679/00) strain (p = 0.0353 and p = 0.00058), as shown in Figure 23A. Further, the lineage level analysis revealed similar results where L4 strains had a mean phagocytosis rate of 33.93% compared to 2.8% in L6 strains (p = 0.0001), as shown in Figure 23D. Compared to the H37Rv ATCC strain, the A4 strains of *Mbv* (specifically 751/01, 7540/01, and 9564/00) showed significantly lower rates of phagocytosis (Figure 23B). This difference was particularly evident in the L4 Vs. A4 lineages strains, with a mean of 33.93% and 2.95%, respectively (p = 0.0002, as shown in Figure 23E). However, there were no significant differences in phagocytosis rates between A4 and L6 strains and lineages (2.95% and 2.87%, p=0.8794), which had similar phagocytosis rates (as shown in Figure 23C and Figure 23F).

Moreover, the intracellular survival of L6 strains was clearly lower in comparison to the H37Rv ATCC reference strain (4 hpi: p = 0.0019, p = 0.0013, p = 0.0025; 24 hpi: p = <0.0001, p = <0.0001, p = <0.0001; 168 hpi: p = <0.0001, p = <0.0001, p = <0.0001, Figure 24A). In addition, significant differences within L4 strains were observed, for example, LAM (7968/03) strain at 4 hpi, 96 hpi, and 168 hpi (p = 0.0018, p = 0.0236, p = 0.0236, respectively). Growth kinetics are noticeably different among lineages; L4 strains successfully outgrew L6 strains dramatically performed at all time points investigated (Figure 24B). Furthermore, *Mbv* strains grew significantly less than the H37Rv ATCC strain at 4 hpi, 24 hpi, and 168 hpi. Strains of L4 also outgrew A4 strains significantly when comparing on a lineage level at 4 hpi, 24 hpi, and 168 hpi (p = <0.0001, Figure 24C-D). It is also worth noting that the growth of L4 strains increased greatly after 96 hpi, while A4 strains stayed almost at the same rate. In contrast to African hMDMs, European hMDMs do not seem to be able to promote the infection of *Mbv* (Figure 24D). Although A4 and L6 strains have a similar intracellular growth pattern, L6 strains exceptionally exhibited more reduced

growth than A4 strains upon infection (Figure 24E-F). The attenuation pattern was also found on a lineage level at 24 hpi (p = <0.0001), 96 hpi (p = <0.0001), and 168 hpi (p = <0.0001) (Figure 24F).

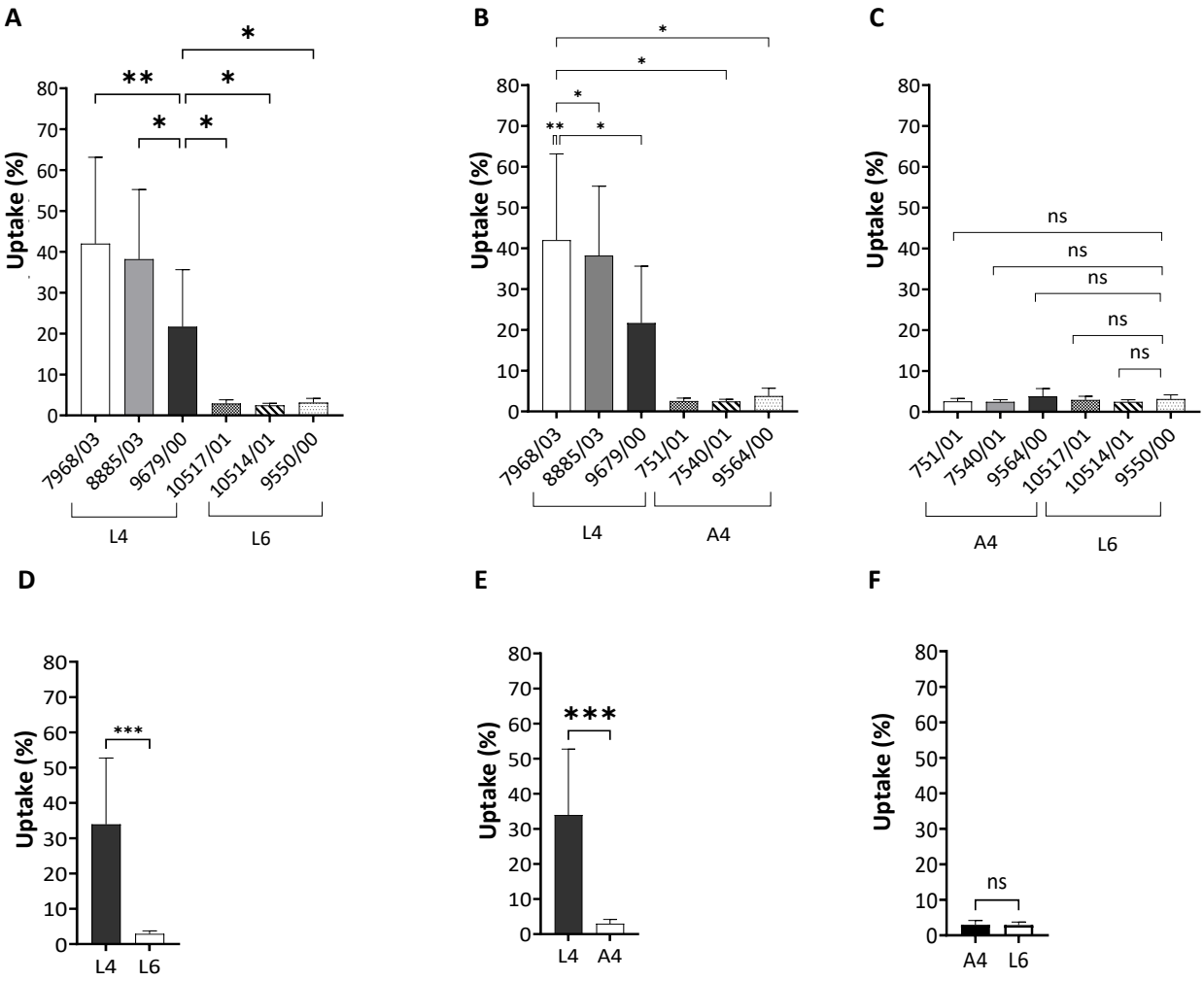
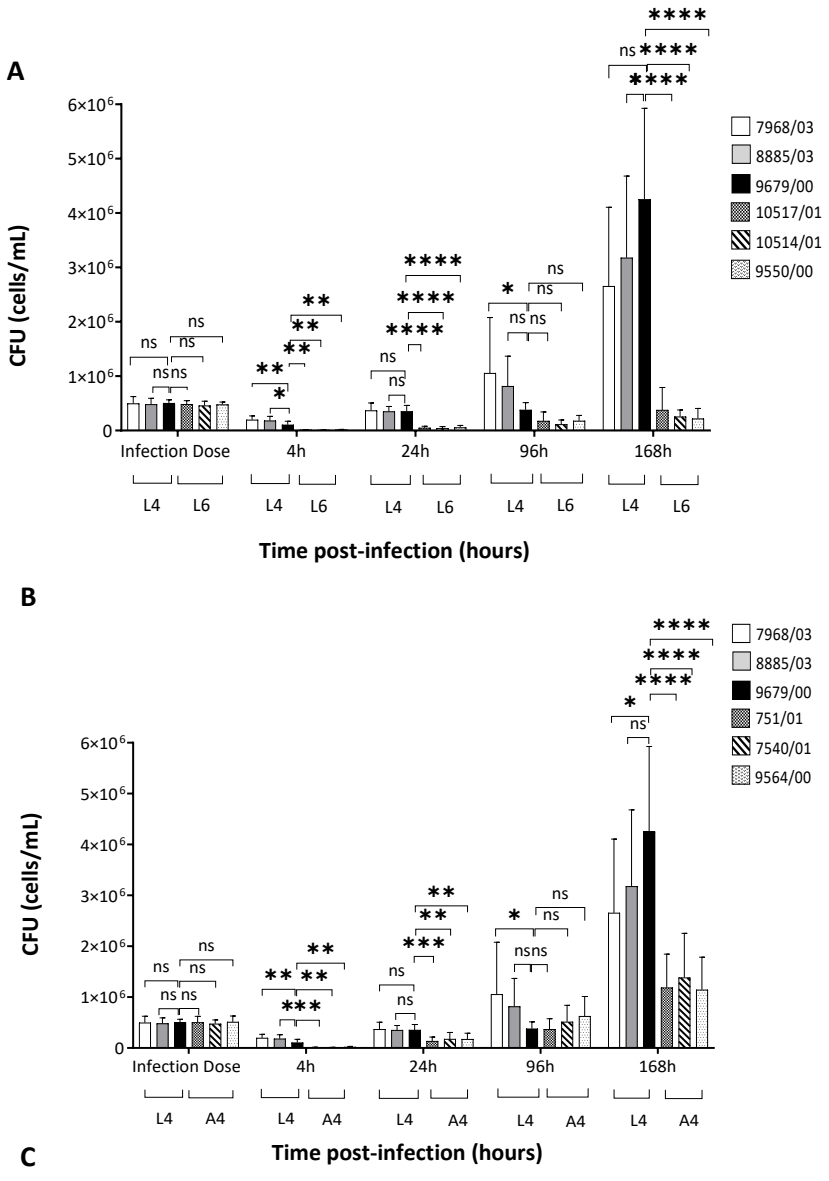


Figure 23: Phagocytosis of distinct *Mycobacterium tuberculosis* complex strains in European human monocyte-derived macrophages.

Comparative analysis per strain (**A**-**C**) and lineage (**D**-**F**) of the uptake of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (**B** and **E**), and A4 Vs. L6 strains (**C** and **F**), respectively. Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately after uptake at 4 hours post-infection (hpi). Figures represent mean and standard deviations comprising a total of three independent experiments. Figures

A-**C** show the mean and standard deviation per *Mtbc* strain (n=9 technical replicates of infection well, n=27 CFU replicates). Figures **D**-**F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=27 technical replicates) and CFU replicates per lineage (n=81). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant; *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; h, hours.



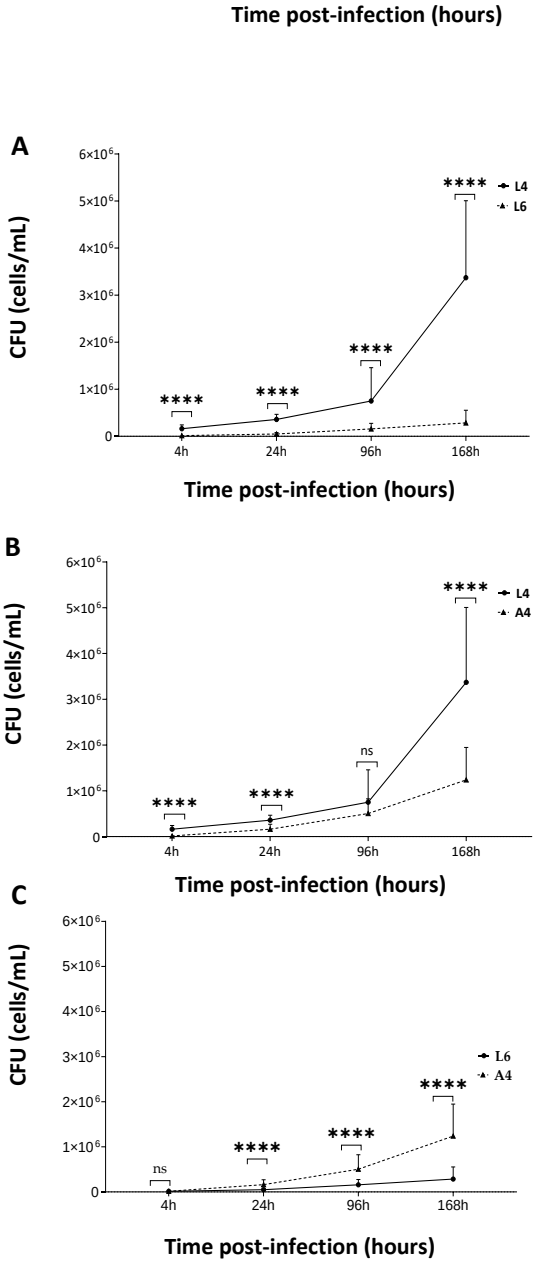


Figure 24: Intracellular growth of *Mycobacterium tuberculosis* complex strains in European human monocyte-derived macrophages.

Comparative analysis of intracellular survival and growth kinetics per strain (**A**-**C**) and lineage (**D**-**F**) of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (F**B** and **E**), and A4 Vs. L6 (**C** and **F**) strains, respectively. Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli). Survival and intracellular growth of *Mtbc* strains were determined immediately at 4 h, 24 h, 96 h, and 168 h post-infection. Macrophage infections comprise three independent experiments. Figures **A**, **B**, and **C** show the mean and standard deviation per *Mtbc* strain (n=9 technical replicates of infection, n=27 CFU replicates). Figures **D**, **E**, and **F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=27 technical replicates of infection, n=81 CFU replicates). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant (p > 0.05)*;* L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; h, hours.

6.3.5 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex in Bovine Monocyte-Derived Macrophages

As detailed in Figure 25A and Figure 25D, no significant differences in uptake were found among L4 vs. L6 single strains, neither at the lineage level in blood bMDMs. However, it is interesting to notice that the clinical LAM strain 8885/03 was significantly phagocytized less than the reference strain *Mtb* H37Rv ATCC (p = 0.0126), in contrast to what was observed in the African hMDMs (Figure 21A). On the other hand, bacterial uptake of A4 strains (751/01, 7540/01, 9564/00) was about 20 times higher than H37Rv ATCC reference strain (p = 0.0002, p = <0.0001, p = <0.0001), similarly at the lineage level A4 vs L4 strains (mean = 28.24% vs mean = 9%, p = <0.0001, unpaired T-test), (Figure 25B and Figure 25E). Furthermore, it is also observed in Figure 25C and Figure 25F that A4 strains were phagocytized at higher levels by the bovine cells, followed by the ancient human-adapted L6 strains (10517/01, 10514/01 and 9550/00) which were approximately 15% less phagocytized than *Mbv* 9564/00 strains (p = <0.0001, p = <0.0001, p = <0.0001 supporting the tropism of the A4 strains for the animal host.

Furthermore, as clearly illustrated in Figure 26A, there was significantly more growth of L6 strains as compared to L4 strains when infected at 24 hpi and 168 hpi (p=<0.0001, p=0.0005, p=<0.0001) and L6 clinical strain 10517/01 at 96 hpi (p=0.0410). Additionally, clinical LAM strains exhibited less growth than the reference strain H37Rv at 4 hpi (p=0.0098 and p=0.0027). Moving on to Figure 26D, it is evident that L6 strains grew exponentially after phagocytosis, from 4 hpi to 24 hpi (p=<0.0001), but decreased by one-fold at 96 hpi (p=0.0005) followed by a two-fold increase at 168 hpi (p=<0.0001). On the other hand, the L4 strains showed no significant growth and remained almost unchanged. Based on these findings, it can be concluded that the infection caused by L4 strains was not successfully maintained in the bovine host, unlike human hosts (Figure 24E). Furthermore, A4 strains grew twice as much as L4 strains until 96 hpi (p=0.0032), as depicted in Figure 26B and Figure 26E. While the infection with L6 strains was successfully

sustained over A4 strains upon infection, A4 strains grew twice as much as L6 strains at 4 hpi (p=<0.0001, Figure 26C and Figure 26F).

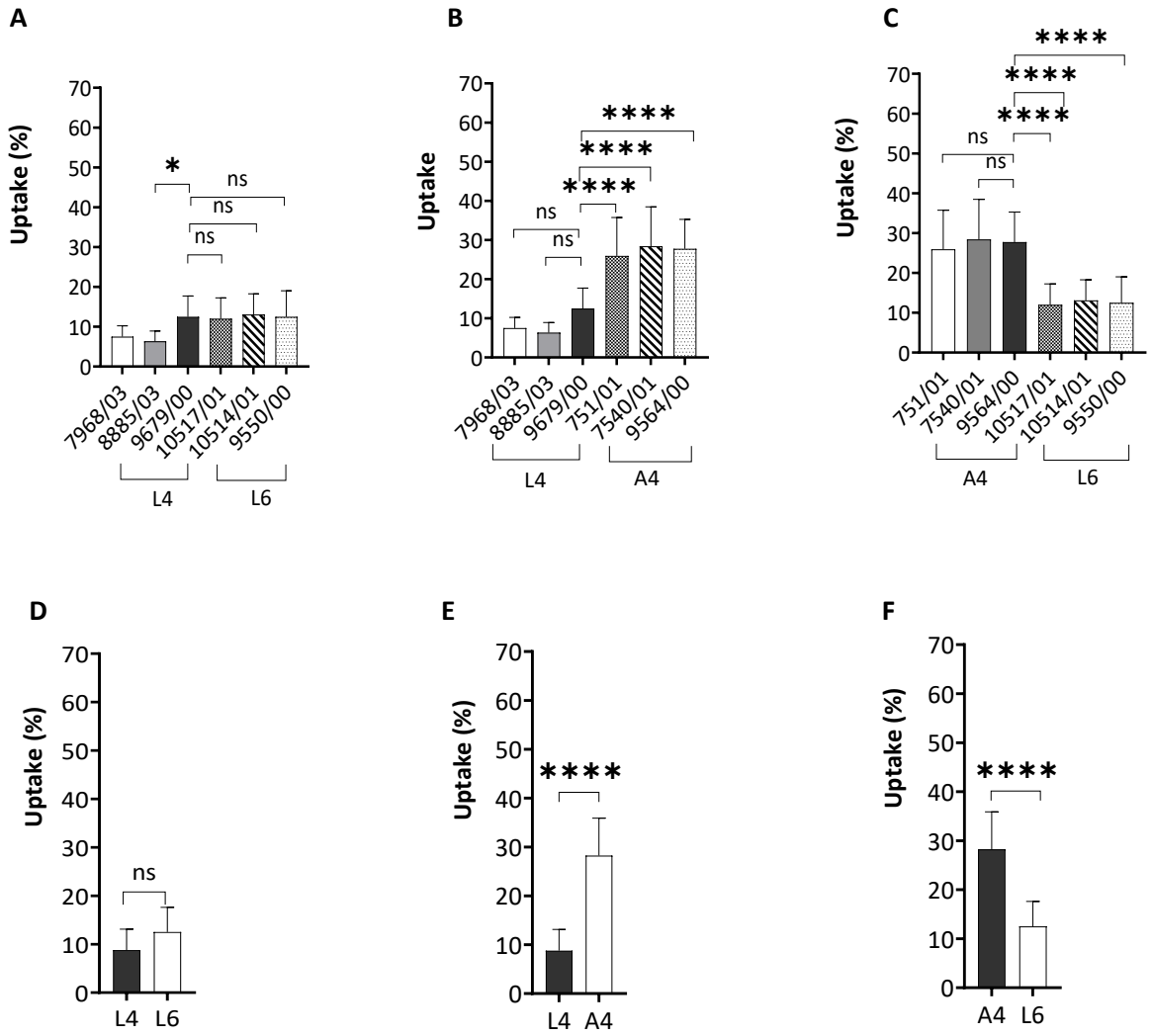
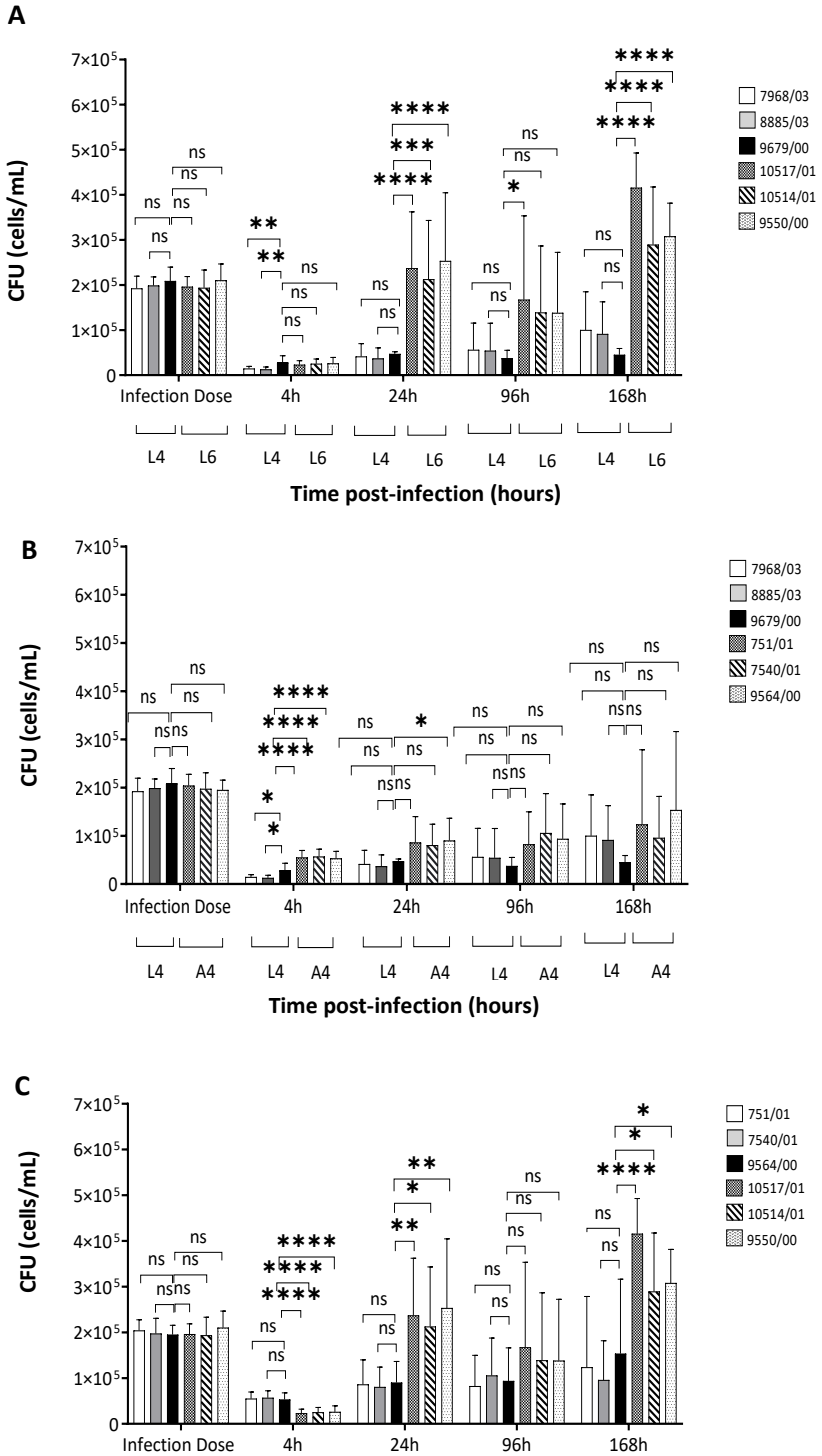


Figure 25: Phagocytosis of distinct *Mycobacterium tuberculosis* complex strains in bovine monocyte-derived macrophages.

Comparative analysis per strain (**A**-**C**) and lineage (**D**-**F**) of the uptake of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (**B** and **E**), and A4 Vs. L6 strains (**C** and **F**), respectively. Bovine monocyte-derived macrophages (bMDMs)

were infected at a MOI ~1:1 (0.2 x10^6 cells: 0.2 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately after uptake at 4 hours post-infection (hpi). Figures represent mean and standard deviations comprising four independent experiments, of which two compromised two donors (see Figure S9). Figures **A**-**C** show the mean and standard deviation per *Mtbc* strain (n=12 technical replicates, n= 36 CFU replicates). Figures **E**-**F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=36 technical replicates, n= 108 CFU replicates). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant results (p > 0.05); *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection



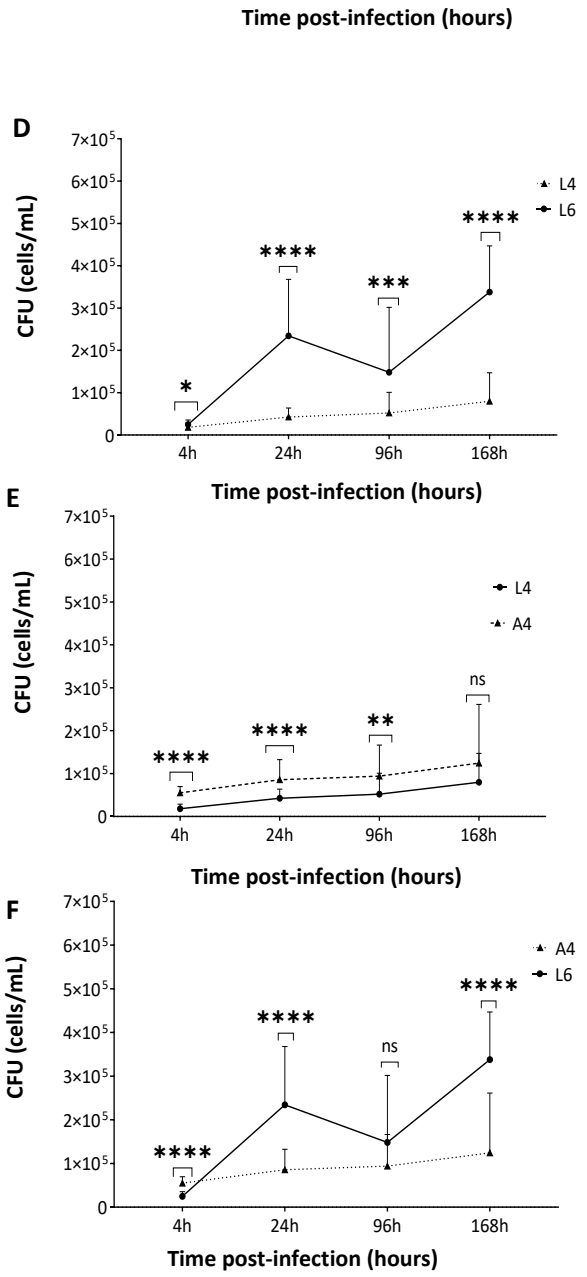


Figure 26: Intracellular survival and growth kinetics of distinct *Mycobacterium tuberculosis* complex strains in bovine monocyte-derived macrophages.

Comparative analysis of intracellular survival and growth kinetics per strain (**A**-**C**) and lineage (**D**-**F**) of L4 Vs. L6 (**A** and **D**), L4 Vs. A4 (**B** and **E**), and A4 Vs. L6 (**C** and **F**) strains, respectively. Bovine monocyte-

derived macrophages (bMDMs) were infected at a MOI ~1:1 (0.2 x10^6 cells: 0.2 x10^6 *Mtbc* bacilli). Survival and intracellular growth of *Mtbc* strains were determined immediately at 4 h, 24 h, 96 h, and 168 h post-infection. Figures **A**, **B**, and **C** show the mean and standard deviation per *Mtbc* strain (n=12 technical replicates of infection, n=36 CFU replicates). Figures **D**, **E**, and **F** show the mean and standard deviation of three *Mtbc* strains representing a given lineage (n=36 technical replicates on infection, n=108 CFU replicates). Mean, standard deviation, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. ns, not significant; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; h, hours.

6.3.6 Phagocytosis and Intracellular Growth Profiles of *Mycobacterium tuberculosis* complex Strains Among Human Ancestries.

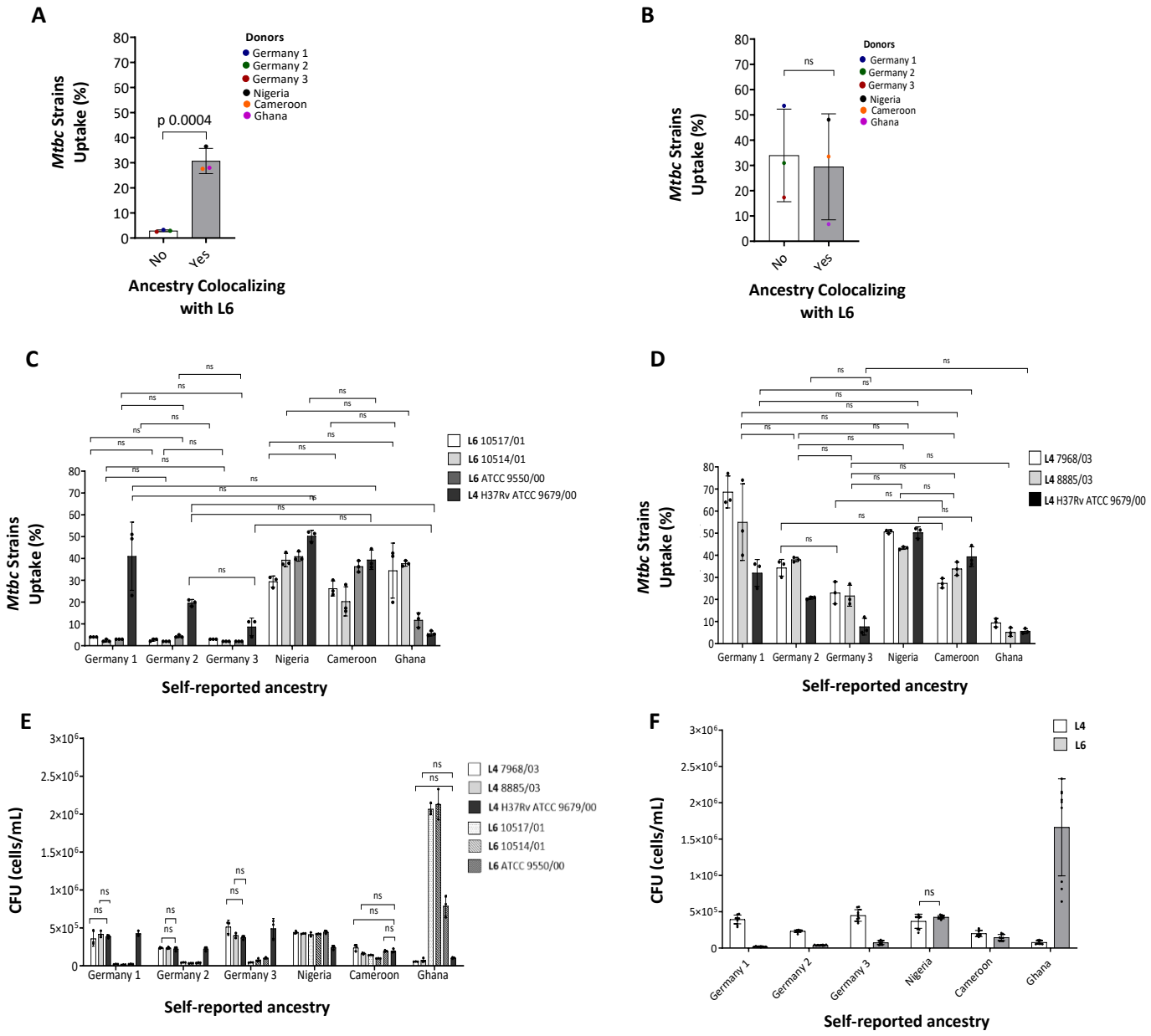
To further investigate whether hMDMs by self-reported ancestry differ in susceptibility to infection with the human-adapted strains of L4 and L6, *in vitro* macrophage infection experiments were analyzed concerning the ancestry of the human donors (German, Nigerian, Cameroonian, or Ghanaian). Phagocytosis/uptake of *Mtbc* strains comprising two clinical isolates of the geographically widespread L4 LAM was compared with two clinical strains of the geographically restricted L6 WAII while including the ATCC reference strains of each lineage in each group. Colony-forming unit readouts of the inoculum ensured equal infection doses. Macrophages of donors of self-reported ancestry co-localizing with the endemicity of L6 phagocytosed strains of L6 on average at 30.4% compared to 2.9% among macrophages of donors whose self-reported ancestry was Germany, not co-localizing with L6 ($p < 0.001$, Figure 27A). There was no significant difference in the growth of L4 strains in macrophages of donors of ancestry that do or do not co-localize with L6 (33.9% Vs. 29.5%, $p = 0.8$, Figure 27B). This difference in uptake of L6 strains between L6 co-localizing and non-co-localizing donors was driven by all three L6 lineages strains (all comparisons $p < 0.001$, Figure 27C). Intra-host and inter-host variability could be found in uptake efficiency of the three L4 and L6 strains (Figure 27D, Figure S10, and Figure S11).

Furthermore, L4 or L6 strains differed in their intracellular growth kinetics after uptake into hMDMs of different ancestries over time. In macrophages from donors with self-reported ancestry of Nigeria, Cameroon, or Ghana, there was no significant difference in growth between strains of L4 and L6 at two of the three time points (Figure 27E-F and Figure S12). However, growth in macrophages from these donors was higher for L6 compared to L4 strains at 24 hpi (comparing average growth across the three strains per lineage for the Ghanaian donor, $p < 0.001$, Cameroonian donor $p < 0.05$, no difference in macrophages of the Nigerian donor, Figure 27D-E). Conversely, in macrophages from donors of self-reported German ancestry, L4 strains grew at higher rates than L6 strains at all the time points tested ($p < 0.05$, Figure S12). Overall, these results evidence that differential transmissibility between allopatric and sympatric pairs may relate to the observed weaker uptake and growth of geographically restricted lineages in allopatric macrophages than increased growth and uptake of these lineages in sympatric macrophages. Thus, suggesting that sympatry predominantly manifests in diminished infectivity of geographically restricted *Mtbc* lineages in allopatric hosts.

Figure 27: Comparative analysis of phagocytosis and intracellular growth of *Mycobacterium tuberculosis* complex strains in human monocyte-derived macrophages by self-reported donor ancestry.

Average of the uptake of *Mycobacterium tuberculosis* complex strains (*Mtbc*) of L6 (**A**) and L4 (**B**) in human monocyte-derived macrophages (hMDM) from donors of self-reported ancestry that co-localizes with the endemicity of L6. Uptake and intracellular growth of *Mtbc* strains of the L6 and L4 (**C**, **D**, and **E**). Average of the intracellular growth of *Mtbc* strains of L4 and L6 at 24 hpi in hMDMs of different self-reported ancestries (**F**). The infection was carried out at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the uptake of *Mtbc* strains was determined immediately at 4 hpi. Statistical results are based on unpaired T-test (**A**, **B**, **F**) or One-way ANOVA with Bonferroni post hoc test correction among donors (**C** and **D**) or between strains and the H37Rv ATCC strain (**E**). Bars show the mean and standard deviation of the average colony-forming units (CFU) values from three infection wells per strain and the mean of nine infection wells per lineage, respectively. For the data on 96 hpi and 168 hpi, refer to Figure S12. Mean, standard deviation, and only statistical results that were not significant (ns >0.05) are depicted in the figures. ns, not significant; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; hpi, hours post-infection; MOI, Multiplicity of Infection; ATCC, American Type Culture Collection.

Figure 27 (Continued)



6.3.7 Inflammatory Profiling Response of Human Macrophages to Distinct *Mycobacterium tuberculosis* complex

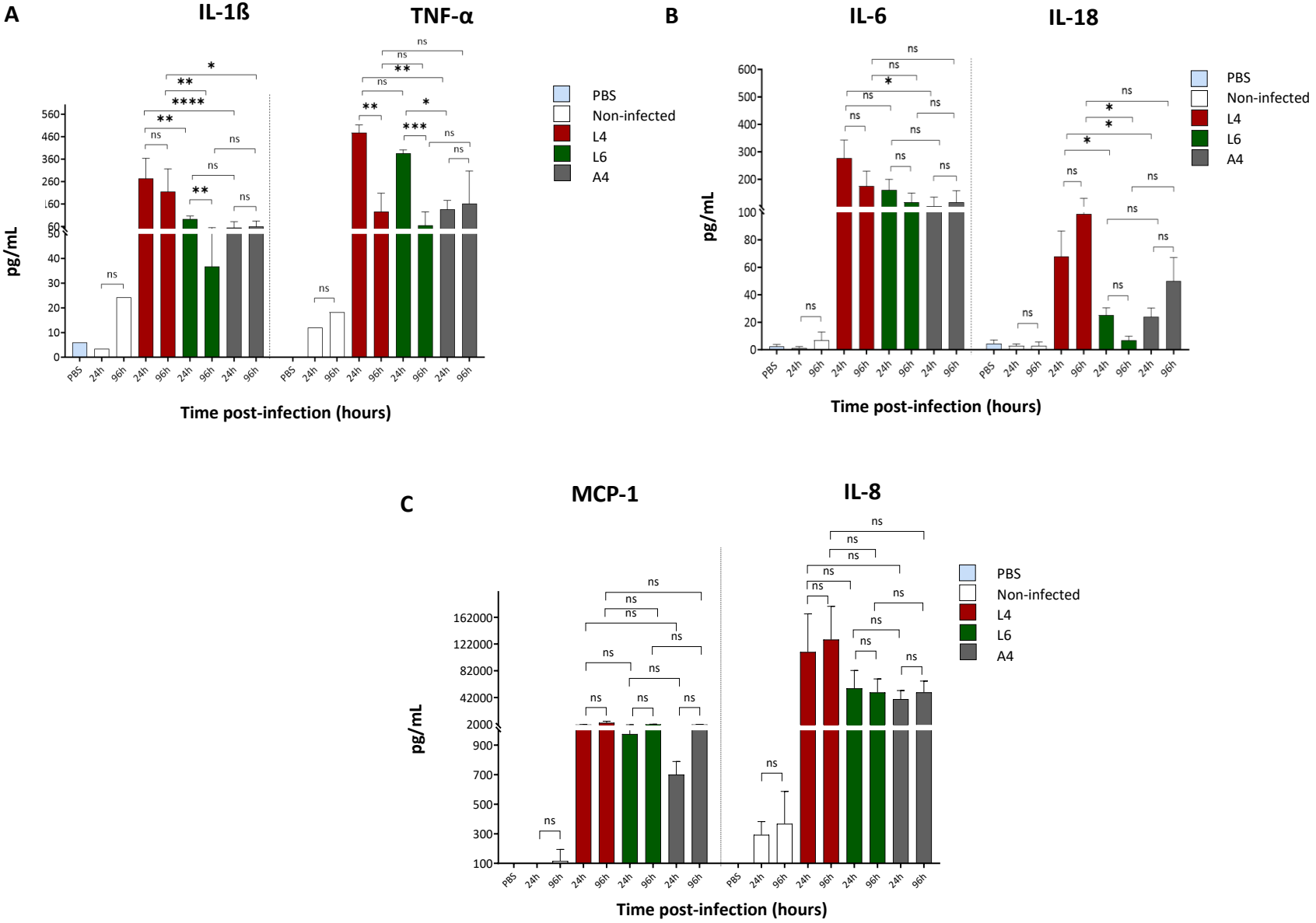
To investigate whether L4, L6, and A4 strains differ in the induction of inflammatory cytokines and chemokines in human macrophages, supernatants of the previous *in vitro* macrophage infection experiments were assessed with a human inflammatory panel of 13 cytokines and chemokines (see 5.3.5.2 Multiplex Cytokine-Chemokine Profiling in 5. METHODS). IL1-ß level induced in macrophages infected with L4 strains was significantly different compared to L6 and A4 strains at 24 hpi and 96 hpi.

However, within macrophages infected with the L4 strains, the levels of IL1-ß did not differ between 24 hpi and 96 hpi in contrast to L6 strains (Figure 28A). TNF-α production was significantly higher at 24 hpi and decreased at 96 hpi in macrophages infected by L4 and L6 strains; however, it remains similar at both time points in those infected with A4 strains (Figure 28A). IL-6 was produced in a similar range among macrophages infected with L4, L6, and A4 strains at 24 and 96 hpi, except in those with L4 strains, which differed with A4 strains at 96 hpi (One- way ANOVA, $p<0.05$). No differences in IL-6 production between 24 hpi and 96 hpi were found within macrophages infected by L4, L6, and A4 strains (Figure 28B). Higher IL-18 levels were found in macrophages infected with L4 than with L6 strains at both time points and with A4 strains at 24 hpi (Figure 28B). MCP-1 and IL-8 were consistently induced at both time points (Figure 28C). These data suggest that modern L4 strains induce a dominant inflammatory response compared to the ancient L6 and animal A4 strains.

Figure 28: Inflammatory response of human monocyte-derived macrophages to distinct *Mycobacterium tuberculosis* complex strains.

This assay was conducted on cell culture supernatants collected from human monocyte-derived macrophages (hMDMs) that were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli) with three representative strains of L4, L6, A4 and not infected hMDMs. Supernatants were collected at 24 and 96 hours post-infection (hpi). Three infection macrophage wells were tested per strain, time point, and donor. Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The production of six detected cytokines-chemokines is depicted in the figure, namely IL-1ß and TNF-α (**A**), IL-6 and IL-18 (**B**), and MCP-1 and IL-8 (**C**). The protein concentration in pg/mL was plotted on the y-axis. The controls (PBS and not infected [NI] at 24 hpi and 96 hpi) and experimental conditions (infected with L4, L6, and A4 strains at 24 hpi and 96 hpi) were represented on the x-axis. The stacked bars compared the mean production of these cytokines-chemokines at 24 hpi and 96 hpi within a lineage and across lineages. Three strains represent each lineage. Each strain comprises the averaged values of six donors. PBS, Macrophage Infection Media (MIM), and cells were used as controls. The MIM values were subtracted from non-infected and infected hMDMs. The mean, standard error of the mean, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. Data were obtained from six independent infection experiments. The statistical results shown in the figures are based on an unpaired T-test among both time points within the same lineage strain and on One-way ANOVA with Bonferroni post hoc test correction-multiple comparisons across distinct lineages at 24 hpi and 96 hpi. ns, not significant; MDMs, Monocyte-Derived Macrophages; *Mtbc* for *Mycobacterium tuberculosis* complex; NI, Non-infected; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection.

Figure 28 (Continued)



6.3.8 Human Ancestry-specific Inflammatory Profiling of Macrophages in Response to Varied *Mycobacterium tuberculosis* complex Strains.

In this study, macrophages of African origin infected with L4 strains showed significantly higher levels of IL1-β than those infected with L6 and A4 strains at 24 hpi and 96 hpi (Figure S13A). hMDMs infected with L4 strains also showed a higher production of IL-6 and TNF-α than A4 strains at 24 hours post-infection, although levels were similar at 96 hpi. In contrast, there were no notable differences in IL-6 production in hMDMs infected with L6 strains or across different lineages (Figure S13B and Figure S13D). While levels of IL-8, MCP-1, and IL-18 were higher in infected hMDMs compared to uninfected hMDMs, there were no significant differences in levels of these cytokines between single or multiple lineage strains at both 24 hpi and 96 hpi (as shown in Figure S13C, Figure S13E, and Figure S13F).

Similarly, macrophages of European (German) origin infected with L4 strains had higher levels of IL1-β than those infected with L6 and A4 strains at 24 hpi and 96 hpi. However, IL1-ß production decreased at 96 hpi in hMDMs infected with L6 strains (Figure S14A). As shown in Figure S14B, infection with L4, L6, and A4 strains induced a higher production of TNF-α at 24 hpi, which decreased significantly at 96 hpi (T-test, L4, and L6 $p = < 0.001$ and A4 $p = < 0.01$). No further differences in the production of MCP-1, IL-6, IL-8, and IL-18 levels were found in hMDMs infected by single lineage strains or across lineages at both time points (t-test, $p = > 0.05$ and One-way ANOVA, $p => 0.05$ respectively), (Figure S14C-F). Finally, the study compared the immune response induced hMDMs from African and European (German) donors. The findings revealed no significant differences in IL-1ß, TNF-α, MCP-1, IL-6, and IL-18 cytokine production between the two groups (Figure 29, Figure 30, and

Figure 31), except for IL-8. Notably, IL-8 levels were significantly higher in African hMDMs when infected with L4 strains at 96 hpi and with L6 strains at 24 hpi and 96 hpi, as demonstrated in Figure 29 and Figure 30. This analysis showed that hMDMs exhibited consistent inflammatory responses, irrespective of the donor's ethnicity, except for IL-8 levels.

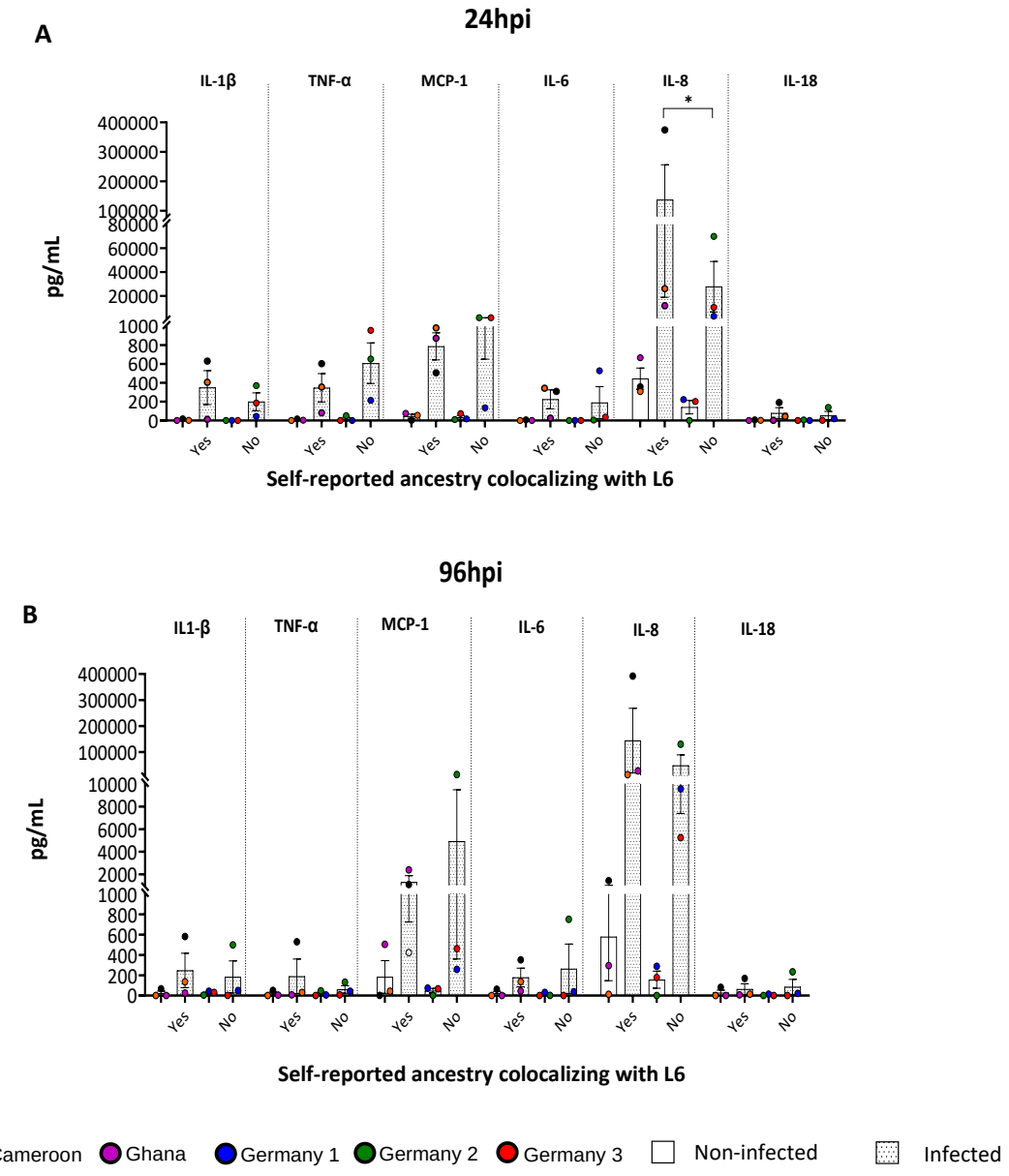


Figure 29: Comparison of the inflammatory response induced by L4 strains in human monocyte-derived macrophages based on self-reported ancestry colocalizing with L6 at 24 (A) and 96 (B) hours post-infection.

Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The stacked bars represent the mean production of six detected cytokines-chemokines: IL-1ß, TNF-α, MCP-1, IL-6, IL-8, and IL-18. Each bar represents three donors colocalizing with Lineage 6 (Yes [Nigerian, Cameroonian, Ghanaian]) and no colocalizing with Lineage 6 (No [Germans]). PBS, Macrophage Infection Media (MIM), and supernatants from non-infected hMDMs were used as controls. The MIM values were subtracted from the non-infected and infected hMDMs. Mean, standard error of the mean, and significant statistical results (*P < 0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are shown. Statistical results were calculated based on Two-way ANOVA multiple comparisons with Bonferroni correction. Data were obtained from six independent infection experiments (Self-reported African donors, n=3 and Self-reported European donors = 3). L4, Lineage 4; hpi, hours post-infection.

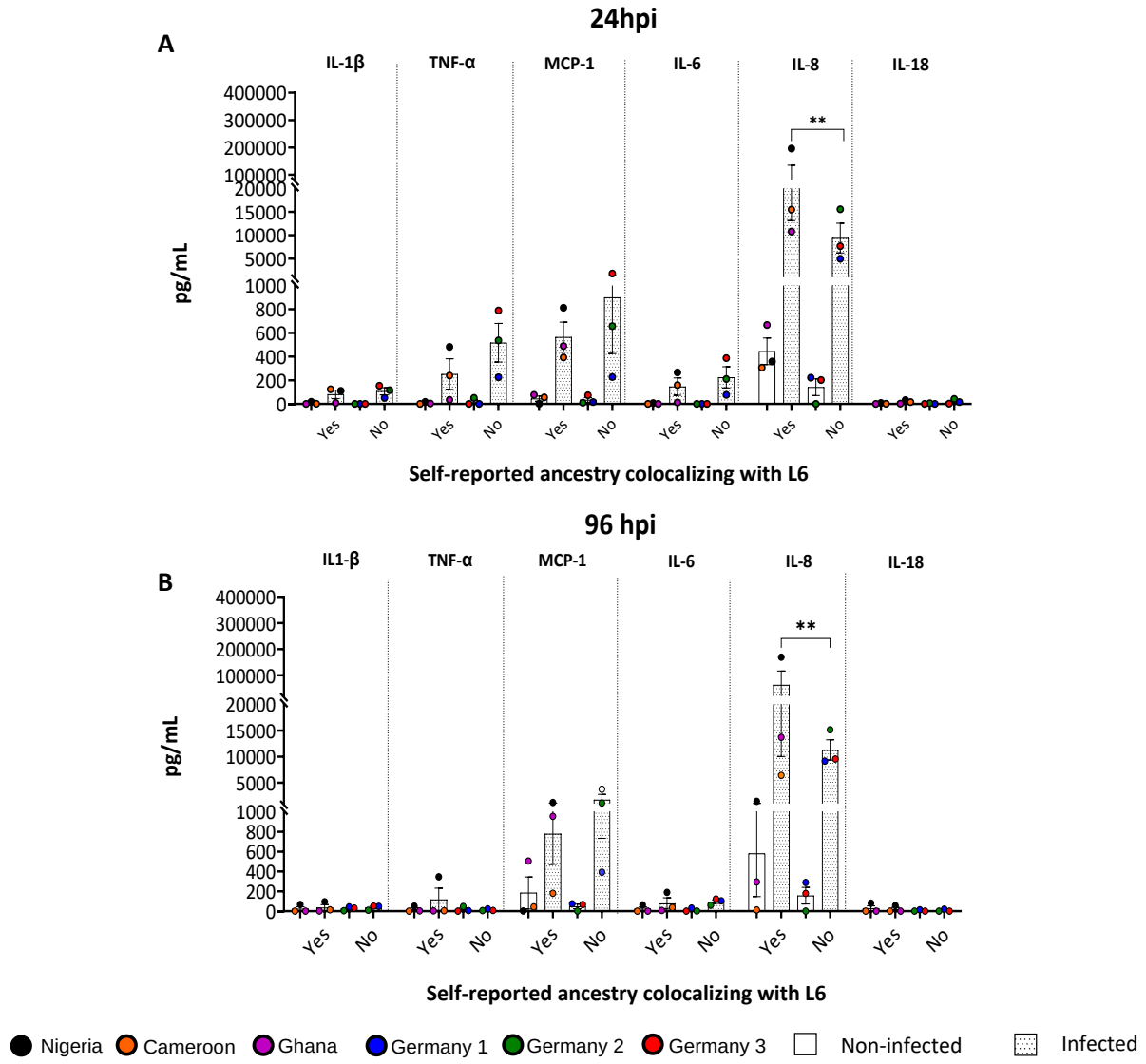


Figure 30: Comparison of the inflammatory response induced by L6 strains in human monocyte-derived macrophages based on self-reported ancestry colocalizing with L6 at 24 (A) and 96 (B) hours post-infection.

Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The stacked bars represent the mean production of six detected cytokines-chemokines: IL-1ß, TNF-α, MCP-1, IL-6, IL-8, and IL-18. Each bar represents three donors colocalizing with Lineage 6 (Yes [Nigerian, Cameroonian, Ghanaian]) and no colocalizing with Lineage 6 (No [Germans]). PBS, Macrophage Infection Media (MIM), and supernatants from non-infected hMDMs were used as controls. The MIM values were subtracted from the non-infected and infected hMDMs. Mean, standard error of the mean, and significant statistical results (*P < 0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are shown. Statistical results were calculated based on Two-way ANOVA multiple comparisons with Bonferroni correction. Data were obtained from six independent infection experiments (Self-reported African donors, n=3 and Self-reported European donors = 3). L6, Lineage 6; hpi, hours post-infection.

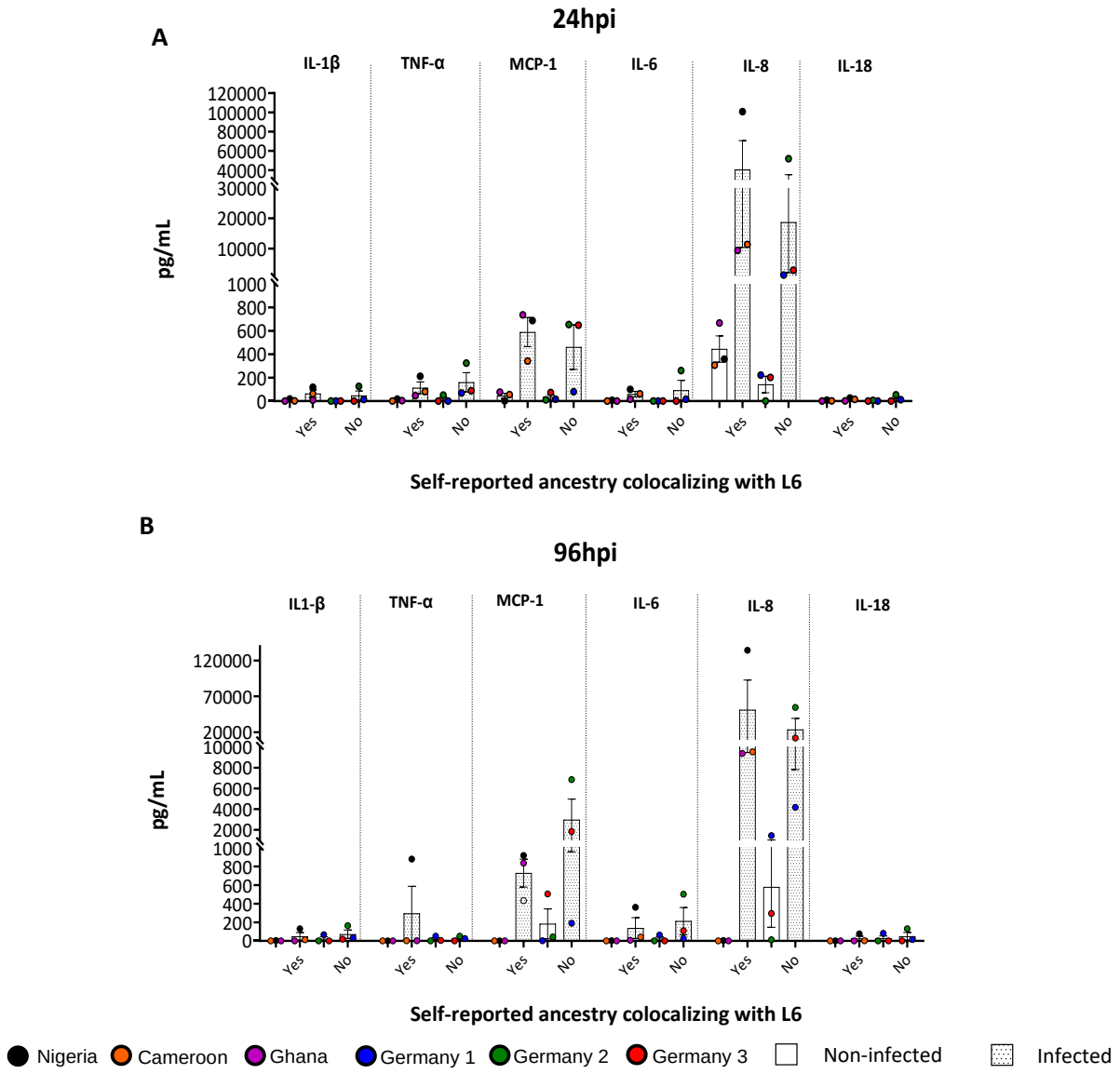


Figure 31: Comparison of the inflammatory response induced by A4 strains in human monocyte-derived macrophages based on self-reported ancestry colocalizing with L6 at 24 (A) and 96 (B) hours post-infection.

Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The stacked bars represent the mean production of six detected cytokines-chemokines: IL-1ß, TNF-α, MCP-1, IL-6, IL-8, and IL-18. Each bar represents three donors colocalizing with Lineage 6 (Yes [Nigerian, Cameroonian, Ghanaian]) and no colocalizing with Lineage 6 (No [Germans]). PBS, Macrophage Infection Media (MIM), and supernatants from non-infected hMDMs were used as controls. The MIM values were subtracted from the non-infected and infected hMDMs. Mean, standard error of the mean, and significant statistical results (*P < 0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are shown. Statistical results were calculated based on Two-way ANOVA multiple comparisons with Bonferroni correction. Data were obtained from six independent infection experiments (Self-reported African donors, n=3 and Self-reported European donors = 3). A4, Animal Clade 4; hpi, hours post-infection.

7. DISCUSSION

Members of the *Mtbc* are spread worldwide, especially strains of the modern lineages that are successfully driving the TB endemic in Europe and the Region of the Americas [121]. Genomics-based approaches such as WGS have provided a deep characterization of TB outbreaks in the last decade, revealing the circulation of strains of certain lineages and dangerous multidrug resistance variants [42]. However, these approaches have been mainly applied in high-income countries and are poorly implemented in low-income countries facing high disease burdens [43]. Thus, their public health authorities have limited knowledge about the nature of the ongoing TB transmission and might not control the spread of *Mtbc* strains efficiently, especially in remote settings inhabited by vulnerable populations such as the indigenous populations in which TB is a current public health threat with devastating consequences [26].

Furthermore, the nature of these isolated settings and the lifestyle of their populations favor the fast spread of virulent *Mtbc* strains that tend to form clonal outbreaks[212], [237], [255], [256], [257]. Considering the low mutation rates and genetic variability of the *Mtbc* strains in clonal outbreaks, analysis of transmission chains can be quite challenging with the default variant calling parameters used by the WGS bioinformatic pipelines [258]. Furthermore, previous work showed that *Mtbc* strains of distinct genetic lineages tend to co-localize with the human host [259]. However, it remains unclear whether co-localization has any biological effect on infection and immunological responses. This thesis work performed an extensive genomic investigation of a clonal TB outbreak affecting an indigenous population settled in the Colombian Amazon Region by using short-long read sequencing and customized variant calling approaches and further conducted an experimental examination of the causative outbreak lineage strain by using a characterized *Mtbc* panel of three distinct lineages in an *in vitro* infection macrophage model. The main findings, conclusions, and outlook of this thesis work are discussed in the following sections.

7.1 Tuberculosis Outbreak Investigation

7.1.1 The Clonal Expansion of a Lineage 4.3.3 Strain in an Indigenous Population.

The indigenous populations of the Amazon region are facing the highest TB burden in the region of the Americas and Colombia [26]. However, the lack of access to these patients limits our understanding of *Mtbc* population structure and transmission dynamics in these remote settings [21], [22], [23], [26], [260]. Therefore, genomic investigations are urgently needed, e.g., to determine the lineages of

circulating strains and to take proper actions to stop the spread of the disease. In 2016, a TB outbreak affected an indigenous population located in Puerto Nariño, a remote setting with a rainforest climate located on the shore of the Amazon River in Colombia [212]. This TB outbreak raised the alarms among national and international public health authorities due to the increase in TB prevalence and the potential risk to cross borders and affect other indigenous settings in Brazil and Peru [235]. A recent study based on classical molecular typing reported that the outbreak was caused by L4 strains with virtually no diversity [212].

In this study, WGS analysis confirmed the clonal expansion of one lineage 4 *Mtbc* strain that infected most patients in this remote Amazon setting in the short sampling time from March to October 2016. It has been suggested that L4 strains have been introduced into the indigenous pre-Columbian societies, presumably by European colonization [132]. Particularly, the finding that LAM 4.3.3 *Mtbc* strains accounted for 97% of the isolates in this study is in line with previous studies reporting LAM *Mtbc* strains as the most common sub-lineage in Northern South America (Ecuador, Peru, Brazil, and Venezuela), followed by Haarlem, Orphan, and T [134].

The SNP-based phylogenetic analysis revealed that the LAM 4.3.3 outbreak comprises two clusters defined by distinct mutations. Patients within these clusters originate from different settlements, and there was no apparent infection hotspot. However, it can be speculated that many infections may have occurred in the municipality of Puerto Nariño, which is a major hub in the region and frequently visited by people from different settlements. 24 out of 42 cases of the main transmission chain (Cluster I, node A) were from Puerto Nariño and the nearby settlements such as San Francisco, Ticoya, Veinte de Julio, and Puerto Esperanza, which comprises the main gathering area (i.e., fishing, shopping, and schooling). It is located in the Amazon River port that connects the indigenous settlements with the modern world. Indeed, standard *Mtbc* genomic analysis did not provide a significantly higher resolution than the classical 24-loci MIRU-VNTR genotyping method in this study setting. Here, only an extended variant calling approach could further increase the resolution of this outbreak and identify new unrecognized transmission chains.

7.1.2 An Alternative High-Resolution Approach Can Increase Resolution of Clonal Tuberculosis Outbreaks.

A newly assembled outbreak reference genome and alternative variant calling parameters enhanced resolution for analysis of the *Mtbc* outbreak compared to classical comparative *Mtbc* genomic analysis to the H37Rv reference genome. This alternative high-resolution approach revealed additional putative transmission events that can improve epidemiological investigations and interventions in *Mtbc*

outbreaks or settings with a clonal *Mtbc* population structure, which is especially needed in the highest TB burden. Additionally, including repetitive and resistance-associated regions of the outbreak reference genome increased resolution and suggested that multiple patients in this remote setting were likely infected with more than one clone.

Repetitive regions such as *PE/PPE* genes were included in the analysis. Of note, *PE/PPE* genes account for nearly 10% of the genes of *Mtbc* and might play a role in virulence mechanisms of host-pathogen interaction [261], [262]. However, their high GC content and sequence repetitions challenge standard WGS pipelines [133], [263], [264]. Nevertheless, recent work has shown that including genes with strict variant calling thresholds does not negatively impact the analysis and can provide novel and reliable information on mutations [264], [265], [266].

Furthermore, a new reference genome was assembled from an outbreak isolate of the largest node A using long-read sequencing (LRS) data. Molecular epidemiological approaches have typically used H37Rv as a mapping reference. However, its use as a sole reference genome has been discussed, showing many gaps and not considering the genetic diversity across all nine distinct human-adapted *Mtbc* lineages and sub-lineages [267], [268]. Thus, it is intriguing whether using of local reference genomes enables an improved resolution power, especially those caused by lineages other than lineage 4. Along these lines, Lee *et al*. have suggested deep sequencing in combination with the assembly of an outbreak-specific reference genome based on LRS as an alternative approach to enhance resolution and identify super-spreaders in highly clonal outbreaks [255]. However, this method is limited by high costs. Here, both approaches, i.e., including repetitive regions for variant calling and a new outbreak reference genome, led to an increased outbreak resolution and several differentiating SNPs.

These findings are in line with a recent study that reported a better resolution in repetitive regions of *M. tuberculosis (Mtb*) clinical strains with LRS over Short-Read Sequencing (SRS) [266]. This study also suggests that combining LRS and SRS increases resolution. Furthermore, these findings add evidence to the statement by De Coster *et al.* that, indeed, many reference genomes based on short-read data might be missing critical information leading to false positive and negative variants [269], for instance, in the genome of *Pseudomonas koreensis,* LRS data have revealed that several identical repeat pairs comprising genes involved in advantageous fitness are missing [270].

With this high resolution obtained in this study, evidence was also found that multiple patients are potentially infected with more than one strain, confirming previous studies showing a considerable intra-patient heterogeneity, however, often related to one infecting clone [271]. Previous work has documented that mixed infections might account for 30% of TB cases, of which the mixed infection rates, i.e., with distinct strains, are estimated to be from 19% in pulmonary samples to 51% in extra-pulmonary

and pulmonary combined samples [272], [273], [274], [275]. However, assessing mixed infections is extremely difficult in settings with a clonal population structure and with classical comparative *Mtbc* genomic pipelines [276]. This scenario is also particularly true for MDR-TB high-incidence regions, often dominated by a few major clades with drastically reduced genetic diversity [76], [277]. Here, it is crucial for molecular diagnostics to also identify minority *Mtbc* populations in one patient, which could have a huge impact on diagnosis, treatment decisions, and clinical outcomes. This study was limited to the short study time and excluded deletions and insertions. Future work should be based on longitudinal sampling, enabling prospective molecular surveillance and allowing tracing of the evolution of drug resistance in locally dominating clades/strains.

7.2 Pathobiology of *Mycobacterium tuberculosis* complex Strains

7.2.1 *Mycobacterium tuberculosis* complex Strains Exhibit Lineage-Specific Growth Patterns

The outcome of TB infection is defined by the interaction of the pathogen with the host based on the lineage-specific variation of the genetic background of *Mtbc* strains and the human genetic diversity [259]. However, little has been documented about the impact of the *Mtbc* genomic diversity on infection, concretely the pathobiology characterization behind the variable success of *Mtbc* strains. To help fill this gap, this investigation delves deeper into the pathobiology of the *Mtbc* strains observed in different host populations. Specifically, I investigated the *Mtbc* strains from the human (L4, L6) and animal (A4) adapted lineages[130], [166], [170], [198], [259]. The modern and widespread L4, concretely the LAM sub-lineage, is particularly prevalent in South America [278]. The ancient and geographically restricted L6 represented by West African II sub-lineage strains are dominant in the West Africa region [151]. The animal-adapted strains comprised the Animal Clade 4 represented by *Mbv* strains.

This study showed different lineage-specific growth profiles in Mycobacteria Liquid Media. L4 strains grew more rapidly than L6 and A4 strains. Concretely, The L4 strains exhibited the most rapid growth compared to the L6 and A4 strains, with the highest C_{max}. In contrast, the L6 strains had the lowest $C_{max,}$ and the A4 strains had an intermediate C_{max}. These findings align with the observations made by Sarkar *et al.* [164], [279], supporting the hypothesis that the increased *in vitro* growth of the L4 strains may indicate a more virulent trait. Thus, these observations suggest that these strains have an increased capacity to cause active disease in the host with higher bacterial burdens than the slower L6 strains. Although the slow growth tendency of L6 and A4 strains can be somewhat explained by a variant in the

pykA gene, which encodes for a pyruvate kinase, L6 strains have been linked to a less virulent trait and thus having a lower progression to active disease in the host [280], [281], [282].

7.2.2 *Mycobacterium tuberculosis* complex Lineages Strains Have Their Transcriptomic Signatures.

Extensive research using WGS has revealed the genetic diversity of *Mtbc* strains and their allocation in distinct lineages, which have co-evolved with humans[283]. However, little is known about how evolutionary forces have impacted the transcriptomic landscape of these lineages. This study characterizes the transcriptomic landscape of three strains of *Mtbc* in detail: modern (L4), ancient (L6), and animal (A4, *Mbv*) lineages. Analysis revealed distinct lineage-specific transcriptomic signatures *in vitro*. For example, L4 and L6 strains had higher expression levels of several genes related to cell wall and cell processes, conserved hypothetical proteins, intermediary metabolism, PE/PPE and respiration, virulence, detoxification, and adaptation, compared to A4 strains. In contrast, L6 Vs. A4 strains had fewer genes up-regulated for these biological functions.

RNA-Seq data analysis showed differences in gene expression patterns between L6 and L4 strains that reflect their distinct virulence. For instance, L6 strains had, in general, fewer genes up-regulated in metabolism, cell wall, and processes in comparison to L4 strains, which resonates with the slow *in vitro* growth of L6 strains in liquid media and their tendency of a slower disease progression in patients unlike the L4 strains[284]. Another exceptional difference was the upregulation of esxO (Rv234C) and esxV (Rv3619c), which are among the most immunodominant antigens recognized by the human system[285] and prophage-like elements such as phiRv1 and phiRv2 only in the L4 strains, being another clear indication of the pathogenicity of the L4 as these elements are relevant in virulence, toxin biosynthesis and secretion, fitness cost, genomic variation, and evolution[286]. On the other hand, strains of L6 strains had fewer genes up-regulated related to virulence response, particularly the Toxin-antitoxin (TA) system genes VapC3 (Rv0549c) and VapB3 (Rv0550c), which were also reported in L6 strains by Chiner-Oms *et al.* [287]. The TA loci are exceptionally present in the genome of *Mtbc* strains compared to other microbial genomes [288] and comprise several operons involved in cell synthesis, DNA replication, cellular division, and the ability to cope with stressful environments such as the phagosome. These environments are characterized by low – moderate pH, limited nutrients or oxygen, and oxygen and nitrogen radicals [289].

Furthermore, TA systems genes have been markedly up-regulated in *Mtbc* persisters´s cells, which are a crucial element in the slower progression of the disease and its elimination [48], [49]. Moreover, this study identified four genes of the mbt operon down-regulated in L6 strains, e.g., *mbtI* (Rv2386c), *mbtB*

(Rv2383c), and *mbtD* (Rv2381c). However, unlike Chiner-Oms and colleagues[287], the *mbtC*, *mbtH*, *mbtE*, *mbtG*, and *mbtF* genes were not identified. The *mbt* operon is involved in synthesizing the siderophore Mycobactin, which is involved in acquiring iron in restrictive environments such as the macrophages [290], [291]. Indeed, the lack of this siderophore has been associated with reduced growth in macrophages [292]. Overall, the marked transcriptomic signatures associated with a slow metabolism and survival in stressing environments are consistent with a limited virulence of the L6 strains that might be required for a slower progression from latent to active disease and their variation in transmission in the human host[280].

On the other hand, L4 strains displayed several transcriptomic signatures associated with metabolism, growth, and virulence, which were especially marked in the clinical strains. Among these signatures are those involved in nutritional virulence, such as the *ctpC* gene (Rv3270). This gene encodes for a P_{1B}-ATPase metal exporter that allows the pathogen to compete for transition metals with the host, who, in an attempt to avoid the proliferation of intracellular pathogens, constantly restricts microbial access to nutrient supply [293]. Consequently, the higher production levels of these transporters might outcompete the host efforts, leading to a nutritional imbalance in favor of the pathogen [294]. Another interesting signature was the upregulation of the *eis* gene (Rv2416) in L4 strains, remarkedly in the LAM clinical strains. The *eis* or "enhance intracellular survival" gene is part of the SigA regulon. It has been associated with promoting intracellular subsistence of mycobacterium strains in macrophages [295], [296] and in conferring kanamycin resistance when overexpressed [297]. High activity of the *eis* gene has been reported in other members of the modern lineages like L2.

In contrast *eis* fostered the intracellular growth of W-Beijing clinical isolates significantly compared to H37Rv in monocytes [298]. These observations confirm and extend those of Homolka and colleagues. They suggest that specific alliterated expression patterns in clinical strains might benefit their functional pathogenicity in the host over the laboratory strains [136].

In contrast to the A4 (*Mbv*) strains, it was observed the transcriptomic signatures of the L4 strains were once again dominant in pathogenesis by the upregulation of encoding genes for the prophage-like elements, phospholipase C, metabolic processes, and the PE/PPE family. These genes were also recently found lineage-related in differential expression among L2 and L4 strains [63]. Exceptionally the PE/PPE family was up-regulated in the L4 strains, this family represents 7-10% of the *Mtb* coding genome [299], and it is involved in virulence mechanisms of host-pathogen interaction such as the modulation of the host´s innate immune system [300], aiding the pathogen's ability to withstand host-induced stress[301], helping with the formation of pore-like channels to resist nutrient starvation [299], mimicking host transcription factors to promote intracellular growth and manipulating host cell fates [302]. Among the

PE/PPE genes identified in this study highlight *PPE59* (Rv3429), *PPE51* (Rv3136), *PPE67* (Rv3739c), *PE_PGRS38* (Rv2162c), and *PPE38* (Rv2352c) genes.

The expression of these key genes linked to host-pathogen interaction provides evidence that L4 strains evolved likely to specifically adapt to humans. For example, the PPE*51* gene is essential to acquire nutrients to survive the limited nutrient environment of its niche and, therefore, provide nutrient starvation resistance [303], [304]. The *PPE38* gene is crucial for producing the PE-PGRS and PPE-MPTR proteins [305] virulence attenuation for long-term intracellular survival. Indeed, it has been documented that PPE38 mutations in the genome of the modern Beijing (L2) strains increase their virulence [306]. Furthermore, expression of essential metabolic genes was observed, e.g., the phospholipase C (PLCs): *plcA* (Rv2351c), *plcC* (Rv2349c), and *plcB* (Rv2350c). These PLCs genes encode essential virulent factors involved in several cellular tasks that weaken the host to secure the infection [307], [308] for example, in membrane lysis, lipid metabolism [309], intracellular signaling [310], cytotoxicity, and other cell death mechanisms [311], [312].

Moreover, it was observed that *narGHJ* genes that are critical in non-replicative states during persistence were exclusively up-regulated in L4 strains but absent in *Mbv* [313]; *narGHJ* is essential to adapt to hypoxic conditions by the regulation of nitrate reduction of *Mtb*, an ability that is absent *in Mbv strains*[314], [315], [316]*,* another critical gene absent in *Mbv* strains due their lack of RD12 was the *CYP130 gene* whose mutations have shown reduced growth of *Mtb* and pathology in macrophage cells and mice[74] [317]. This observation, therefore, provides evidence of the A4 (*Mbv*) strains being not evolutionary shaped to persist and sustain the infection in the human host; on the contrary, L4 strains have exceptionally preserved and presumably extended through reductive evolution the host-pathogen and metabolic genes as essential elements for the pathogenic intracellular success within the human host.

On the other hand, this study investigated the *in vitro* differential expression profiles between L6 and A4 (*Mbv*) strains. Although both lineages differ in their host tropism, they have low genetic diversity. It is known that L6 strains share a common ancestor with *Mbv* strains[170]. Also, further investigations have revealed that both lineages are phylogenetically closely related and are less virulent in humans in comparison with other *Mtbc* lineages strains [117], [170], [318], [319]. For instance, they share substantial pseudogenes, genetic deletions, e.g., RD7, RD8, RD10, RD702, and RD900 [320] [215] [321], and phenotypical characteristics such as the negative nitrate reductase activity[313], [314], [315], low niacin production[144], [322], dysgonic morphology[144], [145], [322], [323], slower growth[148], [324] and no

preference for glycerol as a sole carbon source[325]. These findings have led to the speculation that the original reservoir of L6 strains could be animal and not human [281].

This study found that L6 and A4 strains have a similar transcriptomic landscape overall, with half of the genes being differentially expressed compared to L4 strains. This finding confirms that strains of two lineages are genetically close. However, minor differences between the two suggested that L6 strains have transcriptomic signatures that benefit a preference for the human host. These strains displayed dominant gene expression in encoding genes associated with virulence factors in host-pathogen interaction, intracellular survival, redox properties, sugar metabolism, and hypoxia (Table S6). Some of these genes have also been detected with higher expression levels in L4 strains compared to A4 (*Mbv*) strains [326]. Among these genes, Bacterioferritin BfrB and WhiB3 play key roles in iron homeostasis[327]and adaptation to the intracellular environment[328], respectively. Additionally, *esxO*, which functions as an immunodominant antigen recognized by the human immune system[285], and yrbE4A, part of the operon mce4, which is involved in the import of cholesterol for the long-term survival of the pathogen during LTI [329], were highlighted. Phospholipases and important biliverdin reductases whose encoding genes are missing in the *Mbv* genome were also identified [32] [330].

On the other hand, A4 strains showed marked up-regulated expression levels of encoding genes for the immunodominant antigens mpt83 and mpt70 and several transcriptional regulators. This observation aligns with previous research based on microarrays suggesting that the higher production of these immunodominant antigens is associated with the higher virulence of *Mbv* strains compared to *Mtb* strains [326]. Therefore, they have been proposed as diagnostic markers to detect bovine TB in cattle [331], [332]. This study also observed the increased virulence of *Mbv* strains compared to the attenuated virulence of L4 strains upon infection in bMDMs, which has also been observed in alveolar bovine macrophage infections [185]. These minor differences between the expression repertories of L4 and L6 might cause distinctive physiological and adaptive behaviors based on the host immunological environment. It suggests that through evolution, L6 strains have been shaped with advantageous pathogenic equipment to adapt to the human host compared to the *Mbv* strains, which seem to have a strong adaptation with the bovine host.

Overall, this study provides a transcriptomics framework of the L6, L4, and A4 lineages strains comprising clinical and ATCC reference strains, which is the first of its kind. However, the analysis did not include repetitive regions and structural variations such as deletions, inversions, translocations, and insertions; thus, additional information may have been missed. Therefore, more extensive research based on LRS may reveal further differences in the transcriptomic landscape of *Mtbc* strains when considering the effects of these regions. In summary, strains investigated in this study representing different *Mtbc*

lineages strains showed specific growth rates and a distinct transcriptomic landscape. This study moved forward to investigate further the infection and immunological responses that could explain their specialization and adaptation to distinct hosts.

7.2.3 Sympatric and Allopatric *in Vitro* Macrophage Infection Model to Study Infection Characteristics of *Mycobacterium tuberculosis* complex Strains

Numerous studies conducted around the world using WGS on thousands of *Mtbc* strains in various human populations have indicated that *Mtbc* strains of distinct lineages are specific to certain geographic regions and may be adapted to particular human hosts, which could explain differences in transmission rates, disease outcomes, and treatment responses [259]. When two species overlap geographically, it is referred to as sympatry, which is more favorable for pathogen infection, unlike allopatry, which describes a non-fitting host-pathogen association related to restricted pathogen transmission [333] [334]. A recent study found that the transmission of *Mtbc* strains between sympatric and allopatric contact-pathogen pairs of widespread and restricted *Mtbc* strains resulted in a 38% decrease in infection among contacts with allopatric compared to the sympatric exposures[129].

Macrophages are the cells that *Mtbc* bacilli rely on to survive, replicate, and colonize, making them a primary focus for studying the pathobiology of *Mtbc*.[335] Over the past 71 years, scientists have developed various models and tools to better understand the pathogenic dynamics of TB [336]. Animal and immortal cell lines, such as guinea pigs, mice, murine bone marrow-derived macrophages, and THP-1 cell lines, have been the most commonly used models [337]. While these models are practical and provide a homogeneous cell population and diverse knockouts, they cannot replicate the complexity of the human host. Therefore, extrapolating infection results from these models to humans is a major challenge.

For example, cell lines and animal models differ in inflammatory and metabolic aspects and infection susceptibility, leading to different intracellular environments the *Mtbc* bacilli must face. Therefore, the laboratory “disease outcomes “are depending on the selected model [239] [338]. Some studies have shown no significant difference in cytokine production, mycobacterial uptake, and cellular viability between THP1 cells and primary human-derived macrophages [339] [111]. Recently, human organoids have been introduced as a novel approach to understanding the biology of diseases in humans. These organoids are created from pluripotent cells but they encounter significant obstacles, such as difficulties in establishing universal protocols, scaling up to whole organs, and, notably, higher costs [340]. Although

human alveolar macrophages are the ideal cells for an *in vitro* cell infection model, procuring them from humans is a complex process, which makes using them in the laboratory challenging.[341].

Despite these challenges, studying the human context in experimental models is critical to fully understanding the host-pathogen dynamics of TB, especially in different human populations. These populations may be more or less adapted to specific strains of *Mtbc*, which can affect disease susceptibility and outcomes [259]. To investigate these dynamics closely, I optimized an *in vitro* macrophage infection cell model based on blood monocyte-derived macrophages from healthy human and bovine donors to study the L4, L6, and A4 strains in their respective adapted and preferred host contexts. In this study, I investigated the biological impact of sympatric exposures of modern-widespread L4 and ancient-restricted L6 strains in human hosts of different ethnicities (African and European) in detail. Additionally, I examined the ability of A4 (*Mbv*) strains to establish and sustain infection in the human and bovine hosts.

The results obtained indicate that the phagocytosis rate of all *Mtbc* strains investigated was similar in hMDMs from donors of African ancestry, for instance, 29.47%, 30.43%, and 24.66% for the L4, L6, and A4 strains, respectively. However, although the phagocytosis rate of L4 strains was similar in hMDMs from donors of European ancestry, differences for the L6 and A4 strains could be identified with 2.87% and 2.95% uptake rates, respectively. These observations align with results published by Reiling *et al.* investigating clinical isolates of L6 in European hMDMs. In contrast, the phagocytosis rate of L6 strains by African hMDMs differed from the previous studies, showing a considerably lower phagocytosis rate in general [199], [282]. The substantial reduction in uptake of L6 strains in European hMDMs could be related to the host ancestry. Indeed, there is evidence of the potential impact of the human ancestral genetic makeup on TB infection. For instance, Thye *et al.* showed that the variant G57E of the Mannose-binding lectin (MBL) is associated with TB protection in Ghanaian patients [102]. MBL is an essential element in phagocytosis of pathogens. The host cell´s genetic variability may manipulate infection fate [131]. As described above, the intracellular growth of L4 and A4 Vs. L6 strains progressed almost similarly in African hMDMs. The uptake and growth of L4 strains by African hMDMs might reflect the ongoing success of L4 strains causing TB in African patients; indeed, there is current evidence that L4 strains in West Africa are gradually outcompeting L6 strains [131].

This study found low efficiency in phagocytic uptake in macrophages from donors who did not co-localize with the infecting strain, supporting previous epidemiological observations of sympatric combinations, indicating that the reduced transmissibility of geographically restricted strains may be due to lower uptake and growth of L6 strains in allopatric macrophages after initial exposure. This indication contrasts with the geographically widespread L4 strains that appear to have high compatibility in sympatric and

allopatric hosts [120], [333]. Therefore, it is suggested that the limited success of geographically restricted *Mtbc* strains, such as L6, in driving the global spread of TB compared to modern strains like L4 and L2 may be due to their reduced infectivity in allopatric populations [129]. However, this study did not examine the hypothesis that host ancestry-related genes might be responsible for phenotypic differences in the uptake of *Mtbc* bacilli lineages. That was because human sequence data was not obtained. However, recent research has indicated that up to 9.3% of genes expressed in macrophages show ancestry-related differences in response to infection.[342]. Further research is needed to study more combinations of TB strains and host genetics. For this study, L6 and L4 strains were chosen based on their phylogenetic distance and the availability of host macrophage donors. Despite causing a significant proportion of TB cases in West Africa, L6 strains have been associated with reduced virulence in animal models[125], HIV[162], and diabetes[343] and may represent a lineage of TB with generally lower virulence.

7.2.4 Human Monocyte-Derived Macrophages Induce Homogenous Inflammatory Immunological Responses Irrespective of the Infective *Mycobacterium tuberculosis* complex lineage Strain.

Previous work has documented that the human immune response and TB disease severity might be influenced by the genetic diversity of the *Mtbc* strains implying significant effects in diagnostics, treatment, and vaccine development [98], [113], [116], [119], [139], [344], [199], [345], [125], [131], [189]. To determine the impact of *Mtbc* strain diversity on the host innate response, this study analyzed the global human inflammatory responses induced in hMDMs by the modern-widespread (L4), ancient-restricted (L6), and the animal-adapted (A4) strains at different stages of the infection.

In this study, I found that the cytokines IL-1ß, TNF-α, IL-6, IL-18, and the chemokines MCP-1/ CCL2 and IL-8/ CXCL8 represented the human inflammatory response. Specific cytokines were observed to be induced by strains of specific *Mtbc* lineage strains during infection. For instance, L4 strains consistently induced IL-1ß and IL-18 more than L6 and A4 strains, while L4 strains induced IL-6 more than A4 strains. No differences were found in the production of TNF-α by hMDMs infected with L4 and L6 strains in the early stages of infection. Studies have shown that IL-1ß, IL-18, and TNF-α play a crucial role in protecting the host against TB by mediating innate and adaptive immune defense programs and controlling the pathogen [346], [347] [348] [349], [350]. Mice lacking these cytokines have demonstrated high susceptibility to the infection, marked pathobiology, rapid pathogen dissemination, and poor host survival with higher mortality rates [351], [352]. Similar effects have also been observed with other infectious agents[353].

As IL-1ß, IL-18, IL-6, and TNF-α are known to be protective against virulent mycobacteria [354] and considering the observed higher virulence of the L4 strains over the L6 and A4 strains, a reduced inflammatory response would be expected for the less virulent strains and vice versa as this dynamics could offer advantage to progress to active disease suggested in other studies [139] [344] [355], [125], [131], [189]. However, the data obtained in this study clearly show that, although the induced protective inflammatory response does control the bacterial load at the early stages of the infection, it seems to be subverted at later stages as the intracellular growth of the L4 strains is no longer restrictive and controlled by the host. Thus, a fast and progressive infection succeeded in hMDMs; this was not generally the case for the L6 and A4 strains, as their growth was more restrictive upon infection. These results are in line with previous findings by Romagnoli, Petruccioli, and colleagues[356], showing enhanced intracellular replication in parallel with increased production of TNF-α, IL-1β, and IL-6 in hMDMs infected with L4 strains, in contrast to ancient lineage strains (L1 and L5). Also, they reported higher levels of autophagy induced by IL-1β; this also can explain the restrictive replication of L4 strains early during the infection[357].

Furthermore, It has been reported even higher levels of IL-1β induced by hypervirulent strains known to cause disseminated TB in humans, such as the L2 modern Beijing strains in murine macrophages [131]; however, it is uncertain whether this resembles the human context [344] thus it imperative to investigate in future work. Furthermore, the data generated in this study shows that IL-1β accumulated along with IL-6 and TNF- α in hMDMs irrespective of the lineage strain; however, IL-6 levels induced by L4 strains were again higher compared to A4 strains early in the infection. While IL-6 is not considered an essential protective cytokine in humans, it works together with IL-1ß and TNF-α as inducers of the acute phase response[358] and to control the intracellular growth of the pathogen[359], which can explain the general restrictive growth in hMDMs at 24 hpi. Indeed, it has been observed that knockout mice for IL-6 show a higher disease susceptibility reflected by higher bacterial load and delayed induction of TNF-α early in the infection[360].

Moreover, copious amounts of the chemokines MCP-1 and especially of IL-8 were noted compared to their respective non-infected hMDMs; however, their levels did not differ between lineage strains. This observation confirms the cell activation by the pathogen and the potent acuate response, as these chemokines play a key role in attracting acute and chronic inflammatory cells to contain the pathogen at the site of the infection [135] [348] [361]. Furthermore, the copious amount of these chemokines can indicate the development of active TB disease in some individuals [136], which might have been related to the host genetics and environment, e.g., cells harboring a polymorphism in the promoter of the MCP-

1 gene (2518G) and the encoding gene of IL-8 produce high levels of MCP-1 that seems to inhibit the expression of IL-12 increasing TB disease susceptibility [362], [363], [364].

A thorough data analysis indicates that the genetic diversity of strains of different *Mtbc* lineages studied in this research leads to a consistent cytokine response. This finding partially contradicts the previous report by Portevin and colleagues [122], which described a variable inflammatory response in hMDMs, showing reduced and delayed immune responses in modern lineages strains compared to ancient lineages strains. For example, L6 strains induced significantly higher levels of TNF-α, IL-12p40/p70, and IL-6 compared to L4 strains at 24 hpi, although some L6 strains showed similar responses to L4 strains in the early stages of infection. However, the responses induced by strains of different lineages were nearly identical later during infection. Similar results were observed in hMDMs and whole human blood from TB patients infected with other modern and ancestral lineages such as L2 (East Asia) and L3 (India and East Africa) [345], [365]. Possible explanations for this inconsistency with the findings of Portevin and colleagues [122] could be that some of the human monocytes used in that study were from frozen PBMCs, it has been reported cryopreservation can affect their functionality [366], [367], differentiated in the presence of GM-CSF and FCS, and the clinical isolates used were different to the ones used in the present study, it has been reported that clinical isolates from the same lineage may affect the immune response differently[368].

Conversely, the current study's results align with and expand upon those of Reiling [199]and Romagnoli [356]and colleagues, where no lineage-specific inflammatory responses or correlation with virulence were observed in hMDMs and animal infection models. For instance, the levels of TNF-α and IL-6 released by hMDMs infected with L4 and L6 strains were similar, and the levels of IL-1ß were higher in L4 than in L6 strains. Moreover, the higher levels of TNF-α induced by L4 strains in hMDMs reported in this study coincide with the findings of Sakar et al. [279] and Wang et al.[369]. Overall, the observations of this study suggest that the genetic diversity of the *Mtbc* strains is a potent innate immunological stimulus in hMDMs, which is led by the genetic diversity of the *Mtbc* strains. These host cells react with an acute and consistent inflammatory response to contain the pathogen after phagocytosis and early during infection. However, the immune response is not strongly influenced by this pathogen's genetic diversity and, therefore, does not impact the infection outcome at the assessed time points in hMDMs.

7.2.5 Inflammatory Immune Response of Human Sympatric/Allopatric Monocyte-Derived Macrophages-*Mycobacterium tuberculosis* complex strains combinations

The genetic diversity of *Mtbc* strains and their host populations have a significant impact on the progression and transmission of TB [139], [365] [259]. Studies have shown that the genetic contribution

of the host to the risk of developing TB varies across different continents [150], [151], highlighting the importance of including distinct human ancestries in TB research and experimental studies using human cells. To investigate the impact of *Mtbc* lineage strain diversity on the inflammatory immune response of their human hosts, this study analyzed the combination of hMDMs from African and European ancestries infected with sympatric and allopatric strains.

The combination of sympatric hMDMs of African origin (Ghanaian, Nigerian, and Cameroonian donors) did not elicit any significant immunological response when infected with L6, neither L4 nor A4 strains. These findings are consistent with the results reported by Tientcheu and colleagues [370], who observed similar immunological and metabolic blood gene expression profiles in untreated sympatric African populations infected with L6, L5, and L4 strains. Although a few cytokines were differentially induced, such as higher levels of IL-12p70, IL12A, and LR9 induced by L6 strains and higher levels of IL-15, IL-8, and MIP-1α induced by L4 strains [152], levels of IL-12p70 were not detectable in the present study, possibly due to inhibition by high levels of MCP-1. In addition, the present study observed that L4 strains induced higher levels of IL-1ß and TNF-α compared to L6 and A4 strains. This observation is consistent with the global human inflammatory response and reinforces the role of IL-1ß in restricting the growth of L4 strains early in infection, although this was not observed in the case of L6 strains. It is tempting to speculate that due to the introduction of modern L4 strains in recent years after years of co-existence with ancient L6 strains, L4 strains could act as a solid immunological stimulus for populations that might be susceptible to this modern lineage[371] [243], [372].

Upon analysis, it was found that the virulent L4 strains induced higher levels of TNF-α in sympatric hMDMs of European origin (German Donors), consistent with the cytokine trend in the global human inflammatory response. However, these TNF-α levels did not affect the replication rate of L4 strains after 24 hours, suggesting that a robust immune response does not necessarily correspond to reduced virulence of the pathogen. Additionally, a cytokine expression trend was observed towards lower cytokine production of TNF-α and IL-1ß in the European hMDMs later in the infection by all lineages strains and only for the L6 strains, respectively. This indicates that the immune response was not delayed, regardless of the *Mtbc* lineage strain. Nevertheless, the consistency of the induced immune response across lineages strains does not explain the significant restrictive uptake and poor ability of the L6 strains to replicate in the European hMDMs. It is essential to consider the host genetics and other aspects related to the bacilli that could impact the restrictive phagocytosis, such as differences in the surface composition of lipids[373], sugar, and proteins[374] among *Mtbc* lineage strains[375]. Therefore, further investigation is needed to fully understand the mechanisms behind the varying intracellular

pathogen replication and immune response of different *Mtbc* lineage strains in a given ancestral hMDMs.

Overall, when comparing ancestries colocalizing (African) and no colocalizing (European) with L6 strains on the immune response, specifically about the ability of hMDMs to contain the *Mtbc* lineage strains, it was found that the host ancestry does not have a significant impact on the induced inflammatory immune response. However, hMDMs from donors of African ancestry infected with L6 strains produced a higher amount of IL-8, an early response of macrophages after phagocytosis of the TB pathogen[376]. This observation indicates that African hMDMs are better at containing the *Mtbc* strains at higher rates than European hMDMs. However, this did not limit the intracellular growth of the L6 strains upon infection in comparison to L4 and A4 strains.

Currently, there is limited knowledge about the impact of human ancestry genetics on the immune response. Previous studies have reported that some populations of African ancestry might be high IL-8 producers compared to Caucasian individuals [161], and genetic variation in MCP-1 and IL-8 can affect the outcome of various diseases in different ethnic populations [161]. For instance, this variation in African ancestries confers a protective effect against asthma, while in Caucasian ancestries, it increases their risk of asthma [161]. Additionally, defects in phagocytosis and oxidative burst have been identified in patients infected with HIV, which supports the critical role of the IL-8 response in the proper ability of hMDMs to conduct efficient phagocytosis.

The findings of this study highlight the importance of considering host ancestry when developing targeted therapies for tuberculosis. Despite inducing a homogeneous inflammatory response in hMDMs, regardless of donor ancestry or *Mtbc* lineage strains, this inflammatory response cannot be used as a reliable predictor of the outcome of the infection. This study is significant as it is one of the first to examine the host responses of sympatric and allopatric combinations in an *in vitro* infection cell model. However, it does have some limitations that need to be considered. Future research should consider human ancestry genomics to better understand the interactions between sympatric and allopatric combinations in TB. The recent publication of the first draft of the human pangenome reference, which includes genetically diverse individuals from different ancestries, presents an opportunity to study the differences between ancestries. For example, they have sequenced chromatin precipitation from hMDMs, revealing a more significant number of structural variants in African ancestry genomes than in European ancestry genomes[377]. It would be interesting to investigate their relevance to TB disease susceptibility.

In addition, future studies should evaluate other *Mtbc* lineage strains of the modern and ancient clades, such as L2 (East Asia) and L3 (India and East Africa), and their respective ancestral hosts. The ethical

considerations, blood donor availability, and time and resources in the BSL-3 setting limited this study´s selection of lineages and donors. A larger cohort of blood donors from various ancestries should be examined to build upon these findings. Overall, this study emphasizes the importance of considering host ancestry when creating targeted therapies for tuberculosis.

7.2.6 Conclusion

This study found that the transmission of a single outbreak strain of the L4, specifically the LAM sub-lineage strain, is the main driver of TB epidemiology in the indigenous setting of Puerto Nariño in the Southwest Colombian Amazon. A WGS-based outbreak investigation was used to gain new insights into transmission dynamics in this remote region, including identifying well-supported single nucleotide polymorphisms and a de novo-assembled local reference genome. This approach provided a clearer picture of the circulating outbreak strain and revealed new transmission chains. Almost one-third of the patients may have been infected with at least two clones of the local outbreak strain, highlighting the need for interventions to break transmission chains and improve TB control. The study also demonstrated that customized variant calling pipelines and a new local reference genome could moderately increase the resolution of *Mtbc* outbreaks and have the potential to improve molecular surveillance studies in other high-burden settings. However, further investigation is required to determine the potential benefits of these approaches in other outbreaks, which may vary depending on the strain's genetic background.

Furthermore, this study found biological evidence to support the success of the causative L4 strains of this TB clonal outbreak. In contrast, depending on the host cell ancestry, L6 strains were less infectious. The study also confirmed the host tropism of A4 (*Mbv*) strains for bovine hosts and found that L6 strains could infect and sustain infection in bMDMs, similar to A4 in bovine hosts. This study suggests that the genetic diversity of *Mtbc* strains and their interaction with sympatric and allopatric ancestral host cells play a significant role during the infection, sustainability, and transmission of the pathogen in specific populations. These findings have implications for treatment outcomes and effective vaccines against TB. Overall, the study advances the concept of evolutionary medicine in TB disease and highlights the need for a precision host-pathogen strategy to target, control, and eliminate the disease in a given population.

8. SCIENTFIC CONTRIBUTION

8.1 Whole-genome sequencing-based outbreak investigation

I contributed as the first author to the research article titled “Transmission Dynamics of a *Mycobacterium tuberculosis Complex* Outbreak in an Indigenous Population in the Colombian Amazon Region.” The article was published in the American Society of Microbiology Spectrum Journal on May 24, 2023 (https://journals.asm.org/doi/10.1128/spectrum.05013-22). My contribution included conceiving, designing, and executing the experiments, managing, curating, and analyzing the data, and writing and critically reviewing the manuscript. This work laid the foundation for scientific collaboration between Germany and Colombia to implement and expand NGS in public health practices for TB control in Colombia and positively impacts other countries in Latin America.

Importance of this publication: “The Colombian Amazon around Puerto Nariño has a high tuberculosis burden with a prevalence of 1,267/100,000 people in 2016. Recently, an outbreak of *Mycobacterium tuberculosis* complex (*Mtbc*) bacteria among the indigenous populations was identified with classical *Mtbc* genotyping methods. In this work, whole genome sequencing-based outbreak investigation was conducted in order to improve the phylogenetic resolution and gain new insights into the transmission dynamics in this remote Colombian Amazon Region. The inclusion of well-supported single nucleotide polymorphisms in repetitive regions and a de novo-assembled local reference genome provided a more granular picture of the circulating outbreak strain and revealed new transmission chains. Multiple patients from different settlements were possibly infected with at least two different clones in this high-incidence setting. Thus, these results have the potential to improve molecular surveillance studies in other high-burden settings, especially regions with few clonal multidrug-resistant (MDR) *Mtbc* lineages/clades”[237].

International scientific cooperation: Germany (Molecular and Experimental Mycobacteriology-Research Center Borstel, Borstel) and **Colombia** (Mycobac UN-Microbiology department-School of Medicine-Universidad Nacional de Colombia, Bogota).

8.2 Host-Pathogen in disease susceptibility investigation

I contributed equally as shared first author to the research article titled "Host-pathogen sympatry and differential transmissibility of *Mycobacterium tuberculosis* complex." The article was accepted by nature microbiology on March 13, 2024, in the form of a research article. The article is currently undergoing publication preparation. My role in the research involved the investigation of infection and immune responses of human monocyte-derived macrophages infected with $L4_{widespread}$ and $L6_{restricted}$ strains in host-pathogen allopatric and sympatric combinations. Through my contribution, evidence was provided

for the biological impact of sympatry between *Mtbc* strains and their human hosts. Specifically, my contribution included conceiving, designing, and executing the experiments; data management, curation, and analysis; and supporting the writing and critical reviewing of the manuscript. (https://www.medrxiv.org/content/10.1101/2022.08.04.22278337v2).

Importance of this work: "*Mycobacterium tuberculosis* complex (*Mtbc*) strains separate genetically into nine lineages several of which demonstrate sympatry with their human host, i.e., have distinct and restricted patterns of geographical distribution globally. Geographically restricted *Mtbc* lineages have been hypothesized to be adapted to infect and/or transmit among sympatric human hosts, but this is yet to be confirmed while controlling for exposure, social networks, and risk of disease after exposure. In this work is shown that strains of geographically restricted *Mtbc* lineages L1, $L2_{restricted}$, L3, $L4_{restricted}$, L5, $L6_{restricted}$ are intrinsically less transmissible than widespread *Mtbc* lineages $L2_{widespread}$, $L4_{widespread}$ across Western European and North American cosmopolitan populations. Comparing transmissibility between sympatric and allopatric contact-pathogen pairs, this work found controlled evidence for a biological impact of sympatry between *Mtbc* strains and their human hosts; allopatric host-pathogen exposures have a 38% decrease in the odds of infection among contacts compared with sympatric exposures. Furthermore, this study measured 10- fold lower phagocytosis and growth rates of $L6_{restricted}$ strains compared to $L4_{widespread}$ strains in *in vitro* allopatric macrophage infections. Long-term co-existence of *Mtbc* strains and humans has resulted in differential transmissibility between allopatric and sympatric hosts for strains of geographically restricted lineages. Understanding the specific genetic and immunological underpinnings of sympatry in TB may inform rational vaccine design and TB control" [129].

International scientific cooperation: Germany (Molecular and Experimental Mycobacteriology-Research Center Borstel, Borstel; Institute for Epidemiology, University Medical Hospital Schleswig-Holstein, Kiel; Lungenclinic Grosshansdorf, Airway Research Center North (ARCN), Grosshansdorf; Department of Electrical Engineering and Computer Science, Technical University Berlin and Department of Artificial Intelligence, Fraunhofer Heinrich Hertz Institute, Berlin; German Center for Infection Research, Partner Site Hamburg-Lübeck-Borstel-Riems, Borstel). **USA** (Department of Biomedical Informatics, Harvard Medical School, Boston, MA; Wadsworth Center, New York State Department of Health, Albany and New York City Department of Health and Mental Hygiene, Bureau of Tuberculosis Control, Queens, New York and Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston and the Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston). **The Netherlands** (National Tuberculosis Reference Laboratory, Center for Infectious Disease Control,

National Institute for Public Health, and the Environment (RIVM), Bilthoven and Department of Infectious Diseases, Public Health Service, Amsterdam).

9. REFERENCES

[1] T. Wirth *et al.*, "Origin, Spread and Demography of the Mycobacterium tuberculosis Complex," *PLOS Pathogens*, vol. 4, no. 9, p. e1000160, Sep. 2008, doi: 10.1371/journal.ppat.1000160.

[2] R. J. Dubos and J. Dubos, *The White Plague: Tuberculosis, Man, and Society*. Rutgers University Press, 1987.

[3] I. Comas *et al.*, "Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans," *Nat Genet*, vol. 45, no. 10, pp. 1176–1182, Oct. 2013, doi: 10.1038/ng.2744.

[4] T. Paulson, "Epidemiology: A mortal foe," *Nature*, vol. 502, no. 7470, pp. S2–S3, Oct. 2013, doi: 10.1038/502S2a.

[5] World Health Organization, "Global Tuberculosis Report 2019," World Health Organization, Geneva, 2019. [Online]. Available: https://www.who.int/publications/i/item/9789241565714

[6] World Health Organization, "Global Tuberculosis Report 2022," World Health Organization, Geneva, 2022. [Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022

[7] Robert Koch Institut, "Robert Koch: One of the founders of microbiology." Accessed: Dec. 28, 2021. [Online]. Available: https://www.rki.de/EN/Content/Institute/History/rk_node_en.html

[8] C. J. Cambier, S. Falkow, and L. Ramakrishnan, "Host Evasion and Exploitation Schemes of Mycobacterium tuberculosis," *Cell*, vol. 159, no. 7, pp. 1497–1509, Dec. 2014, doi: 10.1016/j.cell.2014.11.024.

[9] J. Furin, H. Cox, and M. Pai, "Tuberculosis," *The Lancet*, vol. 393, no. 10181, pp. 1642–1656, Apr. 2019, doi: 10.1016/S0140-6736(19)30308-3.

[10] World Health Organization, "Global Tuberculosis Report 2021," Geneva, 2021. [Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021

[11] "Revisiting the timetable of tuberculosis | The BMJ." Accessed: Jan. 01, 2022. [Online]. Available: https://www.bmj.com/content/362/bmj.k2738.long

[12] "Tuberculosis (TB)." Accessed: Jan. 01, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/tuberculosis

[13] "Latent Mycobacterium tuberculosis infection - PubMed." Accessed: Oct. 22, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26017823/

[14] World Health Organization, "Global Tuberculosis Report 2023," Geneva, 2023. Accessed: Nov. 12, 2023. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240083851

[15] A. B. Hogan *et al.*, "Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study," *Lancet Glob Health*, vol. 8, no. 9, pp. e1132–e1141, Sep. 2020, doi: 10.1016/S2214-109X(20)30288-6.

[16] Stop TB Partnership, “Stop TB Partnership’s Global Plan to End TB, 2018–2022,” Stop TB Partnership, Geneva, Switzerland, 2019. Accessed: Jan. 01, 2022. [Online]. Available: https://stoptb.org/assets/documents/global/plan/GPR_2018-2022_Digital.pdf

[17] World Health Organization, “Resolution adopted by the General Assemblyon 5 October 2023. Political declaration of the high-level meeting on the fight against tuberculosis. document WHA73/2020/REC.” Sep. 22, 2023. [Online]. Available: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N23/306/91/PDF/N2330691.pdf?OpenElement

[18] Pan American Health organization, “Tuberculosis in the Americas. Regional Report 2021,” PAHO, Washington, D.C., 2023. [Online]. Available: https://iris.paho.org/handle/10665.2/57001

[19] Pan American Health Organization, “Tuberculosis in the Americas. 2019 Regional Report,” Pan American Health Organization ; Editorial El Manual Moderno, Washington, D.C.; Mexico City [Mexico, 2019. [Online]. Available: https://iris.paho.org/bitstream/handle/10665.2/53026/9789275122747_eng.pdf?sequence=6&isAllowed=y

[20] World Health Organization, “2019 antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline,” Geneva, 2019. [Online]. Available: https://www.who.int/publications/i/item/9789240000193

[21] O. T. Ranzani, J. M. Pescarini, L. Martinez, and A. L. Garcia-Basteiro, “Increasing tuberculosis burden in Latin America: an alarming trend for global control efforts,” *BMJ Glob Health*, vol. 6, no. 3, p. e005639, Mar. 2021, doi: 10.1136/bmjgh-2021-005639.

[22] I. V. Kolte, L. Pereira, A. Benites, I. M. C. de Sousa, and P. C. Basta, “The contribution of stigma to the transmission and treatment of tuberculosis in a hyperendemic indigenous population in Brazil,” *PLoS One*, vol. 15, no. 12, p. e0243988, Dec. 2020, doi: 10.1371/journal.pone.0243988.

[23] J. E. Polanco-Pasaje, I. Rodríguez-Márquez, K. Y. Tello-Hoyos, P. Torres-Pereda, B. L. Guzmán-Salazar, and F. Pérez, “Tuberculosis care cascade for the indigenous population in Colombia: an operational research study*,” *Rev Panam Salud Publica*, vol. 45, p. e20, Feb. 2021, doi: 10.26633/RPSP.2021.20.

[24] D. Tollefson, E. Bloss, A. Fanning, J. T. Redd, K. Barker, and E. McCray, “Burden of tuberculosis in indigenous peoples globally: a systematic review,” p. 36, 2018.

[25] World Bank Group, “Indigenous Latin America in the twenty-first century: the first decade,” Washington, D.C., 2015. [Online]. Available: https://openknowledge.worldbank.org/bitstream/handle/10986/23751/Indigenous0Lat0y000the0first0decade.pdf?sequence=1&isAllowed=y

[26] Pan American Health Organization, “Tuberculosis in the Americas, 2018,” Washington, D.C., 2018. [Online]. Available: https://iris.paho.org/handle/10665.2/49510

[27] M. Cormier *et al.*, “Proximate determinants of tuberculosis in Indigenous peoples worldwide: a systematic review,” *The Lancet Global Health*, vol. 7, no. 1, pp. e68–e80, Jan. 2019, doi: 10.1016/S2214-109X(18)30435-2.

[28] J. D. Ramírez *et al.*, “SARS-CoV-2 in the Amazon region: A harbinger of doom for Amerindians,” *PLoS Negl Trop Dis*, vol. 14, no. 10, p. e0008686, Oct. 2020, doi: 10.1371/journal.pntd.0008686.

[29] United Nations, “Report of the Secretary-General. Progress towards achieving global tuberculosis targets and implementation of the UN political declaration on tuberculosis. Agenda Item 132. Global health and foreign policy.,” 2020. [Online]. Available: https://undocs.org/en/A/75/236

[30] World Health Organization, *WHO operational handbook on tuberculosis. Module 3: diagnosis - rapid diagnostics for tuberculosis detection*. Geneva, 2020.

[31] M. L. Moro, S. Nascetti, F. Morsillo, M. Morandi, and Italian TB-SORV Project Working Group, “Laboratory procedures for the diagnosis of tuberculosis: a survey in ten Italian Regions,” *Ann Ist Super Sanita*, vol. 46, no. 2, pp. 178–184, 2010, doi: 10.4415/ANN_10_02_12.

[32] M. Matee *et al.*, “Sputum microscopy for the diagnosis of HIV-associated pulmonary tuberculosis in Tanzania,” *BMC Public Health*, vol. 8, p. 68, Feb. 2008, doi: 10.1186/1471-2458-8-68.

[33] R. Rimal *et al.*, “Diagnostic performance of GeneXpert MTB/RIF in detecting MTB in smear-negative presumptive TB patients,” *BMC Infect Dis*, vol. 22, no. 1, p. 321, Apr. 2022, doi: 10.1186/s12879-022-07287-5.

[34] “WHO consolidated guidelines on tuberculosis. Module 3: Diagnosis - Rapid diagnostics for tuberculosis detection 2021 update.” Accessed: Jan. 01, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240029415

[35] J. C. Palomino, “Current developments and future perspectives for TB diagnostics,” *Future Microbiol*, vol. 7, no. 1, pp. 59–71, Jan. 2012, doi: 10.2217/fmb.11.133.

[36] “Public call for data on targeted Next-Generation Sequencing solutions for detection of drug resistance among people diagnosed with tuberculosis.” Accessed: Jul. 08, 2023. [Online]. Available: https://www.who.int/news-room/articles-detail/public-call-for-data-on-targeted-next-generation-sequencing-solutions-for-detection-of-drug-resistance-among-people-diagnosed-with-tuberculosis

[37] A. M. Cabibbe *et al.*, “Application of Targeted Next-Generation Sequencing Assay on a Portable Sequencing Platform for Culture-Free Detection of Drug-Resistant Tuberculosis from Clinical Samples,” *Journal of Clinical Microbiology*, vol. 58, no. 10, p. 10.1128/jcm.00632-20, Sep. 2020, doi: 10.1128/jcm.00632-20.

[38] F. Getnet, M. Demissie, N. Assefa, B. Mengistie, and A. Worku, “Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: systematic review and meta-analysis,” *BMC Pulmonary Medicine*, vol. 17, no. 1, p. 202, Dec. 2017, doi: 10.1186/s12890-017-0551-y.

[39] R. McNerney, “Diagnostics for Developing Countries,” *Diagnostics*, vol. 5, no. 2, Art. no. 2, Jun. 2015, doi: 10.3390/diagnostics5020200.

[40] S. Grandjean Lapierre *et al.*, “Complexities and benefits of adopting next-generation sequencing-based tuberculosis diagnostics: a qualitative study among stakeholders in low and high-income countries,” *BMJ Open*, vol. 13, no. 4, p. e066651, Apr. 2023, doi: 10.1136/bmjopen-2022-066651.

[41] M. Vogel *et al.*, “Implementation of whole genome sequencing for tuberculosis diagnostics in a low-middle income, high MDR-TB burden country,” *Sci Rep*, vol. 11, no. 1, Art. no. 1, Jul. 2021, doi: 10.1038/s41598-021-94297-z.

[42] C. J. Meehan *et al.*, “Whole genome sequencing of Mycobacterium tuberculosis :

current standards and open issues," *Nature Reviews Microbiology*, vol. 17, no. 9, Art. no. 9, Sep. 2019, doi: 10.1038/s41579-019-0214-5.

[43] World Health Organization, "The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide.," Geneva, 2018. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/274443/WHO-CDS-TB-2018.19-eng.pdf?sequence=1&isAllowed=y

[44] M. Merker, T. A. Kohl, S. Niemann, and P. Supply, "The Evolution of Strain Typing in the Mycobacterium tuberculosis Complex," *Adv Exp Med Biol*, vol. 1019, pp. 43–78, 2017, doi: 10.1007/978-3-319-64371-7_3.

[45] "The Lancet Respiratory Medicine Commission: 2019 update: epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable tuberculosis - PubMed." Accessed: Jul. 08, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31486393/

[46] "WHO announces updated definitions of extensively drug-resistant tuberculosis." Accessed: Jul. 08, 2023. [Online]. Available: https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis

[47] P. F. Barnes and M. D. Cave, "Molecular epidemiology of tuberculosis," *N Engl J Med*, vol. 349, no. 12, pp. 1149–1156, Sep. 2003, doi: 10.1056/NEJMra021964.

[48] P. Supply *et al.*, "Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis," *J Clin Microbiol*, vol. 44, no. 12, pp. 4498–4510, Dec. 2006, doi: 10.1128/JCM.01392-06.

[49] S. T. Cole *et al.*, "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence," *Nature*, vol. 393, no. 6685, pp. 537–544, Jun. 1998, doi: 10.1038/31159.

[50] T. A. Kohl *et al.*, "MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates," *PeerJ*, vol. 6, p. e5895, 2018, doi: 10.7717/peerj.5895.

[51] T. M. Walker *et al.*, "Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study," *Lancet Infect Dis*, vol. 13, no. 2, pp. 137–146, Feb. 2013, doi: 10.1016/S1473-3099(12)70277-3.

[52] S. Niemann and P. Supply, "Diversity and Evolution of Mycobacterium tuberculosis: Moving to Whole-Genome-Based Approaches," *Cold Spring Harb Perspect Med*, vol. 4, no. 12, p. a021188, Dec. 2014, doi: 10.1101/cshperspect.a021188.

[53] R. Vargas *et al.*, "In-host population dynamics of Mycobacterium tuberculosis complex during active disease," *eLife*, vol. 10, p. e61805, doi: 10.7554/eLife.61805.

[54] C. Nimmo *et al.*, "Dynamics of within-host Mycobacterium tuberculosis diversity and heteroresistance during treatment," *EBioMedicine*, vol. 55, p. 102747, May 2020, doi: 10.1016/j.ebiom.2020.102747.

[55] "England world leaders in the use of whole genome sequencing to diagnose TB," GOV.UK. Accessed: Jul. 25, 2021. [Online]. Available: https://www.gov.uk/government/news/england-world-leaders-in-the-use-of-whole-genome-sequencing-to-diagnose-tb

[56] J. Shea *et al.*, "Comprehensive Whole-Genome Sequencing and Reporting of Drug

Resistance Profiles on Clinical Cases of Mycobacterium tuberculosis in New York State," *J Clin Microbiol*, vol. 55, no. 6, pp. 1871–1882, Jun. 2017, doi: 10.1128/JCM.00298-17.

[57] "Overview of Next-Generation Sequencing Technologies - PubMed." Accessed: Oct. 22, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29851291/

[58] G. A. Logsdon, M. R. Vollger, and E. E. Eichler, "Long-read human genome sequencing and its applications," *Nat Rev Genet*, vol. 21, no. 10, pp. 597–614, Oct. 2020, doi: 10.1038/s41576-020-0236-x.

[59] "BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes - PubMed." Accessed: Nov. 06, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26801360/

[60] World Health Organization, "WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-susceptible tuberculosis treatment," World Health Organization, Geneva, 978-92-4-004812–6, 2022. [Online]. Available: file:///C:/Users/jperez-llanos/Zotero/storage/HU3SZF8T/9789240048126.html

[61] F. Conradie *et al.*, "Treatment of Highly Drug-Resistant Pulmonary Tuberculosis," *New England Journal of Medicine*, vol. 382, no. 10, pp. 893–902, Mar. 2020, doi: 10.1056/NEJMoa1901814.

[62] K. Dheda *et al.*, "The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis," *Lancet Respir Med*, pp. S2213-2600(17)30079–6, Mar. 2017, doi: 10.1016/S2213-2600(17)30079-6.

[63] World Health Organization, "WHO consolidated guidelines on tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update 15 December 2022 | Guideline," World Health Organization, Geneva, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240063129

[64] World Health Organization, "WHO consolidated guidelines on drug resistant tuberculosis treatment," Geneva, 2019. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/311389/9789241550529-eng.pdf?ua=1

[65] A. Van Deun *et al.*, "Short, highly effective, and inexpensive standardized treatment of multidrug-resistant tuberculosis," *Am J Respir Crit Care Med*, vol. 182, no. 5, pp. 684–692, Sep. 2010, doi: 10.1164/rccm.201001-0077OC.

[66] A. J. Nunn *et al.*, "A Trial of a Shorter Regimen for Rifampin-Resistant Tuberculosis," *N Engl J Med*, vol. 380, no. 13, pp. 1201–1213, Mar. 2019, doi: 10.1056/NEJMoa1811867.

[67] "Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis." Accessed: Oct. 22, 2022. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240018662

[68] "Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance." Accessed: Oct. 22, 2022. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240028173

[69] R. Singh, S. P. Dwivedi, U. S. Gaharwar, R. Meena, P. Rajamani, and T. Prasad, "Recent updates on drug resistance in Mycobacterium tuberculosis," *J Appl Microbiol*, vol. 128, no. 6,

pp. 1547–1567, Jun. 2020, doi: 10.1111/jam.14478.

[70] K. J. Seung, S. Keshavjee, and M. L. Rich, "Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis," *Cold Spring Harb Perspect Med*, vol. 5, no. 9, p. a017863, Apr. 2015, doi: 10.1101/cshperspect.a017863.

[71] N. Dookie, S. Rambaran, N. Padayatchi, S. Mahomed, and K. Naidoo, "Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care," *Journal of Antimicrobial Chemotherapy*, vol. 73, no. 5, pp. 1138–1151, May 2018, doi: 10.1093/jac/dkx506.

[72] S. M. Gygli, S. Borrell, A. Trauner, and S. Gagneux, "Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives," *FEMS Microbiology Reviews*, vol. 41, no. 3, pp. 354–373, May 2017, doi: 10.1093/femsre/fux011.

[73] T. M. Walker *et al.*, "The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis," *Lancet Microbe*, vol. 3, no. 4, pp. e265–e273, Apr. 2022, doi: 10.1016/S2666-5247(21)00301-3.

[74] O Neill J, "The review on antimicrobial resistance. Tackling drug-resistant infections globally: final report and recommendations." 2016. Accessed: Jul. 09, 2023. [Online]. Available: https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf

[75] A. S. Dean *et al.*, "25 years of surveillance of drug-resistant tuberculosis: achievements, challenges, and way forward," *The Lancet Infectious Diseases*, vol. 22, no. 7, pp. e191–e196, Jul. 2022, doi: 10.1016/S1473-3099(21)00808-2.

[76] M. Merker *et al.*, "Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage," *Nat Genet*, vol. 47, no. 3, pp. 242–249, Mar. 2015, doi: 10.1038/ng.3195.

[77] M. Merker *et al.*, "Transmission patterns of rifampicin resistant Mycobacterium tuberculosis complex strains in Cameroon: a genomic epidemiological study," *BMC Infectious Diseases*, vol. 21, no. 1, p. 891, Aug. 2021, doi: 10.1186/s12879-021-06593-8.

[78] G. Y. Lesher, E. J. Froelich, M. D. Gruett, J. H. Bailey, and R. P. Brundage, "1,8-NAPHTHYRIDINE DERIVATIVES. A NEW CLASS OF CHEMOTHERAPEUTIC AGENTS," *J Med Pharm Chem*, vol. 91, pp. 1063–1065, Sep. 1962, doi: 10.1021/jm01240a021.

[79] P. Venkatesan, "Worrying lack of funding for tuberculosis," *Lancet Infect Dis*, vol. 22, no. 3, p. 318, Mar. 2022, doi: 10.1016/S1473-3099(22)00073-1.

[80] "Market penetration of Xpert MTB/RIF in high... | Gates Open Research." Accessed: Jan. 01, 2022. [Online]. Available: https://gatesopenresearch.org/articles/2-35/v2

[81] V. Eldholm *et al.*, "Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain," *Nat Commun*, vol. 6, no. 1, p. 7119, May 2015, doi: 10.1038/ncomms8119.

[82] T. Kostyanev *et al.*, "The Innovative Medicines Initiative's New Drugs for Bad Bugs programme: European public-private partnerships for the development of new strategies to tackle antibiotic resistance," *J Antimicrob Chemother*, vol. 71, no. 2, pp. 290–295, Feb. 2016, doi: 10.1093/jac/dkv339.

[83] E. Tacconelli *et al.*, "Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and

tuberculosis," *Lancet Infect Dis*, vol. 18, no. 3, pp. 318–327, Mar. 2018, doi: 10.1016/S1473-3099(17)30753-3.

[84] V. A. Dartois and E. J. Rubin, "Anti-tuberculosis treatment strategies and drug development: challenges and priorities," *Nat Rev Microbiol*, vol. 20, no. 11, Art. no. 11, Nov. 2022, doi: 10.1038/s41579-022-00731-y.

[85] D. G. Russell, H. C. Mwandumba, and E. E. Rhoades, "Mycobacterium and the coat of many lipids," *J Cell Biol*, vol. 158, no. 3, pp. 421–426, Aug. 2002, doi: 10.1083/jcb.200205034.

[86] C. J. Cambier, S. Falkow, and L. Ramakrishnan, "Host evasion and exploitation schemes of Mycobacterium tuberculosis," *Cell*, vol. 159, no. 7, pp. 1497–1509, Dec. 2014, doi: 10.1016/j.cell.2014.11.024.

[87] "Ecology and evolution of Mycobacterium tuberculosis - PubMed." Accessed: Jan. 20, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29456241/

[88] S. H. Kaufmann, "How can immunology contribute to the control of tuberculosis?," *Nat Rev Immunol*, vol. 1, no. 1, pp. 20–30, Oct. 2001, doi: 10.1038/35095558.

[89] H. Clay, J. M. Davis, D. Beery, A. Huttenlocher, S. E. Lyons, and L. Ramakrishnan, "Dichotomous role of the macrophage in early Mycobacterium marinum infection of the zebrafish," *Cell Host Microbe*, vol. 2, no. 1, pp. 29–39, Jul. 2007, doi: 10.1016/j.chom.2007.06.004.

[90] D. G. Russell, "Who puts the tubercle in tuberculosis?," *Nat Rev Microbiol*, vol. 5, no. 1, pp. 39–47, Jan. 2007, doi: 10.1038/nrmicro1538.

[91] L. Ramakrishnan, "Revisiting the role of the granuloma in tuberculosis," *Nat Rev Immunol*, vol. 12, no. 5, Art. no. 5, May 2012, doi: 10.1038/nri3211.

[92] A. M. Cadena, S. M. Fortune, and J. L. Flynn, "Heterogeneity in tuberculosis," *Nat Rev Immunol*, vol. 17, no. 11, Art. no. 11, Nov. 2017, doi: 10.1038/nri.2017.69.

[93] D. F. Warner, A. Koch, and V. Mizrahi, "Diversity and disease pathogenesis in Mycobacterium tuberculosis," *Trends Microbiol*, vol. 23, no. 1, pp. 14–21, Jan. 2015, doi: 10.1016/j.tim.2014.10.005.

[94] "BioRender," BioRender. Accessed: Jan. 20, 2022. [Online]. Available: https://biorender.com/

[95] S. Gordon and P. R. Taylor, "Monocyte and macrophage heterogeneity," *Nat Rev Immunol*, vol. 5, no. 12, pp. 953–964, Dec. 2005, doi: 10.1038/nri1733.

[96] D. M. Mosser and J. P. Edwards, "Exploring the full spectrum of macrophage activation," *Nat Rev Immunol*, vol. 8, no. 12, pp. 958–969, Dec. 2008, doi: 10.1038/nri2448.

[97] "Macrophages in health and disease - PubMed." Accessed: Jul. 18, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36368305/

[98] C. H. Liu, H. Liu, and B. Ge, "Innate immunity in tuberculosis: host defense vs pathogen evasion," *Cell Mol Immunol*, vol. 14, no. 12, pp. 963–975, Dec. 2017, doi: 10.1038/cmi.2017.88.

[99] D. M. E. Bowdish *et al.*, "MARCO, TLR2, and CD14 are required for macrophage cytokine responses to mycobacterial trehalose dimycolate and Mycobacterium tuberculosis," *PLoS Pathog*, vol. 5, no. 6, p. e1000474, Jun. 2009, doi: 10.1371/journal.ppat.1000474.

[100] "Macrophage Receptors for Mycobacterium tuberculosis - PMC." Accessed:

Jul. 20, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC108049/

[101] D. P. Eisen and R. M. Minchinton, "Impact of Mannose-Binding Lectin on Susceptibility to Infectious Diseases," *Clinical Infectious Diseases*, vol. 37, no. 11, pp. 1496–1505, Dec. 2003, doi: 10.1086/379324.

[102] T. Thye *et al.*, "Variant G57E of mannose binding lectin associated with protection against tuberculosis caused by Mycobacterium africanum but not by M. tuberculosis," *PLoS One*, vol. 6, no. 6, p. e20908, 2011, doi: 10.1371/journal.pone.0020908.

[103] J. S. Ferguson *et al.*, "Surfactant Protein D Increases Fusion of Mycobacterium tuberculosis- Containing Phagosomes with Lysosomes in Human Macrophages," *Infection and Immunity*, vol. 74, no. 12, pp. 7005–7009, Dec. 2006, doi: 10.1128/iai.01402-06.

[104] J. S. Ferguson, J. J. Weis, J. L. Martin, and L. S. Schlesinger, "Complement Protein C3 Binding to Mycobacterium tuberculosis Is Initiated by the Classical Pathway in Human Bronchoalveolar Lavage Fluid," *Infect Immun*, vol. 72, no. 5, pp. 2564–2573, May 2004, doi: 10.1128/IAI.72.5.2564-2573.2004.

[105] M. Jackson, "The Mycobacterial Cell Envelope—Lipids," *Cold Spring Harb Perspect Med*, vol. 4, no. 10, p. a021105, Oct. 2014, doi: 10.1101/cshperspect.a021105.

[106] L. Chiaradia *et al.*, "Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane," *Sci Rep*, vol. 7, no. 1, Art. no. 1, Oct. 2017, doi: 10.1038/s41598-017-12718-4.

[107] M. E. Feltcher, J. T. Sullivan, and M. Braunstein, "Protein export systems of Mycobacterium tuberculosis: novel targets for drug development?," *Future Microbiol*, vol. 5, no. 10, pp. 1581–1597, Oct. 2010, doi: 10.2217/fmb.10.112.

[108] T. Palmer and B. C. Berks, "The twin-arginine translocation (Tat) protein export pathway," *Nat Rev Microbiol*, vol. 10, no. 7, pp. 483–496, Jun. 2012, doi: 10.1038/nrmicro2814.

[109] M. I. Gröschel, F. Sayes, R. Simeone, L. Majlessi, and R. Brosch, "ESX secretion systems: mycobacterial evolution to counter host immunity," *Nat Rev Microbiol*, vol. 14, no. 11, pp. 677–691, Nov. 2016, doi: 10.1038/nrmicro.2016.131.

[110] J. Pieters and H. Ploegh, "Microbiology. Chemical warfare and mycobacterial defense," *Science*, vol. 302, no. 5652, pp. 1900–1902, Dec. 2003, doi: 10.1126/science.1092873.

[111] J. Pieters, "Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance," *Cell Host & Microbe*, vol. 3, no. 6, pp. 399–407, Jun. 2008, doi: 10.1016/j.chom.2008.05.006.

[112] K. Pethe, D. L. Swenson, S. Alonso, J. Anderson, C. Wang, and D. G. Russell, "Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation," *Proc Natl Acad Sci U S A*, vol. 101, no. 37, pp. 13642–13647, Sep. 2004, doi: 10.1073/pnas.0401657101.

[113] R. Levin, S. Grinstein, and J. Canton, "The life cycle of phagosomes: formation, maturation, and resolution," *Immunol Rev*, vol. 273, no. 1, pp. 156–179, Sep. 2016, doi: 10.1111/imr.12439.

[114] R. B. Bekale *et al.*, "Mycobacterium Tuberculosis and Interactions with the Host Immune System: Opportunities for Nanoparticle Based Immunotherapeutics and Vaccines," *Pharm Res*, vol. 36, no. 1, p. 8, Nov. 2018, doi: 10.1007/s11095-018-2528-9.

[115] S. Sturgill-Koszycki *et al.*, "Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase," *Science*, vol. 263, no. 5147, pp. 678–681, Feb. 1994, doi: 10.1126/science.8303277.

[116] R. Simeone, F. Sayes, E. Lawarée, and R. Brosch, "Breaching the phagosome, the case of the tuberculosis agent," *Cellular Microbiology*, vol. 23, no. 7, p. e13344, 2021, doi: 10.1111/cmi.13344.

[117] D. Bottai *et al.*, "TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis lineages," *Nat Commun*, vol. 11, no. 1, p. 684, Feb. 2020, doi: 10.1038/s41467-020-14508-5.

[118] S. Gagneux, "Ecology and evolution of Mycobacterium tuberculosis," *Nat Rev Microbiol*, vol. 16, no. 4, Art. no. 4, Apr. 2018, doi: 10.1038/nrmicro.2018.8.

[119] J. C. S. Ngabonziza *et al.*, "A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region," *Nat Commun*, vol. 11, no. 1, Art. no. 1, Jun. 2020, doi: 10.1038/s41467-020-16626-6.

[120] S. Gagneux *et al.*, "Variable host–pathogen compatibility in Mycobacterium tuberculosis," *PNAS*, vol. 103, no. 8, pp. 2869–2873, Feb. 2006, doi: 10.1073/pnas.0511240103.

[121] D. Stucki *et al.*, "Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages," *Nat Genet*, vol. 48, no. 12, pp. 1535–1543, Dec. 2016, doi: 10.1038/ng.3704.

[122] Y. A. Shuaib *et al.*, "Origin and Global Expansion of Mycobacterium tuberculosis Complex Lineage 3," *Genes (Basel)*, vol. 13, no. 6, p. 990, May 2022, doi: 10.3390/genes13060990.

[123] K. E. Wiens *et al.*, "Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis," *BMC Medicine*, vol. 16, no. 1, p. 196, Oct. 2018, doi: 10.1186/s12916-018-1180-x.

[124] T. Netikul, P. Palittapongarnpim, Y. Thawornwattana, and S. Plitphonganphim, "Estimation of the global burden of Mycobacterium tuberculosis lineage 1," *Infect Genet Evol*, vol. 91, p. 104802, Jul. 2021, doi: 10.1016/j.meegid.2021.104802.

[125] D. Yeboah-Manu, B. C. de Jong, and F. Gehre, "The Biology and Epidemiology of Mycobacterium africanum," *Adv Exp Med Biol*, vol. 1019, pp. 117–133, 2017, doi: 10.1007/978-3-319-64371-7_6.

[126] "Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade | PLOS ONE." Accessed: Oct. 20, 2022. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0052841

[127] S. Gagneux, "Host-pathogen coevolution in human tuberculosis," *Philos Trans R Soc Lond B Biol Sci*, vol. 367, no. 1590, pp. 850–859, Mar. 2012, doi: 10.1098/rstb.2011.0316.

[128] D. Brites and S. Gagneux, "Co-evolution of Mycobacterium tuberculosis and Homo sapiens," *Immunol Rev*, vol. 264, no. 1, pp. 6–24, Mar. 2015, doi: 10.1111/imr.12264.

[129] M. I. Gröschel *et al.*, "Host-pathogen sympatry and differential transmissibility of Mycobacterium tuberculosis complex." medRxiv, p. 2022.08.04.22278337, Feb. 15, 2023. doi: 10.1101/2022.08.04.22278337.

[130] S. Niemann, M. Merker, T. Kohl, and P. Supply, "Impact of Genetic Diversity on the Biology of Mycobacterium tuberculosis Complex

Strains," *Microbiology Spectrum*, vol. 4, no. 6, p. 4.6.11, Nov. 2016, doi: 10.1128/microbiolspec.TBTB2-0022-2016.

[131] M. Coscolla and S. Gagneux, "Consequences of genomic diversity in Mycobacterium tuberculosis," *Semin Immunol*, vol. 26, no. 6, pp. 431–444, Dec. 2014, doi: 10.1016/j.smim.2014.09.012.

[132] O. B. Brynildsrud *et al.*, "Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation," *Sci Adv*, vol. 4, no. 10, p. eaat5869, Oct. 2018, doi: 10.1126/sciadv.aat5869.

[133] F. Coll *et al.*, "A robust SNP barcode for typing Mycobacterium tuberculosis complex strains," *Nat Commun*, vol. 5, p. 4812, Sep. 2014, doi: 10.1038/ncomms5812.

[134] M. Woodman, I. Haeusler, and L. Grandjean, "Tuberculosis Genetic Epidemiology: A Latin American Perspective," *Genes*, vol. 10, no. 1, p. 53, Jan. 2019, doi: 10.3390/genes10010053.

[135] I. Mokrousov *et al.*, "Latin-American-Mediterranean lineage of Mycobacterium tuberculosis: Human traces across pathogen's phylogeography," *Mol Phylogenet Evol*, vol. 99, pp. 133–143, Jun. 2016, doi: 10.1016/j.ympev.2016.03.020.

[136] S. Homolka, S. Niemann, D. G. Russell, and K. H. Rohde, "Functional Genetic Diversity among Mycobacterium tuberculosis Complex Clinical Isolates: Delineation of Conserved Core and Lineage-Specific Transcriptomes during Intracellular Survival," *PLOS Pathogens*, vol. 6, no. 7, p. e1000988, Jul. 2010, doi: 10.1371/journal.ppat.1000988.

[137] B. Tizzano *et al.*, "Survival of hypoxia-induced dormancy is not a common feature of all strains of the Mycobacterium tuberculosis complex," *Sci Rep*, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41598-021-81223-6.

[138] D. Portevin *et al.*, "Lipidomics and genomics of Mycobacterium tuberculosis reveal lineage-specific trends in mycolic acid biosynthesis," *Microbiologyopen*, vol. 3, no. 6, pp. 823–835, Dec. 2014, doi: 10.1002/mbo3.193.

[139] D. Portevin, S. Gagneux, I. Comas, and D. Young, "Human macrophage responses to clinical isolates from the Mycobacterium tuberculosis complex discriminate between ancient and modern lineages," *PLoS Pathog*, vol. 7, no. 3, p. e1001307, Mar. 2011, doi: 10.1371/journal.ppat.1001307.

[140] A. Romagnoli *et al.*, "Clinical isolates of the modern Mycobacterium tuberculosis lineage 4 evade host defense in human macrophages through eluding IL-1β-induced autophagy," *Cell Death Dis*, vol. 9, no. 6, p. 624, May 2018, doi: 10.1038/s41419-018-0640-8.

[141] N. Rakotosamimanana *et al.*, "Variation in gamma interferon responses to different infecting strains of Mycobacterium tuberculosis in acid-fast bacillus smear-positive patients and household contacts in Antananarivo, Madagascar," *Clin Vaccine Immunol*, vol. 17, no. 7, pp. 1094–1103, Jul. 2010, doi: 10.1128/CVI.00049-10.

[142] I. Comas *et al.*, "Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved," *Nat Genet*, vol. 42, no. 6, pp. 498–503, Jun. 2010, doi: 10.1038/ng.590.

[143] M. Coscolla *et al.*, "M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens," *Cell Host Microbe*, vol. 18, no. 5, pp. 538–548, Nov. 2015, doi: 10.1016/j.chom.2015.10.008.

[144] A. Zureck, “G. P. Kubica and L. G. Wayne (Editors), The Mycobacteria — a Sourcebook, Part A and Part B (Microbiology Series Volume 15). 1553 S., 225 Abb., 122 Tab. New York-Basel 1984. Marcel Dekker. $ 295.00. ISBN: 0-8247-7009-9 (Part A), 0-8247-1917-4 (Part B),” *Journal of Basic Microbiology*, vol. 25, no. 10, pp. 662–662, 1985, doi: 10.1002/jobm.3620251012.

[145] G. Källenius *et al.*, “Evolution and clonal traits of Mycobacterium tuberculosis complex in Guinea-Bissau,” *J Clin Microbiol*, vol. 37, no. 12, pp. 3872–3878, Dec. 1999, doi: 10.1128/JCM.37.12.3872-3878.1999.

[146] S. Niemann, E. Richter, and S. Rüsch-Gerdes, “Differentiation among members of the Mycobacterium tuberculosis complex by molecular and biochemical features: evidence for two pyrazinamide-susceptible subtypes of M. bovis,” *J Clin Microbiol*, vol. 38, no. 1, pp. 152–157, Jan. 2000, doi: 10.1128/JCM.38.1.152-157.2000.

[147] “Practical handbook for the phenotypic and genotypic identification of mycobacteria. – ScienceOpen.” Accessed: Oct. 16, 2022. [Online]. Available: https://www.scienceopen.com/document?vid=f4807bef-1c8d-431e-be6a-bced8cb452b9

[148] F. Gehre *et al.*, “Deciphering the growth behaviour of Mycobacterium africanum,” *PLoS Negl Trop Dis*, vol. 7, no. 5, p. e2220, 2013, doi: 10.1371/journal.pntd.0002220.

[149] S. Niemann *et al.*, “Mycobacterium africanum subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda,” *J Clin Microbiol*, vol. 40, no. 9, pp. 3398–3405, Sep. 2002, doi: 10.1128/JCM.40.9.3398-3405.2002.

[150] D. Affolabi *et al.*, “First molecular epidemiological study of tuberculosis in Benin,” *Int J Tuberc Lung Dis*, vol. 13, no. 3, pp. 317–322, Mar. 2009.

[151] B. C. de Jong, M. Antonio, and S. Gagneux, “Mycobacterium africanum--review of an important cause of human tuberculosis in West Africa,” *PLoS Negl Trop Dis*, vol. 4, no. 9, p. e744, Sep. 2010, doi: 10.1371/journal.pntd.0000744.

[152] J. M. Grange and M. D. Yates, “Incidence and nature of human tuberculosis due to Mycobacterium africanum in South-East England: 1977-87,” *Epidemiol Infect*, vol. 103, no. 1, pp. 127–132, Aug. 1989, doi: 10.1017/s0950268800030429.

[153] H. Jungbluth, H. Fink, and F. Reusch, “[Tuberculous infection caused by Myco. africanum in black africans resident in the German Federal Republic (author’s transl)],” *Prax Klin Pneumol*, vol. 32, no. 5, pp. 306–309, May 1978.

[154] U. A. Hurtado, J. S. Solano, A. Rodriguez, J. Robledo, and F. Rouzaud, “Draft Genome Sequence of a Mycobacterium africanum Clinical Isolate from Antioquia, Colombia,” *Genome Announc*, vol. 4, no. 3, pp. e00486-16, Jun. 2016, doi: 10.1128/genomeA.00486-16.

[155] E. Desmond *et al.*, “Mycobacterium africanum cases, California,” *Emerg Infect Dis*, vol. 10, no. 5, pp. 921–923, May 2004, doi: 10.3201/eid1005.030016.

[156] D. Bek, M. K. Kjeldsen, Z. Kamper-Jørgensen, N. F. Hansen, and E. M. Rasmussen, “[Tuberculosis caused by Mycobacterium africanum],” *Ugeskr Laeger*, vol. 172, no. 7, pp. 549–550, Feb. 2010.

[157] K. A. Alexander *et al.*, “Novel Mycobacterium tuberculosis complex pathogen, M. mungi,” *Emerg Infect Dis*, vol. 16, no. 8, pp. 1296–1299, Aug. 2010, doi: 10.3201/eid1608.100314.

[158] Z. Rahim *et al.*, "Characterization of Mycobacterium africanum subtype I among cows in a dairy farm in Bangladesh using spoligotyping," *Southeast Asian J Trop Med Public Health*, vol. 38, no. 4, pp. 706–713, Jul. 2007.

[159] M. F. Thorel, "Isolation of mycobacterium africanum from monkeys," *Tubercle*, vol. 61, no. 2, pp. 101–104, Jun. 1980, doi: 10.1016/0041-3879(80)90018-5.

[160] S. I. B. Cadmus, M. K. Yakubu, A. A. Magaji, A. O. Jenkins, and D. van Soolingen, "Mycobacterium bovis, but also M. africanum present in raw milk of pastoral cattle in north-central Nigeria," *Trop Anim Health Prod*, vol. 42, no. 6, pp. 1047–1048, Aug. 2010, doi: 10.1007/s11250-010-9533-2.

[161] A. Gudan *et al.*, "Disseminated tuberculosis in hyrax (Procavia capensis) caused by Mycobacterium africanum," *J Zoo Wildl Med*, vol. 39, no. 3, pp. 386–391, Sep. 2008, doi: 10.1638/06-041.1.

[162] B. C. de Jong *et al.*, "Mycobacterium africanum: a new opportunistic pathogen in HIV infection?," *AIDS*, vol. 19, no. 15, pp. 1714–1715, Oct. 2005, doi: 10.1097/01.aids.0000185991.54595.41.

[163] C. G. Meyer *et al.*, "Pulmonary tuberculosis: virulence of Mycobacterium africanum and relevance in HIV co-infection," *Tuberculosis (Edinb)*, vol. 88, no. 5, pp. 482–489, Sep. 2008, doi: 10.1016/j.tube.2008.05.004.

[164] T. D. Bold, D. C. Davis, K. K. Penberthy, L. M. Cox, J. D. Ernst, and B. C. de Jong, "Impaired fitness of Mycobacterium africanum despite secretion of ESAT-6," *J Infect Dis*, vol. 205, no. 6, pp. 984–990, Mar. 2012, doi: 10.1093/infdis/jir883.

[165] T. Smith, "A COMPARATIVE STUDY OF BOVINE TUBERCLE BACILLI AND OF HUMAN BACILLI FROM SPUTUM," *Journal of Experimental Medicine*, vol. 3, no. 4–5, pp. 451–511, Jul. 1898, doi: 10.1084/jem.3.4-5.451.

[166] K. M. Malone and S. V. Gordon, "Mycobacterium tuberculosis Complex Members Adapted to Wild and Domestic Animals," *Adv Exp Med Biol*, vol. 1019, pp. 135–154, 2017, doi: 10.1007/978-3-319-64371-7_7.

[167] P. Brodin *et al.*, "Bacterial artificial chromosome-based comparative genomic analysis identifies Mycobacterium microti as a natural ESAT-6 deletion mutant," *Infect Immun*, vol. 70, no. 10, pp. 5568–5578, Oct. 2002, doi: 10.1128/IAI.70.10.5568-5578.2002.

[168] S. Mostowy, D. Cousins, and M. A. Behr, "Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the Mycobacterium tuberculosis complex," *J Bacteriol*, vol. 186, no. 1, pp. 104–109, Jan. 2004, doi: 10.1128/JB.186.1.104-109.2003.

[169] S. D. C. Parsons, J. A. Drewe, N. C. Gey van Pittius, R. M. Warren, and P. D. van Helden, "Novel cause of tuberculosis in meerkats, South Africa," *Emerg Infect Dis*, vol. 19, no. 12, pp. 2004–2007, Dec. 2013, doi: 10.3201/eid1912.130268.

[170] D. Brites *et al.*, "A New Phylogenetic Framework for the Animal-Adapted Mycobacterium tuberculosis Complex," *Front Microbiol*, vol. 9, p. 2820, 2018, doi: 10.3389/fmicb.2018.02820.

[171] D. V. Cousins, R. L. Peet, W. T. Gaynor, S. N. Williams, and B. L. Gow, "Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex," *Vet Microbiol*, vol. 42, no. 2–3, pp. 135–145, Nov. 1994, doi: 10.1016/0378-1135(94)90013-2.

[172] S. V. Gordon *et al.*, "Genomics of Mycobacterium bovis," *Tuberculosis (Edinb)*, vol.

81, no. 1–2, pp. 157–163, 2001, doi: 10.1054/tube.2000.0269.

[173] V. W. Lees, S. Copeland, and P. Rousseau, “Bovine tuberculosis in elk (Cervus elaphus manitobensis) near Riding Mountain National Park, Manitoba, from 1992 to 2002,” *Can Vet J*, vol. 44, no. 10, pp. 830–831, Oct. 2003.

[174] M. Pesciaroli *et al.*, “Tuberculosis in domestic animal species,” *Res Vet Sci*, vol. 97 Suppl, pp. S78-85, Oct. 2014, doi: 10.1016/j.rvsc.2014.05.015.

[175] W. R. Waters and M. V. Palmer, “Mycobacterium bovis Infection of Cattle and White-Tailed Deer: Translational Research of Relevance to Human Tuberculosis,” *ILAR J*, vol. 56, no. 1, pp. 26–43, 2015, doi: 10.1093/ilar/ilv001.

[176] P. J. Thompson, D. V. Cousins, B. L. Gow, D. M. Collins, B. H. Williamson, and H. T. Dagnia, “Seals, seal trainers, and mycobacterial infection,” *Am Rev Respir Dis*, vol. 147, no. 1, pp. 164–167, Jan. 1993, doi: 10.1164/ajrccm/147.1.164.

[177] K. I. Bos *et al.*, “Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis,” *Nature*, vol. 514, no. 7523, pp. 494–497, Oct. 2014, doi: 10.1038/nature13591.

[178] J. Thapa, C. Nakajima, B. Maharjan, A. Poudell, and Y. Suzuki, “Molecular characterization of Mycobacterium orygis isolates from wild animals of Nepal,” *Jpn J Vet Res*, vol. 63, no. 3, pp. 151–158, Aug. 2015.

[179] M. De Garine-Wichatitsky *et al.*, “A review of bovine tuberculosis at the wildlife-livestock-human interface in sub-Saharan Africa,” *Epidemiol Infect*, vol. 141, no. 7, pp. 1342–1356, Jul. 2013, doi: 10.1017/S0950268813000708.

[180] D. Park *et al.*, “Tuberculosis due to Mycobacterium bovis in patients coinfected with human immunodeficiency virus,” *Clin Infect Dis*, vol. 51, no. 11, pp. 1343–1346, Dec. 2010, doi: 10.1086/657118.

[181] J. Thapa, S. V. Gordon, C. Nakajima, and Y. Suzuki, “Threat from Mycobacterium orygis-associated tuberculosis in south Asia,” *Lancet Microbe*, vol. 3, no. 9, pp. e641–e642, Sep. 2022, doi: 10.1016/S2666-5247(22)00149-5.

[182] “Economics of Bovine Tuberculosis - Mycobacterium Bovis Infection in Animals and Humans - Wiley Online Library.” Accessed: Jul. 15, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470344538.ch9

[183] J. M. Grange, “Mycobacterium bovis infection in human beings,” *Tuberculosis (Edinb)*, vol. 81, no. 1–2, pp. 71–77, 2001, doi: 10.1054/tube.2000.0263.

[184] B. Villarreal-Ramos *et al.*, “Experimental infection of cattle with Mycobacterium tuberculosis isolates shows the attenuation of the human tubercle bacillus for cattle,” *Sci Rep*, vol. 8, no. 1, p. 894, Jan. 2018, doi: 10.1038/s41598-017-18575-5.

[185] K. M. Malone *et al.*, “Comparative ’omics analyses differentiate Mycobacterium tuberculosis and Mycobacterium bovis and reveal distinct macrophage responses to infection with the human and bovine tubercle bacilli,” *Microb Genom*, vol. 4, no. 3, p. e000163, Mar. 2018, doi: 10.1099/mgen.0.000163.

[186] J. Piercy, D. Werling, and T. J. Coffey, “Differential responses of bovine macrophages to infection with bovine-specific and non-bovine specific mycobacteria,” *Tuberculosis (Edinb)*, vol. 87, no. 5, pp. 415–420, Sep. 2007, doi: 10.1016/j.tube.2007.06.001.

[187] D. A. Magee *et al.*, "Innate cytokine profiling of bovine alveolar macrophages reveals commonalities and divergence in the response to Mycobacterium bovis and Mycobacterium tuberculosis infection," *Tuberculosis (Edinb)*, vol. 94, no. 4, pp. 441–450, Jul. 2014, doi: 10.1016/j.tube.2014.04.004.

[188] Randolph M. Nesse, and George C. Williams, *Why We Get Sick : The New Science of Darwinian Medicine*. New York and Canada: Times Book, 1995.

[189] M. Coscolla and S. Gagneux, "Does M. tuberculosis genomic diversity explain disease diversity?," *Drug Discov Today Dis Mech*, vol. 7, no. 1, pp. e43–e59, 2010, doi: 10.1016/j.ddmec.2010.09.004.

[190] M. Fumagalli, M. Sironi, U. Pozzoli, A. Ferrer-Admettla, L. Pattini, and R. Nielsen, "Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution," *PLOS Genetics*, vol. 7, no. 11, p. e1002355, Nov. 2011, doi: 10.1371/journal.pgen.1002355.

[191] R. Dawkins, J. R. Krebs, J. Maynard Smith, and R. Holliday, "Arms races between and within species," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 205, no. 1161, pp. 489–511, Jan. 1997, doi: 10.1098/rspb.1979.0081.

[192] M. J. Blaser and D. Kirschner, "The equilibria that allow bacterial persistence in human hosts," *Nature*, vol. 449, no. 7164, Art. no. 7164, Oct. 2007, doi: 10.1038/nature06198.

[193] T. D. Bold and J. D. Ernst, "Who Benefits from Granulomas, Mycobacteria or Host?," *Cell*, vol. 136, no. 1, pp. 17–19, Jan. 2009, doi: 10.1016/j.cell.2008.12.032.

[194] N. R. Cabej, "18 - Species and Allopatric Speciation," in *Epigenetic Principles of Evolution*, N. R. Cabej, Ed., London: Elsevier, 2012, pp. 707–723. doi: 10.1016/B978-0-12-415831-3.00018-5.

[195] D. I. Bolnick and B. M. Fitzpatrick, "Sympatric Speciation: Models and Empirical Evidence," *Annual Review of Ecology, Evolution, and Systematics*, vol. 38, no. 1, pp. 459–487, 2007, doi: 10.1146/annurev.ecolsys.38.091206.095804.

[196] J. G. Pasipanodya *et al.*, "Allopatric tuberculosis host–pathogen relationships are associated with greater pulmonary impairment," *Infect Genet Evol*, vol. 16, pp. 433–440, Jun. 2013, doi: 10.1016/j.meegid.2013.02.015.

[197] M. Coscolla, "Biological and Epidemiological Consequences of MTBC Diversity," *Adv Exp Med Biol*, vol. 1019, pp. 95–116, 2017, doi: 10.1007/978-3-319-64371-7_5.

[198] D. Brites and S. Gagneux, "The Nature and Evolution of Genomic Diversity in the Mycobacterium tuberculosis Complex," *Adv Exp Med Biol*, vol. 1019, pp. 1–26, 2017, doi: 10.1007/978-3-319-64371-7_1.

[199] N. Reiling *et al.*, "Clade-Specific Virulence Patterns of Mycobacterium tuberculosis Complex Strains in Human Primary Macrophages and Aerogenically Infected Mice," *mBio*, vol. 4, no. 4, pp. e00250-13, doi: 10.1128/mBio.00250-13.

[200] R. Bellamy, "Susceptibility to mycobacterial infections: the importance of host genetics," *Genes Immun*, vol. 4, no. 1, Art. no. 1, Jan. 2003, doi: 10.1038/sj.gene.6363915.

[201] A. Kone *et al.*, "Differential HLA allele frequency in Mycobacterium africanum vs Mycobacterium tuberculosis in Mali," *HLA*, vol. 93, no. 1, pp. 24–31, Jan. 2019, doi: 10.1111/tan.13448.

[202] M. Salie *et al.*, "Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease," *J Infect Dis*, vol. 209, no. 2, pp. 216–223, Jan. 2014, doi: 10.1093/infdis/jit443.

[203] R. Capparelli *et al.*, "Role Played by Human Mannose-Binding Lectin Polymorphisms in Pulmonary Tuberculosis," *The Journal of Infectious Diseases*, vol. 199, no. 5, pp. 666–672, Mar. 2009, doi: 10.1086/596658.

[204] G. Kerner *et al.*, "Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry," *Proc Natl Acad Sci U S A*, vol. 116, no. 21, pp. 10430–10434, May 2019, doi: 10.1073/pnas.1903561116.

[205] G. Kerner *et al.*, "Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years," *The American Journal of Human Genetics*, vol. 108, no. 3, pp. 517–524, Mar. 2021, doi: 10.1016/j.ajhg.2021.02.009.

[206] S. D. Bentley *et al.*, "The Genome of Mycobacterium Africanum West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the M. tuberculosis Complex," *PLoS Negl Trop Dis*, vol. 6, no. 2, p. e1552, Feb. 2012, doi: 10.1371/journal.pntd.0001552.

[207] K. Winglee *et al.*, "Whole Genome Sequencing of Mycobacterium africanum Strains from Mali Provides Insights into the Mechanisms of Geographic Restriction," *PLoS Negl Trop Dis*, vol. 10, no. 1, p. e0004332, Jan. 2016, doi: 10.1371/journal.pntd.0004332.

[208] M. B. Reed *et al.*, "Major Mycobacterium tuberculosis lineages associate with patient country of origin," *J Clin Microbiol*, vol. 47, no. 4, pp. 1119–1128, Apr. 2009, doi: 10.1128/JCM.02142-08.

[209] S. Asgari *et al.*, "Higher native Peruvian genetic ancestry proportion is associated with tuberculosis progression risk," *Cell Genom*, vol. 2, no. 7, p. 100151, Jul. 2022, doi: 10.1016/j.xgen.2022.100151.

[210] L. Fenner *et al.*, "HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis," *PLoS Genet*, vol. 9, no. 3, p. e1003318, 2013, doi: 10.1371/journal.pgen.1003318.

[211] S. Homolka *et al.*, "High Resolution Discrimination of Clinical Mycobacterium tuberculosis Complex Strains Based on Single Nucleotide Polymorphisms," *PLoS One*, vol. 7, no. 7, p. e39855, Jul. 2012, doi: 10.1371/journal.pone.0039855.

[212] A. V. Marín, N. Rastogi, D. Couvin, V. Mape, and M. I. Murcia, "First approach to the population structure of Mycobacterium tuberculosis complex in the indigenous population in Puerto Nariño-Amazonas, Colombia," *PLoS One*, vol. 16, no. 1, p. e0245084, Jan. 2021, doi: 10.1371/journal.pone.0245084.

[213] Google Maps, "Colombia," 2020. Accessed: Jan. 12, 2020. [Online]. Available: https://www.google.com/maps/d/embed?mid=1DBPmOVnFuSYii8rtHkLi4omKdO8&msa=0&ie=UTF8&t=m&ll=7.166300674106962%2C-74.04785128125002&spn=10.474042%2C14.0625&z=5&output=embed

[214] "Datos Abiertos Colombia | Datos Abiertos Colombia," la plataforma de datos abiertos del gobierno colombiano. Accessed: Feb. 21, 2022. [Online]. Available: https://www.datos.gov.co/

[215] S. D. Bentley *et al.*, "The Genome of Mycobacterium Africanum West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the M. tuberculosis Complex," *PLOS Neglected Tropical Diseases*,

vol. 6, no. 2, p. e1552, Feb. 2012, doi: 10.1371/journal.pntd.0001552.

[216] Brandenburg Julius, "Wnt6: A novel mediator in M. tuberculosis infection linking inflammation and lipid metabolism," University of Lübeck, 2015.

[217] Miltenyi Biotec B.V. & Co. KG, "CD14 MicroBeads human Order no. 130-050-201." [Online]. Available: https://www.miltenyibiotec.com/upload/assets/IM0001260.PDF

[218] P. Bettencourt, N. Carmo, D. Pires, P. Timóteo, and E. Anes, "Mycobacterial infection of macrophages: the effect of the multiplicity of infection," 2017, pp. 651–664.

[219] BioLegend, "LEGENDplextTM Human Inflammation Panel 1 (13-plex) with V-botom Plate Cat. No. 740809." [Online]. Available: https://www.biolegend.com/nl-be/products/legendplex-human-inflammation-panel-1-13-plex-with-v-bottom-plate-16929

[220] "Direct-zol™ RNA Miniprep Plus: TRIzol® In. RNA Out Catalog Numbers: R2070T, R2070, R2071, R2072, R2073. Instruction Manual Ver.2.0.2." Zymo Research, Jun. 20, 2022. [Online]. Available: https://files.zymoresearch.com/protocols/_r2070t_r2070_r2071_r2072_r2073_direct-zol_rna_miniprep_plus_kit.pdf

[221] QIAGEN, "Quick-Start Protocol: DNase Max Kit." QIAGEN, 2017. [Online]. Available: https://www.qiagen.com/fi/resources/resourcedetail?id=dd464b0a-57ee-4219-ac26-47cee8601923&lang=en

[222] M. Baym, S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, and R. Kishony, "Inexpensive multiplexed library preparation for megabase-sized genomes," *PLoS One*, vol. 10, no. 5, p. e0128036, 2015, doi: 10.1371/journal.pone.0128036.

[223] PACBIO, "Procedure & Checklist – Preparing HiFi SMRTbell ® Libraries using the SMRTbell Express Template Prep Kit 2.0. PN 101-853-100 Version 05." Aug. 2021.

[224] Zymoresearch, "Zymo-Seq RiboFree® Total RNA Library Kit: Any Organism. One rRNA Depletion Solution. INSTRUCTION MANUAL Ver.1.3.0." Dec. 01, 2021.

[225] F. Hahne *et al.*, "flowCore: a Bioconductor package for high throughput flow cytometry," *BMC Bioinformatics*, vol. 10, p. 106, Apr. 2009, doi: 10.1186/1471-2105-10-106.

[226] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009, doi: 10.1093/bioinformatics/btp324.

[227] A. McKenna *et al.*, "The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data," *Genome Res*, vol. 20, no. 9, pp. 1297–1303, Sep. 2010, doi: 10.1101/gr.107524.110.

[228] A. J. Page *et al.*, "SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments," *Microbial Genomics*, vol. 2, no. 4, p. e000056, doi: 10.1099/mgen.0.000056.

[229] V. Dreyer *et al.*, "Detection of low-frequency resistance-mediating SNPs in next-generation sequencing data of Mycobacterium tuberculosis complex strains with binoSNP," *Sci Rep*, vol. 10, p. 7874, May 2020, doi: 10.1038/s41598-020-64708-8.

[230] S. T. Cole *et al.*, "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence," *Nature*, vol. 393, no. 6685, pp. 537–544, Jun. 1998, doi: 10.1038/31159.

[231] J.-C. Camus, M. J. Pryor, C. Médigue, and S. T. Y. 2002 Cole, "Re-annotation of the

genome sequence of Mycobacterium tuberculosis H37Rv," *Microbiology*, vol. 148, no. 10, pp. 2967–2973, doi: 10.1099/00221287-148-10-2967.

[232] J. M. Lew, A. Kapopoulou, L. M. Jones, and S. T. Cole, "TubercuList – 10 years after," *Tuberculosis*, vol. 91, no. 1, pp. 1–7, Jan. 2011, doi: 10.1016/j.tube.2010.09.008.

[233] A. C. E. Darling, B. Mau, F. R. Blattner, and N. T. Perna, "Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements," *Genome Res*, vol. 14, no. 7, pp. 1394–1403, Jul. 2004, doi: 10.1101/gr.2289704.

[234] H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov, "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration," *Brief Bioinform*, vol. 14, no. 2, pp. 178–192, Mar. 2013, doi: 10.1093/bib/bbs017.

[235] Clara García, "Situación de la Tuberculosis Pulmonar en Población Indígena con Asentamiento en Puerto Nariño – Amazonas, Año 2016," Universidad Nacional de Colombia, Bogota, 2019.

[236] M. Cerezo-Cortés, J. Rodríguez-Castillo, R. Hernández-Pando, and M. Murcia, "Circulation of M. tuberculosis Beijing genotype in Latin America and the Caribbean," *Pathog Glob Health*, vol. 113, no. 8, pp. 336–351, Jan. 2020, doi: 10.1080/20477724.2019.1710066.

[237] F. J. Pérez-Llanos *et al.*, "Transmission Dynamics of a Mycobacterium tuberculosis Complex Outbreak in an Indigenous Population in the Colombian Amazon Region," *Microbiology Spectrum*, vol. 0, no. 0, pp. e05013-22, May 2023, doi: 10.1128/spectrum.05013-22.

[238] I. Kramnik and G. Beamer, "Mouse models of human TB pathology: roles in the analysis of necrosis and the development of host-directed therapies," *Semin Immunopathol*, vol. 38, no. 2, pp. 221–237, Mar. 2016, doi: 10.1007/s00281-015-0538-9.

[239] V. Vijayan *et al.*, "Human and murine macrophages exhibit differential metabolic responses to lipopolysaccharide - A divergent role for glycolysis," *Redox Biol*, vol. 22, p. 101147, Apr. 2019, doi: 10.1016/j.redox.2019.101147.

[240] P. B. Marko, "Sympatry," in *Encyclopedia of Ecology*, S. E. Jørgensen and B. D. Fath, Eds., Oxford: Academic Press, 2008, pp. 3450–3458. doi: 10.1016/B978-008045405-4.00549-8.

[241] L. Freschi *et al.*, "Population structure, biogeography and transmissibility of Mycobacterium tuberculosis," *Nat Commun*, vol. 12, no. 1, p. 6099, Oct. 2021, doi: 10.1038/s41467-021-26248-1.

[242] L. Baker, T. Brown, M. C. Maiden, and F. Drobniewski, "Silent Nucleotide Polymorphisms and a Phylogeny for Mycobacterium tuberculosis," *Emerg Infect Dis*, vol. 10, no. 9, pp. 1568–1577, Sep. 2004, doi: 10.3201/eid1009.040046.

[243] A. Asante-Poku *et al.*, "Mycobacterium africanum is associated with patient ethnicity in Ghana," *PLoS Negl Trop Dis*, vol. 9, no. 1, p. e3370, Jan. 2015, doi: 10.1371/journal.pntd.0003370.

[244] A. Kapopoulou, J. M. Lew, and S. T. Cole, "The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes," *Tuberculosis (Edinb)*, vol. 91, no. 1, pp. 8–13, Jan. 2011, doi: 10.1016/j.tube.2010.09.006.

[245] L. He *et al.*, "ubiA (Rv3806c) encoding DPPR synthase involved in cell wall synthesis is associated with ethambutol resistance in Mycobacterium tuberculosis," *Tuberculosis*, vol.

95, no. 2, pp. 149–154, Mar. 2015, doi: 10.1016/j.tube.2014.12.002.

[246] J. S. Solano-Gutierrez, C. Pino, and J. Robledo, “Toxin–antitoxin systems shows variability among Mycobacterium tuberculosis lineages,” *FEMS Microbiology Letters*, vol. 366, no. 1, p. fny276, Jan. 2019, doi: 10.1093/femsle/fny276.

[247] M. Chawla *et al.*, “Mycobacterium tuberculosis WhiB4 regulates oxidative stress response to modulate survival and dissemination in vivo,” *Mol Microbiol*, vol. 85, no. 6, pp. 1148–1165, Sep. 2012, doi: 10.1111/j.1365-2958.2012.08165.x.

[248] C. D. Sohaskey and L. Modesti, “Differences in nitrate reduction between Mycobacterium tuberculosis and Mycobacterium bovis are due to differential expression of both narGHJI and narK2,” *FEMS Microbiology Letters*, vol. 290, no. 2, pp. 129–134, Jan. 2009, doi: 10.1111/j.1574-6968.2008.01424.x.

[249] P. Kumar *et al.*, “PapA1 and PapA2 are acyltransferases essential for the biosynthesis of the Mycobacterium tuberculosis virulence factor sulfolipid-1,” *Proc Natl Acad Sci U S A*, vol. 104, no. 27, pp. 11221–11226, Jul. 2007, doi: 10.1073/pnas.0611649104.

[250] T. D. Sirakova, A. K. Thirumala, V. S. Dubey, H. Sprecher, and P. E. Kolattukudy, “The Mycobacterium tuberculosis pks2 Gene Encodes the Synthase for the Hepta- and Octamethyl-branched Fatty Acids Required for Sulfolipid Synthesis*,” *Journal of Biological Chemistry*, vol. 276, no. 20, pp. 16833–16839, May 2001, doi: 10.1074/jbc.M011468200.

[251] R. G. Hewinson, S. Ll. Michell, W. P. Russell, R. A. Mcadam, and W. R. Jacobs Jr, “Molecular Characterization of MPT83: a Seroreactive Antigen of Mycobacterium tuberculosis with Homology to MPT70,” *Scandinavian Journal of Immunology*, vol. 43, no. 5, pp. 490–499, 1996, doi: 10.1046/j.1365-3083.1996.d01-78.x.

[252] A. A. Siromolot, O. S. Oliinyk, D. V. Kolibo, and S. V. Komisarenko, “Mycobacterium tuberculosis antigens MPT63 and MPT83 increase phagocytic activity of murine peritoneal macrophages,” *Ukr Biochem J*, vol. 88, no. 5, pp. 62–70, 2016, doi: 10.15407/ubj88.05.062.

[253] A. J. Harrison *et al.*, “The structure of MbtI from Mycobacterium tuberculosis, the first enzyme in the biosynthesis of the siderophore mycobactin, reveals it to be a salicylate synthase,” *J Bacteriol*, vol. 188, no. 17, pp. 6081–6091, Sep. 2006, doi: 10.1128/JB.00338-06.

[254] S. Mohanty *et al.*, “Mycobacterium tuberculosis EsxO (Rv2346c) promotes bacillary survival by inducing oxidative stress mediated genomic instability in macrophages,” *Tuberculosis (Edinb)*, vol. 96, pp. 44–57, Jan. 2016, doi: 10.1016/j.tube.2015.11.006.

[255] R. S. Lee, J.-F. Proulx, F. McIntosh, M. A. Behr, and W. P. Hanage, “Previously undetected super-spreading of Mycobacterium tuberculosis revealed by deep sequencing,” *Elife*, vol. 9, p. e53245, Feb. 2020, doi: 10.7554/eLife.53245.

[256] R. S. Lee *et al.*, “Reemergence and amplification of tuberculosis in the Canadian arctic,” *J Infect Dis*, vol. 211, no. 12, pp. 1905–1914, Jun. 2015, doi: 10.1093/infdis/jiv011.

[257] R. S. Lee, J.-F. Proulx, D. Menzies, and M. A. Behr, “Progression to tuberculosis disease increases with multiple exposures,” *European Respiratory Journal*, vol. 48, no. 6, pp. 1682–1689, Dec. 2016, doi: 10.1183/13993003.00893-2016.

[258] J. Comín, A. Cebollada, and S. Samper, “Estimation of the mutation rate of

Mycobacterium tuberculosis in cases with recurrent tuberculosis using whole genome sequencing," *Sci Rep*, vol. 12, no. 1, Art. no. 1, Oct. 2022, doi: 10.1038/s41598-022-21144-0.

[259] S. Gagneux, "Host-pathogen coevolution in human tuberculosis," *Philos Trans R Soc Lond B Biol Sci*, vol. 367, no. 1590, pp. 850–859, Mar. 2012, doi: 10.1098/rstb.2011.0316.

[260] D. Tollefson, E. Bloss, A. Fanning, J. T. Redd, K. Barker, and E. McCray, "Burden of tuberculosis in indigenous peoples globally: a systematic review," *Int J Tuberc Lung Dis*, vol. 17, no. 9, pp. 1139–1150, Sep. 2013, doi: 10.5588/ijtld.12.0385.

[261] S. Fishbein, N. van Wyk, R. M. Warren, and S. L. Sampson, "Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity," *Mol Microbiol*, vol. 96, no. 5, pp. 901–916, Jun. 2015, doi: 10.1111/mmi.12981.

[262] L. S. Ates, "New insights into the mycobacterial PE and PPE proteins provide a framework for future research," *Mol Microbiol*, vol. 113, no. 1, pp. 4–21, Jan. 2020, doi: 10.1111/mmi.14409.

[263] M. Ezewudo *et al.*, "Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase," *Sci Rep*, vol. 8, no. 1, p. 15382, Oct. 2018, doi: 10.1038/s41598-018-33731-1.

[264] J. E. Phelan *et al.*, "Recombination in pe/ppe genes contributes to genetic variation in Mycobacterium tuberculosis lineages," *BMC Genomics*, vol. 17, p. 151, Feb. 2016, doi: 10.1186/s12864-016-2467-y.

[265] K. S. Walter *et al.*, "Genomic variant-identification methods may alter Mycobacterium tuberculosis transmission inferences," *Microb Genom*, vol. 6, no. 8, p. mgen000418, Jul. 2020, doi: 10.1099/mgen.0.000418.

[266] F. Di Marco, A. Spitaleri, S. Battaglia, V. Batignani, A. M. Cabibbe, and D. M. Cirillo, "Advantages of long- and short-reads sequencing for the hybrid investigation of the Mycobacterium tuberculosis genome," *Frontiers in Microbiology*, vol. 14, 2023, Accessed: Apr. 22, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1104456

[267] R. F. O'Toole and S. S. Gautam, "Limitations of the Mycobacterium tuberculosis reference genome H37Rv in the detection of virulence-related loci," *Genomics*, vol. 109, no. 5–6, pp. 471–474, Oct. 2017, doi: 10.1016/j.ygeno.2017.07.004.

[268] S. J. Bush *et al.*, "Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines," *Gigascience*, vol. 9, no. 2, p. giaa007, Feb. 2020, doi: 10.1093/gigascience/giaa007.

[269] W. De Coster, M. H. Weissensteiner, and F. J. Sedlazeck, "Towards population-scale long-read sequencing," *Nat Rev Genet*, vol. 22, no. 9, Art. no. 9, Sep. 2021, doi: 10.1038/s41576-021-00367-3.

[270] "Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats - PubMed." Accessed: Apr. 22, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30137508/?dopt=Abstract

[271] M. Merker *et al.*, "Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients," *PLoS One*, vol. 8, no. 12, p. e82551, 2013, doi: 10.1371/journal.pone.0082551.

[272] A. McIvor, H. Koornhof, and B. D. Kana, “Relapse, re-infection and mixed infections in tuberculosis disease,” *Pathogens and Disease*, vol. 75, no. 3, p. ftx020, Apr. 2017, doi: 10.1093/femspd/ftx020.

[273] T. Cohen *et al.*, “Mixed-strain mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control,” *Clin Microbiol Rev*, vol. 25, no. 4, pp. 708–719, Oct. 2012, doi: 10.1128/CMR.00021-12.

[274] I. C. Shamputa *et al.*, “Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia,” *Respir Res*, vol. 7, p. 99, Jul. 2006, doi: 10.1186/1465-9921-7-99.

[275] I. C. Shamputa *et al.*, “Genotypic and phenotypic heterogeneity among Mycobacterium tuberculosis isolates from pulmonary tuberculosis patients,” *J Clin Microbiol*, vol. 42, no. 12, pp. 5528–5536, Dec. 2004, doi: 10.1128/JCM.42.12.5528-5536.2004.

[276] J. E. Galagan, “Genomic insights into tuberculosis,” *Nat Rev Genet*, vol. 15, no. 5, pp. 307–320, May 2014, doi: 10.1038/nrg3664.

[277] N. Casali *et al.*, “Evolution and transmission of drug resistant tuberculosis in a Russian population,” *Nat Genet*, vol. 46, no. 3, pp. 279–286, Mar. 2014, doi: 10.1038/ng.2878.

[278] C. C. Díaz Acosta *et al.*, “Exploring the ‘Latin American Mediterranean’ family and the RDRio lineage in Mycobacterium tuberculosis isolates from Paraguay, Argentina and Venezuela,” *BMC Microbiology*, vol. 19, no. 1, p. 131, Jun. 2019, doi: 10.1186/s12866-019-1479-6.

[279] R. Sarkar, L. Lenders, K. A. Wilkinson, R. J. Wilkinson, and M. P. Nicol, “Modern Lineages of Mycobacterium tuberculosis Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in Human Monocyte-Derived Macrophages,” *PLOS ONE*, vol. 7, no. 8, p. e43170, Aug. 2012, doi: 10.1371/journal.pone.0043170.

[280] B. C. de Jong *et al.*, “Progression to active tuberculosis, but not transmission, varies by M. tuberculosis lineage in The Gambia,” *J Infect Dis*, vol. 198, no. 7, pp. 1037–1043, Oct. 2008, doi: 10.1086/591504.

[281] M. L. Silva, B. Cá, N. S. Osório, P. N. S. Rodrigues, A. R. Maceiras, and M. Saraiva, “Tuberculosis caused by Mycobacterium africanum: Knowns and unknowns,” *PLOS Pathogens*, vol. 18, no. 5, p. e1010490, May 2022, doi: 10.1371/journal.ppat.1010490.

[282] B. Cá *et al.*, “Experimental Evidence for Limited in vivo Virulence of Mycobacterium africanum,” *Frontiers in Microbiology*, vol. 10, 2019, Accessed: May 29, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02102

[283] B. D and G. S, “Co-evolution of Mycobacterium tuberculosis and Homo sapiens,” *Immunological reviews*, vol. 264, no. 1, Mar. 2015, doi: 10.1111/imr.12264.

[284] B. Baya *et al.*, “Association of Mycobacterium africanum Infection with Slower Disease Progression Compared with Mycobacterium tuberculosis in Malian Patients with Tuberculosis,” *Am J Trop Med Hyg*, vol. 102, no. 1, pp. 36–41, Jan. 2020, doi: 10.4269/ajtmh.19-0264.

[285] S. Uplekar, B. Heym, V. Friocourt, J. Rougemont, and S. T. Cole, “Comparative Genomics of esx Genes from Clinical Isolates of Mycobacterium tuberculosis Provides Evidence for Gene Conversion and Epitope Variation ▿,” *Infect Immun*, vol. 79, no. 10, pp. 4042–4049, Oct. 2011, doi: 10.1128/IAI.05344-11.

[286] X. Fan, L. Xie, W. Li, and J. Xie, "Prophage-like elements present in Mycobacterium genomes," *BMC Genomics*, vol. 15, no. 1, p. 243, Mar. 2014, doi: 10.1186/1471-2164-15-243.

[287] Á. Chiner-Oms *et al.*, "Genome-wide mutational biases fuel transcriptional diversity in the Mycobacterium tuberculosis complex," *Nature Communications*, vol. 10, 2019, doi: 10.1038/s41467-019-11948-6.

[288] H. R. Ramage, L. E. Connolly, and J. S. Cox, "Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution," *PLoS Genet*, vol. 5, no. 12, p. e1000767, Dec. 2009, doi: 10.1371/journal.pgen.1000767.

[289] L. Zhu, J. D. Sharp, H. Kobayashi, N. A. Woychik, and M. Inouye, "Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact," *J Biol Chem*, vol. 285, no. 51, pp. 39732–39738, Dec. 2010, doi: 10.1074/jbc.M110.163105.

[290] B. B. D. LaMarca, W. Zhu, J. E. L. Arceneaux, B. Rowe Byers, and M. D. Lundrigan, "Participation of fad and mbt Genes in Synthesis of Mycobactin in Mycobacterium smegmatis," *J Bacteriol*, vol. 186, no. 2, pp. 374–382, Jan. 2004, doi: 10.1128/JB.186.2.374-382.2004.

[291] S. S. Chavadi *et al.*, "Mutational and phylogenetic analyses of the mycobacterial mbt gene cluster," *J Bacteriol*, vol. 193, no. 21, pp. 5905–5913, Nov. 2011, doi: 10.1128/JB.05811-11.

[292] D. V. Jj, R. K, S. Bg, S. H, Z. Y, and B. Ce, "The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, Feb. 2000, doi: 10.1073/pnas.97.3.1252.

[293] Y.-M. Boudehen *et al.*, "Mycobacterial resistance to zinc poisoning requires assembly of P-ATPase-containing membrane metal efflux platforms," *Nat Commun*, vol. 13, no. 1, Art. no. 1, Aug. 2022, doi: 10.1038/s41467-022-32085-7.

[294] Y. Abu Kwaik and D. Bumann, "Microbial quest for food in vivo: 'Nutritional virulence' as an emerging paradigm," *Cellular Microbiology*, vol. 15, no. 6, pp. 882–890, 2013, doi: 10.1111/cmi.12138.

[295] J. Wei *et al.*, "Identification of a Mycobacterium tuberculosis gene that enhances mycobacterial survival in macrophages," *J Bacteriol*, vol. 182, no. 2, pp. 377–384, Jan. 2000, doi: 10.1128/JB.182.2.377-384.2000.

[296] G. Aa, L. Rk, S. R, and S. C, "Thermostable hexameric form of Eis (Rv2416c) protein of M. tuberculosis plays an important role for enhanced intracellular survival within macrophages," *PloS one*, vol. 6, no. 11, 2011, doi: 10.1371/journal.pone.0027590.

[297] M. A. Zaunbrecher, R. D. Sikes, B. Metchock, T. M. Shinnick, and J. E. Posey, "Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis," *Proc Natl Acad Sci U S A*, vol. 106, no. 47, pp. 20004–20009, Nov. 2009, doi: 10.1073/pnas.0907925106.

[298] S. Wu *et al.*, "Activation of the eis gene in a W-Beijing strain of Mycobacterium tuberculosis correlates with increased SigA levels and enhanced intracellular growth," *Microbiology (Reading)*, vol. 155, no. Pt 4, pp. 1272–1281, Apr. 2009, doi: 10.1099/mic.0.024638-0.

[299] "The PE-PPE Family of Mycobacterium tuberculosis: Proteins in Disguise - ScienceDirect." Accessed: Aug. 15, 2023. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0171298522001474

[300] K. Vijay, “Toll-like receptors in immunity and inflammatory diseases: Past, present, and future,” *Int Immunopharmacol*, vol. 59, pp. 391–412, Jun. 2018, doi: 10.1016/j.intimp.2018.03.002.

[301] C. L. Stallings and M. S. Glickman, “Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence,” *Microbes Infect*, vol. 12, no. 14–15, pp. 1091–1101, Dec. 2010, doi: 10.1016/j.micinf.2010.07.014.

[302] Q. Chai, L. Wang, C. H. Liu, and B. Ge, “New insights into the evasion of host innate immunity by Mycobacterium tuberculosis,” *Cell Mol Immunol*, vol. 17, no. 9, pp. 901–913, Sep. 2020, doi: 10.1038/s41423-020-0502-z.

[303] M. R. Babu Sait *et al.*, “PPE51 mediates uptake of trehalose across the mycomembrane of Mycobacterium tuberculosis,” *Sci Rep*, vol. 12, no. 1, Art. no. 1, Feb. 2022, doi: 10.1038/s41598-022-06109-7.

[304] M. Korycka-Machała *et al.*, “PPE51 Is Involved in the Uptake of Disaccharides by Mycobacterium tuberculosis,” *Cells*, vol. 9, no. 3, p. 603, Mar. 2020, doi: 10.3390/cells9030603.

[305] L. S. Ates *et al.*, “Mutations in ppe38 block PE_PGRS secretion and increase virulence of Mycobacterium tuberculosis,” *Nat Microbiol*, vol. 3, no. 2, pp. 181–188, Feb. 2018, doi: 10.1038/s41564-017-0090-6.

[306] C. R. E. McEvoy, P. D. van Helden, R. M. Warren, and N. C. Gey van Pittius, “Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic Mycobacterium tuberculosis PPE38 gene region,” *BMC Evol Biol*, vol. 9, p. 237, Sep. 2009, doi: 10.1186/1471-2148-9-237.

[307] C. Raynaud *et al.*, “Phospholipases C are involved in the virulence of Mycobacterium tuberculosis,” *Mol Microbiol*, vol. 45, no. 1, pp. 203–217, Jul. 2002, doi: 10.1046/j.1365-2958.2002.03009.x.

[308] F. Le Chevalier *et al.*, “Revisiting the role of phospholipases C in virulence and the lifecycle of Mycobacterium tuberculosis,” *Sci Rep*, vol. 5, no. 1, Art. no. 1, Nov. 2015, doi: 10.1038/srep16918.

[309] “Mycobacterial lipolytic enzymes: A gold mine for tuberculosis research - ScienceDirect.” Accessed: May 04, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0300908412002908

[310] D. H. Schmiel and V. L. Miller, “Bacterial phospholipases and pathogenesis,” *Microbes Infect*, vol. 1, no. 13, pp. 1103–1112, Nov. 1999, doi: 10.1016/s1286-4579(99)00205-1.

[311] J. C. Bakala N’goma, M. Schué, F. Carrière, A. Geerlof, and S. Canaan, “Evidence for the cytotoxic effects of Mycobacterium tuberculosis phospholipase C towards macrophages,” *Biochim Biophys Acta*, vol. 1801, no. 12, pp. 1305–1313, Dec. 2010, doi: 10.1016/j.bbalip.2010.08.007.

[312] P. A. Assis *et al.*, “Mycobacterium tuberculosis expressing phospholipase C subverts PGE2 synthesis and induces necrosis in alveolar macrophages,” *BMC Microbiology*, vol. 14, no. 1, p. 128, May 2014, doi: 10.1186/1471-2180-14-128.

[313] L. G. Wayne and L. G. Hayes, “Nitrate reduction as a marker for hypoxic shiftdown of Mycobacterium tuberculosis,” *Tuber Lung Dis*, vol. 79, no. 2, pp. 127–132, 1998, doi: 10.1054/tuld.1998.0015.

[314] S. Virtanen, “A study of nitrate reduction by mycobacteria. The use of the nitrate reduction test in the identification of

mycobacteria," *Acta Tuberc Scand Suppl*, vol. 48, pp. 1–119, 1960.

[315] L. G. Wayne and J. R. Doubek, "CLASSIFICATION AND IDENTIFICATION OF MYCOBACTERIA. II. TESTS EMPLOYING NITRATE AND NITRITE AS SUBSTRATE," *Am Rev Respir Dis*, vol. 91, pp. 738–745, May 1965, doi: 10.1164/arrd.1965.91.5.738.

[316] C. D. Sohaskey and L. G. Wayne, "Role of narK2X and narGHJI inHypoxic Upregulation of Nitrate Reduction byMycobacteriumtuberculosis," *Journal of Bacteriology*, vol. 185, no. 24, pp. 7247–7256, Dec. 2003, doi: 10.1128/jb.185.24.7247-7256.2003.

[317] M. Kj, D. Aj, N. R, D. Md, and M. Aw, "Structure, function and drug targeting in Mycobacterium tuberculosis cytochrome P450 systems," *Archives of biochemistry and biophysics*, vol. 464, no. 2, Aug. 2007, doi: 10.1016/j.abb.2007.03.026.

[318] T. R. Gompo *et al.*, "Risk factors of tuberculosis in human and its association with cattle TB in Nepal: A one health approach," *One Health*, vol. 10, p. 100156, Aug. 2020, doi: 10.1016/j.onehlt.2020.100156.

[319] J. Gonzalo-Asensio *et al.*, "Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator," *Proceedings of the National Academy of Sciences*, vol. 111, no. 31, pp. 11491–11496, Aug. 2014, doi: 10.1073/pnas.1406693111.

[320] M. S, C. D, B. J, A. A, and B. Ma, "Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex," *The Journal of infectious diseases*, vol. 186, no. 1, Jul. 2002, doi: 10.1086/341068.

[321] S. Mostowy *et al.*, "Genomic Analysis Distinguishes Mycobacterium africanum," *J Clin Microbiol*, vol. 42, no. 8, pp. 3594–3599, Aug. 2004, doi: 10.1128/JCM.42.8.3594-3599.2004.

[322] S. E. Hoffner, S. B. Svenson, R. Norberg, F. Dias, S. Ghebremichael, and G. Källenius, "Biochemical heterogeneity of Mycobacterium tuberculosis complex isolates in Guinea-Bissau," *J Clin Microbiol*, vol. 31, no. 8, pp. 2215–2217, Aug. 1993, doi: 10.1128/jcm.31.8.2215-2217.1993.

[323] "Numerical Taxonomy Analysis of Mycobacterium africanum | Microbiology Society." Accessed: Aug. 18, 2023. [Online]. Available: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-28-4-464?crawler=true

[324] D. J. V. Beste, M. Espasa, B. Bonde, A. M. Kierzek, G. R. Stewart, and J. McFadden, "The Genetic Requirements for Fast and Slow Growth in Mycobacteria," *PLoS One*, vol. 4, no. 4, p. e5349, Apr. 2009, doi: 10.1371/journal.pone.0005349.

[325] S. Chavadi *et al.*, "Global Effects of Inactivation of the Pyruvate Kinase Gene in the Mycobacterium tuberculosis Complex," *J Bacteriol*, vol. 191, no. 24, pp. 7545–7553, Dec. 2009, doi: 10.1128/JB.00619-09.

[326] G. Rehren, S. Walters, P. Fontan, I. Smith, and A. M. Zárraga, "Differential gene expression between Mycobacterium bovis and Mycobacterium tuberculosis," *Tuberculosis (Edinb)*, vol. 87, no. 4, pp. 347–359, Jul. 2007, doi: 10.1016/j.tube.2007.02.004.

[327] G. Khare, P. Nangpal, and A. K. Tyagi, "Differential Roles of Iron Storage Proteins in Maintaining the Iron Homeostasis in Mycobacterium tuberculosis," *PLoS One*, vol. 12, no. 1, p. e0169545, Jan. 2017, doi: 10.1371/journal.pone.0169545.

[328] A. Singh *et al.*, "Mycobacterium tuberculosis WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response," *PLoS Pathog*, vol. 5, no. 8, p. e1000545, Aug. 2009, doi: 10.1371/journal.ppat.1000545.

[329] N. Rathor, K. Garima, N. K. Sharma, A. Narang, M. Varma-Basil, and M. Bose, "Expression profile of mce4 operon of Mycobacterium tuberculosis following environmental stress," *International Journal of Mycobacteriology*, vol. 5, no. 3, pp. 328–332, Sep. 2016, doi: 10.1016/j.ijmyco.2016.08.004.

[330] M. Berney and G. M. Cook, "Unique Flexibility in Energy Metabolism Allows Mycobacteria to Combat Starvation and Hypoxia," *PLOS ONE*, vol. 5, no. 1, p. e8614, Jan. 2010, doi: 10.1371/journal.pone.0008614.

[331] H. G. Wiker, "MPB70 and MPB83 – Major Antigens of Mycobacterium bovis," *Scandinavian Journal of Immunology*, vol. 69, no. 6, pp. 492–499, 2009, doi: 10.1111/j.1365-3083.2009.02256.x.

[332] H. G. Wiker, S. Nagai, R. G. Hewinson, W. P. Russell, and M. Harboe, "Heterogenous expression of the related MPB70 and MPB83 proteins distinguish various substrains of Mycobacterium bovis BCG and Mycobacterium tuberculosis H37Rv," *Scand J Immunol*, vol. 43, no. 4, pp. 374–380, Apr. 1996, doi: 10.1046/j.1365-3083.1996.d01-61.x.

[333] M. E. J. Woolhouse, J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin, "Biological and biomedical implications of the co-evolution of pathogens and their hosts," *Nat Genet*, vol. 32, no. 4, Art. no. 4, Dec. 2002, doi: 10.1038/ng1202-569.

[334] Marko, P.B., *Sympatry*. in In Encyclopedia of Ecology. Cambridge, MA, USA,: Academic Press: Cambridge, 2008.

[335] D. G. Russell *et al.*, "Mycobacterium tuberculosis Wears What It Eats," *Cell Host & Microbe*, vol. 8, no. 1, pp. 68–76, Jul. 2010, doi: 10.1016/j.chom.2010.06.002.

[336] E. Suter, "The multiplication of tubercle bacilli within normal phagocytes in tissue culture," *J Exp Med*, vol. 96, no. 2, pp. 137–150, Aug. 1952, doi: 10.1084/jem.96.2.137.

[337] B. K. Johnson, S. M. Thomas, A. J. Olive, and R. B. Abramovitch, "Macrophage Infection Models for Mycobacterium tuberculosis," *Methods Mol Biol*, vol. 2314, pp. 167–182, 2021, doi: 10.1007/978-1-0716-1460-0_6.

[338] "On the killing of mycobacteria by macrophages - Jordao - 2008 - Cellular Microbiology - Wiley Online Library." Accessed: May 13, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1462-5822.2007.01067.x

[339] "Comparison of human monocyte derived macrophages and THP1-like macrophages as in vitro models for M. tuberculosis infection," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 67, p. 101355, Dec. 2019, doi: 10.1016/j.cimid.2019.101355.

[340] J. Kim, B.-K. Koo, and J. A. Knoblich, "Human organoids: model systems for human biology and medicine," *Nat Rev Mol Cell Biol*, vol. 21, no. 10, Art. no. 10, Oct. 2020, doi: 10.1038/s41580-020-0259-3.

[341] N. Hadjesfandiari, M. Khorshidfar, and D. V. Devine, "Current Understanding of the Relationship between Blood Donor Variability and Blood Component Quality," *Int J Mol Sci*, vol. 22, no. 8, p. 3943, Apr. 2021, doi: 10.3390/ijms22083943.

[342] Y. Nédélec *et al.*, "Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens," *Cell*, vol.

167, no. 3, pp. 657-669.e21, Oct. 2016, doi: 10.1016/j.cell.2016.09.025.

[343] A. Asante-Poku *et al.*, "TB-diabetes co-morbidity in Ghana: The importance of Mycobacterium africanum infection," *PLoS One*, vol. 14, no. 2, p. e0211822, 2019, doi: 10.1371/journal.pone.0211822.

[344] N. Krishnan *et al.*, "Mycobacterium tuberculosis Lineage Influences Innate Immune Response and Virulence and Is Associated with Distinct Cell Envelope Lipid Profiles," *PLOS ONE*, vol. 6, no. 9, p. e23870, Sep. 2011, doi: 10.1371/journal.pone.0023870.

[345] P. Ranaivomanana, M. S. Rabodoarivelo, M. D. B. Ndiaye, N. Rakotosamimanana, and V. Rasolofo, "Different PPD-stimulated cytokine responses from patients infected with genetically distinct Mycobacterium tuberculosis complex lineages," *Int J Infect Dis*, vol. 104, pp. 725–731, Mar. 2021, doi: 10.1016/j.ijid.2021.01.073.

[346] A. Mantovani, C. A. Dinarello, M. Molgora, and C. Garlanda, "Interleukin-1 and Related Cytokines in the Regulation of Inflammation and Immunity," *Immunity*, vol. 50, no. 4, pp. 778–795, Apr. 2019, doi: 10.1016/j.immuni.2019.03.012.

[347] A. C. Bohrer, C. Tocheny, M. Assmann, V. V. Ganusov, and K. D. Mayer-Barber, "IL-1R1 mediates host resistance to Mycobacterium tuberculosis by trans-protection of infected cells," *J Immunol*, vol. 201, no. 6, pp. 1645–1650, Sep. 2018, doi: 10.4049/jimmunol.1800438.

[348] R. Domingo-Gonzalez, O. Prince, A. Cooper, and S. A. Khader, "Cytokines and Chemokines in Mycobacterium tuberculosis Infection," *Microbiol Spectr*, vol. 4, no. 5, Oct. 2016, doi: 10.1128/microbiolspec.TBTB2-0018-2016.

[349] A. Olsen *et al.*, "Targeting Mycobacterium tuberculosis Tumor Necrosis Factor Alpha-Downregulating Genes for the Development of Antituberculous Vaccines," *mBio*, vol. 7, no. 3, pp. e01023-15, Mar. 2016, doi: 10.1128/mBio.01023-15.

[350] E. Guirado, L. S. Schlesinger, and G. Kaplan, "Macrophages in Tuberculosis: Friend or Foe," *Semin Immunopathol*, vol. 35, no. 5, pp. 563–583, Sep. 2013, doi: 10.1007/s00281-013-0388-2.

[351] H. Yamada, S. Mizumo, R. Horai, Y. Iwakura, and I. Sugawara, "Protective role of interleukin-1 in mycobacterial infection in IL-1 alpha/beta double-knockout mice," *Lab Invest*, vol. 80, no. 5, pp. 759–767, May 2000, doi: 10.1038/labinvest.3780079.

[352] B. E. Schneider *et al.*, "A role for IL-18 in protective immunity against Mycobacterium tuberculosis," *Eur J Immunol*, vol. 40, no. 2, pp. 396–405, Feb. 2010, doi: 10.1002/eji.200939583.

[353] Z. B. Zalinger, R. Elliott, and S. R. Weiss, "Role of the inflammasome-related cytokines Il-1 and Il-18 during infection with murine coronavirus," *J Neurovirol*, vol. 23, no. 6, pp. 845–854, 2017, doi: 10.1007/s13365-017-0574-4.

[354] I. Dambuza *et al.*, "Efficacy of membrane TNF mediated host resistance is dependent on mycobacterial virulence," *Tuberculosis (Edinb)*, vol. 88, no. 3, pp. 221–234, May 2008, doi: 10.1016/j.tube.2007.08.011.

[355] K. C. Wong *et al.*, "Molecular characterization of clinical isolates of Mycobacterium tuberculosis and their association with phenotypic virulence in human macrophages," *Clin Vaccine Immunol*, vol. 14, no. 10, pp. 1279–1284, Oct. 2007, doi: 10.1128/CVI.00190-07.

[356] A. Romagnoli *et al.*, "Clinical isolates of the modern Mycobacterium tuberculosis lineage 4 evade host defense in human macrophages through eluding IL-1β-induced autophagy," *Cell Death Dis*, vol. 9, no. 6, Art. no. 6, May 2018, doi: 10.1038/s41419-018-0640-8.

[357] M. Pilli *et al.*, "TBK-1 promotes autophagy-mediated antimicrobial defense by controlling autophagosome maturation," *Immunity*, vol. 37, no. 2, pp. 223–234, Aug. 2012, doi: 10.1016/j.immuni.2012.04.015.

[358] F. G. Boni, I. Hamdi, L. M. Koundi, K. Shrestha, and J. Xie, "Cytokine storm in tuberculosis and IL-6 involvement," *Infection, Genetics and Evolution*, vol. 97, p. 105166, Jan. 2022, doi: 10.1016/j.meegid.2021.105166.

[359] A. Gidon, C. Louet, L. M. Røst, P. Bruheim, and T. H. Flo, "The Tumor Necrosis Factor Alpha and Interleukin 6 Auto-paracrine Signaling Loop Controls Mycobacterium avium Infection via Induction of IRF1/IRG1 in Human Primary Macrophages," *mBio*, vol. 12, no. 5, p. 10.1128/mbio.02121-21, Oct. 2021, doi: 10.1128/mbio.02121-21.

[360] B. M. Saunders, A. A. Frank, I. M. Orme, and A. M. Cooper, "Interleukin-6 induces early gamma interferon production in the infected lung but is not required for generation of specific immunity to Mycobacterium tuberculosis infection," *Infect Immun*, vol. 68, no. 6, pp. 3322–3326, Jun. 2000, doi: 10.1128/IAI.68.6.3322-3326.2000.

[361] C. G. Larsen *et al.*, "The delayed-type hypersensitivity reaction is dependent on IL-8. Inhibition of a tuberculin skin reaction by an anti-IL-8 monoclonal antibody," *J Immunol*, vol. 155, no. 4, pp. 2151–2157, Aug. 1995.

[362] X. Ma *et al.*, "Association between Interleukin-8 Gene Alleles and Human Susceptibility to Tuberculosis Disease," *The Journal of Infectious Diseases*, vol. 188, no. 3, pp. 349–355, Aug. 2003, doi: 10.1086/376559.

[363] C. Kewcharoenwong, W. Saenwongsa, S. J. Willcocks, G. J. Bancroft, H. A. Fletcher, and G. Lertmemongkolchai, "Glibenclamide alters interleukin-8 and interleukin-1β of primary human monocytes from diabetes patients against Mycobacterium tuberculosis infection," *Tuberculosis (Edinb)*, vol. 123, p. 101939, Jul. 2020, doi: 10.1016/j.tube.2020.101939.

[364] H. L. Van Epps, "Chemokine drives tuberculosis," *J Exp Med*, vol. 202, no. 12, p. 1614, Dec. 2005, doi: 10.1084/jem20212iti2.

[365] "Innate Immune Response to Mycobacterium tuberculosis Beijing and Other Genotypes | PLOS ONE." Accessed: May 28, 2023. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013594

[366] B. Li *et al.*, "Comprehensive evaluation of the effects of long-term cryopreservation on peripheral blood mononuclear cells using flow cytometry," *BMC Immunology*, vol. 23, no. 1, p. 30, Jun. 2022, doi: 10.1186/s12865-022-00505-4.

[367] A. Koryakina, E. Frey, and P. Bruegger, "Cryopreservation of human monocytes for pharmacopeial monocyte activation test," *Journal of Immunological Methods*, vol. 405, pp. 181–191, Mar. 2014, doi: 10.1016/j.jim.2014.01.005.

[368] L. D. Tientcheu, A. Koch, M. Ndengane, G. Andoseh, B. Kampmann, and R. J. Wilkinson, "Immunological consequences of strain variation within the Mycobacterium tuberculosis complex," *Eur J Immunol*, vol. 47, no. 3, pp. 432–445, Mar. 2017, doi: 10.1002/eji.201646562.

[369] C. Wang *et al.*, "Innate Immune Response to Mycobacterium tuberculosis

Beijing and Other Genotypes," *PLOS ONE*, vol. 5, no. 10, p. e13594, Oct. 2010, doi: 10.1371/journal.pone.0013594.

[370] L. D. Tientcheu *et al.*, "Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment," *Genes Immun*, vol. 16, no. 5, Art. no. 5, Jul. 2015, doi: 10.1038/gene.2015.21.

[371] M. Zwyer *et al.*, "Back-to-Africa introductions of Mycobacterium tuberculosis as the main cause of tuberculosis in Dar es Salaam, Tanzania," *PLOS Pathogens*, vol. 19, no. 4, p. e1010893, Apr. 2023, doi: 10.1371/journal.ppat.1010893.

[372] M. Laamarti, Y. El Fathi Lalaoui, R. Elfermi, R. Daoud, and A. El Allali, "Afro-TB dataset as a large scale genomic data of Mycobacterium tuberuclosis in Africa," *Sci Data*, vol. 10, no. 1, Art. no. 1, Apr. 2023, doi: 10.1038/s41597-023-02112-3.

[373] M. B. Reed *et al.*, "A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response," *Nature*, vol. 431, no. 7004, pp. 84–87, Sep. 2004, doi: 10.1038/nature02837.

[374] I. Palucci *et al.*, "PE_PGRS33 Contributes to Mycobacterium tuberculosis Entry in Macrophages through Interaction with TLR2," *PLOS ONE*, vol. 11, no. 3, p. e0150800, Mar. 2016, doi: 10.1371/journal.pone.0150800.

[375] S. A. Yimer *et al.*, "Lineage-Specific Proteomic Signatures in the Mycobacterium tuberculosis Complex Reveal Differential Abundance of Proteins Involved in Virulence, DNA Repair, CRISPR-Cas, Bioenergetics and Lipid Metabolism," *Frontiers in Microbiology*, vol. 11, 2020, Accessed: Sep. 03, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2020.550760

[376] Y. Zhang *et al.*, "Enhanced interleukin-8 release and gene expression in macrophages after exposure to Mycobacterium tuberculosis and its components.," *J Clin Invest*, vol. 95, no. 2, pp. 586–592, Feb. 1995.

[377] W.-W. Liao *et al.*, "A draft human pangenome reference," *Nature*, vol. 617, no. 7960, Art. no. 7960, May 2023, doi: 10.1038/s41586-023-05896-x.

[378] V. Dreyer *et al.*, "Detection of low-frequency resistance-mediating SNPs in next-generation sequencing data of Mycobacterium tuberculosis complex strains with binoSNP," *Scientific Reports*, vol. 10, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41598-020-64708-8.

[379] T. Weniger, J. Krawczyk, P. Supply, S. Niemann, and D. Harmsen, "MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria," *Nucleic Acids Res*, vol. 38, no. Web Server issue, pp. W326-331, Jul. 2010, doi: 10.1093/nar/gkq351.

[380] T. A. Kohl *et al.*, "MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates," *PeerJ*, vol. 6, p. e5895, Nov. 2018, doi: 10.7717/peerj.5895.

10. SUPPLEMENTARY MATERIAL

10.1 Consumables

Product	Catalog Number	Supplier
Aluminum Adhesive Film 96well	10716943	Fisher Scientific
Biosphere Filter Tips 10 µL	70.1114.210	Sarstedt
Biosphere Filter Tips 100 µL	70.760.212	Sarstedt
Biosphere Filter Tips 200 µL	70.760.211	Sarstedt
Biosphere Filter Tips 1250 µL	70.1186.210	Sarstedt
Cannula 25G x1”- Nr. 18 0.5-25 mm	300400	Becton Dickinson
Cuvettes 10x4x45 Mm	67.742	Sarstedt
Cuvette Single-Use Lids	XK25.1	Roth
Dispenser-Tips II 12.5	705714	BRAND. For Lab. For Life.®
Disposable Cell Scraper	04396-53	Cole-Parmer
Dye Box According to Coplin (Coplin Staining Jar)	YC09.1	Roth
Entellan	1079610100	Merck
G-Tube	10145	Covaris
Handy Step	1302-0034	BRAND. For Lab. For Life.®
Inoculation Loops (sterile)	612-2495	VWR
8-Lid Chain, Flat	65.989.002	Sarstedt
Lysing Matrix B	116911100	MP Biomedicals
Macs Ls Column	130-042-401	MiItenyi Biotec
Microscope Slides	03-0060	R. Langenbrinck
MULTIPLY-µstripe 0.2 ML CHAIN	72.985.002	Sarstedt
Nonstick, RNase-free Tubes, 1.5 mL	AM12450	Thermo Fisher
Oil Tube Septa	100-667-700	PacBio
Pack 5 Backings 4 Arrays per Slide	G2534-60011	Agilent Technologies
Petri Dishes 92 x 16 mm	82.1472	Sarstedt
Polystyrene Round-Bottom Tube 5 mL 12x75 mm style	352054	FALCON
Reagent Reservoir (55 mL, sterile)	E831.2	Roth
Reagent Reservoir (55 mL, sterile)	E831.2	Roth
RNase Free Tubes (1.5 and 2 mL)	72.705.200	Sarstedt
Roller Bottle (490cm^2)	430195	Corning
Rotilabo® Liquid Reservoirs	E831.2	Roth
Sealmat Für 96 Well Plates	710889	Biozym
Sequel Sample Plate Foil	AB-0757	Thermo Scientific
Sequel Pipet Tips V2	100-709-301	PacBio
Slides Superfrost	03-0060	Langenbrinck
Small Cell Scraper	3010	Corning Inc.
Square Media Bottle, Sterile, PETG, 30 mL	2019-0030	Nalgene™
Square Media Bottle, Sterile, PETG, 60 mL	2019-0060	Nalgene™

Stripette® Serological Pipette 10 mL	4488	Corning Inc.
Stripette® Serological Pipette 25 mL	4489	Corning Inc.
Stripette® Serological Pipette 5 mL	4487	Corning Inc.
Syringe 1 mL	303172	Becton Dickinson
Syringe 50 mL	300865	Becton Dickinson
Transfer pipette 3.5 mL	861171010	Sarstedt
Tube 50 mL 114x28mm DNase, RNase-free Sterile	62547254	Sarstedt
Tube 15 mL 120x17mm DNase, RNase-free Sterile	62554502	Sarstedt
Tubes 1.5 mL	72.690.001	Sarstedt
Tubes 0.5 mL (Qubit)	732-0675	Axygen
96 Well Cell Culture Plate, Flat Bottom with Lid Tissue Culture Treated	3596	Corning Inc.
96 Well PCR Plate	72.1979.202	Sarstedt
6 Well Cell Culture Plate, Flat Bottom with Lid Tissue Culture Treated	3516	Corning Inc.
96-Well Plate, Flat Bottom, Sterile	833.924	Sarstedt
96 Well Plate Hard Shell, Skirted	HSP9601	BioRad
48 Well Plates. Nunclon ™ Delta Surface	150687	Thermo Scientific
24 Well Plates. Nunclon ™ Delta Surface	142475	Thermo Scientific
96-Well Plate, U-Shape Bottom, Sterile	833.925	Sarstedt
Vi-CELL™ XR 4 Bags of Sample Vials	383721	Beckman Coulter

10.2 Reagents

Product	**Catalog Number**	**Supplier**
Accutase Cell Detachment Solution	561527	Becton Dickinson
AMPure® PB Beads	100-265-900	PacBio
Auto MACS™ Rising solution	130-091-222	Miltenyi Biotec
Bovine M-CSF Macrophage Colony Stimulating Factor Recombinant Bovine	RP1353B-025	Kingfisher Biotech, Inc.
Brain Heart Infusion Broth	X916.2	Roth
Carbolic fuchsine solution	A130.1	Roth
CD14 MicroBeads, Human	130-050-201	Miltenyi Biotec
cOmplete™, Mini, EDTA-free Protease-Inhibitor-Cocktail Tablets	04693159001	Sigma-Aldrich
Cycloheximide Biochemica	A0879,005	PanReac AppliChem
DEPC H_2O	2317912	Roth
Difco™ Middlebrook 7H10 Agar	262710	Becton Dickinson
Difco™ Middlebrook 7H9 Broth 500g	271310	Becton Dickinson
Dimethyl Sulfoxide, sterile filtered (DMSO)	A7248,0010	PanReac AppliChem
Elution Buffer	101-633-500	PacBio
Entellan	1079610100	Merck
Ethanol 100% (RNA work)	108543	Merck

Ethanol 100% (Microbiology work)	2005786	Roth
Ethidium Bromide Dye	2218.1	Roth
GeneScan-1200 LIZ	4379950	Thermo Fisher Scientific
Hanks Balanced Salt Solution (HBSS)	L2055	Merck Millipore
HCL- Alcohol 3% ready to use	6477.2	Roth
Heat-inactivated Fetal Calf Serum (FCS)	S0115	Biochrom
Heparin Natrium 25.000 I.E./ 5 mL	1708.00.00	BRAUN
LB-Broth	L3022	Sigma-Aldrich
L-Glutamin 200Mm low endotoxin	K0283	Biochrom
Löfflers Methylen Blue 0,5% ready to use	AE64.1	Roth
MACS BSA Stock Solution	130-091-376	Miltenyi Biotec
MagSi-NGS Prep Plus	SL-MDKT01075	Steinbrenner
OADC Middlebrook Enrichment	212351	Becton Dickinson
Pancoll	P04-60500	PAN Biotec
Paraformaldehyde	P6148-500G	Sigma-Aldrich
Penicillin Streptomycin 10000 µL /mL	A2213	Biochrom
peqGOLD Universal Agarose	35-1020	peqlab
Phenol; 250 mL	0038.2	Roth
10 x Phosphate-Buffered Saline (PBS)	L1835	Biochrom
2-Propanol	T910.1	Roth
Pursept® A Xpress	230131	Schülke & Mayr
rhM-CSF Macrophage Colony-Stimulating Factor Recombinant Human (E. coli derived)	216-M	R&D systems
RNase-ExitusPlus™	A7153,0500	Applichem
Saponin	S7900-25G	Sigma-Aldrich
Sequel SMRT Cell Oil	100-621-300	PacBio
Sodium Dodecyl Sulfate (SDS)	2326.2	Roth
Triton X® 100	3051.3	Roth
TRIzol® Reagent 200 mL	15596-018	Life Technologies
Tween^{20}	1.09280.0100	Merck Millipore
Tween^{80}	P8074	Sigma-Aldrich
VLE RPMI 1640	F1415	Biochrom
Water	0082479E	Braun
Xylol	601-022-00-9	CMP GmbH & Co. KG

10.3 Kits

Kit	Catalog Number	Supplier
Barcoded Overhang Adapter Kit- 8A	101-628-400	PacBio
Direct-zol™ RNA MiniPrep Plus (50rxns) DNase Max® Kit	R2070 79254	Zymo Research Qiagen
DNF-472 High Sensitivity RNA Analysis Kit, 15nt	DNF-472-0500	Advanced Analytical Technologies, Inc.
DNF-473 Standard Sensitivity NGS Fragment Analysis Kit, (1bp – 6000 bp)	DNF-473-1000	Advanced Analytical Technologies, Inc.
DNF-474 Standard Sensitivity NGS Fragment Analysis Kit, (1bp – 6000 bp)	DNF-474-1000	Advanced Analytical Technologies, Inc.
MIRU-VNTR Typing kit QUADRUPLEX *M.tuberculosis* complex	785-60VQuad	Genoscreen
Mo Bio DNase Max Kit	15200-50	Qiagen
Next Seq v2.5 Mid Output 300 Cycles	20024905	Illumina
Next Seq Output 150 Cycles	20024904	Illumina
Nextera XT DNA Sample Preparation Kit (96 Samples)	FC-131-1096	Illumina
Nextera XT Index Kit v2 Set A (96 indices, 384 samples)	FC-131-2001	Illumina
Nextera XT Index Kit v2 Set B (96 indices, 384 samples)	FC-131-2002	Illumina
Nextera XT Index Kit v2 Set C (96 indices, 384 samples)	FC-131-2003	Illumina
Nextera XT Index Kit v2 Set D (96 indices, 384 samples)	FC-131-2004	Illumina
PureLinkGenomic DNA Mini Kit	K182001	Invitrogen
Qubit® dsDNA HS Assay kit	Q32854	Life Technologies
Qubit® RNA BR Assay kits RNA Clean & Concentrator™-5 kit	Q32852 R1015	Life Technologies Zymo Research
RNase-ExitusPlus™	A7153,0500	Applichem
RNeasy Protect Bacteria Mini Kit	74524	Qiagen
Sequel II Binding Kit 2.0	101-789-500	PacBio
Sequel II DNA Internal Control Complex 1.0	650-521-8000	PacBio
Sequencing Primer v2	101-847-900	PacBio
Sequel II Sequencing Kit 2.0 (4 rxn)	101-820-200	PacBio
SMRTbell Enzyme Clean Up Kit	101-746-400	PacBio
Sequencing Primer v2	101-847-900	PacBio
SMRTbell Express Template Prep Kit 2.0	100-938-900	PacBio
SMRTbell Enzyme Clean Up Kit	101-746-400	PacBio
Taq DNA Polymerase with Standard Taq Buffer	M0273S	New England Biolabs
Trizol: peqGOLD TriFast™	30-2010	VWR (Peqlab)
Vi-CELL ®Reagent Park	383722	Beckman Coulter
Zymo-Seq RiboFree™ Total RNA Library kit	R3000	Zymo Research

10.4 Antibodies

Marker	Fluorophore	Catalogue	Supplier
CD11b Mouse Ani-Human	PE-Cy7	557743	BIO-RAD
CD14 Anti-human, REAfinity™	PerCP-Vio 700	130-110-578	Miltenyi Biotec
CD16 Mouse Anti Human	FITC	MCA5665F	BIO-RAD
CD68 Mouse Anti -Human	FITC	562117	BIO-RAD
CD163 Mouse Anti-Human	BV421	562643	BIO-RAD
CD172 Ma/b Mouse Anti-Human	BV510	743563	BIO-RAD
CD206 Mouse Anti-Human	APC	550889	BIO-RAD

10.5 Primers

Oligo Name	Sequence (5’ -> 3’)	OD	µg	nmol	TM (°C)	GC content
rrs_F_1202	CTTATGTCCAGGGCTTCA	15.72	463	84,7	53,7	50%
rrs_R_1565	CAGTTGGGGCGTTTTC	13.43	405	82,2	51,7	56.3%

10.6 Laboratory Instruments and Facilities

Device	Supplier
Analytical Balance EMB 2000-2	KERN
Axiocam 202 Mono	ZEISS
Beckman J-6M/E Centrifuge	BeckMan Coulter
Bench top refrigerated centrifuge with swing buckets with Maximum 5000 rpm.	Thermo Scientific
BD BACTEC™ MGIT™ 320	Becton Dickinson
Biometra Compact Multi-Wide	Analytikjena
CASY Cell Counter Analyzer System Model TT	Innovatis
Cell Culture Incubator C 170 CO2	Binder
Cell Density Meter 40	Fisher Scientific
Centrifuge 5430	Eppendorf
Combi-Spin FVL-2400N	Kisker
DS-11FX	DeNovix
DynaMag™-PCR Magnet	Thermo Fisher Scientific
FastPrep 24-5G	MP Biomedicals
Fragment Analyzer	Advanced Analytica
Gel Doc XR + Gel Documentation System	BIO-RAD
Genetic Analyzer 3500XL	Applied Biosystems™
Heat Sealer 96-well	Thermo Fisher Scientific
Heating Block	Biozym Scientific GmbH
MACSQuant Analyzer 10 benchtop Flow Cytometer	Miltenyi Biotec

MACS® MultiStand	Miltenyi Biotec
MidiMACS ™ Manual MACS® Magnetic Separators for Cell Separation.	Miltenyi Biotec
Multi-Channel pipettes (100 µL, 300 µL)	Eppendorf
Multifuge X Pro Series Centrifuge	Thermo Scientific
NextSeq 500	illumina
NanoDrop TM 2000	Thermo Fisher Scientific Inc
Light Microscope Primovert	ZEISS
LSR II Flow Cytometer	Becton Dickinson
Pipettes (10 µL, 100 µL, 200 µL, 1000 µL)	Eppendorf
PowerPacTM Basic 300v|400 mA|75W	BIO-RAD
Pipet boy	Integra Biosciences
Precision Water Bath GP 10	Thermo Scientific
Qubit 3.0	life technologies
Roller Apparatus	Wheaton
Rotilabo®-Mini-Zentrifuge	Roth
Sequel II System	PacBio
Spectrophotometer DS-11 FX	DeNovix
Spectrofotometer / Fluorometer DS-11 FX	DeNovix
Thermocycler C1000 Touch	Biorad
Vi-CELL Cell Viability Counter	Beckman
Vacuum System	Vacuubrand
Vortex Genie2	Eppendorf PCR Cooler

10.7 Software and Coding Tools

Software	**Type**	**Supplier**	**Source**
binoSNP	OSS	V. Dreyer *et al.* [378]	https://github.com/ngs-fzb/binoSNP
BioNumerics Software Version 7.6	CSS	Applied Maths	https://www.applied-maths.com/news/bionumerics-version-76-released
BioRender	CSS	BioRender	https://BioRender.com/
Conda management system	OSS	Anaconda.org	https://anaconda.org/anaconda/conda
DESeq2	OSS	Bioconductor project	https://bioconductor.org/packages/release/bioc/htmL/DESeq2.htmL
FCS Express V. 7	CSS	De Novo Software	https://denovosoftware.com/
Gene Mapper ™ Software 6	CSS	Applied Biosystems™	https://www.thermofisher.com/order/catalog/product/4475073

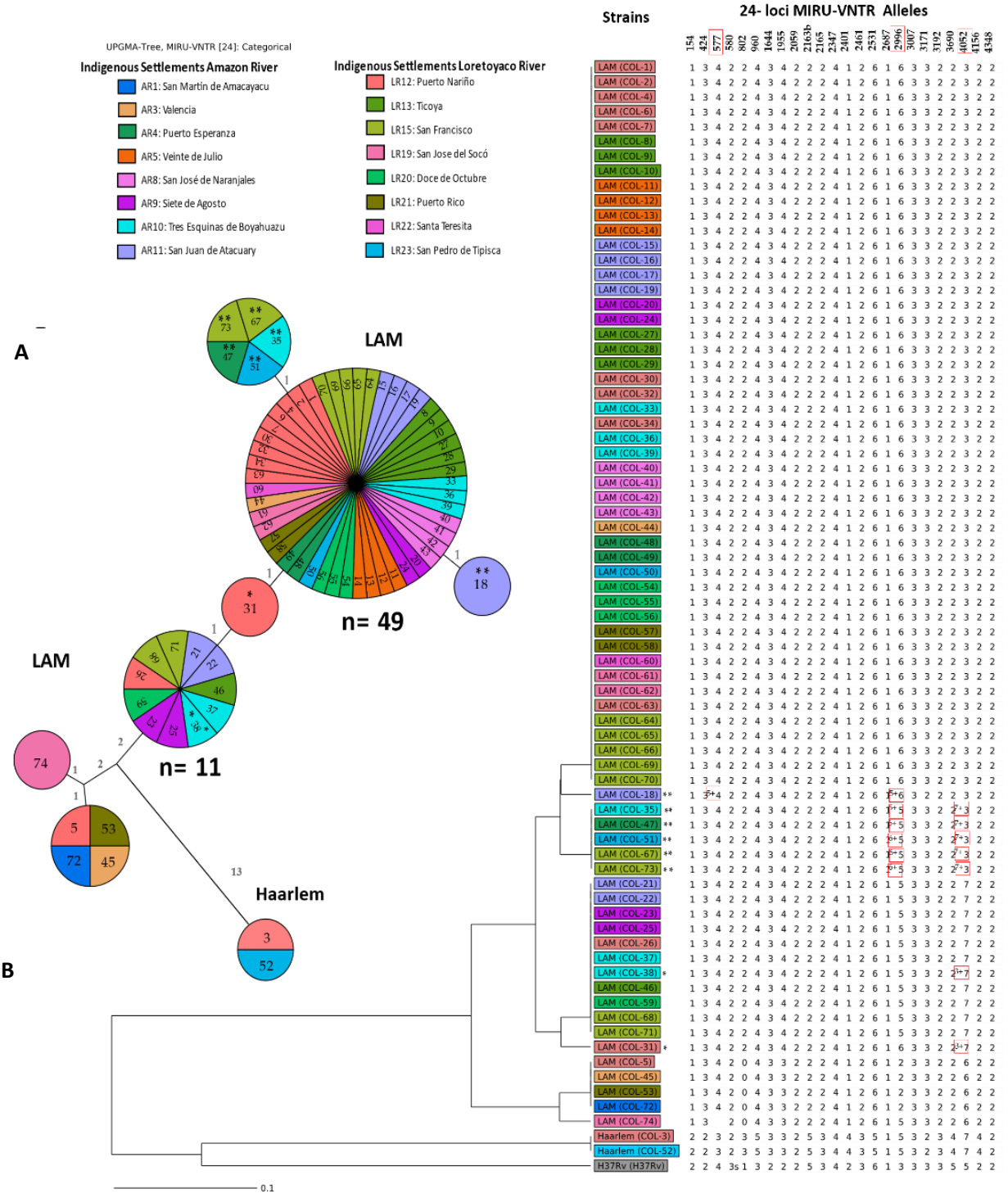
Geneious Prime	CSS	Geneious	https://www.geneious.com
GraphPad Prism V. 9	CSS	GraphPad	https://www.graphpad.com/
Image Lab Software	CSS	BIO-RAD	https://www.bio-rad.com/de-de/product/image-lab-software?ID=KRE6P5E8Z
Integrative genome viewer (IGV)	CSS	Broad Institute	https://software.broadinstitute.org/software/igv/
In house Bioinformatics scripts: -RNA sense counts -Mixed Infections Detector	CSS	RCB	Bioinformatics core
In house Flow Cytometry script: -Multiplex bead array	CSS	RCB	Flow Cytometry core
LEGENDplex V. 8.0	CSS	BioLegend	http://www.BioLegend.com/LEGENDplex/
MIRU-VNTR plus	OSS	miru-vntrplus.org[379]	https://www.miru-vntrplus.org
MTBseq	OSS	Kohl *et al.* [380]	https://github.com/ngs-fzb/MTBseq_source
PROSize 2.0	CSS	Advanced Analytical Technologies, Inc.	https://www.agilent.com/
R Studio	OSS	RStudio	https://www.rstudio.com/
Sequencesphere v5.9	CSS	Ridom	https://www.ridom.de/sequencesphere/
SMRT® Link software version 9	CSS	PacBio	http://www.pacb.com/support/software-downloads

11. SUPPLEMENTARY RESULTS

11.1 Supplementary Figures

Figure S1: Phylogenetic analysis based on 24-loci MIRU-VNTR data of 74 isolates recovered from the tuberculosis outbreak in 16 indigenous settlements in Puerto Nariño during March-October 2016.
A Maximum Parsimony Tree (MPT). Eight nodes are proportional in size to the number of isolates with identical Mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) genotypes. Branches are labeled with the number of MIRU-VNTR alleles that differ between nodes. **B** Dendrogram based on the Unweight Pair Group Method with Arithmetic mean algorithm (UPGMA) using www.miru-vntrplus.org. Strains are color-coded based on the indigenous settlement from which these were recovered. DNA from the H37Rv ATCC reference strain was used as a control. * Samples with one mixed allele: COL-31 (4052, 3+7) and COL -38 (4052, 3+7). ** Samples with two mixed alleles: COL-18 (2996, 5+6) (577, 4+5), COL-35 (2996, 5+6) (4052, 3+7), COL-47 (2996, 5+6) (4052, 3+7), COL-51 (2996, 5+6) (4052, 3+7), COL-67 (2996, 5+6) (4052, 3+7), and COL-73 (2996, 5+6) (4052, 3+7). Red squares indicate the mixed alleles in a given strain. LAM, Latin American Mediterranean; AR, Amazon River; LR, Loretoyaco River.

Figure S1(Continued)



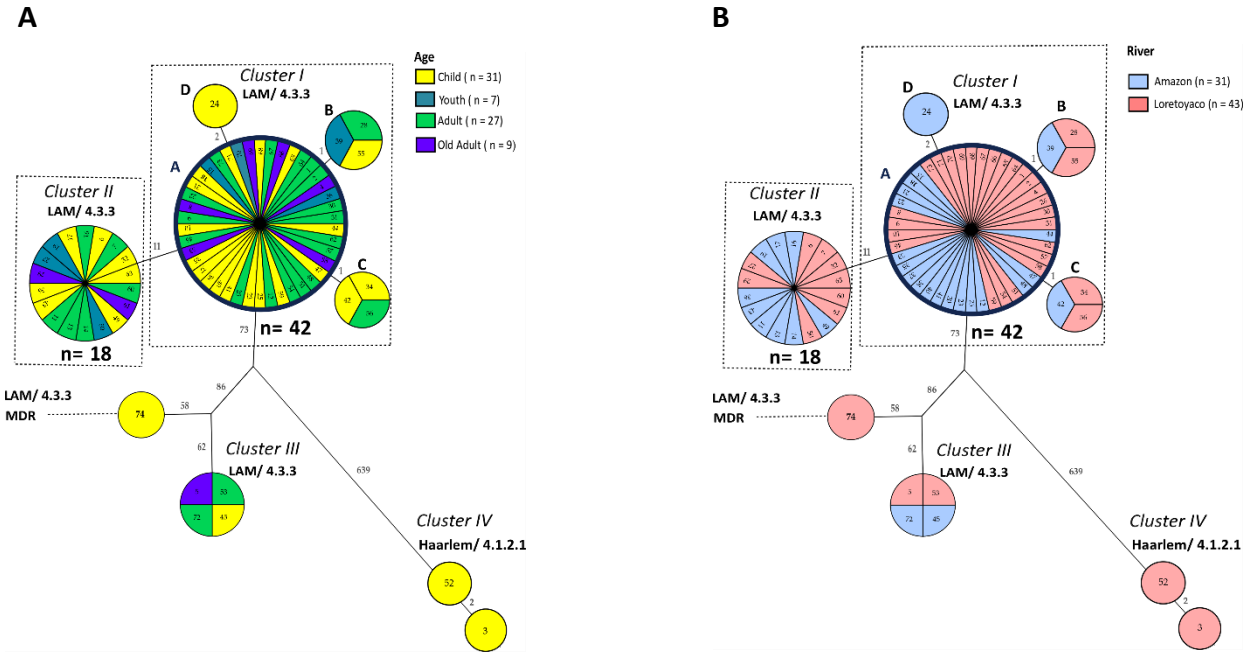


Figure S2: Genetic relationship of patient isolates based on age category and river location.
Standard whole genome sequencing analysis using H37Rv as the reference genome. Maximum Parsimony trees (MPT) based on 890 concatenated Single Nucleotide Polymorphisms (SNPs) of 74 *Mycobacterium tuberculosis* complex (*Mtbc*) isolates from 16 indigenous settlements located on the shore of the Amazon River (AR) or Loretoyaco River (LR). **A** Samples are color-coded based on age category (Child, < 15 years old; Youth, 16-26 years old; Adult, 27-59 years old; Old Adult, ≥ 60 years old). **B** Samples are color-coded based on the patient's river location. Outbreak clusters are named I, II, III, and IV. The size of each node is proportional to the number of isolates. The genetic distance is indicated on branches as the number of SNPs that differ between nodes. Outbreak clusters are named I, II, III, and IV. Clusters comprising the clonal events are shown with dashed boxes, and cluster I nodes were termed A, B, C, and D. LAM, Latin American Mediterranean; MDR, Multidrug-resistant.

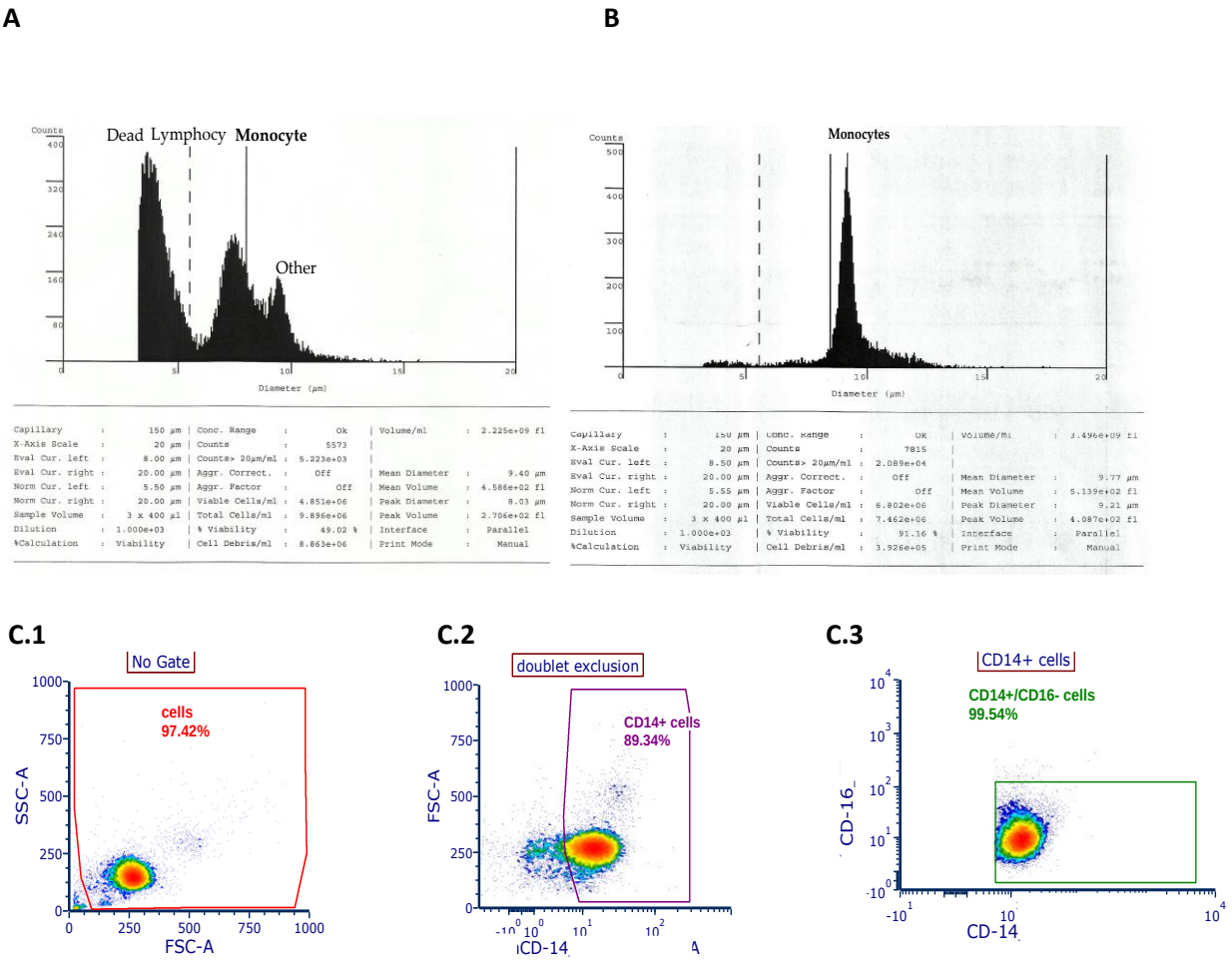


Figure S3: Cell parameters of the bovine cells isolated.

A Peripheral blood mononuclear cells (PBMCs) before magnetic isolation. **B** CD14 positive selected cells after magnetic isolation from a representative experiment. Cell counts (y-axis) are plotted against cell diameter (x-axis). Cell parameters results are depicted below the graphs and were calculated with electronic current exclusion technology with the CASY instrument (capillary of 150µm). The results of cell parameters are depicted below the graphs. Cell counts are given per mL. Monocyte cells have a diameter of 8-9.99 µm. Viable and dying/dead cells are detected based on their volume. Viable cells with a polarized membrane generate a high resistance signal, which correlates with the whole volume of the cell (>µm), while dying/ dead cells generate a much lower resistance due to a depolarized or completely disrupted membrane (< 5 µm). **C.1-C.3** Immunophenotyping Profiling of CD14+ Isolated Bovine Monocytes**.** Density plots are colored based on the number of events/cells. e.g., from red color, highest density, to blue color, lowest density. **C.1** Side Scatter (SSC) and FSC (Forward Scatter) patterns of the monocyte population. **C.2** Gating strategy doublet exclusion showing the percentage of cells expressing the monocyte differentiation antigen CD14+. **C.3** Monocytes CD14+ were further gated on the expression of CD16. Unstained and stained controls were included (not shown). CD, Cluster of Differentiation.

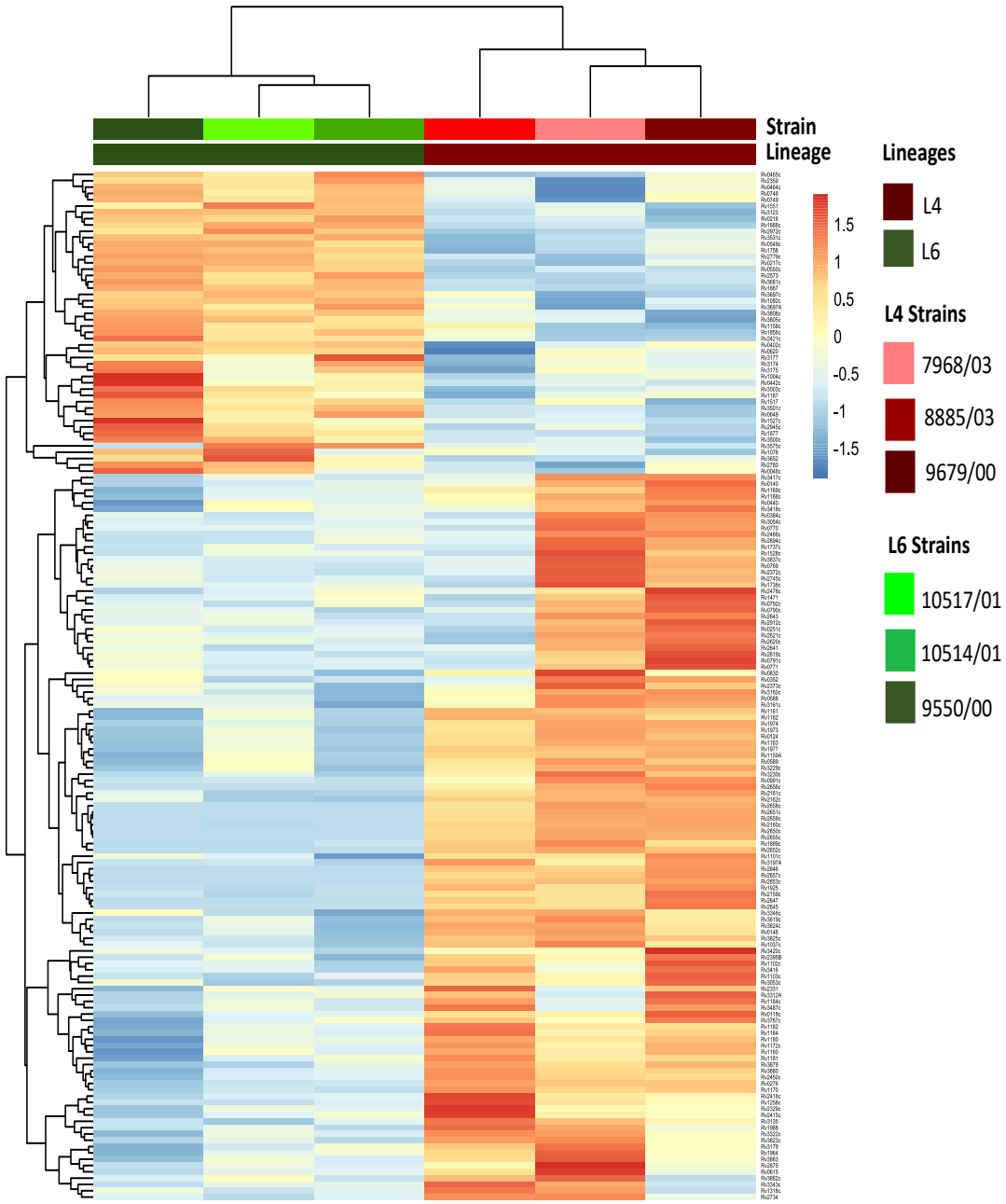


Figure S4: Differentially expressed genes in the L6 Vs. L4 strains pairwise comparison.
The heat map displays the expression profile of 163 differential expressed genes (DEGs) for each strain. Hierarchical clustering among strains is depicted in the figures. Normalized read counts underwent row scaling calculation with the function named scale of the pheatmap R package for visualization purposes and the comparison across rows. Heatmap is based on the feature counts with an adjusted p-value of 0.05, Log Fold Change (logFC) >1.5 for up-regulated (orange color spectrum) and < - for 1.5 down-regulated (blue color spectrum). The R function pheatmap was used to scale the RPKM (Average Reads Per Kilobase of transcript per Million reads mapped) value for each lineage strain into Z-scores and to draw the heatmap. L6, Lineage 6; L4, Lineage 4.

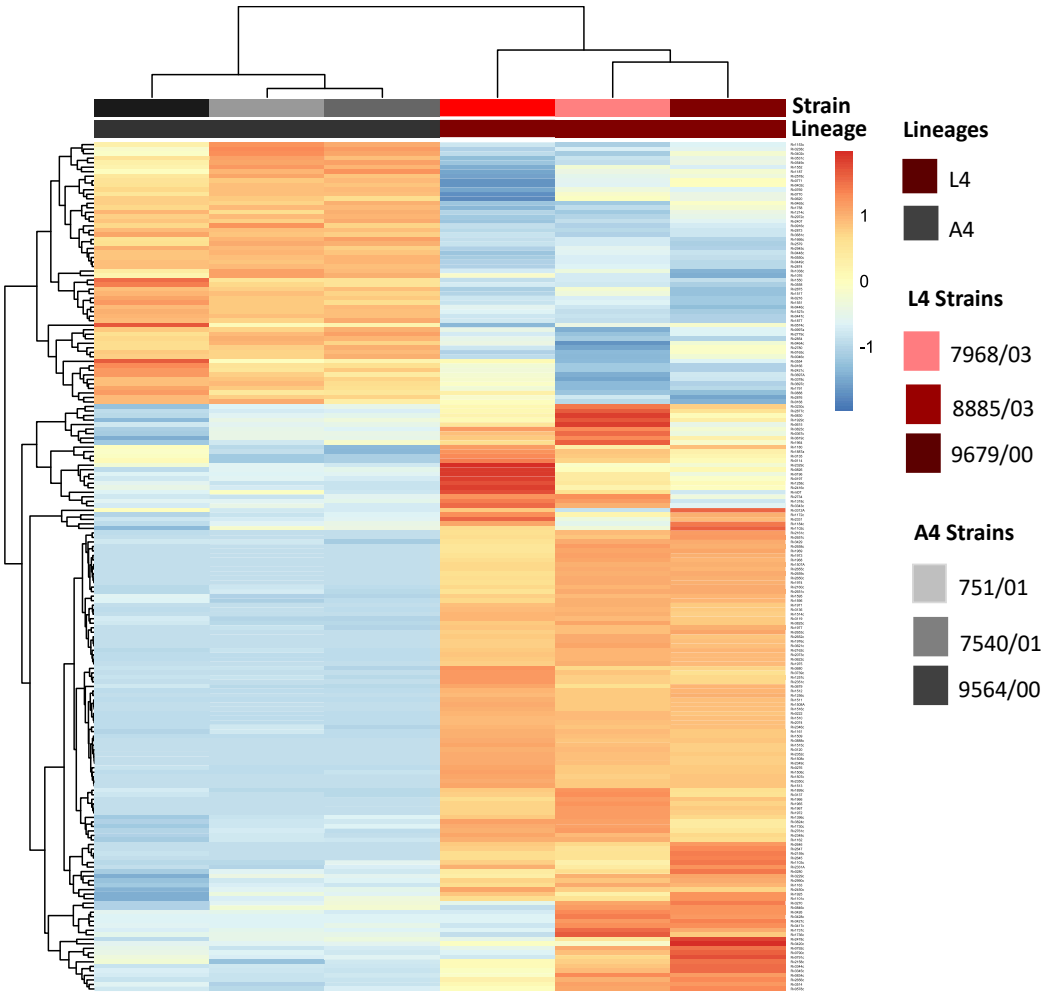


Figure S5: Differentially expressed genes in the L4 Vs. A4 strains pairwise comparison.

The heat map displays the expression profile of 188 differential expressed genes (DEGs) for each strain. Hierarchical clustering among strains is depicted in the figures. Normalized read counts underwent row scaling calculation with the function named scale of the pheatmap R package for visualization purposes and the comparison across rows. Heatmap is based on the feature counts with an adjusted p-value of 0.05, Log Fold Change (logFC) >1.5 for up-regulated (orange color spectrum) and < - 1.5 down-regulated (blue color spectrum). The R function pheatmap was used to scale the RPKM (Average Reads Per Kilobase of transcript per Million reads mapped) value for each lineage strain into Z-scores and to draw the heatmap. L4, Lineage 4; A4, Animal Clade 4.

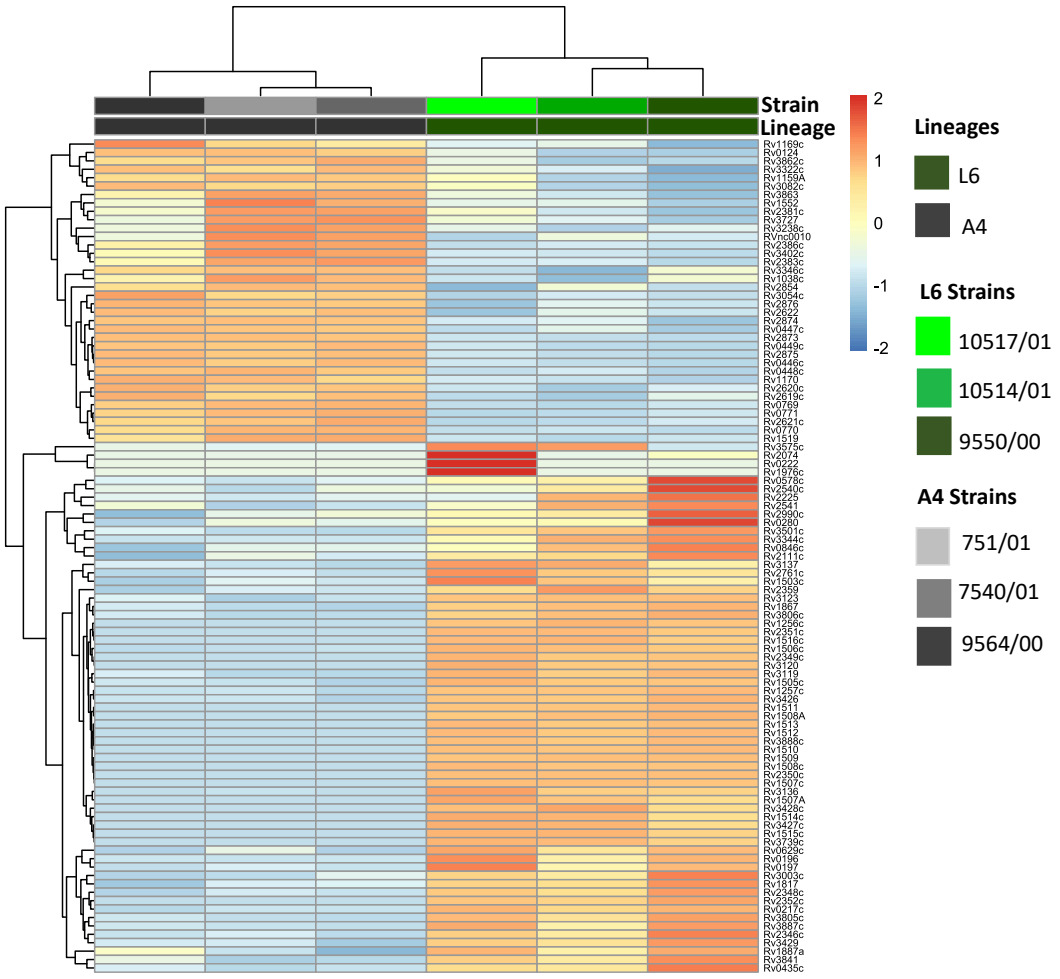


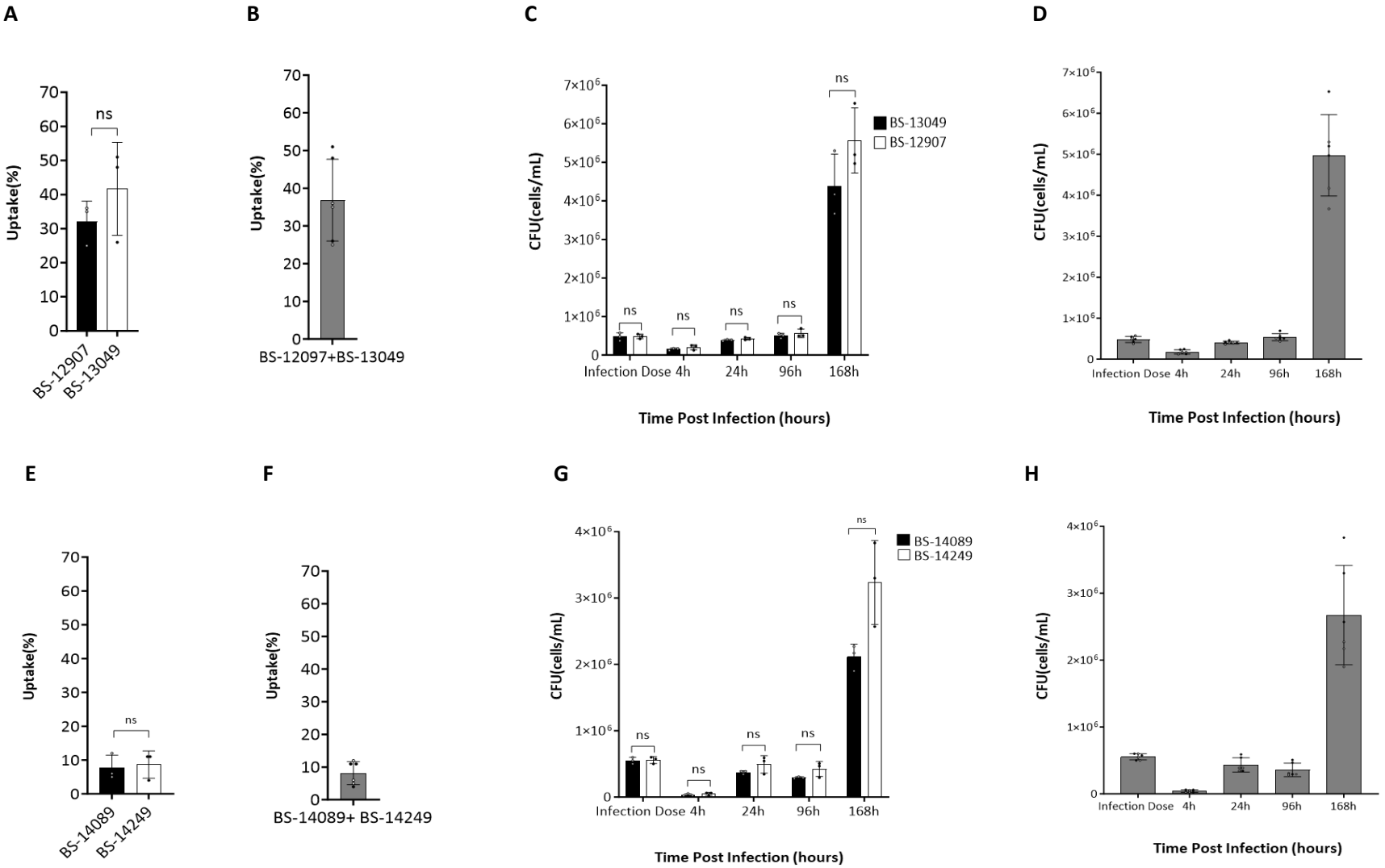
Figure S6: Differentially expressed genes in the L6 Vs. A4 strains pairwise comparison.

The heat map displays the expression profile of 99 differential expressed genes (DEGs) for each strain. Each column represents a strain, and each cell displays normalized gene expression values. Hierarchical clustering among strains is depicted in the figures. Normalized read counts underwent row scaling calculation with the function named scale of the pheatmap R package for visualization purposes and the comparison across rows. Heatmap is based on the feature counts with an adjusted p-value of e 0.05, Log Fold Change (logFC) >1.5 for up-regulated (orange color spectrum) and < - for 1.5 down-regulated (blue color spectrum). The R function pheatmap was used to scale the RPKM (Average Reads Per Kilobase of transcript per Million reads mapped) value for each lineage strain into Z-scores and to draw the heatmap. L6, Lineage 6; A4, Animal Clade 4.

Figure S7: Comparison of the uptake and intracellular growth of *Mycobacterium tuberculosis* H37Rv ATCC (9679/00) laboratory reference strain among two macrophage infection experiments comprising Germany 1 and Germany 3.

The top panel (**A**-**D**) displays results from experiments with human German donors BS-12907 (experiment 1) and BS-13049 (experiment 2) for Germany 1, including uptake per donor (**A**), uptake average among donors (**B**), intracellular growth upon infection per donor (**C**), and intracellular growth average among donors upon infection (**D**). The bottom panel (**E-H**) represents experiments with the human German donors BS-14089 (experiment 4) and BS-14249 (Experiment 5) for Germany 3, including uptake per donor (**E**), uptake average among donors (**F**), intracellular growth upon infection per donor (**G**), and intracellular growth average among donors upon infection (**H**). Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 ($0.5x10^6$ cells: $0.5x10^6$ *Mtb* bacilli), and the number of intracellular bacteria (CFU) was determined at 4 hpi, 24 hpi, 96 hpi, and 168 hpi. Three infection wells per strain (n=3) and per time point were assayed per donor, and three CFU measurements from each infection well were performed. Statistical results were determined by unpaired T-test for uptake experiments and Two-way ANOVA multiple comparison with Bonferroni correction for the growth experiments. No significant statistical differences among donors are depicted in the figure (ns >0.05). Error bars represent the mean with a standard deviation of nine CFU technical replicates for each time point and three CFU technical replicates for the infection dose. CFU, Colony-forming units; MOI, Multiplicity of Infection; hpi, hours post-infection.

Figure S7(Continued)



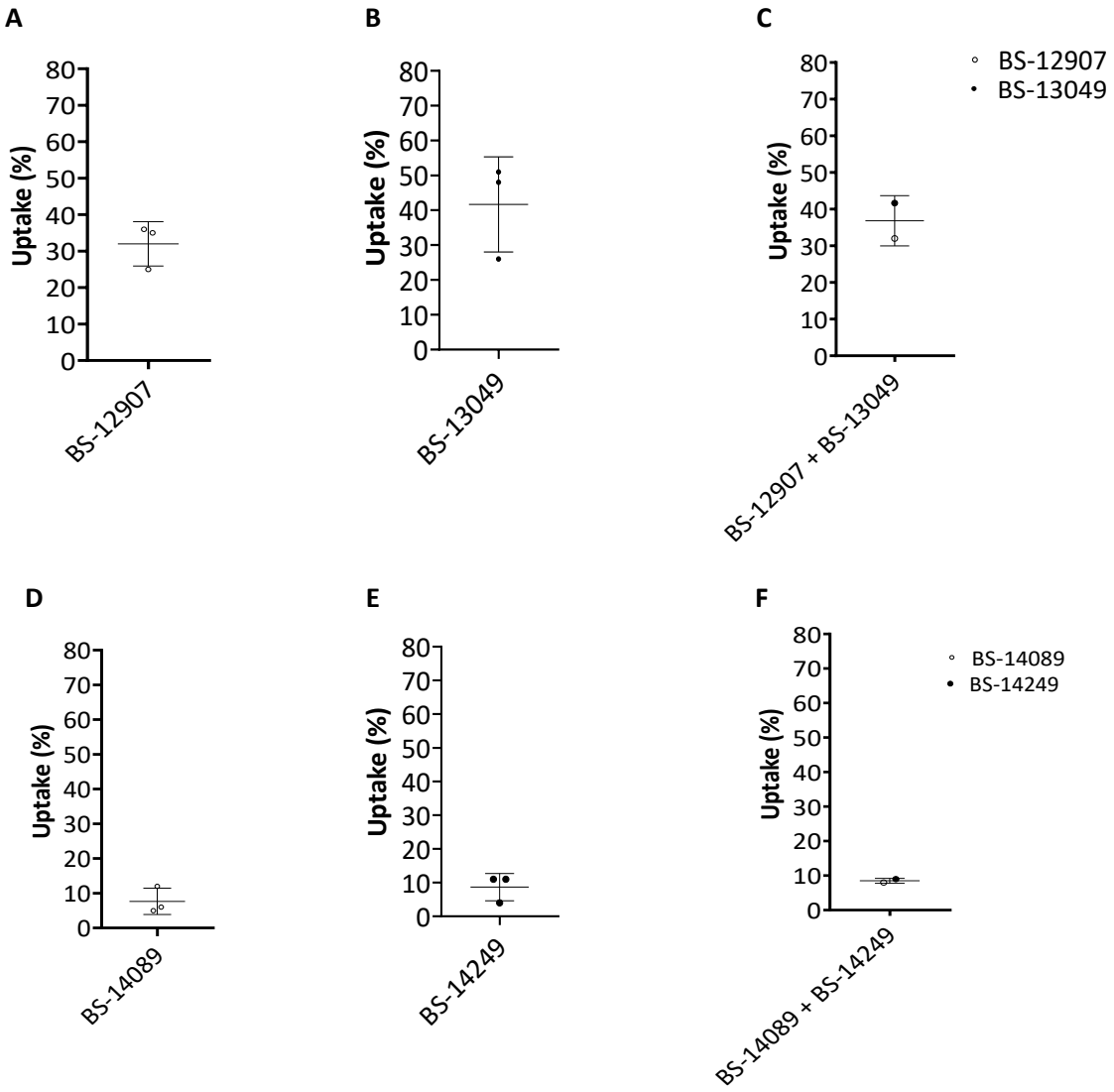


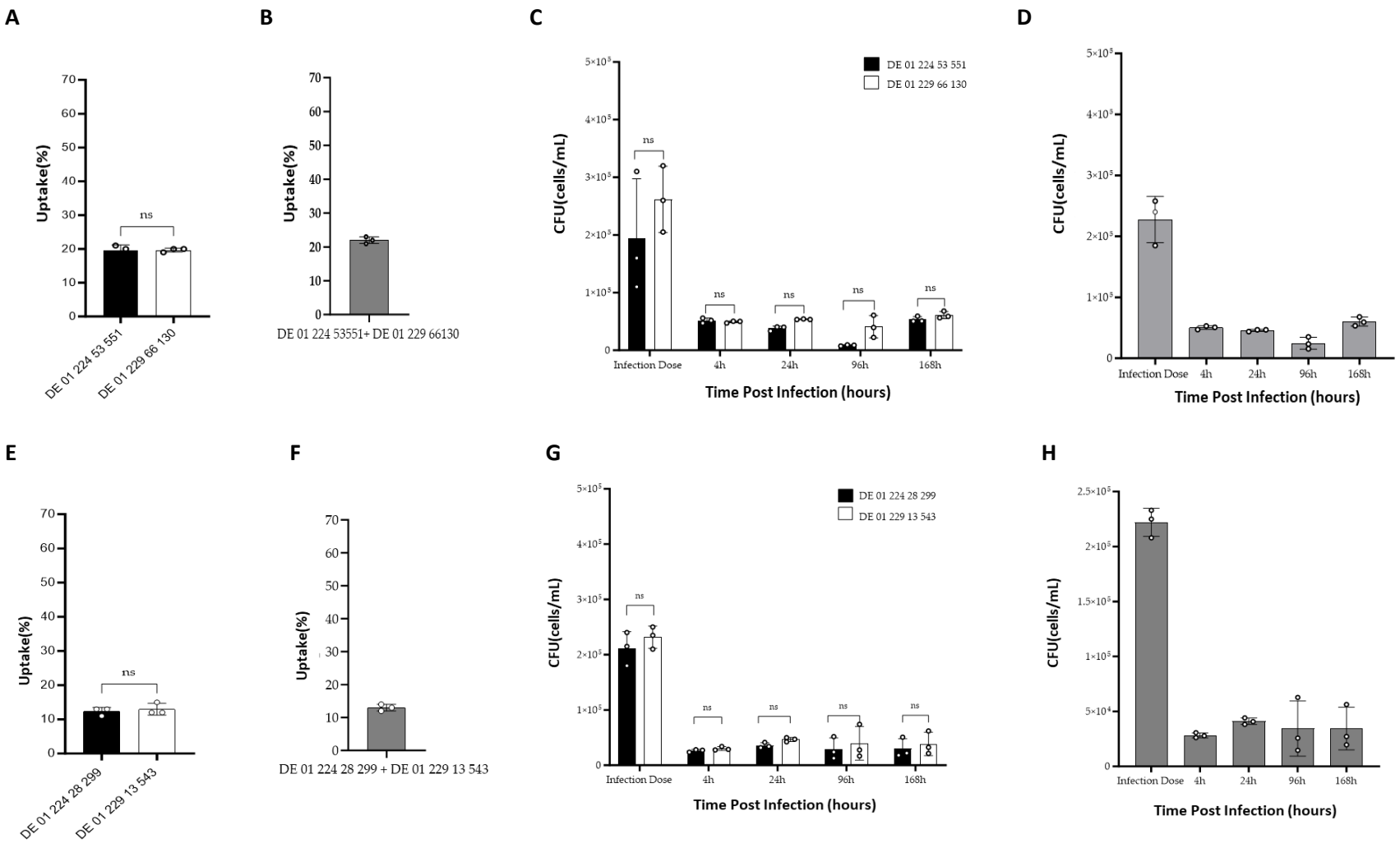
Figure S8: Host variation analysis of *Mycobacterium tuberculosis* H37Rv ATCC (9679/00) uptake among two macrophage infection experiments comprising Germany 1 and Germany 3.

The top panel (**A-C**) displays results from experiments with human German donors BS-12907 (experiment 1) and BS-13049 (experiment 2) for Germany 1, including Intra-host variation (**A-B**) and inter-host variation (**C**). The bottom panel (**D**-**F**) represents experiments with the human German donors BS-14089 (experiment 4) and BS-14249 (Experiment 5) for Germany 3, including Intra-host variation (**D**-**E**) and inter-host variation (**F**). Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 ($0.5x10^6$ cells: $0.5x10^6$ *Mtb* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately after uptake at 4 hours post-infection (hpi). A single dot represents the mean of three CFU measurements per infection well. Therefore, three dots represent three infections well replicates (n=3) per strain and per donor in figures **A**, **B**, **D**, and **E**. On the other hand, figures **C** and **F** represent the mean of three infection wells per experiment. Mean and standard deviation are depicted in the figures. CFU, Colony-forming units; MOI, Multiplicity of infection.

Figure S9: Comparison of the uptake and intracellular growth of *Mycobacterium tuberculosis* H37Rv ATCC (9679/00) laboratory reference strain among two macrophage infection experiments comprising Bovine 1 and Bovine 2.

The top panel (**A**-**D**) displays results from experiments with bovine donors DE 01 224 53 551 (experiment 1) and DE 01 229 66 130 (Experiment 2) for Bovine 1, including uptake per donor (**A**), uptake average among donors (**B**), intracellular growth upon infection per donor (**C**), and intracellular growth average among donors upon infection (**D**). The bottom panel (**E-H**) represents experiments with the bovine donors DE 01 224 28 299 (experiment 3) and DE 01 229 13 543 (Experiment 4) for Bovine 2, including uptake per donor (**E**), uptake average among donors (**F**), intracellular growth upon infection per donor (**G**), and intracellular growth average among donors upon infection (**H**). Bovine monocyte-derived macrophages (bMDMs) were infected at a MOI ~1:1 ($0.2x10^6$ cells: $0.2x10^6$ *Mtb* bacilli), and the number of intracellular bacteria (CFU) was determined at 4 hpi, 24 hpi, 96 hpi, and 168 hpi. Three infection wells per strain (n=3) and per time point were assayed per donor, and three CFU measurements from each infection well were performed. Statistical results were determined by unpaired T-test for uptake experiments and Two-way ANOVA multiple comparison with Bonferroni correction for the growth experiments. No significant statistical results among donors are depicted in the figure (ns >0.05). Error bars represent the mean with a standard deviation of nine CFU technical replicates for each time point and three CFU technical replicates for the infection dose. Ns, not significant; CFU, Colony-forming units; MOI, Multiplicity of infection; hpi, hours post-infection.

Figure S10 (Continued)



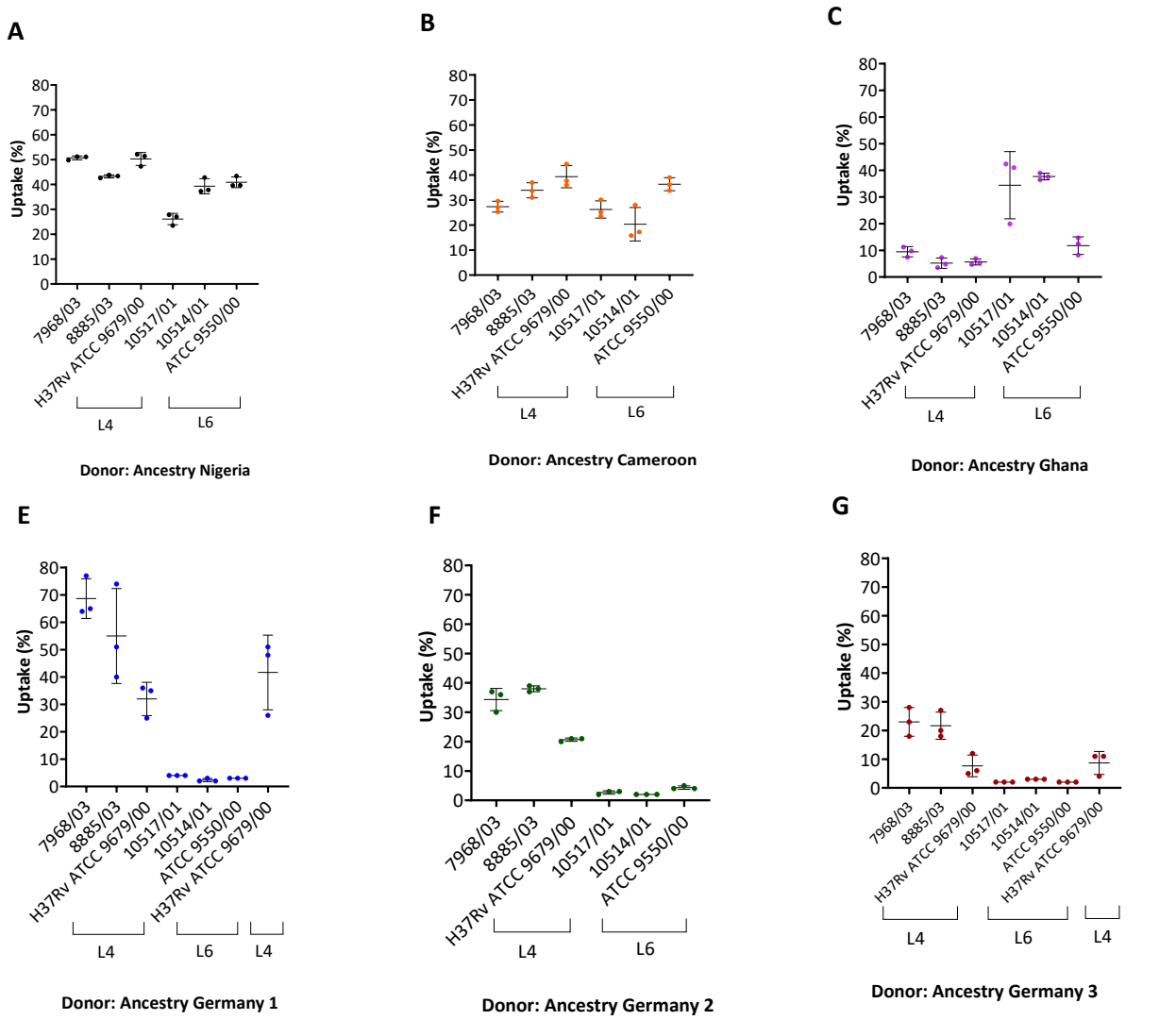


Figure S10: *Mycobacterium tuberculosis* complex strains uptake intra-host variation analysis.
Human donor self-reported ancestries: Nigeria (**A**), Cameroon (**B**), Ghana (**C**), Germany 1 (**D**), Germany 2 (**E**), and Germany 3 (**F**). Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately after uptake at 4 hour post-infection (hpi). Uptake percentage (y-axis) and *Mtbc* strain (x-axis). A single dot represents the mean of three CFU measurements per infection well. Therefore, three dots represent three infection well replicates per strain assayed (n=3). Error bars represent the mean with standard deviation. *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; CFU, Colony-forming units; MOI, Multiplicity of Infection.

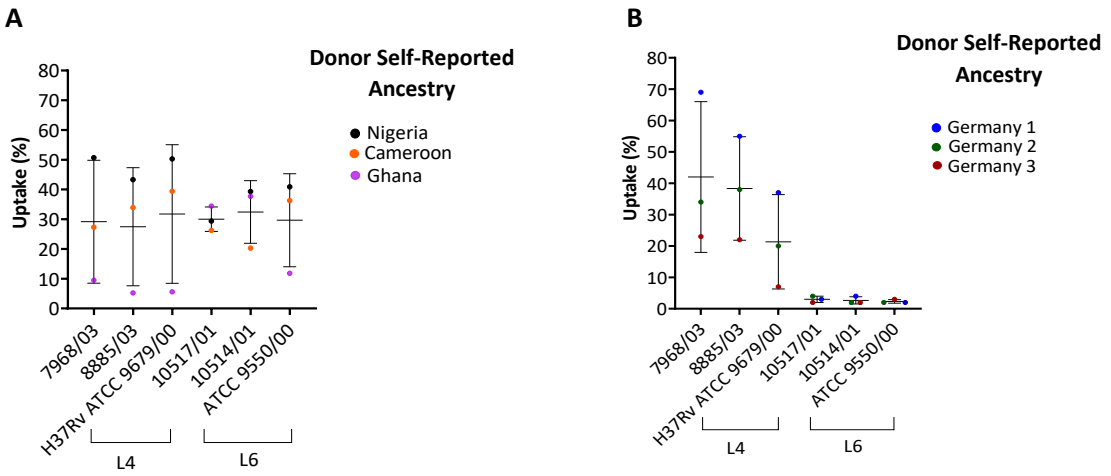


Figure S11: *Mycobacterium tuberculosis* complex strains uptake inter-host variation analysis.
Human donor self-reported ancestries: Nigeria, Cameroon, and Ghana(**A**). Germany 1, Germany 2, and Germany 3 (**B**). Human monocyte-derived macrophages (hMDMs) were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately at 4 hours post-infection (hpi). A single dot represents the mean of three infection macrophage wells (n=3) per strain per donor. Error bars represent the mean with standard deviation. *Mtbc*, *Mycobacterium tuberculosis* complex; L4, Lineage 4; L6, Lineage 6; CFU, Colony-forming units; MOI, Multiplicity of infection.

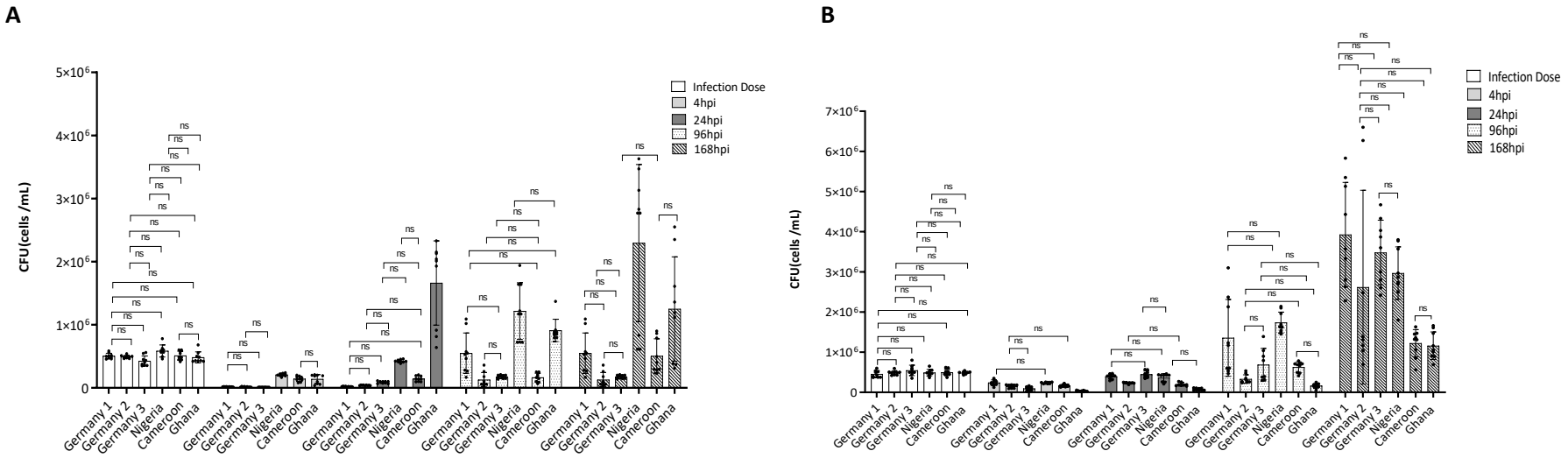


Figure S12: Comparative analysis of intracellular growth *Mycobacterium tuberculosis* complex strains in human monocyte-derived macrophages.

Intracellular growth of *Mycobacterium tuberculosis* complex (*Mtbc*) strains of the L6 (**A**) and L4 (**B**) for seven days in human monocyte-derived macrophages (hMDMs). Each lineage represents the average of three representative strains. hMDMs were infected at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the numbers of intracellular bacteria (CFU) were determined immediately at 4 hpi, 24 hpi, 96 hpi, and 168 hpi. Figures representing the strains and lineage show the mean and standard deviation of CFU values from three infection wells (n=3) per strain and nine infection wells (n=9) per lineage and donor, respectively (see Figures S7-S8). Three CFU measurements from each infection well were performed. CFU values of H37Rv ATCC (9679/00) strain are the average of two infection experiments performed for Germany 1 and Germany 3 (see Statistical results shown in the figures are based on One- way ANOVA multiple comparisons with Bonferroni post hoc test among donors. Mean, standard deviation and, only statistical results that are not significant are shown (ns >0.05). ns, not significant; L4, Lineage 4; L6, Lineage 6; CFU, Colony-forming units; MOI, Multiplicity of Infection; hpi, hours post-infection.

Figure S13: Inflammatory response of African human monocyte-derived macrophages to distinct *Mycobacterium tuberculosis* complex strains.

This assay was conducted on cell culture supernatants collected from human African monocyte-derived macrophages (hMDMs) infected with three representative strains of L4, L6, and A4 and non-infected hMDMs. The infection was carried out at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the supernatants were collected 24 and 96 hours post-infection (hpi). Three infection macrophage wells were tested per strain, time point, and donor. Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The production of the six detected cytokines-chemokines is depicted in the figure, namely IL-1ß (**A**), TNF-α (**B**), MCP-1 (**C**), IL-6 (**D**), IL-8 (**E**), and IL-18 (**F**). The protein concentration in pg/mL was plotted on the y-axis. The controls (PBS and non-infected [NI] at 24 hpi and 96 hpi) and experimental conditions (infected with L4, L6, and A4 strains at 24 hpi and 96 hpi) were represented in the x-axis. The stacked bars compared the mean production of these cytokines-chemokines at 24 hpi and 96 hpi within a lineage and across lineages. Three strains represent each lineage, and each strain comprises the averaged values of 3 donors. PBS, Macrophage Infection Media (MIM), and non-infected cells were used as controls. The MIM values were previously subtracted from non-infected and infected hMDMs. The mean, standard error of the mean, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. Data were obtained from three independent infection experiments. The statistical results shown in the figures are based on an unpaired T-test among both time points within the same lineage strain and One-way ANOVA with Bonferroni post hoc test correction-multiple comparisons across distinct lineages at 24 hpi and 96 hpi. *Mtbc*, *Mycobacterium tuberculosis* complex; ns, no significant; NI, Non-infected; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of Infection; hpi, hours post-infection.

Figure S13 (Continued)

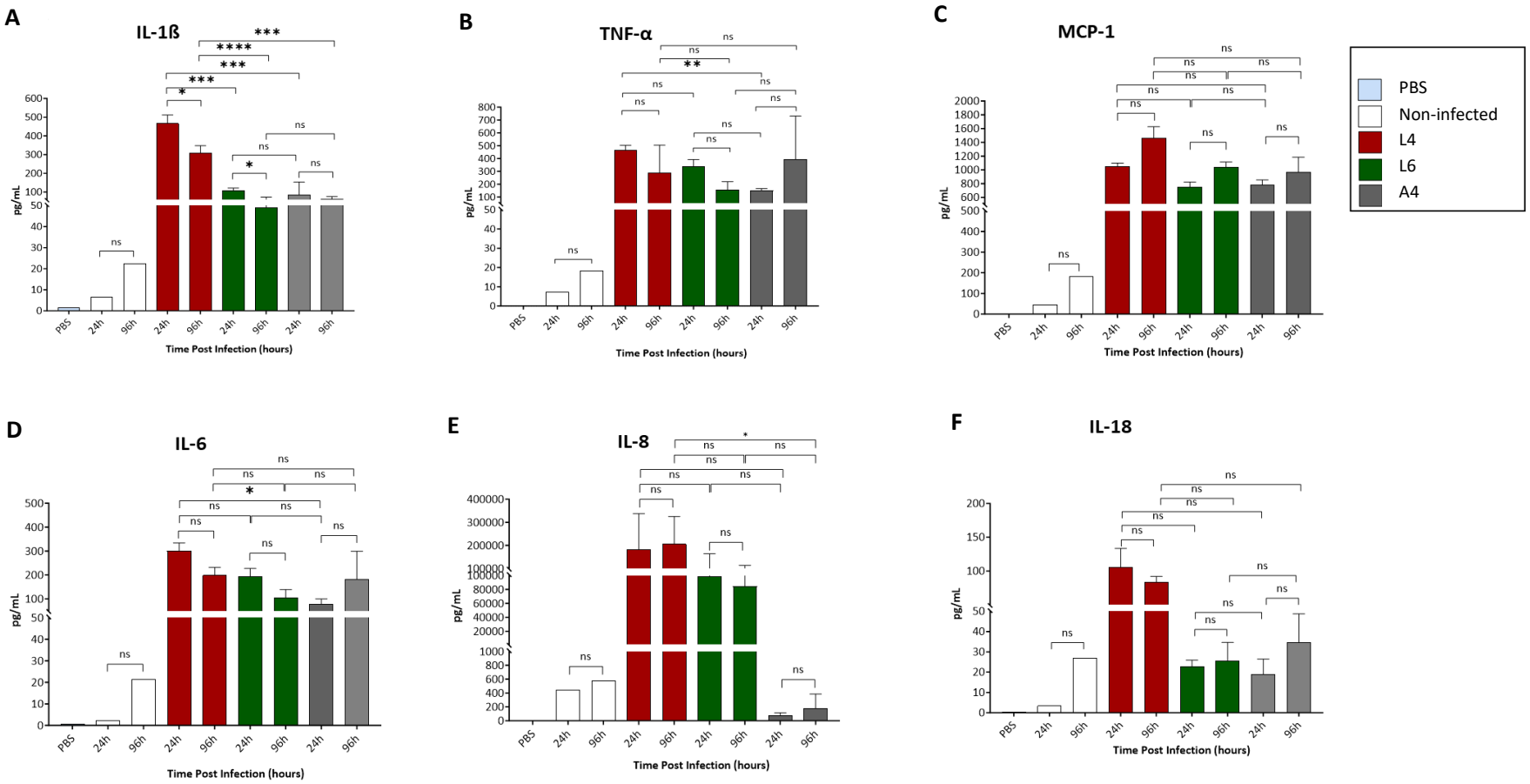
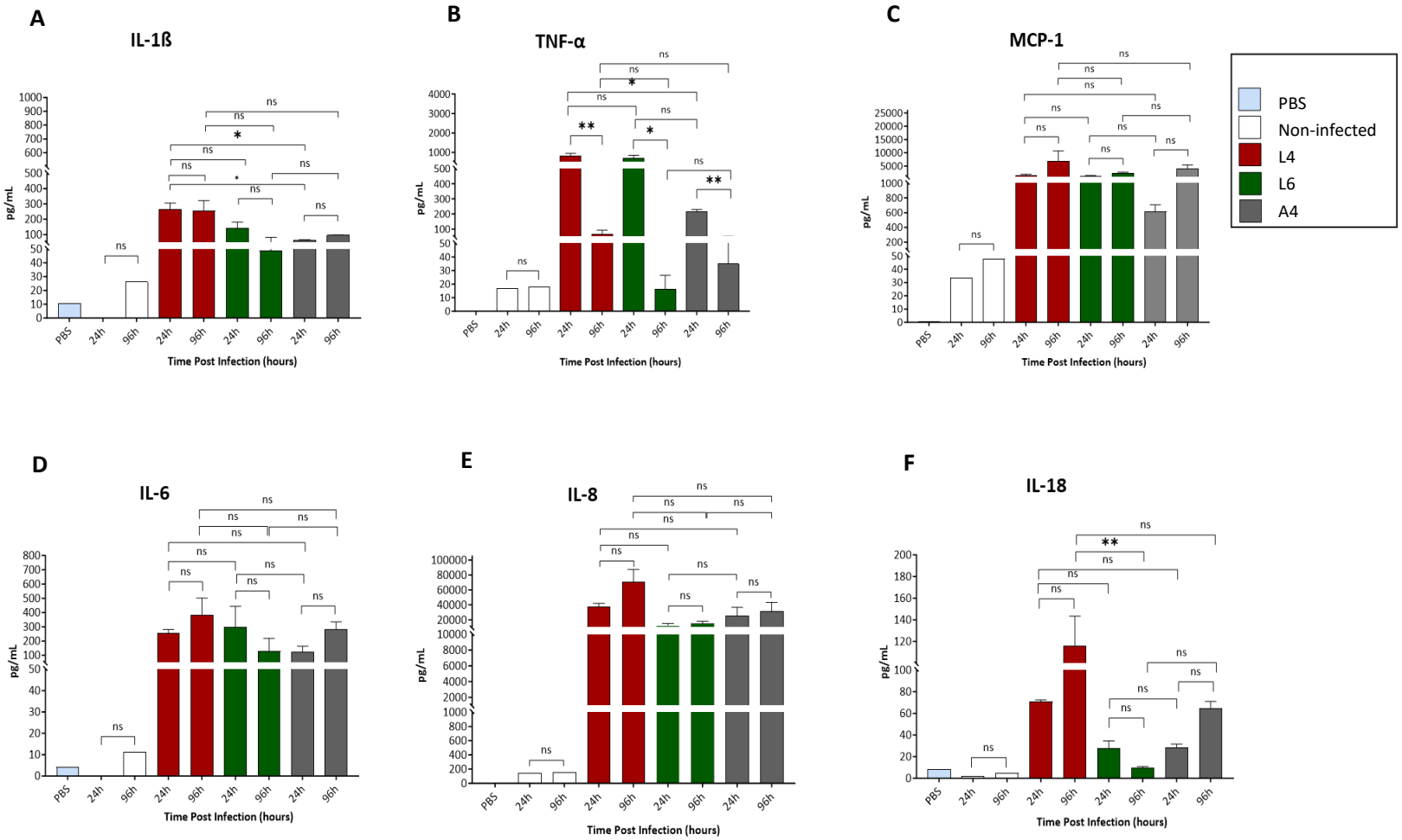


Figure S14: Inflammatory response of European human monocyte-derived macrophages to distinct *Mycobacterium tuberculosis* complex strains.

This assay was conducted on cell culture supernatants collected from human European monocyte-derived macrophages (hMDMs) that were infected with three representative strains of L4, L6, and A4 and also non-infected hMDMs. The infection was carried out at a MOI ~1:1 (0.5 x10^6 cells: 0.5 x10^6 *Mtbc* bacilli), and the supernatants were collected at 24 and 96 hours post-infection (hpi). Three infection macrophage wells were tested per strain, time point, and donor. Thirteen human inflammatory cytokines-chemokines were screened using LEGENDplex™. The production of six detected cytokines-chemokines is depicted in the figure, namely IL-1ß (**A**), TNF-α (**B**), MCP-1 (**C**), IL-6 (**D**), IL-8 (**E**) and IL-18 (**F**). The protein concentration in pg/mL was plotted on the y-axis. The controls (PBS and non-infected [NI] at 24 hpi and 96 hpi) and experimental conditions (infected with L4, L6, and A4 strains at 24 hpi and 96 hpi) were represented in the x-axis. The stacked bars compared the mean production of these cytokines-chemokines at 24 hpi and 96 hpi within a lineage and across lineages. Each lineage is represented by three strains, and each strain comprises the averaged values of three donors. PBS, Macrophage Infection Media (MIM), and non-infected cells were used as controls. The MIM values were previously subtracted from non-infected and infected hMDMs. The mean, standard error of the mean, and statistical results (ns >0.05; *P <0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001) are depicted in the figures. Data were obtained from three independent infection experiments. The statistical results shown in the figures are based on an unpaired T-test among both time points within the same lineage strain and One-way ANOVA with Bonferroni post hoc test correction-multiple comparisons across distinct lineages at 24 hpi and 96 hpi. *Mtbc, Mycobacterium tuberculosis* complex; ns, no significant; NI, Non-infected; L4, Lineage 4; L6, Lineage 6; A4, Animal Clade 4; CFU, Colony-forming units; MOI, Multiplicity of infection; hpi, hours post-infection.

Figure S14 (Continued)



11.2 Supplementary Tables

Table S1: Overview of genotyping, microbiological, and patient data of the outbreak isolates.

ID Sequencing	Coll L4 Sublineage	GDST	PDTS	AFB	Indigenous Settlement	Age Category	No. People Per Household	Health Insurance
COL-1	LAM/ 4.3.3	WT	S	Positive	Puerto Nariño	Adult	2-7	Uncovered
COL-2	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Adult	8-14	Subsidized
COL-3	Haarlem/4.1.2.1	rpsA (Thr210Ala), embB (Gly406Ala)	S	Negative	Puerto Nariño	Child	2-7	Subsidized
COL-4	LAM/ 4.3.3	WT	S	Positive	Puerto Nariño	Old Adult	2-7	Other
COL-5	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Old Adult	8-14	Subsidized
COL-6	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Child	8-14	Subsidized
COL-7	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Adult	8-14	Employer-based
COL-8	LAM/ 4.3.3	WT	S	Positive	Ticoya	Old Adult	2-7	Subsidized
COL-9	LAM/ 4.3.3	WT	S	Negative	Ticoya	Adult	8-14	Subsidized
COL-10	LAM/ 4.3.3	WT	S	Negative	Ticoya	Child	2-7	Subsidized
COL-11	LAM/ 4.3.3	WT	S	Positive	Veinte De Julio	Adult	2-7	Subsidized
COL-12	LAM/ 4.3.3	WT	S	Negative	Veinte De Julio	Adult	2-7	Subsidized
COL-13	LAM/ 4.3.3	WT	S	Negative	Veinte De Julio	Adult	8-14	Subsidized
COL-14	LAM/ 4.3.3	WT	S	Negative	Veinte De Julio	Adult	2-7	Subsidized
COL-15	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Youth	8-14	Uncovered
COL-16	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Adult	8-14	Subsidized

COL-17	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Child	2-7	Subsidized
COL-18	LAM/ 4.3.3	rpoB (Ser431Thr); rpsA (Thr210Ala); embB (Gly406Ala)	S	Negative	San Juan De Atacuary	Child	8-14	Subsidized
COL-19	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Youth	2-7	Subsidized
COL-20	LAM/ 4.3.3	WT	S	Negative	Siete De Agosto	Adult	8-14	Subsidized
COL-21	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Child	8-14	Subsidized
COL-22	LAM/ 4.3.3	WT	S	Negative	San Juan De Atacuary	Adult	8-14	Subsidized
COL-23	LAM/ 4.3.3	WT	S	Positive	Siete De Agosto	Child	8-14	Subsidized
COL-24	LAM/ 4.3.3	WT	S	Negative	Siete De Agosto	Child	8-14	Uncovered
COL-25	LAM/ 4.3.3	WT	S	Negative	Siete De Agosto	Child	8-14	Uncovered
COL-26	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Youth	2-7	Uncovered
COL-27	LAM/ 4.3.3	WT	S	Negative	Ticoya	Youth	2-7	Subsidized
COL-28	LAM/ 4.3.3	WT	S	Negative	Ticoya	Adult	2-7	Subsidized
COL-29	LAM/ 4.3.3	WT	S	Negative	Ticoya	Old Adult	2-7	Subsidized
COL-30	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Adult	2-7	Employer-based
COL-31	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Adult	8-14	Subsidized
COL-32	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Child	8-14	Subsidized
COL-33	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Old Adult	2-7	Subsidized

COL-34	LAM/ 4.3.3	WT	S	Negative	Puerto Nariño	Child	8-14	Subsidized
COL-35	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Child	2-7	Subsidized
COL-36	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Child	8-14	Subsidized
COL-37	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Child	2-7	Subsidized
COL-38	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Child	2-7	Subsidized
COL-39	LAM/ 4.3.3	WT	S	Negative	Tres Esquinas De Boyahuazu	Youth	2-7	Employer-based
COL-40	LAM/ 4.3.3	WT	S	Negative	San José De Naranjales	Child	8-14	Subsidized
COL-41	LAM/ 4.3.3	WT	S	Negative	San José De Naranjales	Child	8-14	Subsidized
COL-42	LAM/ 4.3.3	WT	S	Negative	San José De Naranjales	Child	2-7	Subsidized
COL-43	LAM/ 4.3.3	WT	S	Negative	San José De Naranjales	Child	2-7	Subsidized
COL-44	LAM/ 4.3.3	WT	S	Negative	Valencia	Child	8-14	Subsidized
COL-45	LAM/ 4.3.3	WT	S	Negative	Valencia	Child	8-14	Subsidized
COL-46	LAM/ 4.3.3	WT	S	Positive	Ticoya	Adult	2-7	Subsidized
COL-47	LAM/ 4.3.3	WT	S	Negative	Puerto Esperanza	Child	8-14	Subsidized
COL-48	LAM/ 4.3.3	WT	S	Positive	Puerto Esperanza	Adult	2-7	Subsidized
COL-49	LAM/ 4.3.3	WT	S	Negative	Puerto Esperanza	Child	2-7	Subsidized

COL-50	LAM/ 4.3.3	WT	S	Negative	San Pedro De Tipisca	Youth	2-7	Subsidized
COL-51	LAM/ 4.3.3	WT	S	Negative	San Pedro De Tipisca	Adult	2-7	Subsidized
COL-52	Haarlem/4.1 .2.1	WT	S	Negative	San Pedro De Tipisca	Child	8-14	Subsidized
COL-53	LAM/ 4.3.3	WT	S	Negative	Puerto Rico	Adult	2-7	Subsidized
COL-54	LAM/ 4.3.3	WT	S	Negative	Doce De Octubre	Adult	8-14	Subsidized
COL-55	LAM/ 4.3.3	WT	S	Negative	Doce De Octubre	Child	8-14	Subsidized
COL-56	LAM/ 4.3.3	WT	S	Negative	Doce De Octubre	Adult	2-7	Subsidized
COL-57	LAM/ 4.3.3	WT	S	Negative	Puerto Rico	Adult	2-7	Subsidized
COL-58	LAM/ 4.3.3	WT	S	Negative	Puerto Rico	Old Adult	2-7	Subsidized
COL-59	LAM/ 4.3.3	WT	S	Positive	Doce De Octubre	Child	2-7	Subsidized
COL-60	LAM/ 4.3.3	WT	S	Positive	Santa Teresita	Adult	8-14	Subsidized
COL-61	LAM/ 4.3.3	WT	S	Positive	San Juan Del Socó	Old Adult	2-7	Subsidized
COL-62	LAM/ 4.3.3	WT	S	Negative	San Juan Del Socó	Adult	2-7	Subsidized
COL-63	LAM/ 4.3.3	WT	S	Positive	Puerto Nariño	Child	2-7	Subsidized
COL-64	LAM/ 4.3.3	WT	S	Negative	San Francisco	Adult	2-7	Subsidized
COL-65	LAM/ 4.3.3	WT	S	Positive	San Francisco	Child	8-14	Subsidized
COL-66	LAM/ 4.3.3	WT	S	Positive	San Francisco	Old Adult	2-7	Subsidized
COL-67	LAM/ 4.3.3	WT	S	Negative	San Francisco	Adult	2-7	Subsidized
COL-68	LAM/ 4.3.3	WT	S	Negative	San Francisco	Child	8-14	Subsidized

COL-69	LAM/ 4.3.3	WT	S	Positive	San Francisco	Old Adult	8-14	Subsidized
COL-70	LAM/ 4.3.3	WT	S	Negative	San Francisco	Youth	8-14	Subsidized
COL-71	LAM/ 4.3.3	WT	S	Positive	San Francisco	Child	8-14	Subsidized
COL-72	LAM/ 4.3.3	WT	S	Positive	Amacayacu	Adult	8-14	Subsidized
COL-73	LAM/ 4.3.3	WT	S	Positive	San Francisco	Adult	2-7	Uncovere d
COL-74	LAM/ 4.3.3	rpoB (Ser431Thr); katG (Ser315Thr; pncA (His82Asp); embB (Met306lle) and rpsL (Lys88Thr)	R	Positive	San Juan Del Socó	Child	2-7	Subsidized

Clusters were defined based on a pairwise SNP distance of 5 ≤ SNP.

LAM, Latin-American Mediterranean; COL, Colombia; GDS, Genotypic Drug Susceptibility Testing; PDS, Phenotypic Drug Susceptibility Testing WT, Wild Type; S, Susceptible; R, Resistant; AFB, Acid-Fast Bacilli

Table S2: Genotyping of outbreak strains by MIRU VNTR 24 alleles.

580	2996	802	1955	960	2163b	3192	1644	424	577	2165	4052	154	2531	4348	2401	2059	2687	3007
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	3	3	5	5	3	3	2	3	3	7	2	5	2	4	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	2	0	3	4	2	2	3	3	4	2	6	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3

2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5+6	2	4	4	2	2	3	3	5+4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3+7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6+5	2	4	4	2	2	3	3	4	2	7+3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3

2	5	2	4	4	2	2	3	3	4	2	3+7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	2	0	3	4	2	2	3	3	4	2	6	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	6+5	2	4	4	2	2	3	3	4	2	7+3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6+5	2	4	4	2	2	3	3	4	2	7+3	1	6	2	1	2	1	3
2	5	3	3	5	5	3	3	2	3	3	7	2	5	2	4	2	1	3
2	2	0	3	4	2	2	3	3	4	2	6	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3

2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6+5	2	4	4	2	2	3	3	4	2	7+3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	6	2	4	4	2	2	3	3	4	2	3	1	6	2	1	2	1	3
2	5	2	4	4	2	2	3	3	4	2	7	1	6	2	1	2	1	3
2	2	0	3	4	2	2	3	3	4	2	6	1	6	2	1	2	1	3
2	6+5	2	4	4	2	2	3	3	4	2	7+3	1	6	2	1	2	1	3
2	5	0	3	4	2	2	3	3		2	6	1	6	2	1	2	1	3

MIRU-VNTR, Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat

Table S2 (Continued)

4156	2347	2461	3171	3690
2	4	2	3	2
2	4	2	3	2
4	4	3	2	4
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	2	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2

2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
4	4	3	2	4

2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2
2	4	2	3	2

Table S3: Comparison of COL-2 outbreak genome and H37Rv reference genome.

Locus ID	COL-2	H37Rv	Gene ID	Gene name	Annotation
NP_216037.1	LOF_DEL		Rv1520	*Rv1520*	Sugar transferase
NP_216036.3	LOF_DEL		Rv1521	*fadD25*	Fatty-acid-AMP synthetase (fatty-acid-AMP synthase)
NP_214891.1	LOF_DEL		Rv0377	*Rv0377*	HTH-type transcriptional regulator
NP_216867.1	new CDS	*INS - mobile element*	Rv2351c	*plcA*	Membrane-associated phospholipase C 1 PlcA (MTP40 antigen)
NP_215071.1	new CDS		Rv0557	*mgtA*	Mannosyltransferase MgtA
MF_01206	new CDS		NA	*msrP*	Rotein-methionine-sulfoxide reductase catalytic subunit MsrP
NP_215190.1	new CDS		Rv0676c	*mmpL5*	Probable conserved transmembrane transport protein MmpL5
NP_216411.1	LOF		Rv1895	*Rv1895*	Zinc-binding alcohol dehydrogenase
NP_216793.1	new CDS	*INS - mobile element*	Rv2277c	*Rv2277c*	Glycerolphosphodiesterase
NP_214568.1	new CDS	*INS - mobile element*	Rv0054	*ssb*	Single-strand DNA-binding protein
NP_217152.1	LOF_INS		Rv2636	*Rv2636*	O-phosphotransferase
NP_217629.1	LOF_INS		Rv3113	*Rv3113*	Phosphatase
YP_177926.1	new CDS	*INS - mobile element*	Rv3110	*moaB1*	Pterin-4-alpha-carbinolamine dehydratase
YP_177925.1	new CDS		Rv3109	*moaA1*	Cyclic pyranopterin monophosphate synthase
NP_215783.1	new CDS		Rv1267c	*embR*	Transcriptional regulator EmbR
NP_217990.1	new CDS	*INS - mobile element*	Rv3473c	*bpoA*	Peroxidase BpoA

COL, Colombia; ID, Identification; DEL, Deletion; CDS, Coding Sequence; INS, Insertion; LOF, putative Loss of Function

Table S4: Frequencies of outbreak-specific SNPs and putative mixed infections.

SNP#	SNP1_ Rv1194c	SNP2_ PPE65	SNP3_ PPE60	SNP4_ PPE60	SNP5_ Rv1945	SNP6_ glyA2	SNP 7_ iniC	SNP 8_ pitA	SNP 9_ atsD	SNP10_ Rv0862c
Genome Pos.	**1337784.0**	**4061132.0**	**3895400.0**	**3895403.0**	**2196969.0**	**78383.0**	**413981.0**	**637055.0**	**758066.0**	**960798.0**
Cluster No_node_ID	**Cluster I, Node G SNPs**						**Cluster II specific SNPs**			
Cluster I_Node A_COL-1	0.0	0.0	0.0	0.0	0.0	1.0	0.9	0.0	2.3	2.1
Cluster I_Node A_COL-2	0.0	0.0	0.0	0.0	0.0	3.6	4.0	2.0	4.1	0.0
Cluster I_Node A_COL-4	0.0	0.0	0.0	0.0	0.0	15.3	14.2	14.3	12.1	16.2
Cluster I_Node A_COL-8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
Cluster I_Node A_COL-12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-18	0.0	0.0	13.0	13.1	11.5	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0
Cluster I_Node	11.1	14.0	8.9	9.8	7.5	3.9	2.0	3.7	3.2	7.3

A_COL-30										
Cluster I_Node A_COL-31	23.5	20.2	20.7	19.4	12.8	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-33	3.5	1.5	0.0	0.0	1.9	2.4	1.2	4.1	1.4	3.2
Cluster I_Node A_COL-40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-41	8.3	8.4	5.7	5.5	1.8	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-44	0.0	0.0	0.0	0.0	0.0	7.4	10.0	4.3	4.3	6.8
Cluster I_Node A_COL-48	6.6	2.0	0.0	0.0	5.4	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-54	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-57	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-58	18.4	17.5	11.9	10.1	13.7	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-62	0.0	0.0	16.7	15.6	8.1	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-64	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.8	0.0	6.9
Cluster I_Node A_COL-65	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node A_COL-66	0.0	0.0	0.0	0.0	0.0	12.1	7.5	11.6	8.4	3.2
Cluster I_Node	43.4	DEL	33.3	35.7	25.0	0.0	0.0	0.0	0.0	0.0

A_COL-67										
Cluster I_Node A_COL-69	12.4	18.8	7.3	11.7	4.1	6.3	7.1	4.1	0.0	2.0
Cluster I_Node A_COL-70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.7
Cluster I_Node B_COL-28	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0
Cluster I_Node B_COL-39	7.4	2.9	4.4	4.5	1.3	0.0	0.0	0.0	0.0	0.0
Cluster I_Node B_COL-55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node C_COL-34	8.2	13.1	3.9	3.8	6.1	0.0	0.0	0.0	0.0	0.0
Cluster I_Node C_COL-42	21.7	16.0	13.0	15.8	17.4	0.0	0.0	0.0	0.0	0.0
Cluster I_Node C_COL-56	16.7	13.6	11.4	10.9	6.6	0.0	0.0	0.0	0.0	0.0
Cluster I_Node D_COL-24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node E_COL-51	55.7	56.9	29.3	27.7	24.6	0.0	0.0	2.2	0.0	0.0
Cluster I_Node F_COL-35	62.6	59.5	55.2	53.4	36.5	0.0	0.0	0.0	0.0	0.0
Cluster I_Node F_COL-47	57.4	53.9	59.0	54.8	35.1	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-21	99.1	100.0	100.0	100.0	100.0	0.0	0.0	2.2	0.0	0.0
Cluster I_Node	100.0	100.0	97.7	100.0	100.0	0.0	0.0	0.0	0.0	0.0

G_COL-22										
Cluster I_Node G_COL-23	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-25	100.0	100.0	97.9	100.0	98.1	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-26	98.6	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-37	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-38	94.7	89.6	87.2	87.8	86.7	1.2	1.0	0.0	2.0	1.5
Cluster I_Node G_COL-46	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-59	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-68	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-71	98.2	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
Cluster I_Node G_COL-73	81.7	80.2	76.2	75.9	62.7	0.0	0.0	0.0	0.0	0.0
Cluster II_ Node_COL-6	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-7	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	98.8
Cluster II_ Node_COL-11	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0

Node_COL-13										
Cluster II_ Node_COL-14	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-16	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-17	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-19	0.0	0.0	0.0	0.0	0.0	100.0	100.0	98.8	100.0	100.0
Cluster II_ Node_COL-27	0.0	0.0	0.0	0.0	0.0	70.6	72.4	71.9	63.2	72.9
Cluster II_ Node_COL-29	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	99.5	100.0
Cluster II_ Node_COL-32	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-36	0.0	2.0	0.0	0.0	0.9	94.9	96.4	98.8	97.7	99.5
Cluster II_ Node_COL-43	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-49	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-50	0.0	DEL	0.0	0.0	0.0	98.9	100.0	100.0	100.0	100.0
Cluster II_ Node_COL-60	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0
Cluster II_ Node _COL-61	0.0	0.0	0.0	0.0	0.0	86.6	82.1	75.4	76.4	78.9
Cluster II_	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	99.0

Node_COL-63

COL, Colombia; SNPs, Single Nucleotide Polymorphisms; PE, Proline Glutamic acid; PPE, Proline-Proline-Glutamic acid

Table S4 (Continued)

SNP11_ sucC	SNP12_ trcR	SNP13_ narG	SNP14_ cyp130	SNP15_ Rv1985c	SNP16_ Rv3093c
1062381.0	**1158574.0**	**1290447.0**	**1403781.0**	**2229203.0**	**3461997.0**
Cluster II specific SNPs					
4.0	1.8	0.9	0.0	1.5	2.8
3.6	2.7	3.4	4.2	0.0	4.0
11.7	14.2	13.9	14.5	15.6	9.9
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
5.0	0.0	0.0	0.0	0.0	3.1
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	4.3	2.1	0.0	0.0
0.0	8.0	3.1	6.3	7.1	4.8
0.0	0.0	0.0	0.0	0.0	0.0
5.8	1.5	3.1	3.1	2.9	1.3
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
5.3	4.6	7.7	8.0	9.2	3.7
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	98.5	DEL	4.3	2.4	0.0
0.0	0.0	0.0	0.0	0.0	0.0
8.1	7.7	9.4	6.7	7.9	9.6
0.0	0.0	0.0	0.0	0.0	0.0
0.0	2.9	4.3	5.4	3.2	3.8
0.0	0.0	2.7	0.0	3.4	6.8
0.0	0.0	DEL	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0

0.0	0.0	0.0	2.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	1.5	1.7	1.7	1.6	2.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0
100.0	100.0	100.0	100.0	100.0	100.0
100.0	97.8	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	98.9
100.0	97.8	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	DEL	100.0
100.0	100.0	100.0	100.0	100.0	100.0
100.0	98.6	100.0	100.0	100.0	100.0
69.0	76.9	71.1	63.0	64.9	71.4
100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0
97.0	98.6	99.6	99.4	99.5	98.1
100.0	100.0	98.8	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0
98.5	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0
76.1	83.0	68.8	72.3	79.7	81.3
100.0	98.5	100.0	100.0	100.0	100.0

Table S5: Top 50 differential expressed genes L4 Vs. L6 strains.

Locus	Gene Name	Functional Category	Product
Rv1092c	coaA	intermediary metabolism and respiration	Probable pantothenate kinase CoaA (pantothenic acid kinase)
Rv3531c	Rv3531c	conserved hypotheticals	Hypothetical protein
Rv0465c	Rv0465c	regulatory proteins	Probable transcriptional regulatory protein
Rv3501c	yrbE4A	virulence, detoxification, adaptation	Conserved integral membrane protein YrbE4A. Possible ABC transporter.
Rv3806c	ubiA	cell wall and cell processes	Decaprenylphosphoryl-5-phosphoribose (DPPR) synthase (decaprenyl-phosphate 5-phosphoribosyltransferase)
Rv3123	Rv3123	conserved hypotheticals	Hypothetical protein
Rv0216	Rv0216	intermediary metabolism and respiration	Double hotdog hydratase
Rv1666c	cyp139	intermediary metabolism and respiration	Probable cytochrome P450 139 Cyp139
Rv1877	Rv1877	cell wall and cell processes	Probable conserved integral membrane protein
Rv0550c	vapB3	virulence, detoxification, adaptation	Possible antitoxin VapB3
Rv2573	Rv2573	conserved hypotheticals	Conserved hypothetical protein
Rv3681c	whiB4	regulatory proteins	Probable transcriptional regulatory protein WhiB-like WhiB4
Rv1867	Rv1867	lipid metabolism	Conserved protein
Rv2779c	Rv2779c	regulatory proteins	Possible transcriptional regulatory protein (probably Lrp/AsnC-family)
Rv0217c	lipW	intermediary metabolism and respiration	Possible esterase LipW
Rv0549c	vapC3	virulence, detoxification, adaptation	Possible toxin VapC3
Rv3054c	Rv3054c	conserved hypotheticals	Conserved hypothetical protein
Rv0991c	Rv0991c	conserved hypotheticals	Conserved serine rich protein
Rv3824c	papA1	lipid metabolism	Conserved polyketide synthase associated protein PapA1
Rv3322c	Rv3322c	intermediary metabolism and	Possible methyltransferase

		respiration	
Rv2416c	eis	virulence, detoxification, adaptation	Enhanced intracellular survival protein Eis, GCN5-related N-acetyltransferase
Rv1258c	Rv1258c	cell wall and cell processes	Probable conserved integral membrane transport protein
Rv1182	papA3	lipid metabolism	Probable conserved polyketide synthase associated protein PapA3
Rv3679	Rv3679	cell wall and cell processes	Probable anion transporter ATPase
Rv3680	Rv3680	cell wall and cell processes	Probable anion transporter ATPase
Rv0276	Rv0276	conserved hypotheticals	Conserved hypothetical protein
Rv1170	mshB	intermediary metabolism and respiration	N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase MshB (GlcNAc-Ins deacetylase)
Rv1169c	lipX	PE/PPE	PE family protein. Possible lipase LipX.
Rv1161	narG	intermediary metabolism and respiration	Respiratory nitrate reductase (alpha chain) NarG
Rv1162	narH	intermediary metabolism and respiration	Probable respiratory nitrate reductase (beta chain) NarH
Rv0124	PE_PGRS2	PE/PPE	PE-PGRS family protein PE_PGRS2
Rv1163	narJ	intermediary metabolism and respiration	Probable respiratory nitrate reductase (delta chain) NarJ
Rv1977	Rv1977	conserved hypotheticals	Conserved protein
Rv3825c	pks2	lipid metabolism	Polyketide synthase Pks2
Rv3197A	whiB7	regulatory proteins	Probable transcriptional regulatory protein WhiB-like WhiB7
Rv1159A	Rv1159A	conserved hypotheticals	Unknown protein
Rv2646	Rv2646	insertion sequences and phages	Probable integrase
Rv2657c	Rv2657c	insertion sequences and phages	Probable PhiRv2 prophage protein
Rv2653c	Rv2653c	insertion sequences and phages	Possible PhiRv2 prophage protein
Rv2159c	Rv2159c	conserved hypotheticals	Conserved protein
Rv1925	fadD31	lipid metabolism	Probable acyl-CoA ligase FadD31 (acyl-CoA synthetase) (acyl-CoA synthase)
Rv2161c	Rv2161c	intermediary metabolism and respiration	Conserved protein

Rv2162c	PE_PGRS38	PE/PPE	PE-PGRS family protein PE_PGRS38
Rv1899c	lppD	cell wall and cell processes	Possible lipoprotein LppD
Rv2652c	Rv2652c	insertion sequences and phages	Probable PhiRv2 prophage protein
Rv2651c	Rv2651c	insertion sequences and phages	Possible PhiRv2 prophage protease
Rv2659c	Rv2659c	insertion sequences and phages	Probable PhiRv2 prophage integrase
Rv2160c	Rv2160c	regulatory proteins	Conserved hypothetical protein
Rv2650c	Rv2650c	insertion sequences and phages	Possible PhiRv2 prophage protein
Rv2655c	Rv2655c	insertion sequences and phages	Possible PhiRv2 prophage protein

Table S6: Top 50 differential expressed genes L4 Vs. A4 strains.

Locus	Gene Name	Functional Category	Product
Rv1666c	cyp139	intermediary metabolism and respiration	Probable cytochrome P450 139 Cyp139
Rv0448c	Rv0448c	conserved hypotheticals	Conserved hypothetical protein
Rv0550c	vapB3	virulence, detoxification, adaptation	Possible antitoxin VapB3
Rv0449c	Rv0449c	conserved hypotheticals	Conserved hypothetical protein
Rv2874	dipZ	intermediary metabolism and respiration	Possible integral membrane C-type cytochrome biogenesis protein DipZ
Rv2873	mpt83	cell wall and cell processes	Cell surface lipoprotein Mpt83 (lipoprotein P23)
Rv3681c	whiB4	regulatory proteins	Probable transcriptional regulatory protein WhiB-like WhiB4
Rv0446c	Rv0446c	cell wall and cell processes	Possible conserved transmembrane protein
Rv0447c	ufaA1	lipid metabolism	Probable cyclopropane-fatty-acyl-phospholipid synthase UfaA1 (cyclopropane fatty acid synthase) (CFA synthase)
Rv0216	Rv0216	intermediary metabolism and respiration	Double hotdog hydratase
Rv2159c	Rv2159c	conserved hypotheticals	Conserved protein
Rv3429	PPE59	PE/PPE	PPE family protein PPE59
Rv1977	Rv1977	conserved hypotheticals	Conserved protein
Rv2160c	Rv2160c	regulatory proteins	Conserved hypothetical protein

Rv2162c	PE_PGRS38	PE/PPE	PE-PGRS family protein PE_PGRS38
Rv2161c	Rv2161c	intermediary metabolism and respiration	Conserved protein
Rv2651c	Rv2651c	insertion sequences and phages	Possible PhiRv2 prophage protease
Rv2650c	Rv2650c	insertion sequences and phages	Possible PhiRv2 prophage protein
Rv2659c	Rv2659c	insertion sequences and phages	Probable PhiRv2 prophage integrase
Rv1899c	lppD	cell wall and cell processes	Possible lipoprotein LppD
Rv3137	Rv3137	intermediary metabolism and respiration	Probable monophosphatase
Rv2652c	Rv2652c	insertion sequences and phages	Probable PhiRv2 prophage protein
Rv3136	PPE51	PE/PPE	PPE family protein PPE51
Rv1514c	Rv1514c	conserved hypotheticals	Conserved hypothetical protein
Rv3119	moaE1	intermediary metabolism and respiration	Probable molybdenum cofactor biosynthesis protein E MoaE1 (molybdopterin converting factor large subunit) (molybdopterin [MPT] converting factor, subunit 2)
Rv3825c	pks2	lipid metabolism	Polyketide synthase Pks2
Rv2348c	Rv2348c	unknown	Hypothetical protein
Rv1257c	Rv1257c	intermediary metabolism and respiration	Probable oxidoreductase
Rv3739c	PPE67	PE/PPE	PPE family protein PPE67
Rv2351c	plcA	intermediary metabolism and respiration	Membrane-associated phospholipase C 1 PlcA (MTP40 antigen)
Rv3679	Rv3679	cell wall and cell processes	Probable anion transporter ATPase
Rv1512	epiA	intermediary metabolism and respiration	Probable nucleotide-sugar epimerase EpiA
Rv1256c	cyp130	intermediary metabolism and respiration	Probable cytochrome P450 130 Cyp130
Rv1510	Rv1510	cell wall and cell processes	Conserved probable membrane protein
Rv2074	Rv2074	intermediary metabolism and respiration	Possible pyridoxamine 5’-phosphate oxidase (PNP/PMP oxidase) (pyridoxinephosphate oxidase) (PNPOX) (pyridoxine 5’-phosphate oxidase)
Rv1511	gmdA	intermediary metabolism and respiration	GDP-D-mannose dehydratase GmdA (GDP-mannose 4,6 dehydratase) (GMD)
Rv1508A	Rv1508A	conserved hypotheticals	Conserved hypothetical protein

Rv1516c	Rv1516c	intermediary metabolism and respiration	Probable sugar transferase
Rv1509	Rv1509	unknown	Hypothetical protein
Rv3888c	Rv3888c	cell wall and cell processes	Probable conserved membrane protein
Rv1515c	Rv1515c	conserved hypotheticals	Conserved hypothetical protein
Rv3120	Rv3120	conserved hypotheticals	Conserved hypothetical protein
Rv2352c	PPE38	PE/PPE	PPE family protein PPE38
Rv1508c	Rv1508c	cell wall and cell processes	Probable membrane protein
Rv2349c	plcC	intermediary metabolism and respiration	Probable phospholipase C 3 PlcC
Rv0276	Rv0276	conserved hypotheticals	Conserved hypothetical protein
Rv1506c	Rv1506c	unknown	Hypothetical protein
Rv1507c	Rv1507c	conserved hypotheticals	Conserved protein
Rv2350c	plcB	intermediary metabolism and respiration	Membrane-associated phospholipase C 2 PlcB
Rv1513	Rv1513	conserved hypotheticals	Conserved protein

Table S7: Top 50 differential expressed genes L6 Vs. A4 strains.

Locus	Gene Name	Functional Category	Product
Rv0124	PE_PGRS2	PE/PPE	PE-PGRS family protein PE_PGRS2
Rv2386c	mbtI	lipid metabolism	Isochorismate synthase MbtI
Rv1519	Rv1519	conserved hypotheticals	Conserved hypothetical protein
Rv0770	Rv0770	intermediary metabolism and respiration	Probable dehydrogenase/reductase
Rv0769	Rv0769	intermediary metabolism and respiration	Probable dehydrogenase/reductase
Rv0771	Rv0771	intermediary metabolism and respiration	Possible 4-carboxymuconolactone decarboxylase (CMD)
Rv2876	Rv2876	cell wall and cell processes	Possible conserved transmembrane protein
Rv2874	dipZ	intermediary metabolism and respiration	Possible integral membrane C-type cytochrome biogenesis protein DipZ
Rv0447c	ufaA1	lipid metabolism	Probable cyclopropane-fatty-acyl-phospholipid synthase UfaA1

			(cyclopropane fatty acid synthase) (CFA synthase)
Rv2873	mpt83	cell wall and cell processes	Cell surface lipoprotein Mpt83 (lipoprotein P23)
Rv0449c	Rv0449c	conserved hypotheticals	Conserved hypothetical protein
Rv2875	mpt70	cell wall and cell processes	Major secreted immunogenic protein Mpt70
Rv0446c	Rv0446c	cell wall and cell processes	Possible conserved transmembrane protein
Rv0448c	Rv0448c	conserved hypotheticals	Conserved hypothetical protein
Rv1170	mshB	intermediary metabolism and respiration	N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-alpha-D-glucopyranoside deacetylase MshB (GlcNAc-Ins deacetylase)
Rv2074	Rv2074	intermediary metabolism and respiration	Possible pyridoxamine 5'-phosphate oxidase (PNP/PMP oxidase) (pyridoxinephosphate oxidase) (PNPOX) (pyridoxine 5'-phosphate oxidase)
Rv0222	echA1	lipid metabolism	Probable enoyl-CoA hydratase EchA1 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
Rv1976c	Rv1976c	conserved hypotheticals	Conserved hypothetical protein
Rv2349c	plcC	intermediary metabolism and respiration	Probable phospholipase C 3 PlcC
Rv1507A	Rv1507A	unknown	Hypothetical protein
Rv3120	Rv3120	conserved hypotheticals	Conserved hypothetical protein
Rv1256c	cyp130	intermediary metabolism and respiration	Probable cytochrome P450 130 Cyp130
Rv2351c	plcA	intermediary metabolism and respiration	Membrane-associated phospholipase C 1 PlcA (MTP40 antigen)
Rv1516c	Rv1516c	intermediary metabolism and respiration	Probable sugar transferase
Rv1506c	Rv1506c	unknown	Hypothetical protein
Rv1515c	Rv1515c	conserved hypotheticals	
Rv3739c	PPE67	PE/PPE	PPE family protein PPE67
Rv1514c	Rv1514c	conserved hypotheticals	Conserved hypothetical protein
Rv3136	PPE51	PE/PPE	PPE family protein PPE51
Rv2346c	esxO	cell wall and cell processes	Putative ESAT-6 like protein EsxO (ESAT-6 like protein 6)
Rv3429	PPE59	PE/PPE	PPE family protein PPE59
Rv3501c	yrbE4A	virulence, detoxification, adaptation	Conserved integral membrane protein YrbE4A. Possible ABC

			transporter.
Rv2348c	Rv2348c	unknown	Hypothetical protein
Rv2352c	PPE38	PE/PPE	PPE family protein PPE38
Rv1867	Rv1867	lipid metabolism	Conserved protein
Rv0217c	lipW	intermediary metabolism and respiration	Possible esterase LipW
Rv3806c	ubiA	cell wall and cell processes	Decaprenylphosphoryl-5-phosphoribose (DPPR) synthase (decaprenyl-phosphate 5-phosphoribosyltransferase)
Rv3119	moaE1	intermediary metabolism and respiration	Probable molybdenum cofactor biosynthesis protein E MoaE1 (molybdopterin converting factor large subunit) (molybdopterin [MPT] converting factor, subunit 2)
Rv1505c	Rv1505c	conserved hypotheticals	Conserved hypothetical protein
Rv1513	Rv1513	conserved hypotheticals	Conserved protein
Rv1512	epiA	intermediary metabolism and respiration	Probable nucleotide-sugar epimerase EpiA
Rv3888c	Rv3888c	cell wall and cell processes	Probable conserved membrane protein
Rv1510	Rv1510	cell wall and cell processes	Conserved probable membrane protein
Rv1509	Rv1509	unknown	Hypothetical protein
Rv1508c	Rv1508c	cell wall and cell processes	Probable membrane protein
Rv2350c	plcB	intermediary metabolism and respiration	Membrane-associated phospholipase C 2 PlcB
Rv1507c	Rv1507c	conserved hypotheticals	Conserved protein
Rv1257c	Rv1257c	intermediary metabolism and respiration	Probable oxidoreductase
Rv1511	gmdA	intermediary metabolism and respiration	GDP-D-mannose dehydratase GmdA (GDP-mannose 4,6 dehydratase) (GMD)
Rv1508A	Rv1508A	conserved hypotheticals	Conserved hypothetical protein
Locus	Gene Name	Functional Category	Product
Rv0124	PE_PGRS2	PE/PPE	PE-PGRS family protein PE_PGRS2
Rv2386c	mbtI	lipid metabolism	Isochorismate synthase MbtI
Rv1519	Rv1519	conserved hypotheticals	Conserved hypothetical protein
Rv0770	Rv0770	intermediary metabolism and respiration	Probable dehydrogenase/reductase

Rv0769	Rv0769	intermediary metabolism and respiration	Probable dehydrogenase/reductase
Rv0771	Rv0771	intermediary metabolism and respiration	Possible 4-carboxymuconolactone decarboxylase (CMD)
Rv2876	Rv2876	cell wall and cell processes	Possible conserved transmembrane protein
Rv2874	dipZ	intermediary metabolism and respiration	Possible integral membrane C-type cytochrome biogenesis protein DipZ
Rv0447c	ufaA1	lipid metabolism	Probable cyclopropane-fatty-acyl-phospholipid synthase UfaA1 (cyclopropane fatty acid synthase) (CFA synthase)
Rv2873	mpt83	cell wall and cell processes	Cell surface lipoprotein Mpt83 (lipoprotein P23)

12. SCIENTIFIC PUBLICATIONS

March 2024

Matthias I Gröschel*, **Francy J. Pérez-Llanos***, Roland Diel, Roger Vargas Jr, Vincent Escuyer, Kimberlee Musser, Lisa Trieu, Jeanne Sullivan Meissner, Jillian Knorr, Don Klinkenberg, Peter Kouw, Susanne Homolka, Wojciech Samek, Barun Mathema, Dick van Soolingen, Stefan Niemann**, Shama Ahuja, Maha R Farhat**.

* Shared co-first authorship ** Shared co-last authorship

Host-pathogen sympatry and differential transmissibility of *Mycobacterium tuberculosis* complex.

Nature Microbiology. 2024. NMICROBIOL-23102922A

Accepted for publication.

medRxiv 2022.08.04.22278337.
doi: https://doi.org/10.1101/2022.08.04.22278337.

Mai 2023

Pérez-Llanos FJ, Dreyer V, Barilar I, Utpatel C, Kohl TA, Murcia MI, Homolka S, Merker M, Niemann S.

Transmission Dynamics of Mycobacterium tuberculosis complex outbreak in an indigenous population in the Colombian Amazon region.

Microbiol. Spectr. 2023 Jun 15;11(3):e0501322.
doi: 10.1128/spectrum.05013-22.

March 2018

Magda Beltrán-León , **Francy Pérez-Llanos**, Liliana Sánchez, Carlos Parra-López, Myriam Navarrete, Ricardo Sánchez, Carlos Awad, Ana María Granada, Edgardo Quintero, Óscar Briceño, Óscar Cruz, Martha Isabel Murcia

Prevalence and risk factors associated to tuberculosis and non-tuberculous mycobacterial infections in HIV-positive patients in Bogotá.

Biomedica. 2018 Mar 15;38(1):120-127.
doi: 10.7705/biomedica.v38i0.3410.

June 2015

Mongui A, **Pérez-Llanos FJ**, Yamamoto MM, Lozano M, Zambrano MM, Del Portillo P, Fernández-Becerra C, Restrepo S, Del Portillo HA, Junca H.

Development of a genetic tool for functional screening of anti-malarial bioactive extracts in metagenomic libraries.

Malar J. 2015 Jun 4;14:233.
doi: 10.1186/s12936-015-0748-6.

June 2010

Castillo Bohórquez, Martha; Mora Bautista, Ana Isabel; Laiton Donato, Katherine; **Pérez Llanos, Francy**; Tapiero Rodríguez, Milena.

Identification of iron deficiency risk by index soluble transferrin receptor-log ferritin in African descent men living at San Basilio de Palenque, Cartagena de Indias, DT and C., Bolivar, Colombia.

Nova, 2010. Spanish. Repositorio Institucional UNAD.
https://repository.unad.edu.co/handle/10596/29915.

13. SCIENTIFIC PRESENTATIONS

September 2021	Kiel (Germany). Evolution by the Sea (conference). **Pérez F. J.,** Homolka S., and Niemann S**.** *Dissection of the Complex between different Mycobacterium tuberculosis complex strains and host cells in determining bacterial pathogenesis and host susceptibility*.
February 2020	Hamburg (Germany). 7th International Symposium Inflammation Medicine – From Bench to Bedside. **Pérez F. J.,** Utpatel C, Dreyer V, Homolka S and Niemann S. *Application of nanopore sequencing for drug-resistant prediction of Mycobacterium tuberculosis complex strains.*
August 2019	Zürich (Switzerland). The International Society for Evolution, Medicine & Public Health. **Pérez F. J.,** Homolka S and Niemann S. *Impact of host-pathogen mismatch in tuberculosis infections: Insights from a new infection model employing M. africanum and bovine macrophages*.
July 2019	Valencia (Spain). 40th Annual Congress of the European Society of Mycobacteriology. **Pérez F. J.,** Homolka S and Niemann S. *Application of nanopore sequencing for drug-resistant prediction of Mycobacterium tuberculosis complex strains.*
September 2019	Lübeck (Germany). Flash talk and abstract. German Center for Infection and Research Academy 2019. *Application of nanopore sequencing for drug-resistant Prediction of Mycobacterium tuberculosis complex Strains.*

July 2018	Dresden (Germany). 39th Annual Congress of the European Society of Mycobacteriology. *Tuberculosis in indigenous peoples with settlements at Puerto Nariño- Amazonas, Colombia.*
July 2016	Catania, Sicily (Italy). 37th Annual Congress of the European Society of Mycobacteriology. **Pérez F. J.**, Beltrán M, Parra C, Navarrete M, Knudson A, Sánchez R, Murcia M.I *Detection of latent Tuberculosis infection (LTBI) within an HIV-infected population in Bogota, Colombia.* Beltrán M, **Pérez F. J.**, Rodríguez J, Parra C1, Navarrete M, Knudson A, Sánchez R, Murcia M.I *The prevalence of Mycobacterial infections in HIV-infected patients from Bogota, Colombia.*
June to July 2015	Latvia (Riga). 36th Annual Congress of the European Society of Mycobacteriology. **Pérez F. J.,** Beltrán M, Sánchez L, Parra C, Navarrete M, Knudson A, Sánchez R, Murcia M.I *Detection of latent tuberculosis in HIV-infected patients from Simón Bolivar and Santa Clara hospitals in Bogotá, Colombia.*
2014	Cartagena de Indias (Colombia). XXI Latin American Congress of Microbiology ALAM. Alvaro Mongui, **Francy J. Pérez** and Howard Junca. *Design and validation of a gene encoding and eukaryotic peptide for functional analysis of metagenomics libraries to explore for novel antimalarial compounds*.