



# Accurate Identification of Closely Related *Mycobacterium tuberculosis* Complex Species by High Resolution Tandem Mass Spectrometry

## OPEN ACCESS

### Edited by:

Di Xiao,

National Institute for Communicable Disease Control and Prevention (China CDC), China

### Reviewed by:

Motlatso Tiny Hlokwé,

Onderstepoort Veterinary Institute (ARC-SA), South Africa

Keyvan Tadayon,

Razi Vaccine and Serum Research Institute, Iran

### \*Correspondence:

Adam P. Barker  
adam.barker@aruplab.com  
Amol O. Bajaj  
amol.bajaj@aruplab.com

### †Present address:

Suraj Saraswat,  
Pivotal group, Attribute Sciences,  
Amgen, Thousand Oaks, CA, United States

### Specialty section:

This article was submitted to  
Clinical Microbiology,  
a section of the journal  
Frontiers in Cellular and Infection  
Microbiology

Received: 21 January 2021

Accepted: 07 May 2021

Published: 22 June 2021

### Citation:

Bajaj AO, Saraswat S, Knuutila JEA, Freeke J, Stielow JB and Barker AP (2021) Accurate Identification of Closely Related *Mycobacterium tuberculosis* Complex Species by High Resolution Tandem Mass Spectrometry. *Front. Cell. Infect. Microbiol.* 11:656880. doi: 10.3389/fcimb.2021.656880

Amol O. Bajaj<sup>1\*</sup>, Suraj Saraswat<sup>1†</sup>, Juha E. A. Knuutila<sup>2</sup>, Joanna Freeke<sup>3,4</sup>, J. Benjamin Stielow<sup>3,4</sup> and Adam P. Barker<sup>1\*</sup>

<sup>1</sup> Research & Development, Associated Regional and University Pathologists, Inc. (ARUP) Institute for Clinical and Experimental Pathology, Salt Lake City, UT, United States, <sup>2</sup> Research & Development, Thermo Fisher Scientific, Helsinki-Vantaa, Finland, <sup>3</sup> Centre for Infectious Diseases, Radboud University Medical Center (UMC), Nijmegen, Netherlands, <sup>4</sup> Research & Development, Thermo Fisher Scientific, Landsmeer, Netherlands

Rapid and accurate differentiation of *Mycobacterium tuberculosis* complex (MTBC) species from other mycobacterium is essential for appropriate therapeutic management, timely intervention for infection control and initiation of appropriate health care measures. However, routine clinical characterization methods for *Mycobacterium tuberculosis* (*Mtb*) species remain both, time consuming and labor intensive. In the present study, an innovative liquid Chromatography-Mass Spectrometry method for the identification of clinically most relevant *Mycobacterium tuberculosis* complex species is tested using a model set of mycobacterium strains. The methodology is based on protein profiling of *Mycobacterium tuberculosis* complex isolates, which are used as markers of differentiation. To test the resolving power, speed, and accuracy of the method, four ATCC type strains and 37 recent clinical isolates of closely related species were analyzed using this new approach. Using different deconvolution algorithms, we detected hundreds of individual protein masses, with a subpopulation of these functioning as species-specific markers. This assay identified 216, 260, 222, and 201 proteoforms for *M. tuberculosis* ATCC 27294<sup>TM</sup>, *M. microti* ATCC 19422<sup>TM</sup>, *M. africanum* ATCC 25420<sup>TM</sup>, and *M. bovis* ATCC 19210<sup>TM</sup> respectively. All clinical strains were identified to the correct species with a mean of 95% accuracy. Our study successfully demonstrates applicability of this novel mass spectrometric approach to identify clinically relevant *Mycobacterium tuberculosis* complex species that are very closely related and difficult to differentiate with currently existing methods. Here, we present the first proof-of-principle study employing a fast mass spectrometry-based method to identify the clinically most prevalent species within the *Mycobacterium tuberculosis* species complex.

**Keywords:** mass spectrometry, species delimitation, clinical mycobacteriology, *Mycobacterium tuberculosis*, clinical diagnostics

## INTRODUCTION

*Mycobacterium tuberculosis* is responsible for one of the most devastating and chronic infectious diseases known to science and is an important and formidable human pathogen that has claimed nearly 1.2 million lives in 2019 (Andersen and Brennan, 1994). Approximately 10 million people were infected in 2019 and the pathogen can persist in a hidden form if undiagnosed for a long period of time within the human host (Bloom and Murray, 1992; WHO, 2020). Almost 10% of an infected population will progress to disease (active tuberculosis) following a latent period (from weeks to decades) (Bloom and Murray, 1992). Multidrug-resistant tuberculosis has recently emerged (MDR-TB) and represents an enormous challenge to the public health system worldwide. Worldwide in 2019, close to half a million people developed rifampicin-resistant TB (RR-TB), of which 78% had multidrug-resistant TB (MDR-TB). The three countries with the largest share of the global burden were India (27%), China (14%) and the Russian Federation (8%) (WHO, 2020). The largest increase seen globally in the number of nontuberculosis mycobacterial (NTM) and *MTBC* infections was observed in patients who had already contracted the human immunodeficiency virus (HIV).

*Mycobacterium tuberculosis* complex encompasses a group of organisms of paramount clinical relevance that cause tuberculosis (TB) in humans and animals. TB in humans is caused mainly by *M. tuberculosis* comprising an enormous diversity of genetic lineages (e.g. including its formerly known *M. africanum*) but also by other members of *MTBC*. Zoonotic TB is caused by its less well understood animal-associated subspecies *M. bovis*, *M. canetti*, *M. caprae*, *M. pinnipedii*, *M. suricattae*, *M. orygis*, *M. microti*, and *M. mungi* (Gagneux and Small, 2007; Alexander et al., 2010; Gupta et al., 2018; Riojas et al., 2018; Lipworth et al., 2019). However, not all zoonotic *MTBC* species are nomenclatural validly published and *MTBC* taxonomy has been under constant scientific revision, with future changes being expected. Despite their genetically close relationship, they differ in epidemiology, pathogenicity, geographical range, host preference, and severity of tuberculosis disease in humans. Genetically, all members of this complex are highly conserved, possessing 99.9% similarity at the nucleotide level and identical 16S rRNA sequences (Peters et al., 2020). *M. tuberculosis* species, the most common pathogens in humans, can be further divided into genetic groups that also show differences in their levels of virulence, immunogenicity, and geographical distributions (Gagneux and Small, 2007). *M. bovis* BCG is one of the most common *MTBC* organisms found in clinical laboratory and is used to treat bladder cancer (DeGeorge et al., 2017; Larsen et al., 2020). *M. bovis* BCG is one of the most commonly administered vaccines and its complications include disseminated BCG disease which is rare but increasingly reported in immunodeficient children (Hesseling et al., 2007; Ritz et al., 2009). It is clinically important to differentiate *MTBC* members from other mycobacterium species considering the differences in resistance and treatment response rate to antibiotics (Scorpio and Zhang, 1996).

The ability to distinguish between strict human and zoonotic tuberculosis and to trace source exposure during epidemiological studies is critical for the infection control process (Djelouadji et al., 2008).

Characterization of the proteome of *MTBC* is essential for differentiation at the species level with the proposed approach. It has been shown that during infection some expressed proteins vary during different stages of infection while others are present during the entire infection (Mehaffy et al., 2012). Some of the *Mtb* secreted proteins are essential for pathogenesis of the disease and can be potential targets for vaccine development (Brock et al., 2001). For example *MtbHU* (a small dimeric nucleoid-associated protein), *HspX/14-kDa* ( $\alpha$ -crystalline homologue), the molecular chaperone *GroEL2*, and bacterioferritin (*BfrB*) are essential for growth and pathogenesis of *Mtb* and can serve as target for treatment (Bhowmick et al., 2014). Any delay in correct diagnosis and treatment may lead to sepsis and emergence of resistance to available drugs, this can be one of the driving forces for the search of new drugs that will take years or decades to develop. There are many known mechanisms of resistance such as possible mutations of the *pncA* and *rpoB* gene, hydrophobic and waxy cell wall composition, slow growth rate while dormant, capacity to suppress the host immune response, and an efficient efflux pump system (Scorpio and Zhang, 1996; Somoskovi et al., 2001; de Steenwinkel et al., 2010; Louw et al., 2011). Outside of these known resistance pathways there may be many unknown resistance mechanisms which need to be explored as well. Therefore, a rapid and accurate differentiation of closely related pathogens is crucial. However, most of the recent methods that are based on phenotypes, biochemical characteristics, nucleic acids, and molecular systems (Somoskovi et al., 2008) for *MTBC* species differentiation and identification have suboptimal diagnostic performance. In addition, next generation sequencing is laborious to perform (Hughes et al., 2012). Even MALDI-TOF, which has revolutionized the clinical microbiology identification of mycobacterium (Mather et al., 2014), could not discriminate *MTBC* species (particularly as IVD cleared test) (Chen et al., 2013; Lévesque et al., 2015).

In addition to the need of a more discerning assay, it is also essential to have a pre-analytical method to render the microbes non-viable. A unique incubation with pre-incubation solution (Thermo Fisher Scientific, proprietary) followed by rigorous mixing in buffered solution method was developed for inactivation of *MTBC* to overcome shortcomings of other inactivation methods such as heat treatment (Doig et al., 2002), exposure to lethal irradiation (Murdoch et al., 2012), multiple centrifugation and washing steps etc. In this study, we utilize micro (nano) flow liquid chromatography-mass spectrometry (LC-MS) to separate proteins from microbial extracts and analyze them in a high-resolution accurate mass (HRAM) orbitrap mass spectrometer. We used a workflow consisting of cell sonication, solid-phase extraction (SPE) purification, mass spectrometry, and computational algorithms to achieve our high identification accuracy for these organisms. The goal of the present study was to provide a proof-of-principle experiment employing

orbitrap LC-MS for discrimination of mycobacterium species. This is the first study in which differentiation of the *MTBC* at a species level is achieved by employing a mass spectrometry driven approach with the potential to discriminate all types of closely related *Mtb* pathogens.

## MATERIALS AND METHODS

### Strain Selection and Reference Identifications

*Mycobacterium* strains belonging to the species *M. africanum* (*syn. M. tuberculosis*; Genetic lineage), *M. tuberculosis*, *M. bovis*, *M. bovis* BCG and *M. microti* being of paramount clinical relevance were selected as representatives for the *Mycobacterium tuberculosis* complex). Although *M. canetti*, *M. caprae*, *M. orygis*, *M. suricattae*, *M. pinnipedii*, and *M. mungi* also form the *MTBC*, we excluded them in this study due occurrence as causative agents of rare zoonotic human TB infections and their partially unclarified taxonomic status. Key objective to this study was providing an initial proof-of-principle towards the diagnostic potential of our novel LC-MS assay only. Taxonomic anchors were defined by ATCC type material and selected for the subsequent analysis (*M. tuberculosis* ATCC 27294<sup>TM</sup>, *M. microti* ATCC 19422<sup>TM</sup>, *M. africanum* ATCC 25420<sup>TM</sup>, and *M. bovis* ATCC 19210<sup>TM</sup> (Note: This type strain appears to be *M. bovis* BCG by ARUP internal PCR based validation). In addition, 37 recent US clinical isolates obtained by Associated Regional and University Pathologists (ARUP) laboratories (7 clinical isolate strains of *M. africanum*, 10 *M. tuberculosis*, 10 *M. bovis* and 10 *M. bovis* BCG) were analyzed in order to assess assay robustness. ATCC type material was obtained pre-identified in accordance with ATCC identifications procedures; all other recent clinical isolates were identified by quantitative (real-time) PCR and meltcurve analysis with a multiplex probe set to obtain a reference identification (Pounder et al., 2010) prior to LC-MS data acquisition. All strains of recent clinical origin were collected throughout the year 2019 at ARUP laboratories, with no additional metadata being available (double blinded clinical identifiers).

### Chemicals and Reagents

Optima LC-MS grade Water (H<sub>2</sub>O), Optima LC-MS grade acetonitrile (ACN), and Optima LC-MS grade formic acid (FA) were purchased from Fisher Scientific (Fair Lawn, NJ).

### Sample Preparation

#### Preparation of Whole Cell Extracts (WCE) for *M. tuberculosis* Complex Species

ATCC strains and clinical isolates of *M. tuberculosis* complex were cultured on 7H11 agar plates for approximately 18-24 days in a CO<sub>2</sub> incubator maintained at 37°C in a BSL3 laboratory. A 1 µL loopful of colonies was transferred into sonication vials (Thermo Fisher Scientific, proprietary). Subsequently the pre-incubation solution containing alcohol (Thermo Fisher Scientific, proprietary) was added to the cells at room temperature (RT, 20-25°C) and following a short centrifugation step (12,000 x g for 2

minutes at RT) the supernatant was discarded and then the pellet was suspended into 100 µL of incubation solution containing formic acid and acetonitrile (Thermo Fisher Scientific, proprietary). The cell lysate were incubated for 20 minutes (vortexed once at the 10-minute mark for 2 seconds), followed by sonication for one minute at 50% amplitude. Cells were diluted with 100 µL dilution buffer containing acetonitrile (Thermo Fisher Scientific, proprietary) and centrifuged for 5 minutes at 12,000 × g at RT. The supernatant was collected in low protein binding (LBE) Eppendorf tubes and stored at -80°C if not immediately used for LC-MS analysis.

#### RP4H SPE Protocol

The whole cell extract (WCE) was diluted with SPE buffer-1 (Thermo Fisher Scientific, proprietary) before Solid-Phase Extraction (SPE) cleanup. Reverse Phase monolith (RP4H) SPE tips (Thermo Fisher Scientific, proprietary) were placed in 96 well-plate and conditioned and equilibrated with 50 µL proprietary SPE buffer-2 (Thermo Fisher Scientific, proprietary), respectively. A centrifuge (Laboratory centrifuge 4-15C, Thermo Fisher, Osterode, Germany) capable of handling two 96 well-plates was used for centrifugation at 2000 × g for 2 minutes at RT. Diluted WCE (50 µL) was loaded into the SPE tips and centrifuged. The tips were washed with 50 µL of proprietary SPE buffer-3 and placed in liquid chromatography (modified Easy-1000, Thermo Fisher Scientific, proprietary) autosampler plate for further online elution and mass spectrometric analysis. Blanks were treated as negative controls and were prepared in a similar way as described above, instead of cell lysate water was added to the SPE tips.

#### LC-MS Analysis

All experiments were performed on a micro flow Liquid chromatography-Mass spectrometry (LC-MS) system. The liquid chromatography (LC) system was connected to a Q Exactive<sup>TM</sup> HF Orbitrap<sup>TM</sup> mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Mobile phase A consisted of 0.2% FA, 10% ACN in H<sub>2</sub>O and mobile phase B consisted of 0.2% FA in ACN. Elution was done using gradient elution where after flow stabilization at 4 µL/min for 2%B, percent of B was increased to 33% B in 5 min. The LC and mass spectrometer were controlled by Xcalibur software version 3.0 (Thermo Fisher Scientific). The mass spectrometer was operated in positive electrospray ionization (+ ESI) mode. The general mass spectrometric conditions were: The spray voltage was set at 2.0 kV. The capillary temperature was 325°C and S-lens RF level was set to 65. The maximum injection time was 200 millisecond and 40 microscans were used. Automatic Gain Control (AGC) target was 3e6, resolution was 120,000 for MS. Intact protein MS mode was used with trapping gas pressure set to 0.2 and C-trap charge detection was set OFF. Mass analysis of proteins was performed in the range from 450 to 2000 *m/z* for full scan mode.

#### Data Processing and Species Identification

The raw data was deconvoluted (proprietary algorithm) in real time to monoisotopic masses (i.e. list of proteoforms).

Subsequently, the data was processed with in-house algorithms for building of the database and identification of mycobacterium species. The individual data acquisition and processing steps are depicted in **Figure 1**. Each strain in the analysis was measured in two technical replicates for database construction and subsequent species prediction. Each ATCC strain in the analysis was also analyzed in six replicates (two technical replicates from three biological replicates) for the estimation of technical reproducibility of the sample processing and measurement.

Processing of the mass spectra was performed by Thermo Fisher proprietary software to deconvolute spectra in  $m/z$  space into monoisotopic protein masses between 5 and 40 kDa. Measured monoisotopic masses from each measurement were mass aligned to construct a list of consensus markers. One-way ANOVA was used to select top 200 consensus markers that have the most predictive value for the differentiation of the species. Species prediction was accomplished by a Thermo Fisher Scientific proprietary distance-based clustering algorithm. The robustness of species prediction was simulated by leave-one-out iterations where each sample was tested against a database constructed without data for the same strain (training and test dataset). The data will be made available on reasonable request.

## RESULTS

### Inference and Determination of LC-MS Proteome Patterns

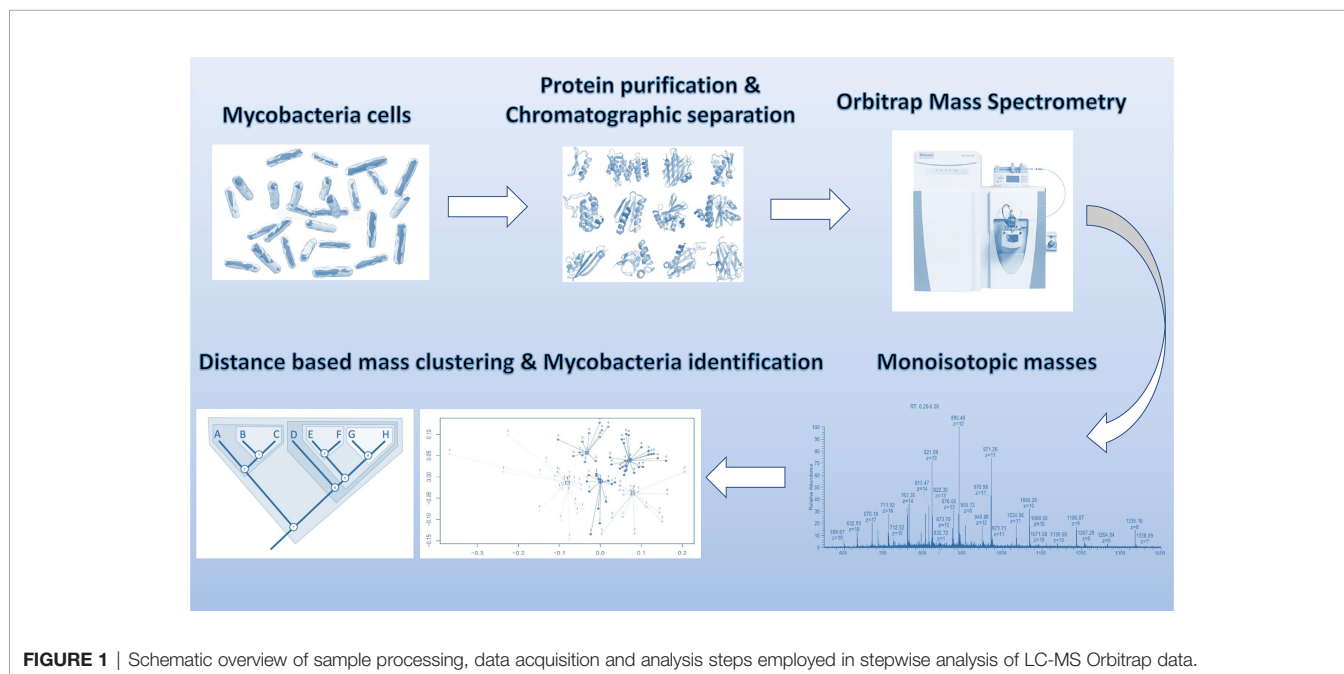
To verify whether different *MTBC* species show unique MS spectrum profiles, the masses of the deconvoluted proteoforms from each species were aligned to construct a list of consensus markers (clustering of monoisotopic masses). Subsequently, we could depict common (shared) proteoforms

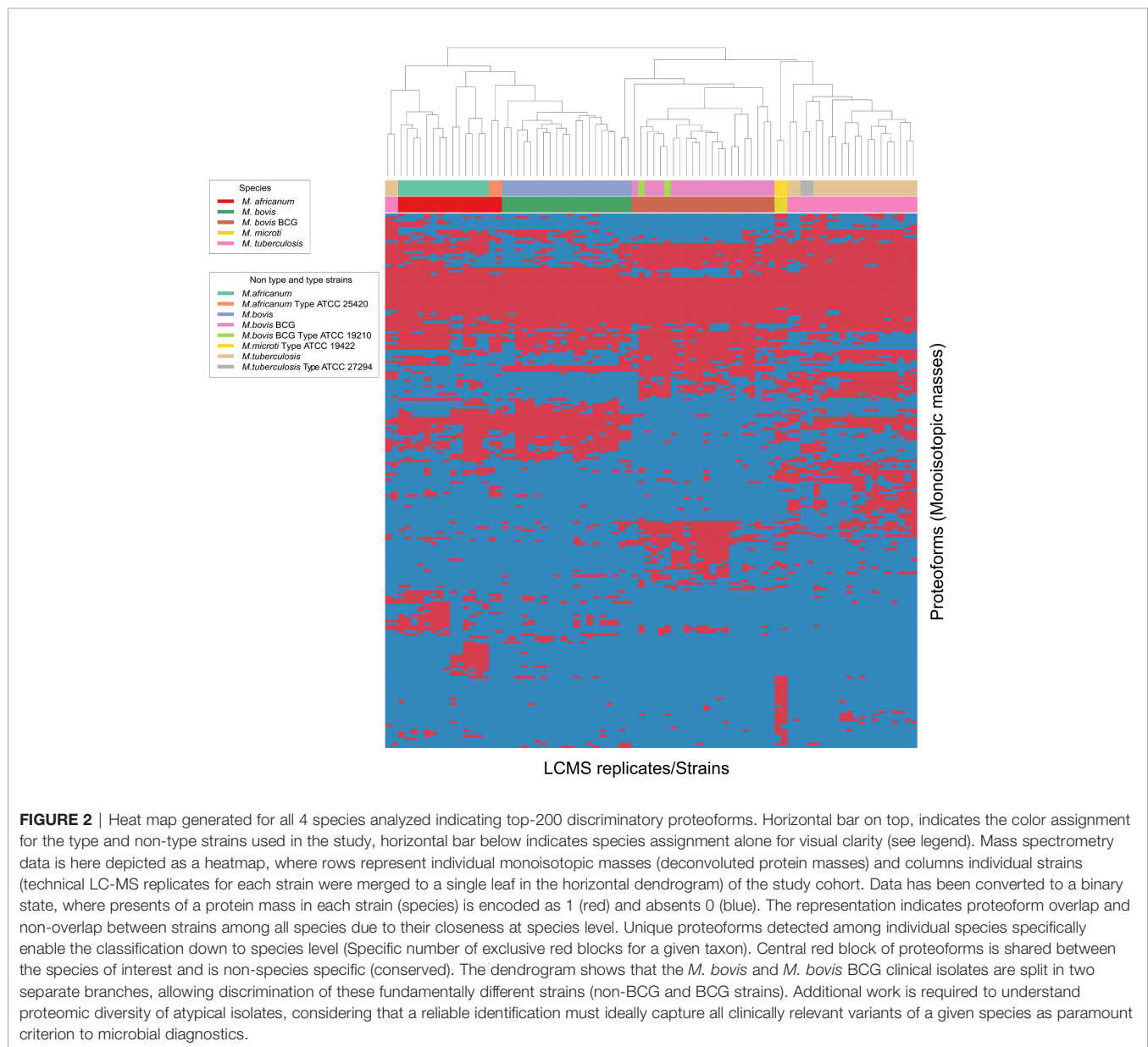
from deconvoluted spectrum profiles in order to determine visually common and non-common proteoforms (**Figure 2**). Unique masses per species were those masses present in all or most replicates of a given species and predominantly absent from replicates from other species. To verify whether different *MTBC* strains show similar mass spectral profiles, we analyzed 37 clinical isolates of *MTBC* with the depicted LC-MS analysis workflow (**Figure 1**). Consistent with the mass spectrum profiles obtained from the ATCC type strains, all clinical isolates exhibited a similar pattern (data not shown), which comprises common proteoforms ranging from 5 to 40 kDa. Accordingly, similar elution profiles and identified proteins, result in consistent species-specific proteome patterns.

### Differentiation of *MTBC* Species

We next evaluated whether the MS spectrum profile and common proteoforms obtained from different species within *MTBC* can differentiate the species within the complex. Based on the different expression of species-specific protein profiles (unique masses), we developed species-specific databases capable of rapidly differentiating *MTBC* at species level with very high identification accuracy.

As represented by the clustered monoisotopic masses which were visually depicted as heatmap in **Figure 2**, the unique protein signatures are successfully able to differentiate *M. bovis* and *M. bovis* BCG from other taxa of the *MTBC* in individual distinct monophyletic clusters. Similarly, *M. tuberculosis* and its genetic lineage syn. *M. africanum* as well *M. microti* are monophyletically differentiated based on unique protein profiles. The results of the mass clustering were in accordance with the alternative identification obtained by qPCR meltcurve analysis as described above under material and methods. However, a single *M. tuberculosis* isolates was falsely clustered





among the *M. africanum* (sensu stricto) isolates. This misplaced isolate is obvious at the very left of the heatmap-depicted clustered monoisotopic masses (Figure 2). This result disagrees with our reference-based identification and can likely be explained by proteoforms which might be rarely expressed by atypical *M. tuberculosis* strains. This inconsistency will require further investigation with additional *M. tuberculosis* and *M. africanum* strains. Subsequently, the classification results returned from the training and test data definitions returned very high success rates for the clinically important species *M. tuberculosis* and *M. bovis* BCG, while *M. africanum* and *M. bovis* returned slightly inferior results (100% correct identification vs. 90% and 88% respectively). Reciprocally, identification success is lowered for *M. africanum*, by false hits against *M. bovis* but not *M. tuberculosis* and for *M. bovis* sensu

stricto against *M. bovis* BCG. The results of the species prediction are given in Table 1. *M. microti* was excluded from the analysis due to representation of the type strain only (eg. training and test dataset as described under material and methods accordingly would be identical).

Based on the different expression of species-specific protein patterns, we were capable to differentiate the clinically most prevailing species within the *Mycobacterium tuberculosis* complex down to species level. Differential proteoforms may serve as potential biomarkers for identifying these microorganisms in patient samples in additional confirmatory studies using clinical specimens. These results outline the potential to the employed LC-MS technology, as a first-line platform in the rapid and accurate identification of MTBC species in routine clinical microbiological laboratories.

**TABLE 1** | Classification success for identification of clinically prevalent *MTBC* species.

Species	N	Identification success (Top match only)	Combined top 3 matches for each data file
<i>M. africanum</i>	16	88%	41 hits to <i>M. africanum</i> , 7 hits to <i>M. bovis</i>
<i>M. bovis</i>	20	90%	56 hits to <i>M. bovis</i> 4 hits to <i>M. bovis</i> BCG
<i>M. tuberculosis</i>	22	100%	64 hits to <i>M. tuberculosis</i> 2 hits to <i>M. africanum</i>
<i>M. bovis</i> BCG	22	100%	66 hits to <i>M. bovis</i> BCG
<b>Total</b>	<b>80</b>	<b>95%</b>	

*N* refers to the total number of mass spectra measured for each species. Total identification success is presented as 'percent success' as a result to predict the target species with the employed algorithm. For each data file, the 3 closest matches were collected (additional and including the top match) if applicable, and the species results for these top 3 matches is displayed in the right-hand column. In all cases the best match was to another strain of the same species, but in some few cases the 2<sup>nd</sup> or 3<sup>rd</sup> best match was for a different species as detailed in the column on the right. In the analysis we used only the type-strain of the *M. microti*, accordingly no species predictions are returned in this case as training and test data would be identical.

## DISCUSSION

In this study, we evaluated the resolution power of LC-MS as a novel method to delineate clinically relevant closely related *MTBC* species as a model system for challenging identifications. Rapid and accurate identification of microbial infections is essential for accurate patient supervision. Especially for *Mtb* infection, differentiation of *MTBC* species is very critical for the accurate disease diagnosis, starting drug therapy, to control the spread of disease, for public health surveillance and appropriate patient case management. *MTBC* infections have been one of the most important and common mycobacterial clinical pathogens, especially in immune deficient adults and pediatric patients. Among *MTBC* species, *Mycobacterium tuberculosis* is more predominant in clinical samples and environmental samples like water and soil which contributes almost exclusively to disseminated *MTBC* disease.

Acid-fast staining and microscopy as a routine technique was the most common and least expensive option for laboratory diagnosis of human tuberculosis. However, this technology has a low sensitivity and specificity and does not provide information on the identity of the pathogen. Comparatively rapid nucleic acid amplification test (NAAT) (Derese et al., 2012) and two-step PCR processes have been used for differentiation of *MTBC* species (Pinsky and Banaei, 2008; Pounder et al., 2010; Reddington et al., 2012; Nikam et al., 2013). However, these techniques fail to differentiate *MTBC* species, is more laborious and complexity of multiplex RT-PCR can be a problem. Routine laboratory culture, chemical testing procedures, and nested PCR targeting (Sinha et al., 2016) assays can differentiate *MTBC* members at the species level but the long identification process is not practical for rapid diagnosis of patients in need of immediate therapy. Alternatively, molecular methods such as GenoType MTBDRplus test (Hain Lifescience, Nehren, Germany) (Richter et al., 2004) or DNA probe assays (Kasai et al., 2000; Niemann et al., 2000) have been able to identify mycobacterial isolates as *MTBC* species, but lack capability to identify the isolate to the species level and limitation of its reliability on bacterial cultures to produce enough amount of bacterial DNA. Differentiation of *MTBC* species is possible based on intergenic spacers analysis in the genotyping of

*M. tuberculosis* such as the exact tandem repeat D (ETR-D) (Frothingham and Meeker-O'Connell, 1998) and aliased mycobacterial interspersed repeat unit 04 (MIRU04) (Supply et al., 1997). And also, based on molecular methods such as the major polymorphism of tandem repeat (MPTR) sequencing (Frothingham, 1995), single nucleotide polymorphisms (SNP) in the *pncA* gene (Djelouadi et al., 2008), the *oxyR* locus (Sreevatsan et al., 1996), the detection of deleted regions (Somoskovi et al., 2008), the restriction fragment length polymorphism of the *hupB* gene (Prabhakar et al., 2004), *gyrB* gene based differentiation (Chimara et al., 2004; Arnold et al., 2005), and mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping (Godreuil et al., 2007). Previous methods could differentiate the *MTBC* at the species level but are often complicated by low sequence variability at the nucleotide level. However, these methods may have some limitations when encountered with rare species or atypical strain variants, and do not have the ability to differentiate all *MTBC* species such as *M. pinnipedii* and *M. microti* (Ayele et al., 2004). Such time-consuming and laborious methods hamper proper treatment of patients with respect to the different antibiotics and supportive treatments and are not practical for surveillance purposes. MALDI-TOF methods have been applied to mycobacterial (Mather et al., 2014) species identification and have shown to be a reliable technique in the routine laboratory. Although, MALDI-TOF for discrimination of mycobacterium is relatively rapid and reasonably accurate, however, this technique also could not differentiate *MTBC* species in clinical laboratories. Recently, a cheap and rapid new High Resolution Melting (HRM) assay for identification and differentiation of *Mycobacterium tuberculosis* complex samples was developed but it still had limitation to differentiate *M. bovis* from *M. bovis* BCG and *M. caprae* as well as *M. africanum* from *M. tuberculosis* (Landolt et al., 2019).

In contrast, in the present study, we demonstrated that ESI-Orbitrap LC-MS could be used for accurate identification of taxonomically complex microbial species. One of the advantages of using ESI is that it ionizes analytes to produce multiple charge states, bringing the mass to charge ratio of larger proteins into the window of the usual mass spectrometer range. These higher

mass proteins show increased discriminatory power for speciation between highly similar species which may be a reason for the improved success of this approach above other Mass Spectrometry approaches. While in this study protein identification was not carried out, the setup would allow the further characterization of the discriminatory proteins through a top-down proteomics workflow. These protein identifications if characterized as related to resistance or virulence could also have possible applications in routine clinical diagnostics (LeDuc et al., 2004). While novel diagnostic approaches employing complex LCMS technology might not be directly suitable and useful in developing countries, great benefit arises to many other countries and reference laboratories which will have access and funds to high resolution MS systems where the incidence of TB infection is still high.

ESI-LC-MS has been applied to study secretome, cell wall proteome, membrane proteome, and PTM profiling of *M. tuberculosis* (Ge et al., 2003; Gunawardena et al., 2013). However, no study has demonstrated the applicability of ESI-Orbitrap LC-MS in mycobacterial identification. Also, Dukik et al. demonstrated the use of ultra-high-resolution mass spectrometry technique for identification of closely related dermatophytes but no other report yet outlined this technique for microbial identification (Dukik et al., 2018). To the best of our knowledge, this is first study showing applicability of this technique for *MTBC* identification and discrimination at species level with very high accuracy in a very short time utilizing unique masses (protein profiles) from each species. Our present study was optimized to have an LC-MS analysis method of 5 minutes and demonstrates its potential for rapid diagnosis which could impact treatment times. This method is universally applicable and can be beneficial in the veterinary and other health care related applications for fast and precise testing. The methodology could serve as a tool in the differentiation of *MTBC* members and detection of transmission in captive and other animal populations. The *MTBC* consists of the closely related organisms *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. bovis BCG*, and animal born zoonotic species rarely identified in patients. The accurate molecular species identification within the *MTBC* is paramount to guide public health and primary care decisions more effectively due to species specific epidemiology, partial specificity in their host spectra (e.g. zoonotic taxa), geographic prevalence and drug susceptibility. *M. tuberculosis* and *M. bovis* may highly differ in contact tracing. Treatment with pyrazinamide can be excluded in case of *M. bovis* or *M. bovis BCG* as they are naturally resistant to the drug. *Mycobacterium bovis* Calmette-Guerin (BCG) is a live, attenuated strain of *M. bovis*. It is widely used as a vaccine against tuberculosis worldwide, and BCG is very effective for the treatment of transitional cell carcinoma of the bladder. Typically, it may require months to years for physicians to identify patients being infected with BCG disease upon exposure. Also, *M. bovis* mostly causes infection in cattle, deer and other mammals but the consumption of unpasteurized infected cow milk or transmission from infectious tuberculosis patient harboring *M. bovis* can cause disseminated infection in humans. The distribution of the various species is as follows and may vary

according to the patient population served: *M. tuberculosis* (95%), *M. bovis* (2%), *M. bovis BCG* (1.5%), and others (1.5%) (ARUP internal communication Mycobacteriology department; Incident rates 2020).

In this study, we show potential for utilizing LC-MS and algorithmic discrimination at the species level for *MTBC* species with high classification success for the clinically most relevant entities. This method offers the ability to generate classification models from large numbers of spectra in a relatively rapid and flexible way. The aim is to determine a common signature among spectra for each of the identified taxa in such a way that spectra of test isolates can be classified accurately. The utilized algorithm uses proteoforms in the approximate mass range of 5 to 40 kDa which were obtained from the mass spectra. A characteristic protein profile in this range provides enough information for the differentiation of various clinically relevant mycobacterial species and clinical isolates at least at species level with accuracy as high as 100% as shown in **Table 1**. While molecular taxonomy of mycobacterium is in general very dynamic, and species and subspecies (lineages) boundaries are evolving with advancement of laboratory technologies, our approach demonstrates the possibility to identify strains to their species by mass spectrometry in a method that matches their current genomic reference.

## CONCLUSION

Whole-protein top-down LC-MS analysis has significant diagnostic potential because of its ability to detect proteins routinely in a wider mass range and the method flexibility which can be optimized to provide the necessary protein coverage, e.g. if needed even to discriminate below the species level i.e. at the lineage or strain level. The accurate determination of protein masses, separation of high numbers of individual proteins, and high-resolution enabling detection of single-amino-acid substitutions are responsible for this high performance. Our algorithms showed a potential to recognize individual strains that could be applied in epidemics or outbreak scenarios. This novel assay is the first to use LC-ESI-MS to accurately identify the *M. tuberculosis* complex to the species level. The assay can rapidly, precisely, and accurately discriminate *M. tuberculosis* complex. Most importantly it can discriminate among *M. tuberculosis* complex isolates to the species level. These species are closely related members of the *M. tuberculosis* complex but very diverse pathogenetically. Having an effective and accurate way to distinguish these taxa will help in understanding the epidemiological behavior of these pathogens with a goal of supporting the development of improved diagnostics and treatment. However, our analysis was limited to 7H11 agar medium only and not yet tested with liquid culture medium as well as Loewenstein-Jensen agar which need to be assessed in the future. Additional research is required in targeting a wider species and strain panel, including the entire *MTBC* and its rarer zoonotic taxa, a large sample size and fine-tuned protocols. Algorithmic improvements to increase and further test classification success and to provide confirmatory experimental data to this first proof of principle study splitting

clinically common species in the *M. tuberculosis* species complex will be required to underline the high potential of Orbitrap LCMS technology.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M., et al. (2010). Novel Mycobacterium Tuberculosis Complex Pathogen, *M. Mungi*. *Emerg. Infect. Dis.* 16 (8), 1296–1299. doi: 10.3201/eid1608.100314
- Andersen, A., and Brennan, P. (1994). “Proteins and Antigens of Mycobacterium Tuberculosis.” in *Tuberculosis* American Society of Microbiology, Wiley Online Library. vol. 307-332. Ed. B. R. Bloom
- Arnold, C., Westland, L., Mowat, G., Underwood, A., Magee, J., and Gharbia, S. (2005). Single-Nucleotide Polymorphism-Based Differentiation and Drug Resistance Detection in Mycobacterium Tuberculosis From Isolates or Directly From Sputum. *Clin. Microbiol. Infect.* 11 (2), 122–130. doi: 10.1111/j.1469-0691.2004.01034.x
- Ayele, W. Y., Neill, S. D., Zinsstag, J., Weiss, M. G., and Pavlik, I. (2004). Bovine Tuberculosis: An Old Disease But a New Threat to Africa. *Int. J. Tuberc. Lung Dis.* 8 (8), 924–937.
- Bhowmick, T., Ghosh, S., Dixit, K., Ganesan, V., Ramagopal, U. A., Dey, D., et al. (2014). Targeting Mycobacterium Tuberculosis Nucleoid-Associated Protein HU With Structure-Based Inhibitors. *Nat. Commun.* 5, 4124. doi: 10.1038/ncomms5124
- Bloom, B. R., and Murray, C. J. (1992). Tuberculosis: Commentary on a Reemergent Killer. *Science* 257 (5073), 1055–1064. doi: 10.1126/science.257.5073.1055
- Brock, I., Munk, M. E., Kok-Jensen, A., and Andersen, P. (2001). Performance of Whole Blood IFN-Gamma Test for Tuberculosis Diagnosis Based on PPD or the Specific Antigens ESAT-6 and CFP-10. *Int. J. Tuberc. Lung Dis.* 5 (5), 462–467.
- Chen, J. H., Yam, W. C., Ngan, A. H., Fung, A. M., Woo, W. L., Yan, M. K., et al. (2013). Advantages of Using Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry as a Rapid Diagnostic Tool for Identification of Yeasts and Mycobacteria in the Clinical Microbiological Laboratory. *J. Clin. Microbiol.* 51 (12), 3981–3987. doi: 10.1128/jcm.01437-13
- Chimara, E., Ferrazoli, L., and Leão, S. C. (2004). Mycobacterium Tuberculosis Complex Differentiation Using GyrB-Restriction Fragment Length Polymorphism Analysis. *Mem Inst Oswaldo Cruz* 99 (7), 745–748. doi: 10.1590/s0074-02762004000700014
- DeGeorge, K. C., Holt, H. R., and Hodges, S. C. (2017). Bladder Cancer: Diagnosis and Treatment. *Am. Fam. Phys.* 96 (8), 507–514.
- Derese, Y., Hailu, E., Assefa, T., Bekele, Y., Mihret, A., Aseffa, A., et al. (2012). Comparison of PCR With Standard Culture of Fine Needle Aspiration Samples in the Diagnosis of Tuberculosis Lymphadenitis. *J. Infect. Dev. Countries* 6 (1), 53–57. doi: 10.3855/jidc.2050
- de Steenwinkel, J. E., de Knecht, G. J., ten Kate, M. T., van Belkum, A., Verbrugh, H. A., Kremer, K., et al. (2010). Time-Kill Kinetics of Anti-Tuberculosis Drugs, and Emergence of Resistance, in Relation to Metabolic Activity of Mycobacterium Tuberculosis. *J. Antimicrob. Chemother.* 65 (12), 2582–2589. doi: 10.1093/jac/dkq374
- Djelouadi, Z., Raoult, D., Daffé, M., and Drancourt, M. (2008). A Single-Step Sequencing Method for the Identification of Mycobacterium Tuberculosis Complex Species. *PLoS Negl. Trop. Dis.* 2 (6), e253. doi: 10.1371/journal.pntd.0000253
- Doig, C., Seagar, A. L., Watt, B., and Forbes, K. J. (2002). The Efficacy of the Heat Killing of Mycobacterium Tuberculosis. *J. Clin. Pathol.* 55 (10), 778–779. doi: 10.1136/jcp.55.10.778

## AUTHOR CONTRIBUTIONS

AOB and SS have drafted the work, AOB, SS, JF, JBS, and APB designed research, AOB performed experimental work, data acquisition, analysis, AOB and JBS conducted review and editing, JK and JBS performed interpretation of data and computational analysis used for the work, and JF and APB provided funding acquisition, project administration, and resources. All authors contributed to the article and approved the submitted version.

- Dukik, K., Freeke, J., Jamalain, A., van den Ende, B. G., Yip, P., Stephenson, J. L., et al. (2018). Ultra-High-Resolution Mass Spectrometry for Identification of Closely Related Dermatophytes With Different Clinical Predilections. *J. Clin. Microbiol.* 56 (7), 1–11. doi: 10.1128/jcm.00102-18
- Frothingham, R. (1995). Differentiation of Strains in Mycobacterium Tuberculosis Complex by DNA Sequence Polymorphisms, Including Rapid Identification of *M. Bovis* BCG. *J. Clin. Microbiol.* 33 (4), 840–844. doi: 10.1128/jcm.33.4.840-844.1995
- Frothingham, R., and Meeker-O’Connell, W. A. (1998). Genetic Diversity in the Mycobacterium Tuberculosis Complex Based on Variable Numbers of Tandem DNA Repeats. *Microbiol. (Reading)* 144 (Pt 5), 1189–1196. doi: 10.1099/00221287-144-5-1189
- Gagneux, S., and Small, P. M. (2007). Global Phylogeography of Mycobacterium Tuberculosis and Implications for Tuberculosis Product Development. *Lancet Infect. Dis.* 7 (5), 328–337. doi: 10.1016/s1473-3099(07)70108-1
- Ge, Y., El-Naggar, M., Sze, S. K., Oh, H. B., Begley, T. P., McLafferty, F. W., et al. (2003). Top Down Characterization of Secreted Proteins From Mycobacterium Tuberculosis by Electron Capture Dissociation Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 14 (3), 253–615. doi: 10.1016/s1044-0305(02)00913-3
- Godreuil, S., Torrea, G., Terru, D., Chevenet, F., Diabougou, S., Supply, P., et al. (2007). First Molecular Epidemiology Study of Mycobacterium Tuberculosis in Burkina Faso. *J. Clin. Microbiol.* 45 (3), 921–927. doi: 10.1128/jcm.01918-06
- Gunawardena, H. P., Feltcher, M. E., Wrobel, J. A., Gu, S., Braunstein, M., and Chen, X. (2013). Comparison of the Membrane Proteome of Virulent Mycobacterium Tuberculosis and the Attenuated Mycobacterium Bovis BCG Vaccine Strain by Label-Free Quantitative Proteomics. *J. Proteome Res.* 12 (12), 5463–5474. doi: 10.1021/pr400334k
- Gupta, R. S., Lo, B., and Son, J. (2018). Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium Into an Emended Genus Mycobacterium and Four Novel Genera. *Front. Microbiol.* 9, 67. doi: 10.3389/fmicb.2018.00067
- Hesseling, A. C., Marais, B. J., Gie, R. P., Schaaf, H. S., Fine, P. E., Godfrey-Faussett, P., et al. (2007). The Risk of Disseminated Bacille Calmette-Guerin (BCG) Disease in HIV-Infected Children. *Vaccine* 25 (1), 14–18. doi: 10.1016/j.vaccine.2006.07.020
- Hughes, R., Wonderling, D., Li, B., and Higgins, B. (2012). The Cost Effectiveness of Nucleic Acid Amplification Techniques for the Diagnosis of Tuberculosis. *Respir. Med.* 106 (2), 300–307. doi: 10.1016/j.rmed.2011.10.005
- Kasai, H., Ezaki, T., and Harayama, S. (2000). Differentiation of Phylogenetically Related Slowly Growing Mycobacteria by Their GyrB Sequences. *J. Clin. Microbiol.* 38 (1), 301–308.
- Landolt, P., Stephan, R., and Scherrer, S. (2019). Development of a New High Resolution Melting (HRM) Assay for Identification and Differentiation of Mycobacterium Tuberculosis Complex Samples. *Sci. Rep.* 9 (1), 1850. doi: 10.1038/s41598-018-38243-6
- Larsen, E. S., Joensen, U. N., Poulsen, A. M., Goletti, D., and Johansen, I. S. (2020). Bacillus Calmette-Guérin Immunotherapy for Bladder Cancer: A Review of Immunological Aspects, Clinical Effects and BCG Infections. *Apmis* 128 (2), 92–103. doi: 10.1111/apm.13011
- LeDuc, R. D., Taylor, G. K., Kim, Y. B., Januszyk, T. E., Bynum, L. H., Sola, J. V., et al. (2004). ProSight PTM: An Integrated Environment for Protein Identification and Characterization by Top-Down Mass Spectrometry. *Nucleic Acids Res.* 32 (Web Server issue), W340–W345. doi: 10.1093/nar/gkh447



- Lévesque, S., Dufresne, P. J., Soualhine, H., Domingo, M. C., Bekal, S., Lefebvre, B., et al. (2015). A Side by Side Comparison of Bruker Biotyper and VITEK MS: Utility of MALDI-TOF MS Technology for Microorganism Identification in a Public Health Reference Laboratory. *PLoS One* 10 (12), e0144878. doi: 10.1371/journal.pone.0144878
- Lipworth, S., Jajou, R., de Neeling, A., Bradley, P., van der Hoek, W., Maphalala, G., et al. (2019). SNP-IT Tool for Identifying Subspecies and Associated Lineages of Mycobacterium Tuberculosis Complex. *Emerg. Infect. Dis.* 25 (3), 482–488. doi: 10.3201/eid2503.180894
- Louw, G. E., Warren, R. M., Gey van Pittius, N. C., Leon, R., Jimenez, A., Hernandez-Pando, R., et al. (2011). Rifampicin Reduces Susceptibility to Ofloxacin in Rifampicin-Resistant Mycobacterium Tuberculosis Through Efflux. *Am. J. Respir. Crit. Care Med.* 184 (2), 269–276. doi: 10.1164/rccm.201011-1924OC
- Mather, C. A., Rivera, S. F., and Butler-Wu, S. M. (2014). Comparison of the Bruker Biotyper and Vitek MS Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry Systems for Identification of Mycobacteria Using Simplified Protein Extraction Protocols. *J. Clin. Microbiol.* 52 (1), 130–138. doi: 10.1128/jcm.01996-13
- Mehaffy, M. C., Kruh-Garcia, N. A., and Dobos, K. M. (2012). Prospective on Mycobacterium Tuberculosis Proteomics. *J. Proteome Res.* 11 (1), 17–25. doi: 10.1021/pr2008658
- Murdoch, L. E., Maclean, M., Endarko, E., MacGregor, S. J., and Anderson, J. G. (2012). Bactericidal Effects of 405 Nm Light Exposure Demonstrated by Inactivation of Escherichia, Salmonella, Shigella, Listeria, and Mycobacterium Species in Liquid Suspensions and on Exposed Surfaces. *Sci. World J.* 2012, 137805. doi: 10.1100/2012/137805
- Niemann, S., Harmsen, D., Rünsch-Gerdes, S., and Richter, E. (2000). Differentiation of Clinical Mycobacterium Tuberculosis Complex Isolates by GyrB DNA Sequence Polymorphism Analysis. *J. Clin. Microbiol.* 38 (9), 3231–3234. doi: 10.1128/jcm.38.9.3231-3234.2000
- Nikam, C., Jagannath, M., Narayanan, M. M., Ramanabhiraman, V., Kazi, M., Shetty, A., et al. (2013). Rapid Diagnosis of Mycobacterium Tuberculosis With Truenat MTB: A Near-Care Approach. *PLoS One* 8 (1), e51121. doi: 10.1371/journal.pone.0051121
- Peters, J. S., Nabila, I., Anzaan, D., Shuyi, M., David, R. S., Robin, M. W., et al. (2020). Genetic Diversity in Mycobacterium Tuberculosis Clinical Isolates and Resulting Outcomes of Tuberculosis Infection and Disease. *Annu. Rev. Genet.* 54 (1), 511–537. doi: 10.1146/annurev-genet-022820-085940
- Pinsky, B. A., and Banaei, N. (2008). Multiplex Real-Time PCR Assay for Rapid Identification of Mycobacterium Tuberculosis Complex Members to the Species Level. *J. Clin. Microbiol.* 46 (7), 2241–2246. doi: 10.1128/jcm.00347-08
- Pounder, J. I., Anderson, C. M., Voelkerding, K. V., Salfinger, M., Dormandy, J., Somoskovi, A., et al. (2010). Mycobacterium Tuberculosis Complex Differentiation by Genomic Deletion Patterns With Multiplex Polymerase Chain Reaction and Melting Analysis. *Diagn. Microbiol. Infect. Dis.* 67 (1), 101–105. doi: 10.1016/j.diagmicrobio.2009.12.014
- Prabhakar, S., Mishra, A., Singhal, A., Katoch, V. M., Thakral, S. S., Tyagi, J. S., et al. (2004). Use of the HupB Gene Encoding a Histone-Like Protein of Mycobacterium Tuberculosis as a Target for Detection and Differentiation of M. Tuberculosis and M. Bovis. *J. Clin. Microbiol.* 42 (6), 2724–2732. doi: 10.1128/jcm.42.6.2724-2732.2004
- Reddington, K., Zumla, A., Bates, M., van Soolingen, D., Niemann, S., Barry, T., et al. (2012). Seektb, a Two-Stage Multiplex Real-Time-PCR-Based Method for Differentiation of the Mycobacterium Tuberculosis Complex. *J. Clin. Microbiol.* 50 (7), 2203–2206. doi: 10.1128/jcm.00718-12
- Richter, E., Weizenegger, M., Fahr, A. M., and Rünsch-Gerdes, S. (2004). Usefulness of the Genotype MTBC Assay for Differentiating Species of the Mycobacterium Tuberculosis Complex in Cultures Obtained From Clinical Specimens. *J. Clin. Microbiol.* 42 (9), 4303–4306. doi: 10.1128/jcm.42.9.4303-4306.2004
- Riojas, M. A., McGough, K. J., Rider-Riojas, C. J., Rastogi, N., and Hazbón, M. H. (2018). Phylogenomic Analysis of the Species of the Mycobacterium Tuberculosis Complex Demonstrates That Mycobacterium Africanum, Mycobacterium Bovis, Mycobacterium Caprae, Mycobacterium Microti and Mycobacterium Pinnipedii are Later Heterotypic Synonyms of Mycobacterium Tuberculosis. *Int. J. Syst. Evol. Microbiol.* 68 (1), 324–332. doi: 10.1099/ijsem.0.002507
- Ritz, N., Tebruegge, M., Connell, T. G., Sievers, A., Robins-Browne, R., and Curtis, N. (2009). Susceptibility of Mycobacterium Bovis BCG Vaccine Strains to Antituberculous Antibiotics. *Antimicrob. Agents Chemother.* 53 (1), 316–318. doi: 10.1128/aac.01302-08
- Scorpio, A., and Zhang, Y. (1996). Mutations in PncA, a Gene Encoding Pyrazinamidase/Nicotinamidase, Cause Resistance to the Antituberculous Drug Pyrazinamide in Tubercle Bacillus. *Nat. Med.* 2 (6), 662–667. doi: 10.1038/nm0696-662
- Sinha, P., Gupta, A., Prakash, P., Anupurba, S., Tripathi, R., and Srivastava, G. N. (2016). Differentiation of Mycobacterium Tuberculosis Complex From Non-Tubercular Mycobacteria by Nested Multiplex PCR Targeting IS6110, MTP40 and 32kd Alpha Antigen Encoding Gene Fragments. *BMC Infect. Dis.* 16, 123. doi: 10.1186/s12879-016-1450-1
- Somoskovi, A., Dormandy, J., Rivenburg, J., Pedrosa, M., McBride, M., and Salfinger, M. (2008). Direct Comparison of the Genotype MTBC and Genomic Deletion Assays in Terms of Ability to Distinguish Between Members of the Mycobacterium Tuberculosis Complex in Clinical Isolates and in Clinical Specimens. *J. Clin. Microbiol.* 46 (5), 1854–1857. doi: 10.1128/jcm.00105-07
- Somoskovi, A., Parsons, L. M., and Salfinger, M. (2001). The Molecular Basis of Resistance to Isoniazid, Rifampin, and Pyrazinamide in Mycobacterium Tuberculosis. *Respir. Res.* 2 (3), 164–168. doi: 10.1186/rr54
- Sreevatsan, S., Escalante, P., Pan, X., Gillies, D. A., Siddiqui, Z. S., Khalaf, C. N., et al. (1996). Identification of a Polymorphic Nucleotide in OxyR Specific for Mycobacterium Bovis. *J. Clin. Microbiol.* 34 (8), 2007–2010. doi: 10.1128/jcm.34.8.2007-2010.1996
- Supply, P., Magdalena, J., Himpens, S., and Loch, C. (1997). Identification of Novel Intergenic Repetitive Units in a Mycobacterial Two-Component System Operon. *Mol. Microbiol.* 26 (5), 991–1003. doi: 10.1046/j.1365-2958.1997.6361999.x
- WHO (2020). “Global Tuberculosis Report” Geneva: World Health Organization. (Accessed October 15). Available at: [https://www.who.int/tb/publications/global\\_report/en/](https://www.who.int/tb/publications/global_report/en/).

**Conflict of Interest:** Authors JF, JBS, and JK were employed by company Thermo Fisher Scientific.

The authors declare that this study received funding from Thermo Fisher Scientific. The funder was involved (equal partner) in the study design, assisting with providing tools and assistance in analyzing the data, and technical interpretation of results. The funder was not further involved in the collection, interpretation of data, the writing of this article (beyond assisting with technical details for the materials and methods and textual review) or the decision to submit it for publication.

Copyright © 2021 Bajaj, Saraswat, Knuutila, Freeke, Stielow and Barker. 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.