



OPEN ACCESS

Sub-Lineage Specific Phenolic Glycolipid Patterns in the *Mycobacterium tuberculosis* Complex Lineage 1

Edited by:

Axel Cloeckaert,
Institut National de Recherche Pour
l'Agriculture, l'Alimentation et
l'Environnement (INRAE), France

Reviewed by:

Brosch Roland,
Université Louis-Pasteur, France
Sanjay Gautam,
The University of Melbourne, Australia
Jesús Gonzalo-Asensio,
University of Zaragoza, Spain

***Correspondence:**

Nicolas Gisch
ngisch@fz-borstel.de
Stefan Niemann
sniemann@fz-borstel.de

† Present address:

Sven Malm,
Karlsruhe Institute of Technology,
Safety and Environment,
Eggenstein-Leopoldshafen, Germany

‡ These authors have contributed
equally to this work and share first
authorship

§ These authors have contributed
equally to this work and share last
authorship

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 09 December 2021

Accepted: 20 January 2022

Published: 08 March 2022

Citation:

Gisch N, Utpatel C,
Gronbach LM, Kohl TA, Schombel U,
Malm S, Dobos KM, Hesser DC,
Diel R, Götsch U, Gerdes S,
Shuaib YA, Ntinginya NE, Khosa C,
Viegas S, Kerubo G, Ali S,
Al-Hajoj SA, Ndung'u PW, Rachow A,
Hoelscher M, Maurer FP,
Schwudke D, Niemann S, Reiling N
and Homolka S (2022) Sub-Lineage
Specific Phenolic Glycolipid Patterns
in the *Mycobacterium tuberculosis*
Complex Lineage 1.
Front. Microbiol. 13:832054.
doi: 10.3389/fmicb.2022.832054

Nicolas Gisch^{1*}, Christian Utpatel^{2†}, Lisa M. Gronbach¹, Thomas A. Kohl²,
Ursula Schombel¹, Sven Malm^{2†}, Karen M. Dobos³, Danny C. Hesser³, Roland Diel⁴,
Udo Götsch⁵, Silke Gerdes⁶, Yassir A. Shuaib^{7,8}, Nyanda E. Ntinginya⁹, Celso Khosa¹⁰,
Sofia Viegas¹⁰, Glennah Kerubo¹¹, Solomon Ali¹², Sahal A. Al-Hajoj¹³,
Perpetual W. Ndung'u¹⁴, Andrea Rachow^{15,16}, Michael Hoelscher^{15,16},
Florian P. Maurer^{17,18}, Dominik Schwudke^{1,19,20}, Stefan Niemann^{2,19*§}, Norbert Reiling^{19,21§}
and Susanne Homolka^{2§}

¹ Bioanalytical Chemistry, Priority Area Infections, Research Center Borstel, Leibniz Lung Center, Borstel, Germany, ² Molecular and Experimental Mycobacteriology, Priority Area Infections, Research Center Borstel, Leibniz Lung Center, Borstel, Germany, ³ Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO, United States, ⁴ Lung Clinic Grosshansdorf, Airway Disease Center North (ARC/N), German Center for Lung Research (DZL), Großhansdorf, Germany, ⁵ Municipal Health Authority Frankfurt am Main, Frankfurt am Main, Germany, ⁶ Municipal Health Authority Hannover, Hanover, Germany, ⁷ College of Veterinary Medicine, Sudan University of Science and Technology, Khartoum, Sudan, ⁸ WHO-Supranational Reference Laboratory of Tuberculosis, Institute of Microbiology and Laboratory Medicine (IML Red), Gauting, Germany, ⁹ National Institute for Medical Research Tanzania – Mbeya Medical Research Center, Mbeya, Tanzania, ¹⁰ Instituto Nacional de Saúde (INS), Marracuene, Mozambique, ¹¹ Department of Medical Microbiology and Parasitology, School of Medicine, Kenyatta University, Nairobi, Kenya, ¹² Department of Microbiology, Immunology, and Parasitology, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia, ¹³ Mycobacteriology Research Section, Department of Infection and Immunity, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, ¹⁴ Institute of Tropical Medicine and Infectious Diseases (ITROMID), Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya, ¹⁵ Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, Munich, Germany, ¹⁶ German Centre for Infection Research (DZIF), Partner Site Munich, Munich, Germany, ¹⁷ National and WHO Supranational Reference Centre for Mycobacteria, Research Center Borstel, Leibniz Lung Center, Borstel, Germany, ¹⁸ Institute of Medical Microbiology, Virology, and Hygiene, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, ¹⁹ German Center for Infection Research (DZIF), Partner Site Hamburg-Lübeck-Borstel-Riems, Borstel, Germany, ²⁰ Airway Research Center North, Member of the German Center for Lung Research (DZL), Borstel, Germany, ²¹ Microbial Interface Biology, Priority Area Infections, Research Center Borstel, Leibniz Lung Center, Borstel, Germany

“Ancestral” *Mycobacterium tuberculosis* complex (MTBC) strains of Lineage 1 (L1, East African Indian) are a prominent tuberculosis (TB) cause in countries around the Indian Ocean. However, the pathobiology of L1 strains is insufficiently characterized. Here, we used whole genome sequencing (WGS) of 312 L1 strains from 43 countries to perform a characterization of the global L1 population structure and correlate this to the analysis of the synthesis of phenolic glycolipids (PGL) – known MTBC polyketide-derived virulence factors. Our results reveal the presence of eight major L1 sub-lineages, whose members have specific mutation signatures in PGL biosynthesis genes, e.g., *pkS15/1* or glycosyltransferases Rv2962c and/or Rv2958c. Sub-lineage specific PGL production was studied by NMR-based lipid profiling and strains with a completely abolished phenolphthiocerol dimycoserolate biosynthesis showed in average a more prominent growth in human macrophages. In conclusion, our results show a diverse population structure of L1 strains that is associated with the presence of specific PGL

types. This includes the occurrence of mycoside B in one sub-lineage, representing the first description of a PGL in an *M. tuberculosis* lineage other than L2. Such differences may be important for the evolution of L1 strains, e.g., allowing adaptation to different human populations.

Keywords: *Mycobacterium tuberculosis* complex, single nucleotide polymorphism (SNPs), NMR, phenolic glycolipids, genetic diversity, phylogeny, whole genome sequencing (WGS), structural chemistry

INTRODUCTION

The genetic diversity of *Mycobacterium tuberculosis* complex (MTBC) strains has been shown to influence transmission and virulence of clinical MTBC isolates as well as the immune response and clinical outcome (reviewed in Coscolla and Gagneux, 2014; Tientcheu et al., 2017). MTBC strains of so called “modern” phylogenetic lineages (clade 1) (Wirth et al., 2008; Barbier and Wirth, 2016) are defined by a genomic deletion known as TbD1 (Brosch et al., 2002). Clade 1 strains that comprises Lineage 2 (L2, Beijing genotype), Lineage 3 [L3, Delhi/Central-Asian (Delhi/CAS)], and Lineage 4 (L4, Euro-American) are responsible for the majority of tuberculosis (TB) cases worldwide (Brites and Gagneux, 2017). In contrast, the distribution of patients infected with clade 2 strains (“ancestral” strains defined by an intact TbD1 region) including Lineage 1 [L1, East African Indian genotype (EAI)], Lineage 5 [L5, *M. africanum* West African 1 genotype (WA1)], and Lineage 6 [L6, *M. africanum* West African 2 genotype (WA2)] (Wirth et al., 2008) are geographically more restricted and show a globally lower prevalence in comparison to “modern” strains (Brites and Gagneux, 2017). Especially, L1 strains are a prominent cause of TB on the Indian subcontinent (Gutierrez et al., 2006; Thomas et al., 2011) and in Malaysia (Ismail et al., 2014). It has been reported that L1 strains display a reduced transmission rate in comparison to “modern” strains of L2 or L4 (Albanna et al., 2011; Nebenzahl-Guimaraes et al., 2015). One of the molecular reasons for that is linked to the presence of the TbD1 region in L1 strains. Loss of TbD1 in “modern” MTBC strains has been associated with a better circumvention of and persistence against oxidative stress and hypoxic conditions within the host cell micro-environments. This may generate an important advantage for those MTBC pathogens especially during prolonged stages of infection (Bottai et al., 2020). In addition, these observations are in line with data from own experiments using *in vitro* and *in vivo* model systems, since we previously showed that “modern” MTBC strains from L2 or L4 exhibit significantly enhanced growth rates in human monocyte-derived and alveolar macrophages as well as in aerogenically infected mice, when compared to L1 strains (Reiling et al., 2013).

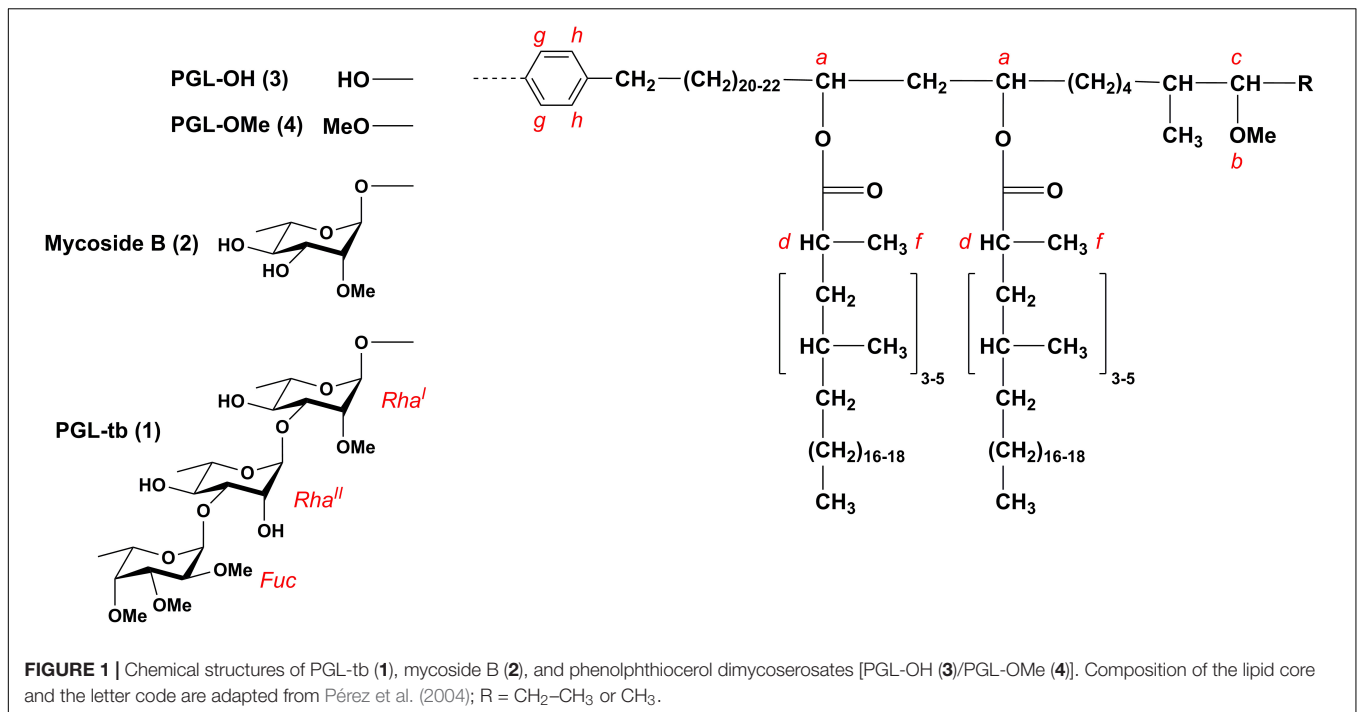
Cell wall components such as lipids and lipoglycans including trehalose dimycolates (TDM), sulfolipids (SL), phthiocerol dimycoserates (PDIM) and phenolic glycolipids (PGL; also termed glycosyl phenolphthiocerol dimycoserates) have been discussed as virulence factors of MTBC strains (Forrellad et al., 2013; Madacki et al., 2019).

Trehalose dimycolates, also known as the mycobacterial cord factor, plays a major role in modulating the acute response of the

host cell upon entry of the MTBC pathogen. It has been shown that TDM modulates the early M1-like macrophage response as seen during initiation of the granulomas of primary pathology (Nguyen et al., 2020). Furthermore, TDM delays phagosomal maturation through MINCLE signaling (Axelrod et al., 2008; Patin et al., 2017) and antagonizes IFN γ -induced expression of MHC-II as well as antimicrobial effector genes such as GBP1 (Huber et al., 2020).

Independent studies have shown that PDIM and also structurally related PGL are lipid virulence factors that are associated with mycobacterial host evasion and modulate the host immune response in many ways (Quigley et al., 2017; Mishra et al., 2019; Bah et al., 2020). *Mycobacterium tuberculosis* mutants showing PDIM transport deficiency induced in THP-1 macrophages lower levels of microtubule-associated protein light chain 3 (LC3), often used as autophagy marker associated with the macrophage membranes, when compared to the respective wild-type strain (CDC1551) (Quigley et al., 2017). Mishra et al. (2019) proposed a multifaceted role of SL, since SL-1 modulates key biophysical properties of the host cell membrane such as fluidity, hydration and lipid domain re-organization. Bah et al. (2020) showed that SL limits TLR/MyD88-dependent and phagosomal damage-independent autophagy by acting as a Toll-Like Receptor 2 (TLR2) antagonist, whereas PDIM prevents this process to a certain extent by favoring Esx-1-dependent phagosomal damage. Furthermore, it was shown that PDIM, but not SLs, limit the acidification of LC3-positive compartments containing *M. tuberculosis*. Finally, PDIM also mask other pathogen associated molecular pattern on the cell wall, thereby dampening immune responses mediated by pattern recognition receptors (Arbues et al., 2014; Cambier et al., 2014).

Phenolic glycolipids are capable of recruiting permissive macrophages via chemokine receptor 2 enabling the pathogen to reach the lower respiratory tract (Cambier et al., 2014, 2017). PGL have further been shown to inhibit the TLR2 response, leading to the reduction of pro-inflammatory cytokine production (Arbués et al., 2016; Blanc et al., 2017). A subset of L2 clinical isolates displayed a hypervirulent phenotype in mice, which was correlated with the presence of PGL-tb [2,3,4-tri-*O*-methyl-L-fucopyranosyl- α -(1 \rightarrow 3)-L-rhamnopyranosyl- α -(1 \rightarrow 3)-2-*O*-methyl-L- α -(1 \rightarrow)-rhamnopyranosyl phenolphthiocerol dimycoserate (**1**); **Figure 1**; Reed et al., 2004]. Although PGL-tb (**1**) is capable of modulating the early host cytokine response (Stadthagen et al., 2006) and contributes to the manipulation of macrophage recruitment by the pathogen as described above, (Cambier et al., 2014) it does not confer hypervirulence itself. A biochemically engineered PGL-tb (**1**)-producing H37Rv strain, a strain normally devoid of PGL-tb (**1**) synthesis, suppressed



the induction of several pro- and anti-inflammatory cytokines *in vitro* in human monocytes, similar to PGL-tb (1)-producing clinical *M. tuberculosis* isolates. However, this did not lead to increased virulence in infected mice and rabbits (Sinsimer et al., 2008). In contrast to L2 strains, all strains of the Euro-American lineage (L4) lack PGLs and phenolphthiocerol dimycoserates (phenolic PDIM) due to a frameshift in the *pks15/1* gene (Constant et al., 2002). Notably, this frameshift does not affect the presence of PDIM (Constant et al., 2002; Siméone et al., 2010). Moreover, *M. tuberculosis* strains that are unable to produce PGL-tb (1) due to this frameshift can still produce glycosylated *para*-hydroxybenzoic acid methyl esters (pHBAD), which are essential building blocks for PGL biosynthesis (Pérez et al., 2004). Such molecules have been shown to dampen the immune response as well (Stadhagen et al., 2006; Bourke et al., 2014). Interestingly, a mutation in Rv2962c (Q294Stop), which deletion leads to an abolishment of PGL biosynthesis at the stage of phenolic PDIM (Pérez et al., 2004), was described in strains of the L1 lineage (Indo-Oceanic lineage) (Krishnan et al., 2011). However, the impact on pathogenicity mediated by the molecular type of PGL (Figure 1) synthesized in the MTBC strains, either as PGL-tb (1), mycoside B (2), phenolphthiocerol dimycoserates PGL-OH (3) and PGL-OMe (4), is not yet fully elucidated (Arbues et al., 2014; Barnes et al., 2017).

This study aimed to link the population structure of L1 strains to the presence of distinct PGL types and to address their potential correlation with differences in virulence. First, we analyzed the global population structure of L1 strains by Whole Genome Sequencing (WGS) analysis of 312 strains from 43 countries. Based on this L1 population structure, we determined major sub-lineages, out of which representative clinical isolates were selected for further analysis of their

lipid composition using NMR-based approaches and performed growth analyses in human macrophages, the natural host cell of the TB pathogen.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Bacterial strains were initially cultured from clinical samples on Löwenstein/Jensen (L/J) media at the National Reference Center for Mycobacteria in Borstel, Germany. The MTBC clinical isolates used in this study were handled to minimize *in vitro* passaging. All cultures used in this study were derived from frozen stocks prepared after a single *in vitro* passage of original archived samples. The virulent lab strain H37Rv ATCC 27294, however, has undergone countless rounds of *in vitro* passaging. All strains were further characterized by WGS as previously described and phenotypic susceptibility testing according to WHO standards (Meehan et al., 2019).

Homogenous bacterial suspensions were prepared from L/J cultures in 10 ml 7H9 media (Difco Becton Dickinson, Sparks, MD, United States) supplemented with 10% OADC (Becton Dickinson GmbH, Heidelberg, Germany), 0.05% Tween 80 (Merck KGaA, Darmstadt, Germany) as well as 0.2% glycerol (Merck KGaA, Darmstadt, Germany) and incubated in 30 ml square media bottles (Nalgene®, Thermo Fisher Scientific, Darmstadt, Germany) at 37°C without shaking (pre-culture). Growth to mid-log phase was monitored by measuring OD₆₀₀ every second day (BioTek Synergy2, Agilent, Santa Clara, CA, United States). Pre-cultures were transferred into a roller bottle system (Corning®, Merck KGaA, Darmstadt, Deutschland) and incubated at 37°C and 5 rpm. 20 ml of fresh media were gradually

added to a final volume of 100 ml (main-culture). Mid-log phase bacteria suspension (OD₆₀₀ 0.2–0.4) were stored in 1 ml aliquots at –80°C for further investigations. Colony forming units were determined for each strain investigated. Furthermore, sterility controls (Ziehl-Neelsen staining, blood agar, LB media, BHI bouillon) were performed for all pre- and main-cultures.

For ¹H, ¹³C-HSQC NMR lipid profiling an aliquot of a given MTBC strains was thawed, centrifuged at 4,500 × g and the supernatant was discarded. The bacteria were resuspended in 1 ml PBS using a G 26 × 1/2''/ø 0.45 mm × 12 mm Sterican® canula (B. Braun, Melsungen, Germany) and inoculated in 25–40 ml 7H9 media with the addition of 10% self-prepared OADC containing ¹³C glucose (D-Glucose U-¹³C₆, Euriso-Top, Gif-sur-Yvette, France), 0.2% ¹³C₃ glycerol (Cambridge Isotope Laboratories, Andover, MA, United States) and 0.05% Tyloxapol (Sigma-Aldrich Chemie GmbH, Steinheim, Germany).

Bacteria were grown in roller bottles at 37°C until an OD₆₀₀ of 0.6–0.8. Subsequently, 1.25 ml of a 20% PFA solution were added for 24 h at 37°C to inactivate the bacteria prior to the investigation of mycobacterial lipid profiles (see below).

Study Population of *Mycobacterium tuberculosis* L1 – East African Indian Strains

Whole genome sequencing datasets for L1 strains were obtained by screening our internal WGS database of 21,000 strains, which have been automatically classified based on the Coll lineage classification system (Coll et al., 2014) for an initial number of 475 genomes. Out of these, low quality datasets (average read coverage ≤ 50 fold and/or coverage breadth ≤ 95%), double sequenced strains and datasets with no metadata were excluded. After a preliminary analysis datasets of low diversity samples originating from possible transmission events with a distance of ≤ 5 SNPs to at least one other strain were excluded. This resulted in a final dataset of 312 L1 strains from 43 countries.

Whole Genome Sequencing

For all WGS datasets, libraries were prepared from gDNA using a modified Nextera protocol (Baym et al., 2015) or the Nextera XT kit (Illumina, San Diego, CA, United States) and were sequenced with 2 bp × 100 bp, 2 bp × 150 bp, or 2 bp × 300 bp paired-end reads on an Illumina HiSeq, MiSeq or NextSeq 500 platform as instructed by the manufacturer (Illumina, San Diego, CA, United States).

Raw sequence data (FASTQ files) were analyzed with MTBseq, a semi-automated bioinformatic pipeline for the analysis of MTBC isolates (Kohl et al., 2018). Briefly, reads were mapped to the *M. tuberculosis* H37Rv reference genome (GenBank ID: NC_000962.3) and alignments were refined with regard to base quality re-calibration and alignment corrections for possible PCR artifacts. Further data processing for variant calling and inference of transmission events was done as previously described (Merker et al., 2018). Variants were considered if covered with at least four reads indicating the variant in both forward and reverse read orientation, four reads showing the variant with at least a Phred score of 20, and found with at least 75% allele frequency. For the

evaluation of variations in the PDIM and PGL pathway (Siméone et al., 2010), the presence of pathway genes was confirmed for all strains. All datasets reached a coverage depth of at least 50× and a coverage breadth of at least 95% of the reference genome.

Phylogenetic Inference

After filtering SNPs within a window of 12 bp in the same genome, or located in repetitive regions and resistance-associated genes, the remaining SNP positions from all compared datasets were combined into a concatenated core-SNP alignment. Maximum likelihood (ML) phylogenetic inference was calculated from this alignment only using polymorphic sites by IQ-TREE v1.6.8 (Nguyen et al., 2015), with automatic model selection by ModelFinder Plus (Kalyaanamoorthy et al., 2017) based on the Akaike information criterion (AIC), corrected Akaike information criterion (AICc), and the Bayesian information criterion (BIC) including ascertainment bias correction (ASC), using ultrafast bootstrap approximation (UFBoot) (Hoang et al., 2018) with 1,000 replicates and the -bnni option to account for the possibility of severe model violations. ML trees were midpoint rooted. We detected sub-lineages within the tree based on deep splits with an UFBoot support ≥ 95 and using an hierarchical Bayesian Analysis of Population Structure (hierBAPS) model as implemented in R (RhierBAPS) with a maximum depth of 2 and 20 initial clusters (Cheng et al., 2013). Results were compared to the lineage classification of Coll et al. (2014).

Sub-Lineage Specific Single Nucleotide Polymorphisms

From the set of detected polymorphic positions, we extracted sets of variants specific for groups of isolates defined either by the phylogenetic analysis or manual inspection, i.e., phylogenetic SNPs characteristic for groups or sub-lineages of the constructed phylogenetic trees. For that, we evaluated the joint list of the polymorphic positions found in the 312 study isolates by filtering the list for SNP alleles specific for each of the targeted groups of isolates. For phylogenetic informative positions for the classification of sub-lineages only SNPs present in 100% of the sub-lineage strains were evaluated. Synonymous SNPs in essential, coding regions were preferred. All strains were reclassified based on the newly selected SNPs.

Phylogeographic Ancestral State Inference

To infer the geographic origin of the ancestors of L1 lineage we used the information of the patient's origin defined by the UN-region as discrete states. When strains were isolated in a low incidence TB country (e.g., Europe), the majority of such cases corresponds to immigrants from high incidence countries and it is highly likely that they were infected in their country of origin. Thus, the native geographic range of the strains is the informative criterion for the phylogeographic reconstruction. As no information on the individual living and health, working, travel and immigration history was available, it cannot be excluded that in some cases samples may have been attributed to an incorrect country if infection happened not in the home

country. To mitigate an influence on the results, countries were binned into UN-regions. We used the origin information off all 312 strains and the calculated ML-phylogeny for the ancestral reconstruction. First the best model of character evolution was selected by calculating the AIC for an equal-rates (ER) model, a symmetric, un-ordered model (SYM), and an all-rates-different (ARD) model with fitMk from the R package phytools (Revell, 2012) and computing their Akaike weights. Scaled likelihoods for each ancestral state were calculated with the best fitting model ER (aic.w = 0.97) and the ace function from the R package ape (Paradis et al., 2004) using joint maximum likelihood estimation and were mapped back to the ML phylogeny.

Total Lipid Extracts

Extraction of total lipids from mycobacterial cells was performed following a previously described procedure (Mahrous et al., 2008) with a few modifications. Briefly, the cells were harvested by centrifugation at $1,900 \times g$ and 4°C for 10 min. The cell pellet was washed three times with D_2O (appr. 1 ml per 0.1 g cells), removing the aqueous wash in between by centrifugation ($3,700 \times g$, 20°C , 10 min). In a final washing step, the material is transferred with 1–2 ml of D_2O into a 2 ml Eppendorf safe-lock tube ($3,700 \times g$, 20°C , 10 min). After careful removal of the remaining D_2O and lyophilization for 30 mins, the semi-dry pellet was weighed and subsequently extracted with a 2:1 (v/v) mixture of CDCl_3 and CD_3OD (3 ml per milligram of cells). The extraction was performed at 37°C for 90 min with continuous shaking at 250 rpm. After extraction, CD_3OD (1 ml per milligram of cells) was added to allow the formation of a single solvent phase. The pellet was then separated by centrifugation ($3,700 \times g$, 20°C , 10 min). However, in most cases the addition of the double volume of CD_3OD was necessary to form a single solvent phase. Finally, the supernatant was dried under a stream of nitrogen gas and subsequently freeze-dried and weighted.

NMR Spectroscopy

Deuterated solvents were purchased from Deutero GmbH (Kastellaun, Germany). NMR spectroscopic measurements were performed in $\text{CDCl}_3/\text{CD}_3\text{OD}/\text{D}_2\text{O}$ 60:35:8 (v/v/v) at 300 K on a Bruker Avance^{III} 700 MHz (equipped with an inverse 5 mm quadruple-resonance Z-grad cryoprobe). Total lipid extracts were dissolved in a concentration of $6 \mu\text{g}/\mu\text{l}$ and ^1H , ^{13}C -HSQC NMR lipid profiles were recorded in 3 or 5 mm NMR tubes, respectively, based on the available amount of total lipid extracts. In the case of insufficient amount of lipids, the complete sample was dissolved in 150 μl of solvent mixture and measured using a 3 mm NMR tube. TMS was used as an external standard for calibration of ^1H ($\delta_{\text{H}} = 0.0$ ppm) and ^{13}C ($\delta_{\text{C}} = 0.0$ ppm) NMR spectra. All data were acquired and processed using Bruker TOPSPIN V 3.0 or higher (Bruker BioSpin Corporation, Billerica, MA, United States).

Isolation of Glycolipids by Preparative Thin-Layer Chromatography

Total lipid extracts were dissolved in $\text{CHCl}_3/\text{CH}_3\text{OH}$ 4:1 (v/v) in a concentration of $20 \mu\text{g}/\mu\text{l}$ and applied to HPTLC plates

(glass-backed 10 cm \times 10 cm silica gel 60 F₂₅₄; Merck KGaA, Darmstadt, Germany, 1.05635.0001; up to five mg per plate). Lipids were separated with $\text{CHCl}_3/\text{CH}_3\text{OH}$ 95:5 (v/v), and a small sidebar of the plate was stained with Hanessian's stain (Pirson et al., 2015). Lipid bands (R_f range was 0.49–0.54 for **1**, 0.54–0.60 for **2**) were isolated from the silica and recovered by extraction with $\text{CHCl}_3/\text{CH}_3\text{OH}$ 2:1 (v/v) for 30 s with thorough vortexing and subsequent centrifugation ($1,080 \times g$, 10 min, 4°C ; three rounds of collecting and replacing the organic supernatant). After drying under a stream of nitrogen gas, the residual silica was removed by filtration in $\text{CHCl}_3/\text{CH}_3\text{OH}$ 2:1 (v/v) using an Acrodisc CR 13 mm syringe filter (0.2 μm PTFE membrane, Pall Corporation, Puerto Rico, United States; washed with 2 ml \times 2 ml $\text{CHCl}_3/\text{CH}_3\text{OH}$ 2:1 (v/v) prior to use).

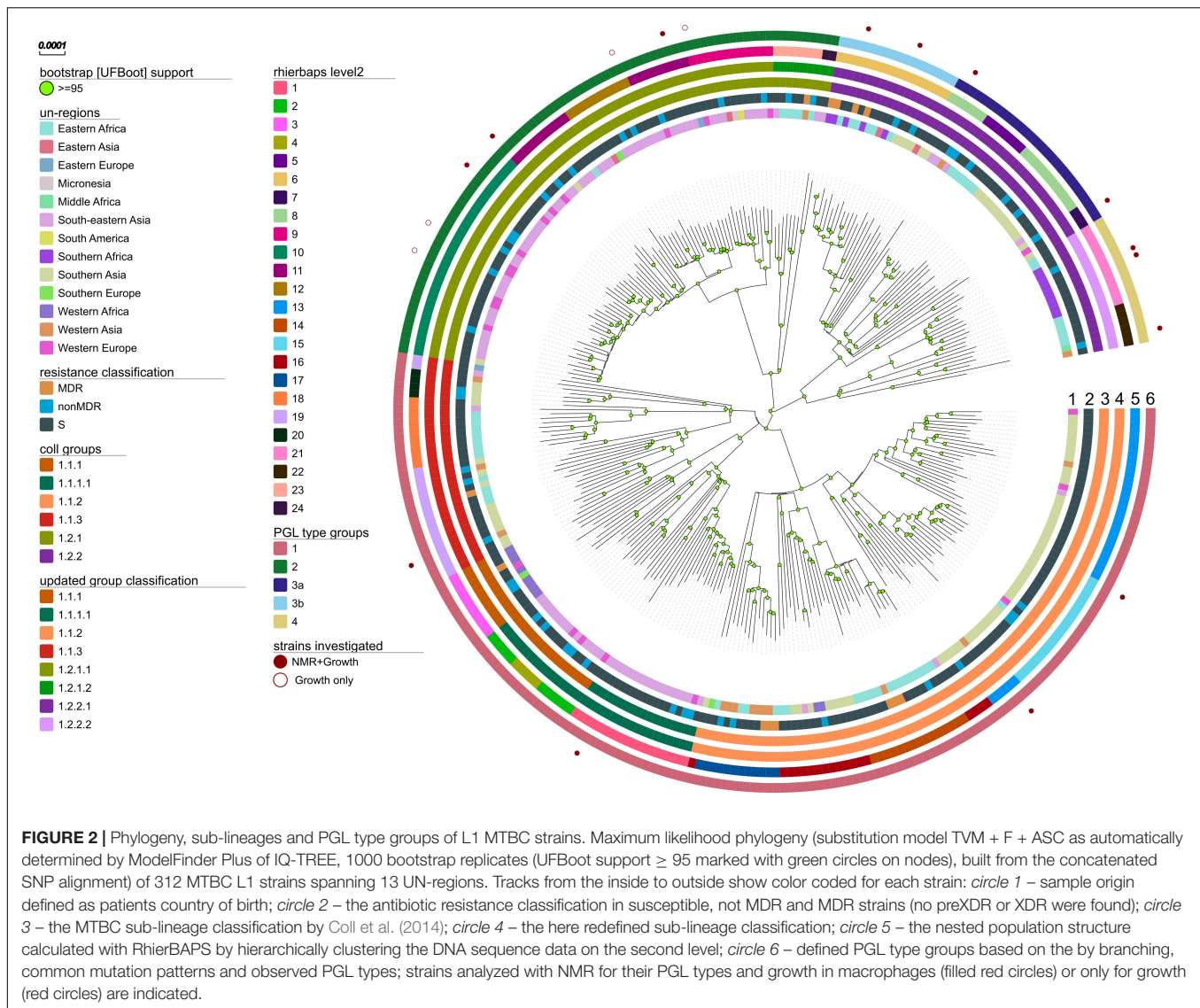
Macrophage Assays

Macrophage infection experiments were performed as described previously (Reiling et al., 2013). In brief, 2×10^5 human monocyte-derived macrophages (hMDMs) were cultured in 500 μl RPMI 1640 (Biochrom, Merck KGaA, Darmstadt, Germany) with 10% FCS and 4 mM L-glutamine (Biochrom, Merck KGaA, Darmstadt, Germany) in 48-well flat-bottom microtiter plates (Nunc, Thermo-Fisher Scientific, Waltham, MA, United States) at 37°C in a humidified atmosphere containing 5% CO_2 . Macrophages were infected with designated MTBC strains with a MOI of 1 bacterium per macrophage (1:1) for 4 h. Subsequently non-phagocytosed bacteria were removed by washing three times with 0.5 ml Hanks' balanced salt solution (HBSS; Biochrom, Merck KGaA, Darmstadt, Germany) at 37°C . After washing and after 3 days of cultivation, 0.5 ml media was added to the macrophage culture. At day 7 supernatants were completely removed. Macrophage cultures were lysed at 4 h and 7 day post infection by adding 10 μl 10% Saponin solution (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) in HBSS at 37°C for 15 min. Inoculated bacteria and macrophage lysates were serially diluted in sterile water containing 0.05% Tween 80 (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) and plated twice on 7H10 agar containing 0.5% glycerol (Carl Roth, Karlsruhe, Germany) and 10% heat-inactivated Bovine Calf Serum (BioWest, Nuaille, France). After 3 weeks at 37°C the CFUs were counted. Strain 1797/03 was used as an internal control in all experiments performed.

RESULTS

Mycobacterium tuberculosis Complex Lineage 1 Phylogeny

To analyze the global population structure of L1 strains we compiled a WGS dataset of 312 L1 strains from 43 countries, spanning 13 UN-regions that comprise all major geographical areas with reported occurrence of L1 strains (Figure 2, Supplementary Figure 1, and Supplementary Table 1). Sub-lineages have been automatically assigned by using the currently most widely used classification system of Coll et al. (2014). Our WGS data analysis revealed 35,517 single nucleotide polymorphisms (SNPs) occurring in the genome of at least one

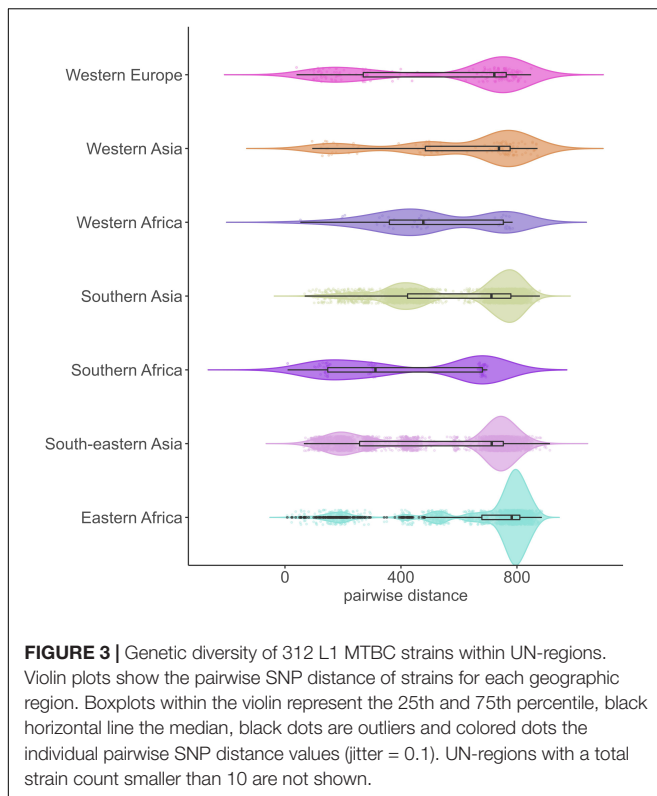


strain in comparison to the reference genome of *M. tuberculosis* H37Rv. All SNP positions were combined into a concatenated sequence alignment and used for downstream data analysis, e.g., phylogenetic tree calculations, definitions of sub-lineages and their signature SNPs. The Maximum likelihood (ML)-phylogeny revealed a diverse population structure with an overall number of eight sub-lineages, out of which six have been described by Coll et al. (2014) (**Figure 2** and **Supplementary Figure 2**). For our high-resolution analysis, we updated the sub-lineage classification based on Coll et al.: Coll lineage 1.2.2 is split into two sub-lineages named 1.2.2.1 and 1.2.2.2, and Coll lineage 1.2.1 is split into 1.2.1.1 and 1.2.1.2. Additionally, we redefined Coll groups 1.1.1 and 1.1.1.1 to reflect a deeper split in the phylogeny. For all sub-lineages, we assigned new signature SNPs that allow rapid L1 sub-lineage classification (**Supplementary Table 2**).

The average pairwise distances (avg. pwd) of distinct SNP positions in the sub-lineages ranged between 205 (median = 199) SNPs for group 1.2.1.1 to 556 (median = 642) SNPs for 1.1.3

(**Supplementary Figure 3**). On the geographic level, the region with the lowest avg. pwd distance of 437 (median = 312) was Southern Africa followed by Western Africa (avg. pwd = 518, median = 477) (**Figure 3**). In both regions only two sub-lineages were identified, 1.2.2.1 and 1.2.2.2 in Southern Africa and 1.1.1 and 1.1.2 in Western Africa, respectively (**Figure 4**). Sub-lineage 1.1.1 of this classification is, apart from two strains from Europe, exclusively found in Western Africa supporting this sub-lineage redefinition. The highest L1 population diversity was found among strains from Eastern Africa (avg. pwd = 694, median = 782) with six L1 sub-lineages present. Strains of six sub-lineages were also found to be present in South-eastern Asia (avg. pwd = 544, median = 713), but in contrast to Eastern Africa two sub-lineages (1.1.1.1 and 1.2.1.1) were dominant, which were nearly exclusively found in this region.

To better understand the global spread and origin of L1 strains, we mapped the geographical regions the strains have been obtained from on the ML-phylogeny and performed



a phylogeographic ancestral state reconstruction analysis (Figure 5). The ancestral origin of L1 strains was predicted to be Southern Asia (89% scaled likelihood), from which the dispersal of strains of particular L1 sub-lineages into other geographical regions likely occurred. For example, our analysis indicates six separate and independent introductions of strains of distinct L1 sub-lineages into Eastern Africa by human migration (Figure 5). From Eastern Africa, strains of particular L1 sub-lineages were then further distributed to Southern Africa (S1, Figure 5). Furthermore, L1 strains were likely introduced from Southern Asia to Western Asia (WA1, Figure 5) and Western Africa (W1, Figure 5) at least once and different sub-lineages twice to South-eastern Asia (SE1-2, Figure 5).

Phenolic glycolipids represent polyketide-derived MTBC virulence factors that require a complex biosynthesis, therefore we screened the WGS data for variants in genes involved in this pathway. Overall, we found 46 variations with 29 non-synonymous mutations also comprising the previously reported Q294Stop in Rv2962c (Krishnan et al., 2011) or a 1 bp insertion in Rv2946c (Supplementary Table 3).

2D ^1H , ^{13}C -Heteronuclear Single Quantum Coherence NMR Lipid Profiling

In the next step of our study, 14 strains of different L1 sub-lineages were investigated to determine the type of synthesized PGLs using 2D ^1H , ^{13}C -heteronuclear single quantum coherence (HSQC) NMR (Figure 6). Strain HN878 (Beijing) was used as a reference for a PGL-tb [(1); Figure 1] producing strain (Mahrous et al., 2008, 2009). The chemical shift region specific for the

aromatic C–H-correlations is depicted for representative strains of each sub-lineage in Figure 6, for all other analyzed strains in Supplementary Figure 4.

The NMR spectrum of strain HN878 (Figure 6) contains the two signals specific for a glycosylated PGL (closed circles), here PGL-tb (1). Consistent with their specific mutation patterns (Supplementary Table 3), all strains of sub-lineages 1.1.1, 1.1.1.1, 1.1.2, 1.1.3 (PGL group 1, red), and 1.2.2.2 (PGL group 4, yellow) according to our updated classification showed signals in their respective NMR lipid profile indicating the presence of PGL-OH [(3); dotted circles] and PGL-OMe [(4); dashed circles]. Strains of sub-lineages 1.2.1.1 and 1.2.1.2 (PGL group 2, green), sharing a 1 bp insertion in Rv2946c, exhibit none of these signals. Very interestingly, some spectra of strains of sub-lineage 1.2.2.1 contain a glycosylated PGL (PGL group 3b, light blue) in addition to PGL-OH and PGL-OMe, which are the molecules present in the spectra of the other members of this sub-lineage (PGL group 3a, dark blue). To the best of our knowledge, these are the first examples of *M. tuberculosis* strains not belonging to L2 producing a glycosylated PGL. However, the shifts of the aromatic C–H-correlations are not sufficient to determine the number of bound glycosyl residues.

Therefore, we isolated PGL-tb (1) and mycoside B (2) from an *Mycobacterium canettii* clinical isolate (3040/99) as reference molecules for the investigation of the PGL from L1 strain 8316/09 (sub-lineage 1.2.2.1). *Mycobacterium canettii* was chosen for PGL-tb (1) and mycoside B (2) isolation, since *M. canettii* is the only member of the MTBC known to produce both molecules at the same time in sufficient amounts. Mycoside B (2), the major PGL present in *M. bovis*, has been shown to be identical to that of *M. canettii* (Daffé et al., 1988). The structure of PGL-tb (1) isolated from *M. canettii* (Daffé et al., 1987) and from *M. tuberculosis* L2 strains (Huet et al., 2009) is almost identical showing equality in the glycosidic part. The reference molecules showed the indicative three [PGL-tb, (1)] or one [mycoside B, (2)] C–H-correlations for anomeric protons in the HSQC spectra, respectively (Figure 7 and Table 1). The spectrum for the PGL isolated from L1 strain 8316/09 showed the presence of mycoside B (2) in this strain (Figure 8) if compared to the reference, suggesting in turn that the SNP in Rv2958c (resulting in Y182D on the amino acid level) might render the glycosyltransferase inactive. The observed correlation between SNP analysis and PGL structures led us to the definition of five “PGL type groups”. Remarkably, sub-lineage 1.2.2.1 comprises PGL type groups 3a and 3b, respectively (Figure 2). The observed PGL types are summarized in association to the crucial genetic variants in PGL biosynthesis associated genes in Table 2.

Growth of L1 Strains in Human Macrophages

Next, we performed infection experiments in human monocyte-derived macrophages to assess the growth behavior of the analyzed L1 strains in the primary host cell of MTBC strains. Cells were infected with a multiplicity of infection (MOI) of 1:1 for 4 h, subsequently washed and analyzed for the cellular bacterial burden on day 7 post infection, as previously described

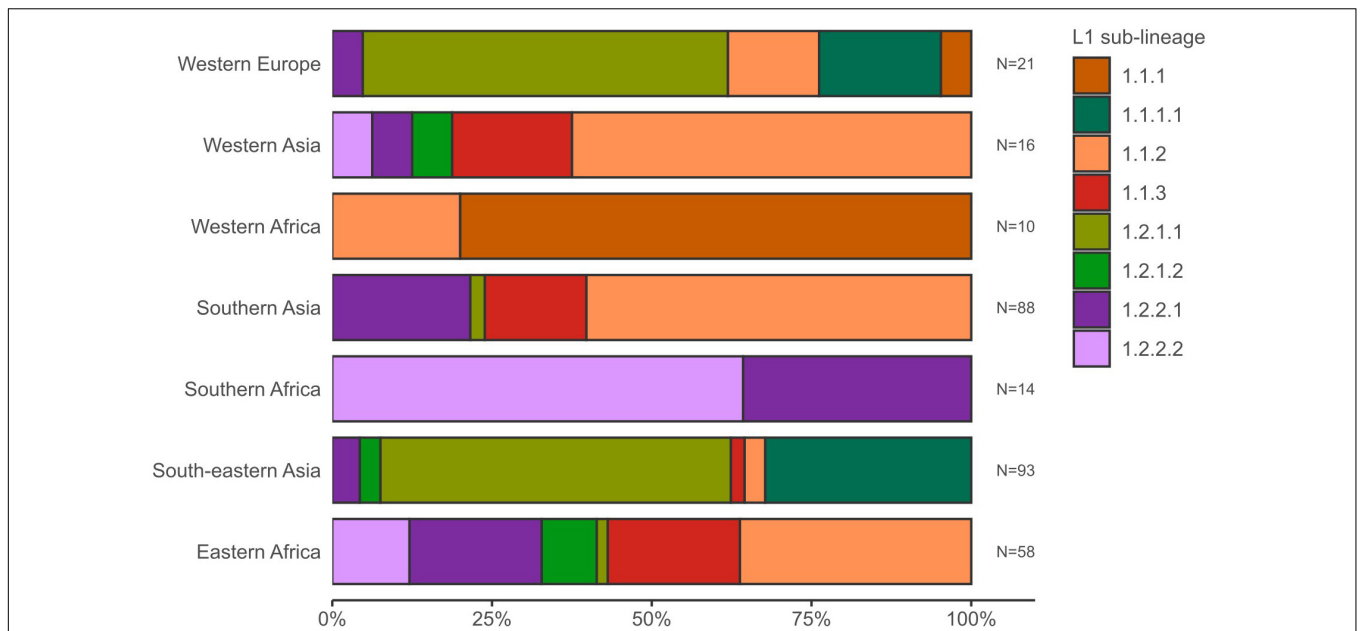


FIGURE 4 | Sub-lineage diversity of L1 MTBC 1 strains within UN-regions. Stacked bars show for each geographic region the sub-lineage presence and proportions of the L1 strain population. UN-regions with a total strain count smaller than 10 are not shown.

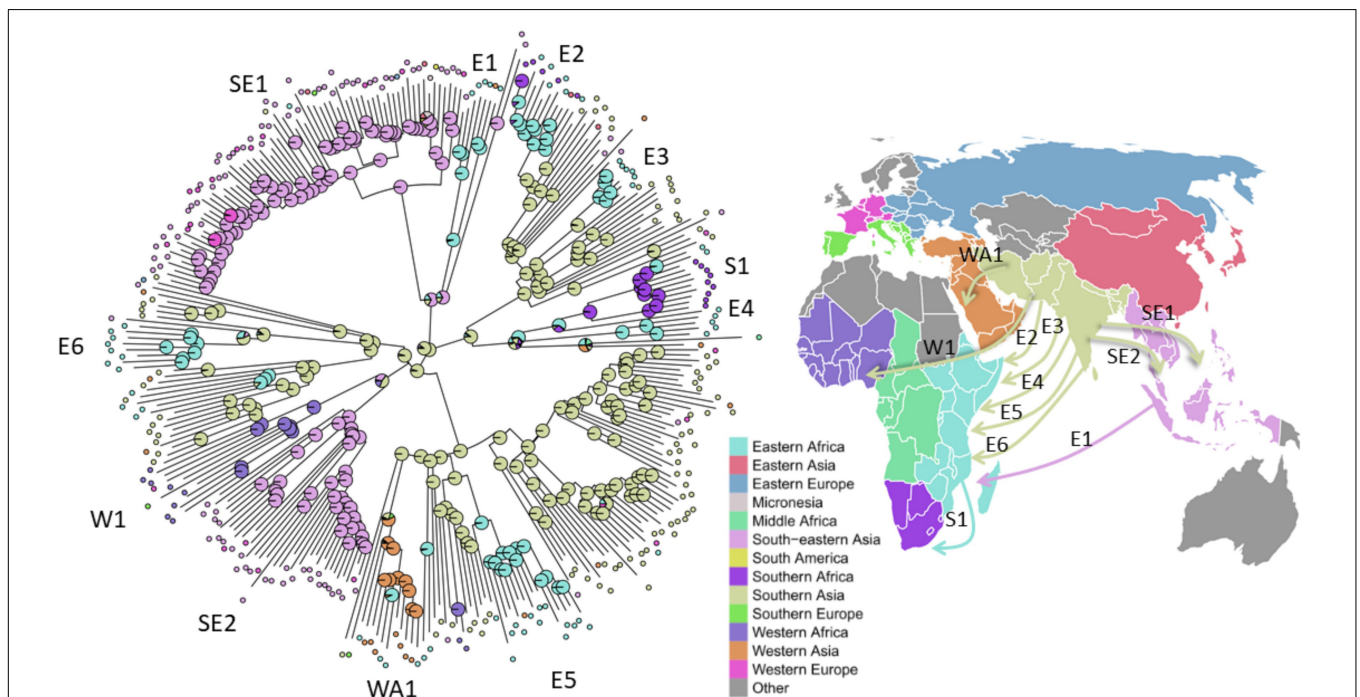
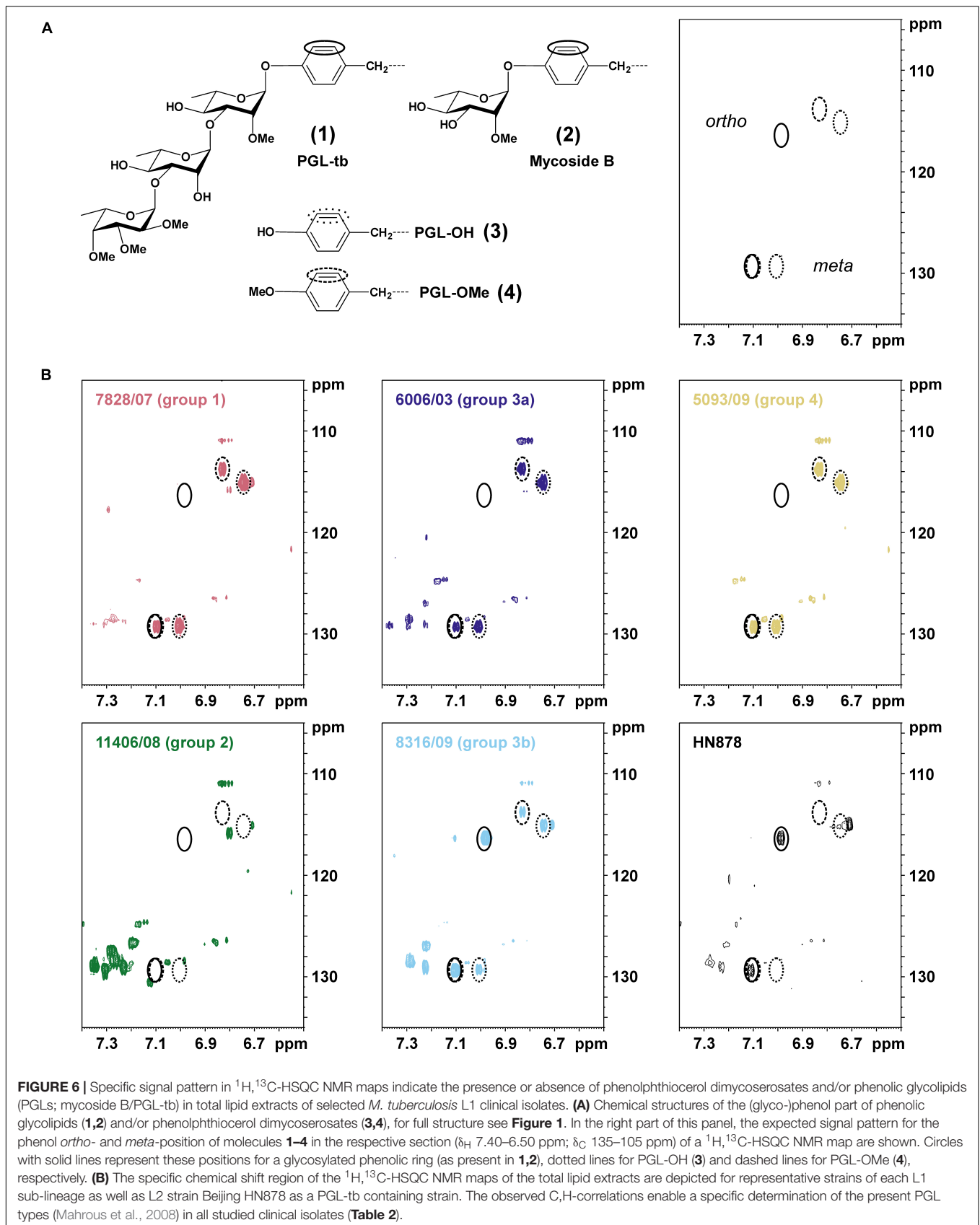
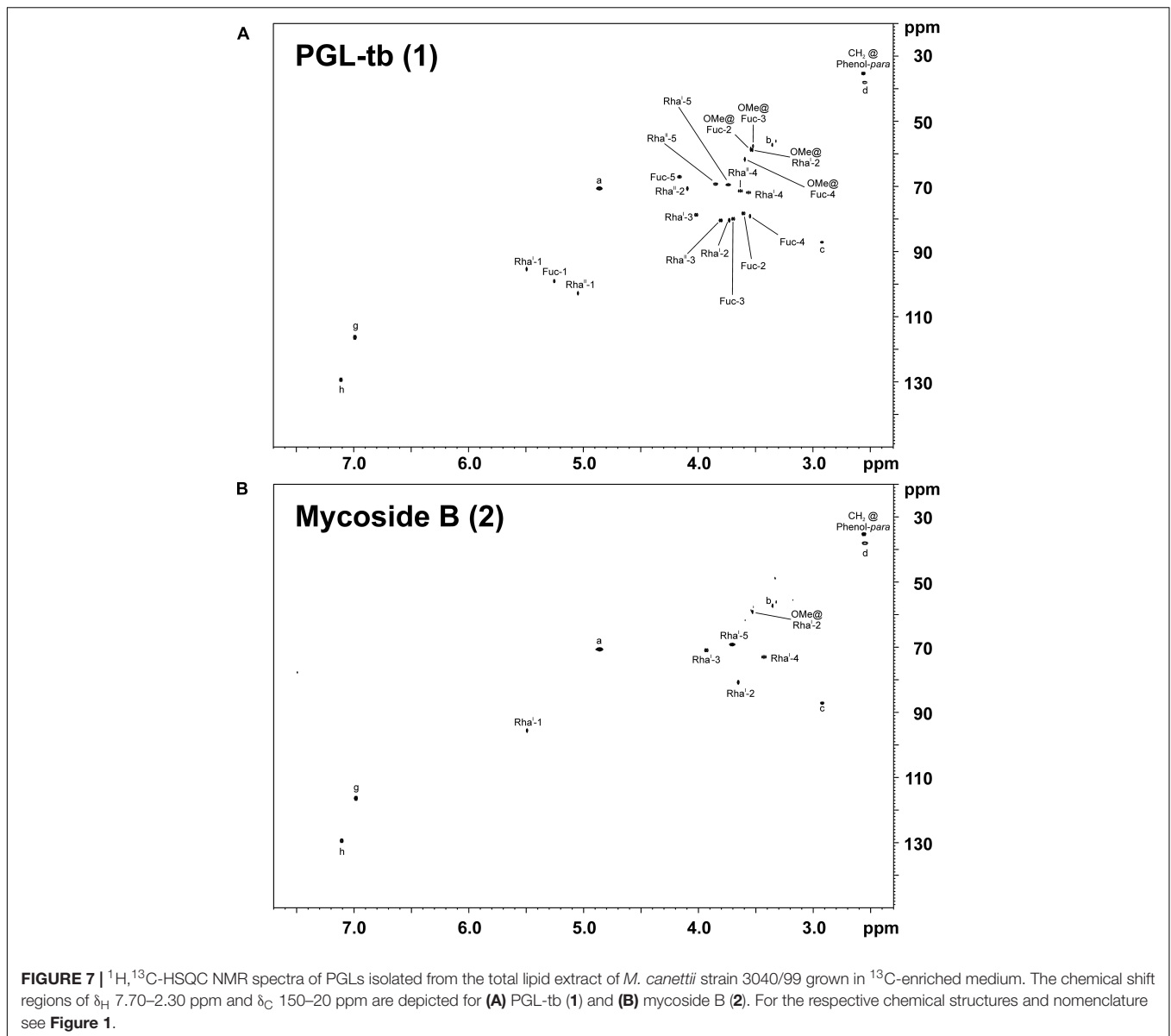


FIGURE 5 | Inference of the L1 strains geographic origin and migration. *Left:* Scaled likelihoods of geographic ancestral states as calculated by joint maximum likelihood estimation mapped onto nodes of the maximum likelihood phylogeny of the 312 L1 strains, colored according to the UN-region (pie charts). Selected migration events of L1 strains to South-eastern Asia (SE), to Eastern Africa (E), to Southern Africa (S), to Western Africa (W), and to Western Asia (WA) are numbered and marked on the tree. *Right:* Directional global movement of the selected multiple migration events of L1 strains.

(Reiling et al., 2013). The L1 strains 1797/03 and 947/01 as well as two other strains belonging to the same genetic group (sub-lineages 1.1.1, 1.1.1.1, 1.1.2, and 1.1.3; PGL group 1,

red), showed a limited growth behavior (reduced to persistent growth). This ranged from a fold increase of 1.5 (1797/03) to 4.7 (7828/07) if the CFU after 7 days was related to bacterial





numbers in the macrophage culture after 4 h (Figure 9A). Similar observations were made for the four strains of sub-lineage 1.2.2.1 (PGL group 3a/b, blue) and the three strains analyzed for sub-lineage 1.2.2.2 (PGL group 4, yellow). While strains of these three groups revealed a comparable and rather similar growth behavior, the L1 strains belonging to sub-lineage 1.2.1.1. (PGL group 2, green), which do not express PGLs and phenolphthiocerol dimycoserates, showed a more diverse pattern. Two isolates (5620/06, 4858/06) matched the growth characteristics of the strains of the other groups, however, five strains showed a significantly increased growth capacity in human macrophages. The fold increase rates ranged from 9.1 (724/11) to 15.8 (11406/08). A comparative analysis of these strains carrying a 1 bp insertion in Rv2946c to all other strains analyzed showed a significantly enhanced average growth rate ($P = 0.0083$; Figure 9B). The differences in macrophage growth

characteristics were not due to reduced or increased uptake of the bacteria after 4 h, since the majority of L1 strains added to the cells in the respective experiment (except 11406/08) were taken up by the macrophages in a comparable range ($6.3 \pm 2.3\%$; mean \pm SD) (Supplementary Figure 5).

DISCUSSION

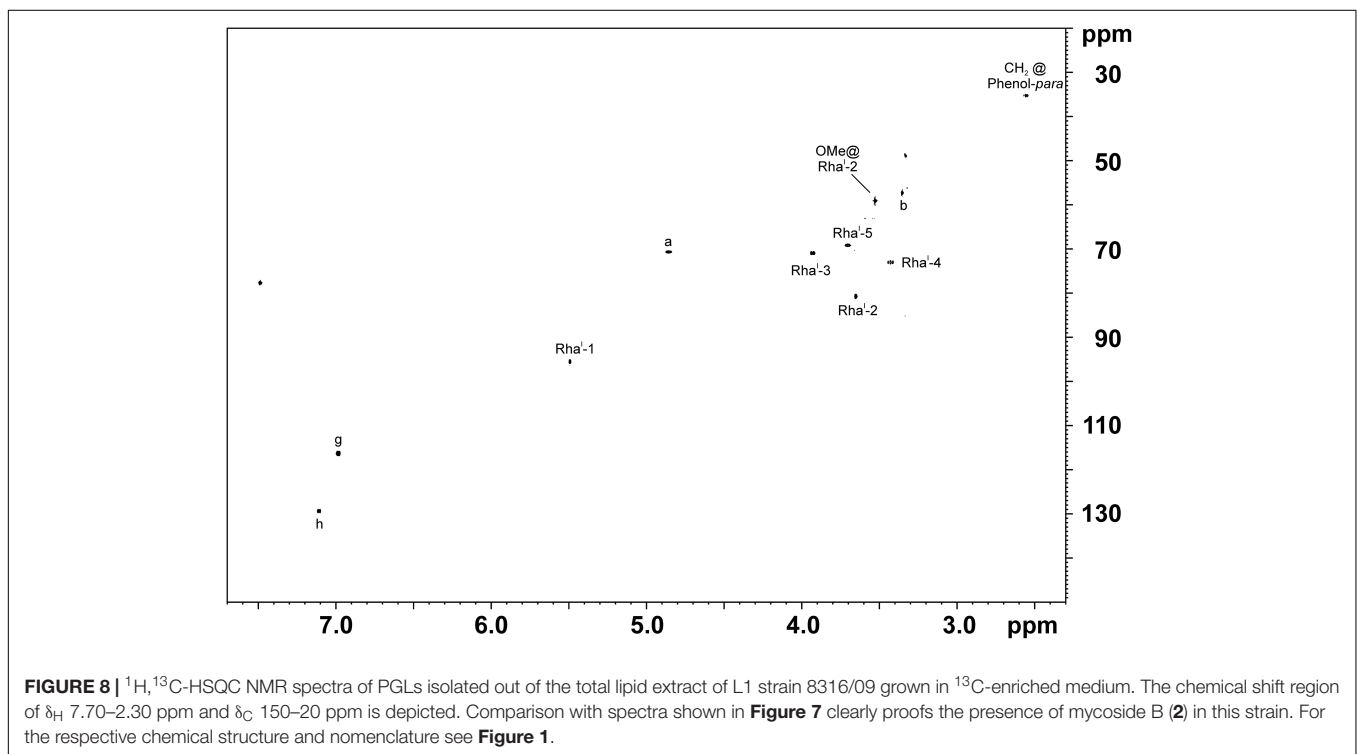
Our study presents a detailed analysis of the phylogeography for L1 strains and correlates the observed genetic diversity with differences in the PGL biosynthesis. These lipids have been shown to be important immunomodulators and their differential production may indeed influence host-pathogen interaction and the outcome of TB infection. Phylogenetic analysis of 312 L1 strains from 43 countries representing 13 UN regions confirmed

TABLE 1 | ^1H (700.4 MHz) and ^{13}C NMR (176.1 MHz) chemical shift data (δ , ppm) [J , Hz] for PGL-tb (**1**) isolated from *Mycobacterium canettii* strain 3040/99, recorded in $\text{CDCl}_3/\text{CD}_3\text{OD}/\text{D}_2\text{O}$ 60:35:8 (v/v/v) at 300 K.

Residue (assignment)	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
2,3,4-tri- <i>O</i> -methyl- α -Fucp-(1→ [Fuc])	5.25 [d, 3.8] 99.2	3.62–3.58* 78.4	3.71–3.67* 80.1	3.56–3.53* 79.3	4.19–4.14* 67.2	1.25 [d, 6.6] 16.3
OCH ₃		3.54 [s] 58.8	3.52 [s] 57.7	3.59 [s] 61.8		
→3)- α -Rhap-(1→ [Rha ^{II}]	5.05 [d, 1.2] 102.9	4.10–4.07* 70.8	3.82–3.78* 80.6	3.66–3.60* 71.5	3.87–3.83* 69.4	1.35 [d, 6.3] 17.8
3)-2- <i>O</i> -methyl- α -Rhap-(1→ [Rha ^I]	5.49 [d, 1.2] 95.4	3.74–3.71* 80.6	4.03–4.00* 78.9	3.58–3.53* 72.0	3.76–3.71* 69.6	1.29–1.25* 17.5
OCH ₃		3.53 [s] 58.9				
	OC	ortho (g[#])	meta (h[#])	para		
→ <i>O</i> -phenol	154.7	6.99 [d, 8.5] 116.4	7.11 [d, 8.5] 129.6	71.2		

Further indicative groups [#]	
a	4.88–4.83* 70.8
b	3.35 [s] 57.5
c	2.94–2.90* 87.3
d	2.57–2.52* 38.1
f	1.17 [d, 7.0] 18.6
CH ₂ (@ <i>para</i> -position of the phenolic ring)	2.58–2.54* 35.4
a – CH ₂ – a	1.82–1.76* 38.7
a – CH ₂ – CH ₂	1.61–1.56*, 1.56–1.48* 34.8

[#]Following the nomenclature of Pérez et al. (2004) (see Figure 1). *Non-resolved multiplet.



a highly diverse phylogeographical population structure that is associated with a specific PGL biosynthesis. Indeed, testing the growth behavior in a macrophage infection model indicates that the different PGL profiles may influence the host-pathogen interaction. Ancestry analysis indicated that L1 strains likely originated from Southern Asia, which is in line with previous studies (Couvin et al., 2019; Menardo et al., 2021). Indeed,

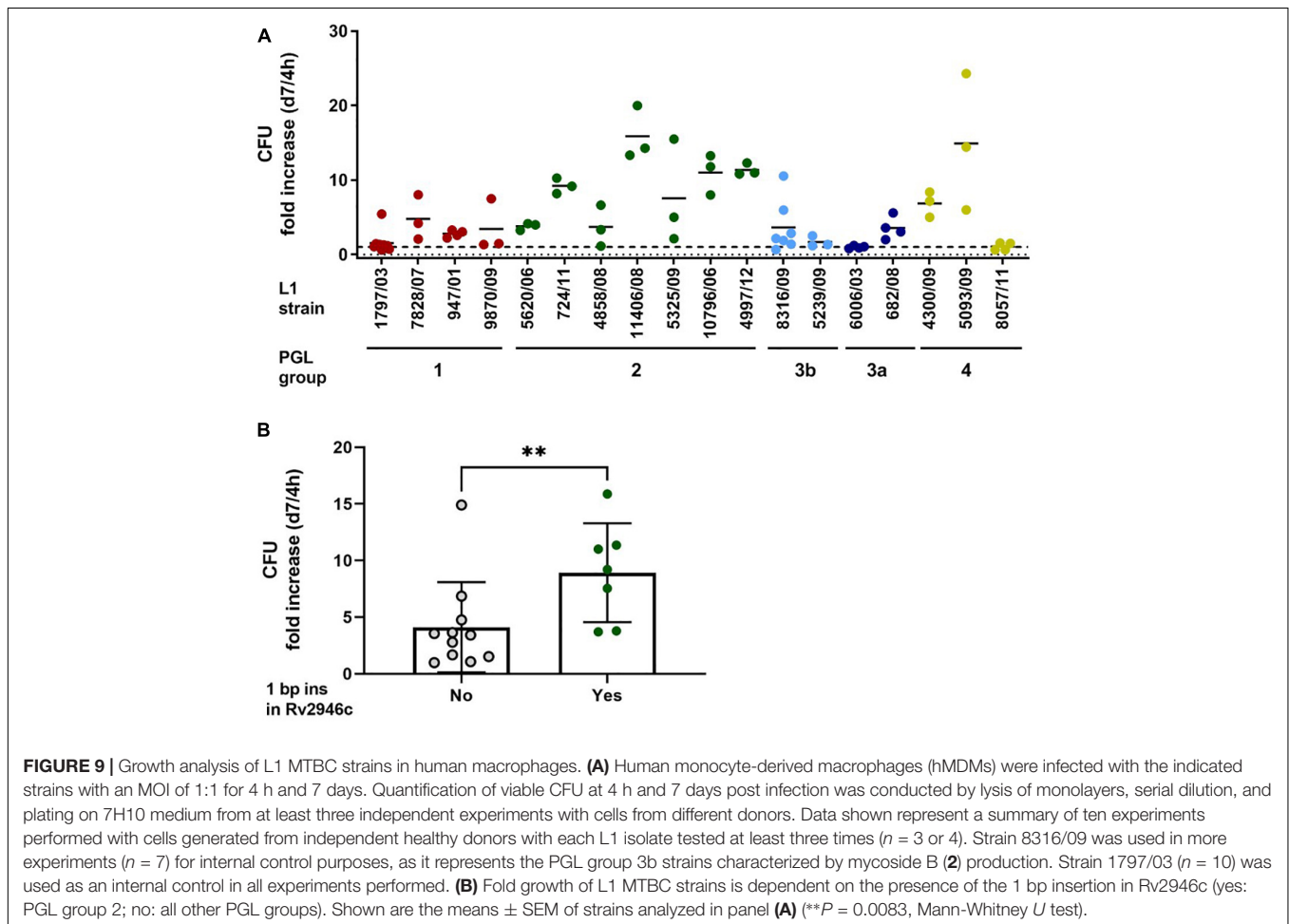
the data obtained here are in accordance with a previous study suggesting an “out of India” scenario for L1 stains with subsequent spread to Eastern Africa and further diversification within Africa (O’Neill et al., 2019).

Based on our phylogenetic analysis, we redefined the L1 sub-lineage classification that can now be linked to distinct geographical occurrence and pathobiology, e.g., variations in

TABLE 2 | Summary of the observed phenolic glycolipids (PGL) types in the 14 selected clinical L1 MTBC strains in correlation to the respective causative variations in PGL biosynthesis genes.

<i>Mtb</i> strain	L1 sub-lineage ^a	PGL group ^a	PGL profile ^b	Genetic variation in PGL biosynthesis associated genes ^c
1797/03	1.1.2	1	PGL-OH, PGL-OMe	Rv2962c Gln294Stop
7828/07	1.1.2	1	PGL-OH, PGL-OMe	Rv2962c Gln294Stop
9870/09	1.1.1.1	1	PGL-OH, PGL-OMe	Rv2962c Gln294Stop
947/01	1.1.3	1	PGL-OH, PGL-OMe	Rv2962c Gln294Stop
4858/08	1.2.1.1	2	No	1 bp insertion in Rv2946c
11406/08	1.2.1.1	2	No	1 bp insertion in Rv2946c
5325/09	1.2.1.1	2	No	1 bp insertion in Rv2946c
8316/09	1.2.2.1	3b	Mycoside B, PGL-OH, PGL-OMe	Rv2958c Tyr182Asp, Rv2962c Leu46Gly
5239/09	1.2.2.1	3b	Mycoside B, PGL-OH, PGL-OMe	Rv2958c Tyr182Asp, Rv2962c Leu46Gly
6006/03	1.2.2.1	3a	PGL-OH, PGL-OMe	Rv2962c Leu46Arg
682/08	1.2.2.1	3a	PGL-OH, PGL-OMe	Rv2962c Leu46Arg
5093/09	1.2.2.2	4	PGL-OH, PGL-OMe	1 bp insertion in Rv2962c
4300/09	1.2.2.2	4	PGL-OH, PGL-OMe	1 bp insertion in Rv2962c
8057/11	1.2.2.2	4	PGL-OH, PGL-OMe	1 bp insertion in Rv2962c

^aAccording to **Figure 2**; ^bFor chemical structures see **Figures 1, 6A**; ^cPosition of amino acid exchange after start codon.



PGL profiles. We provide signature SNPs for the redefined L1 sub-lineages, thus allowing an easy to perform L1 strains sub-lineage classification in future studies. In particular, Coll

sub-lineage 1.2.1 is split in 1.2.1.1, mainly present in South eastern Asia, and 1.2.1.2 mainly occurring in Western Asia and Eastern Africa. Sub-lineage 1.2.2 is split into two sub-lineages

named 1.2.2.1 and 1.2.2.2. Noteworthy, an update to the Coll-nomenclature (Coll et al., 2014) was provided by Napier et al. (2020), recently. However, as in that work sub-lineage names were severely shifted, we here decided to build on the established classification scheme from Coll et al. (2014) for reasons of clarity. Nonetheless, the correlation between sub-lineages defined in our study and both other nomenclature systems (Coll et al., 2014; Napier et al., 2020) is depicted in **Supplementary Figure 2**. Indeed, the presence of multiple sub-lineages within the L1 lineage as shown in our work is in agreement with Napier et al. (2020). Even more, based on the RhierBAPS classification (level 2) performed in this study, a total of 24 L1 sub-lineages are conceivable. These sub-lineages comprise all sub-lineages observed by Napier et al. (2020) and suggest additional ones. Still, we kept the classification in eight sub-lineages as they capture the main subtypes and PGL biosynthesis prototypes.

Interestingly, linking L1 subtypes with geographical occurrence, it became obvious that strains of the two sub-lineages 1.2.2.1 and 1.2.2.2 were identified with the highest per country ratio in Southern Africa, suggesting an adaptation of the pathogen to a specific host population. This is supported by the fact that strains in Southern Africa showed the lowest average pairwise distance of 437 among all regions. However, further data are needed to investigate potential patho-adaptation of strains of particular L1 sub-lineages.

One unique feature of our study is that we could correlate distinct PGL type groups to particular sub-lineages. Lipid compositions differ among mycobacterial species and many of these lipids have been shown or are at least suspected to interfere with the host immune system (Jackson, 2014), among them PGL and PDIM (Forrellad et al., 2013). A first indication of immune modulation or even “super virulence” induction for MTBC strains was obtained by investigation of a subset of clinical strains from the L2 lineage that displayed PGL-tb (**1**; **Figure 1**; Reed et al., 2004). Further studies showed that this lipid is capable of modulating the early host cytokine response (Stadthagen et al., 2006), and contributes to the manipulation of macrophage recruitment (Cambier et al., 2014), but does not confer hypervirulence itself (Sinsimer et al., 2008). Remarkably, all MTBC strains of the globally distributed highly successful L4 do not produce PGLs due to a fixed 7 bp deletion in *pks15/1* compared to L2 strains, leading to a complete termination of phenolphthiocerol dimycoserolate biosynthesis (Constant et al., 2002). Accordingly, different PGL production patterns may rather represent a host-pathogen adaptation mechanism than an overall virulence marker.

With this regard, our data confirm that L1 sub-lineage specific PGL types are highly interesting considering the differences in the geographical spread of strains from different L1 sub-lineages and macrophage growth profiles. Remarkably, all investigated L1 strains have mutations in this particular biosynthesis pathway. This suggests that the alteration of PGLs is potentially important in the evolution of L1 strains. The majority of strains contained PGL-OMe (**3**) and PGL-OH (**4**) due to a stop-codon (Q294Stop; sub-lineages 1.1.1, 1.1.1.1, 1.1.2, and 1.1.3) or a 1 bp insertion in Rv2962c (sub-lineage 1.2.2.2), respectively. Interestingly, strains belonging to sub-lineages 1.2.1.1 and 1.2.1.2 showed a 1 bp

insertion in Rv2946c (*pks1*), leading to a complete abolishment of phenolphthiocerol dimycoserolate biosynthesis. Even more intriguing are the observed Rv2962c SNPs in strains of sub-lineage 1.2.2.1 and the resulting PGL profiles (groups 3a and 3b). While group 3a strains only produce PGL-OMe (**3**) and PGL-OH (**4**), group 3b strains contain mycoside B (**2**) suggesting a restoration of the biosynthetic pathway by the second mutation. This is most likely also influenced by a further group 3b specific SNP in Rv2958c (Tyr182Asp), the next acting glycosyltransferase in PGL biosynthesis (Pérez et al., 2004), which is neither present in group 3a nor in the other groups.

Mycoside B has been described so far in strains of other species of the MTBC like *M. africanum*, *M. pinnepedii*, *M. bovis*, and *M. microti*, but not in *M. tuberculosis* strains. In these species, the occurrence of this PGL type was linked to a frameshift mutation in Rv2958c orthologs (Malaga et al., 2008). Further studies are needed to biochemically validate these assumptions by targeted genetic modifications to substantiate the functional impact of the described SNPs, but a link seems quite likely. In summary, our data give a first insight into PGL diversity and their potential role in host-specific adaptation of L1 strains.

The uptake and the growth characteristics of the analyzed L1 strains in human primary macrophages differs between individual strains. Overall, the investigated L1 strains showed limited to moderate growth in macrophages, which corroborated previous observations and was also mirrored in *in vivo* experiments using aerosol infected mice (Reiling et al., 2013). However, strains belonging to sub-lineages 1.2.1.1 and 1.2.1.2 (PGL group 2) with a 1 bp insertion in Rv2946c (*pks1*), leading to a complete termination of phenolphthiocerol dimycoserolate biosynthesis, which is also a characteristic of L4 strains (Constant et al., 2002), showed in average a more prominent growth in macrophages compared to the average growth of all other analyzed L1 strains ($P = 0.0083$; **Figure 9B**). These data suggest that either full PGL synthesis (PGL-tb), e.g., found in some L2 strains (Reed et al., 2004, Reed et al., 2007), or complete loss of PGL as observed in L4 strains (Constant et al., 2002) and in most L2 strains (Sinsimer et al., 2008) as well as shown here for the newly defined L1 sub-lineage 1.2.1.1 (**Figure 2**) may represent, among other factors, an advantage for MTBC strain replication in the host. Notably, strains that are unable to produce PGL-tb (**1**) due to a frameshift in *pks15/1* are still able to synthesize pHBAD, the glycosidic precursor building blocks for PGL synthesis (Pérez et al., 2004). Since these molecules can suppress the innate immune response of the infected host like the PGL themselves (Stadthagen et al., 2006; Bourke et al., 2014). Thus, their presence is likely favorable for the mycobacterial pathogen. It was shown that deletion of glycosyl transferase Rv2962c leads to the termination of PGL and also pHBAD synthesis (Pérez et al., 2004). Thus, strains with a non-functional glycosyl transferase Rv2962c, comprising all L1 strains investigated here except strains of sub-lineages 1.2.1.1 and 1.2.1.2, are not capable of synthesizing such molecules. Only the population of sub-lineages 1.2.2.1 that belongs to PGL group 3b has at least a partially restored function. Nonetheless, these strains still synthesize the phenolic PDIM PGL-OH (**3**) and PGL-OMe (**4**), which are present in the other members of this

sub-lineage belonging to PGL group 3a as well as sub-lineages 1.1.1, 1.1.1.1, 1.1.2, 1.1.3, and 1.2.2.2. However, the specific impact of PGL-OH (3) and PGL-OMe (4) on MTBC infection is to the best of our knowledge elusive to date.

Taken together, our study emphasizes a high population diversity of L1 strains that correlates to a distinct phylogeographical population structure and associated to distinct sub-lineage specific PGL profiles. Our finding that strains within a given sub-lineage can vary in their synthesized PGL type highlights their importance for host-pathogen coevolution even at the sub-lineage level. In this regard, the current study underscores the importance of genomic diversity of MTBC strains for their pathobiology. Further in-depth analyses, e.g., using cells from donor populations where L1 strains are most abundant, would be of great interest for studies focusing on the function of these molecules. In addition, the use of more elaborated models is urgently needed to decipher the virulence characteristics of selected L1 individual strains to further prove this concept.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories, their accession numbers are listed in **Supplementary Table 1**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the University of Lübeck (14-032). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NG, SN, NR, and SH designed and guided the study. NG, CU, SN, NR, and SH wrote the initial manuscript. NG, CU, and

NR produced the figures. RD, UG, SG, YS, NN, CK, SV, GK, SA, SA-H, PN, AR, MH, and FM collected the mycobacterial isolates. CU, TK, and SH performed the WGS experiments. CU and TK performed the bioinformatics analysis. NG, LG, KD, DH, and US conceived and performed lipid isolations. NG and LG performed and analyzed NMR experiments. NR and SH conceived and supervised the microbiology culturing. NR conceived and supervised macrophage experiments. NG, CU, LG, SM, DS, SN, NR, and SH contributed the ideas and interpreted the results. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the German Research Foundation (DFG, SPP1580 priority program Re 1228/5-1, Re 1228/5-2 to NR; Germanys Excellence Strategy – EXC 2167 Precision Medicine in Inflammation to SN), the Leibniz Science Campus Evolutionary Medicine of the LUNG (EvoLUNG), and the German Center for Infection Research (DZIF). The funders had no role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

We thank all of the collaborators who generously provided their data as well as H. Käßner, T. Niemann, C. Hahn, C. Golin, L. Niwinski, J. Zallet, and V. Mohr for providing excellent technical assistance at the Research Center Borstel, Germany.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.832054/full#supplementary-material>

REFERENCES

- Albanna, A. S., Reed, M. B., Kotar, K. V., Fallow, A., McIntosh, F. A., Behr, M. A., et al. (2011). Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. *PLoS One* 6:e25075. doi: 10.1371/journal.pone.0025075
- Arbues, A., Lugo-Villarino, G., Neyrolles, O., Guilhot, C., and Astarie-Dequeker, C. (2014). Playing hide-and-seek with host macrophages through the use of mycobacterial cell envelope phthiocerol dimycocerosates and phenolic glycolipids. *Front. Cell. Infect. Microbiol.* 4:173. doi: 10.3389/fcimb.2014.00173
- Arbués, A., Malaga, W., Constant, P., Guilhot, C., Prandi, J., and Astarie-Dequeker, C. (2016). Trisaccharides of phenolic glycolipids confer advantages to pathogenic mycobacteria through manipulation of host-cell pattern-recognition receptors. *ACS Chem. Biol.* 11, 2865–2875. doi: 10.1021/acscchembio.6b00568
- Axelrod, S., Oschkinat, H., Enders, J., Schlegel, B., Brinkmann, V., Kaufmann, S. H. E., et al. (2008). Delay of phagosome maturation by a mycobacterial lipid is reversed by nitric oxide. *Cell. Microbiol.* 10, 1530–1545. doi: 10.1111/j.1462-5822.2008.01147.x
- Bah, A., Sanicas, M., Nigou, J., Guilhot, C., Astarie-Dequeker, C., and Vergne, I. (2020). The lipid virulence factors of *Mycobacterium tuberculosis* exert multilayered control over autophagy-related pathways in infected human macrophages. *Cells* 9:666. doi: 10.3390/cells9030666
- Barbier, M., and Wirth, T. (2016). The evolutionary history, demography, and spread of the *Mycobacterium tuberculosis* complex. *Microbiol. Spectr.* 4, 1–21. doi: 10.1128/microbiolspec.TBTB2-0008-2016
- Barnes, D. D., Lundahl, M. L. E., Lavelle, E. C., and Scanlan, E. M. (2017). The emergence of phenolic glycans as virulence factors in *Mycobacterium tuberculosis*. *ACS Chem. Biol.* 12, 1969–1979. doi: 10.1021/acscchembio.7b00394
- Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. doi: 10.1371/journal.pone.0128036
- Blanc, L., Gilleron, M., Prandi, J., Song, O.-r., Jang, M.-S., Gicquel, B., et al. (2017). *Mycobacterium tuberculosis* inhibits human innate immune responses via the production of TLR2 antagonist glycolipids. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11205–11210. doi: 10.1073/pnas.1707840114
- Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., et al. (2020). TbD1 deletion as a driver of the evolutionary success of modern

- epidemic *Mycobacterium tuberculosis* lineages. *Nat. Commun.* 11:684. doi: 10.1038/s41467-020-14508-5
- Bourke, J., Brereton, C. F., Gordon, S. V., Lavelle, E. C., and Scanlan, E. M. (2014). The synthesis and biological evaluation of mycobacterial *p*-hydroxybenzoic acid derivatives (*p*-HBADs). *Org. Biomol. Chem.* 12, 1114–1123. doi: 10.1039/c3ob42277a
- Brites, D., and Gagneux, S. (2017). The nature and evolution of genomic diversity in the *Mycobacterium tuberculosis* complex. *Adv. Exp. Med. Biol.* 1019, 1–26. doi: 10.1007/978-3-319-64371-7_1
- Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., et al. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3684–3689. doi: 10.1073/pnas.052548299
- Cambier, C. J., O'Leary, S. M., O'Sullivan, M. P., Keane, J., and Ramakrishnan, L. (2017). Phenolic glycolipid facilitates mycobacterial escape from microbicidal tissue-resident macrophages. *Immunity* 47, 552–565.e4. doi: 10.1016/j.immuni.2017.08.003
- Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., et al. (2014). Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* 505, 218–222. doi: 10.1038/nature12799
- Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., and Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30, 1224–1228. doi: 10.1093/molbev/mst028
- Coll, F., McEnerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigo, J., Viveiros, M., et al. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* 5:4812. doi: 10.1038/ncomms5812
- Constant, P., Perez, E., Malaga, W., Lanéelle, M.-A., Saurel, O., Daffé, M., et al. (2002). Role of the *pkv15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. evidence that all strains synthesize glycosylated *p*-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pkv15/1* gene. *J. Biol. Chem.* 277, 38148–38158. doi: 10.1074/jbc.M206538200
- Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26, 431–444. doi: 10.1016/j.smim.2014.09.012
- Couvin, D., Reynaud, Y., and Rastogi, N. (2019). Two tales: Worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of *Mycobacterium tuberculosis* underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. *PLoS One* 14:e0219706. doi: 10.1371/journal.pone.0219706
- Daffé, M., Lacave, C., Lanéelle, M.-A., and Lanéelle, G. (1987). Structure of the major triglycosyl phenol-phthiocerol of *Mycobacterium tuberculosis* (strain Canetti). *Eur. J. Biochem.* 167, 155–160. doi: 10.1111/j.1432-1033.1987.tb13317.x
- Daffé, M., Lanéelle, M.-A., Lacave, C., and Lanéelle, G. (1988). Monoglycosyldiacetylphenol-phthiocerol of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Biochim. Biophys. Acta* 958, 443–449.
- Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y García, J., Morbidoni, H. R., de la Paz Santangelo, M., et al. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3–66. doi: 10.4161/viru.22329
- Gutierrez, M. C., Ahmed, N., Willery, E., Narayanan, S., Hasnain, S. E., Chauhan, D. S., et al. (2006). Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg. Infect. Dis.* 12, 1367–1374. doi: 10.3201/eid1209.050017
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Huber, A., Killy, B., Grummel, N., Bodendorfer, B., Paul, S., Wiesmann, V., et al. (2020). Mycobacterial cord factor reprograms the macrophage response to IFN- γ towards enhanced inflammation yet impaired antigen presentation and expression of GBP1. *J. Immunol.* 205, 1580–1592. doi: 10.4049/jimmunol.2000337
- Huet, G., Constant, P., Malaga, W., Lanéelle, M.-A., Kremer, K., van Soolingen, D., et al. (2009). A lipid profile typifies the Beijing strains of *Mycobacterium tuberculosis*: identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J. Biol. Chem.* 284, 27101–27113. doi: 10.1074/jbc.M109.041939
- Ismail, F., Couvin, D., Farakhin, I., Abdul Rahman, Z., Rastogi, N., and Suraiya, S. (2014). Study of *Mycobacterium tuberculosis* complex genotypic diversity in Malaysia reveals a predominance of ancestral East-African-Indian lineage with a Malaysia-specific signature. *PLoS One* 9:e114832. doi: 10.1371/journal.pone.0114832
- Jackson, M. (2014). The mycobacterial cell envelope-lipids. *Cold Spring Harb. Perspect. Med.* 4:a021105. doi: 10.1101/cshperspect.a021105
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jeremiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kohl, T. A., Utpatel, C., Schleusener, V., De Filippo, M. R., Beckert, P., Cirillo, D. M., et al. (2018). MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* 6:e5895. doi: 10.7717/peerj.5895
- Krishnan, N., Malaga, W., Constant, P., Caws, M., Chau, T. T. H., Salmons, J., et al. (2011). *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLoS One* 6:e23870. doi: 10.1371/journal.pone.0023870
- Madacki, J., Mas Fiol, G., and Brosch, R. (2019). Update on the virulence factors of the obligate pathogen *Mycobacterium tuberculosis* and related tuberculosis-causing mycobacteria. *Infect. Genet. Evol.* 72, 67–77. doi: 10.1016/j.meegid.2018.12.013
- Mahrous, E. A., Lee, R. B., and Lee, R. E. (2008). A rapid approach to lipid profiling of mycobacteria using 2D HSQC NMR maps. *J. Lipid Res.* 49, 455–463. doi: 10.1194/jlr.M700440-JLR200
- Mahrous, E. A., Lee, R. B., and Lee, R. E. (2009). Lipid profiling using two-dimensional heteronuclear single quantum coherence NMR. *Methods Mol. Biol.* 579, 89–102. doi: 10.1007/978-1-60761-322-0_5
- Malaga, W., Constant, P., Euphrasie, D., Cataldi, A., Daffé, M., Reyat, J.-M., et al. (2008). Deciphering the genetic bases of the structural diversity of phenolic glycolipids in strains of the *Mycobacterium tuberculosis* complex. *J. Biol. Chem.* 283, 15177–15184. doi: 10.1074/jbc.M710275200
- Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., et al. (2019). Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol* 17, 533–545. doi: 10.1038/s41579-019-0214-5
- Menardo, F., Rutaiwa, L. K., Zwyrer, M., Borrell, S., Comas, I., Conceição, E. C., et al. (2021). Local adaptation in populations of *Mycobacterium tuberculosis* endemic to the Indian Ocean rim. *F1000Res.* 10:60. doi: 10.12688/f1000research.28318.2
- Merker, M., Barbier, M., Cox, H., Rasigade, J.-P., Feuerriegel, S., Kohl, T. A., et al. (2018). Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife* 7:e38200. doi: 10.7554/eLife.38200
- Mishra, M., Adhyapak, P., Dadhich, R., and Kapoor, S. (2019). Dynamic remodeling of the host cell membrane by virulent mycobacterial sulfolipid-1. *Sci. Rep.* 9:12844. doi: 10.1038/s41598-019-49343-2
- Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., et al. (2020). Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med.* 12:114. doi: 10.1186/s13073-020-00817-3
- Nebenzahl-Guimaraes, H., Verhagen, L. M., Borgdorff, M. W., and van Soolingen, D. (2015). Transmission and progression to disease of *Mycobacterium tuberculosis* phylogenetic lineages in the Netherlands. *J. Clin. Microbiol.* 53, 3264–3271. doi: 10.1128/jcm.01370-15
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nguyen, T. K. T., d'Aigle, J., Chinae, L., Niaz, Z., Hunter, R. L., Hwang, S.-A., et al. (2020). Mycobacterial trehalose 6,6'-dimycolate-induced M1-type inflammation. *Am. J. Pathol.* 190, 286–294. doi: 10.1016/j.ajpath.2019.10.006
- O'Neill, M. B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., et al. (2019). Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol. Ecol.* 28, 3241–3256. doi: 10.1111/mec.15120
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

- Patin, E. C., Geffken, A. C., Willcocks, S., Leschczyk, C., Haas, A., Nimmerjahn, F., et al. (2017). Trehalose dimycolate interferes with FcγR-mediated phagosome maturation through Mincle, SHP-1 and FcγRIIB signalling. *PLoS One* 12:e0174973. doi: 10.1371/journal.pone.0174973
- Pérez, E., Constant, P., Lemassu, A., Laval, F., Daffé, M., and Guilhot, C. (2004). Characterization of three glycosyltransferases involved in the biosynthesis of the phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. *J. Biol. Chem.* 279, 42574–42583. doi: 10.1074/jbc.M406246200
- Pirson, C., Engel, R., Jones, G. J., Holder, T., Holst, O., and Vordermeier, H. M. (2015). Highly purified mycobacterial phosphatidylinositol mannosides drive cell-mediated responses and activate NKT cells in cattle. *Clin. Vaccine Immunol.* 22, 178–184. doi: 10.1128/cvi.00638-14
- Quigley, J., Hughitt, V. K., Velikovskiy, C. A., Mariuzza, R. A., El-Sayed, N. M., and Briken, V. (2017). The cell wall lipid PDIM contributes to phagosomal escape and host cell exit of *Mycobacterium tuberculosis*. *mBio* 8:e00148-17. doi: 10.1128/mBio.00148-17
- Reed, M. B., Domenech, P., Manca, C., Su, H., Barczak, A. K., Kreiswirth, B. N., et al. (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 431, 84–87. doi: 10.1038/nature02837
- Reed, M. B., Gagneux, S., DeRiemer, K., Small, P. M., and Barry, C. E. III (2007). The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J. Bacteriol.* 189, 2583–2589. doi: 10.1128/jb.01670-06
- Reiling, N., Homolka, S., Walter, K., Brandenburg, J., Niwinski, L., Ernst, M., et al. (2013). Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *mBio* 4:e00250-13. doi: 10.1128/mBio.00250-13
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x
- Siméone, R., Léger, M., Constant, P., Malaga, W., Marrakchi, H., Daffé, M., et al. (2010). Delineation of the roles of FadD22. *FEBS J.* 277, 2715–2725. doi: 10.1111/j.1742-464X.2010.07688.x
- Sinsimer, D., Huet, G., Manca, C., Tsenova, L., Koo, M.-S., Kurepina, N., et al. (2008). The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not itself confer hypervirulence. *Infect. Immun.* 76, 3027–3036. doi: 10.1128/iai.01663-07
- Stadthagen, G., Jackson, M., Charles, P., Boudou, F., Barilone, N., Huerre, M., et al. (2006). Comparative investigation of the pathogenicity of three *Mycobacterium tuberculosis* mutants defective in the synthesis of *p*-hydroxybenzoic acid derivatives. *Microbes Infect.* 8, 2245–2253. doi: 10.1016/j.micinf.2006.04.008
- Thomas, S. K., Irvatham, C. C., Moni, B. H., Kumar, A., Archana, B. V., Majid, M., et al. (2011). Modern and ancestral genotypes of *Mycobacterium tuberculosis* from Andhra Pradesh. *PLoS One* 6:e27584. doi: 10.1371/journal.pone.0027584
- Tientcheu, L. D., Koch, A., Ndengane, M., Andoseh, G., Kampmann, B., and Wilkinson, R. J. (2017). Immunological consequences of strain variation within the *Mycobacterium tuberculosis* complex. *Eur. J. Immunol.* 47, 432–445. doi: 10.1002/eji.201646562
- Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., et al. (2008). Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog.* 4:e1000160. doi: 10.1371/journal.ppat.1000160

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gisch, Utpatel, Gronbach, Kohl, Schombel, Malm, Dobos, Hesser, Diel, Götsch, Gerdes, Shuaib, Ntinginya, Khosa, Viegas, Kerubo, Ali, Al-Hajoj, Ndung'u, Rachow, Hoelscher, Maurer, Schwudke, Niemann, Reiling and Homolka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.