Peptidomic analysis of mycobacterial secreted proteins enables species identification

Qingbo Shu1 | Meena Rajagopal2 | Jia Fan1 | Lingpeng Zhan1 | Xiangxing Kong1 | Yifan He3 | Suwatchareeporn Rotcheewaphan2,4 | Christopher J. Lyon1 | Wei Sha3 | Adrian M. Zelazny2 | Tony Hu1

1Center for Cellular and Molecular Diagnostics, Department of Biochemistry and Molecular Biology, School of Medicine, Tulane University, New Orleans, Louisiana, USA
2Department of Laboratory Medicine, Clinical Center, National Institutes of Health, Bethesda, Maryland, USA
3Clinic and Research Center of Tuberculosis, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, People’s Republic of China
4Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

Correspondence
Tony Hu, Department of Biochemistry and Molecular Biology, Center for Cellular and Molecular Diagnostics, School of Medicine, Tulane University, New Orleans, LA, USA.
Email: tonyhu@tulane.edu

Funding information
NIH, Grant/Award Numbers: R01HD090927, R01AI122932, R01AI113725, R21AI26361-01; Department of Defense, Grant/Award Number: W81XWH1910926; Arizona Biomedical Research Commission (ABRC) young investigator award; the Weatherhead Presidential Endowment Fund

Abstract
Pulmonary disease arising from slow-growing mycobacterial infections has emerged as an increasingly prevalent clinical concern over the past two to three decades. Proteins belonging to the family of ESAT-6 secretion (Esx) systems play critical roles in the virulence of most pathogenic mycobacterial species and are associated with drug resistance. However, no clinical applications can detect and discriminate the expression of species-specific variants of these proteins in clinical samples, such as early growth cultures, for rapid diagnosis of specific mycobacterial infections, which may require distinct interventions. Conventional immunoassay approaches are not suitable for this purpose due to the significant degree of conservation of Esx proteins among species. Herein we describe the development of a novel immunoprecipitation-coupled mass spectrometry assay that can distinguish Esx proteins that are expressed by slow-growing mycobacterial species commonly detected in clinical isolates. This approach uses custom antibodies raised against single semi-conserved peptide regions in \textit{M. tuberculosis} (\textit{Mtb}) EsxB and EsxN to capture corresponding peptides from protein orthologs of mycobacteria associated with human respiratory infections, including \textit{Mtb}, \textit{M. avium}, \textit{M. intracellulare}, \textit{M. kansasii}, \textit{M. gordonae}, and \textit{M. marinum}, to detect these species in standard clinical cultures at the first sign mycobacterial growth to allow rapid disease diagnosis.

KEYWORDS
antigen peptides, CFP-10, EsxN, mass spectrometry, mycobacterium
INTRODUCTION

*M. tuberculosis* (*Mt*) infections can progress to tuberculosis (*TB*), the leading cause of death from respiratory infection prior to coronavirus disease 2019 and, similar to severe acute respiratory syndrome coronavirus 2 infections, is primarily spread by respiratory droplet exposure through close contact with infected individuals. However, the genus *Mycobacterium* contains more than 190 species, several of which can also cause human infections. Such non-tuberculous mycobacteria (NTM) infections, unlike *Mt* infections, are usually contracted by exposure to environmental sources. NTMs that can function as human pathogens and may cause a variety of disease manifestations, including pulmonary disease (*e.g.*, *M. abscessus*, *M. kansasii*, and *M. avium* complex [MAC] species) that resembles TB, while others may primarily cause skin and other soft tissue infections (*e.g.*, *M. marinum*).[2] Notably, the incidence of NTM-related pulmonary disease has substantially increased in recent decades, with estimates now ranging from 15.5 to 26.7 new cases per 100,000 adults in populations >50 years of age in developed countries. NTM infections are less common than TB in the developing world but may be misdiagnosed as a significant fraction of suspected (4%–30%) or multidrug-resistant TB cases (18%–27%) in countries with high endemic TB rates.

NTM infection rates can exhibit different geographic distributions, but certainly slow-growing NTM (*e.g.*, MAC, *M. gordonae*, *M. xenopi*, and *M. kansasii*) and rapidly growing NTM (*M. abscessus* and *M. fortuitum*) tend to be the most prevalent NTM isolates worldwide,[3] while *M. gordonae*-positive isolates are frequently the result of environmental contamination.[4] In the United States, MAC is the most common cause of pulmonary NTM infection, followed by *M. kansasii*, and *M. abscessus* is responsible for most pulmonary NTM infections caused by rapidly growing NTM species.[5,6]

Effective treatment of *Mt* and NTM infections requires accurate identification of the responsible mycobacterial species or species complexes (*e.g.*, *M. abscessus*, *M. kansasii*, or the *Mt* and MAC complexes) since different species require different interventions. GeneXpert MTB/RIF, which amplifies genome regions that are specific to members of the *Mt* complex, has improved the speed of TB diagnosis, but nucleic amplification assays are not yet available for NTM species. Culture remains the gold standard for clinical confirmation of mycobacterial infection, species identification, and drug susceptibility testing,[7–9] and various automated culture-reading systems are employed for the rapid evaluation of *Mt* and NTM growth in mycobacterial growth indicator tube (MGIT) liquid cultures. NTM species identification is typically performed by polymerase chain reaction (PCR) sequencing[10] or the detection of PCR-amplification of species-specific target regions[11–14] in laboratory-developed tests, or by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis of positive MGIT cultures, or solid-phase subcultures of these samples.[15,16] However, several of these assays have limitations that may reduce their clinical utility, including complexity, the potential for contamination, and/or the need for late-stage or secondary culture material. New laboratory tests that employ nucleic amplification approaches[17] or MALDI-TOF analyses[8,18] have been employed to analyze polymicrobial clinical samples for species identification and antimicrobial susceptibility testing. Liquid chromatography-MS (LC-MS) analysis of virulence factors detectable in early-stage MGIT cultures may offer another alternative for species-specific identification of mycobacteria.

Mycobacteria utilize five homologous secretion systems (ESX-1 to ESX-5) that have resulted from gene duplication of ancestral gene loci to export many of their secreted proteins, and these loci have undergone further duplication events.[19] The *Mt* genome contains 23 ESX genes among these loci that can be classified into specific gene subfamilies.[19,20] The ESX-1 subfamily contains only EsxA (ESAT-6) and EsxB (CFP-10), while the ESX-5 subfamily contains EsxM and EsxN genes located within the ESX-5 locus as well as EsxM and EsxN paralogs at four distant loci that exhibit gene arrangements similar to those in the ESX-5 locus. *Mt* requires ESX-5 for virulence, and the secretion of ESX-5 substrates, including EsxM and EsxN paralogs, is tightly controlled to avoid *Mt* elimination by host immune responses,[21] and it has been proposed that the ESX-5 secretion system can selectively export a subset of the ESX-5 secretome.[22] Further, a Thr2Ala variant of the ESX-5 substrate EsxW has been shown to be under positive selection and may play a role in substrate selection for ESX-5 secretion, and multi-drug resistant TB.[23,24]

Due to their critical importance in mycobacterial virulence, Esx-1 and Esx-5 substrates appear to represent good candidates for biomarkers of species-specific mycobacterial infection. LC coupled tandem mass spectrometry (LC-MS/MS) can be employed to unambiguously identify peptides from conserved protein homologs (*e.g.*, EsxN paralogs/homologs) expressed by distinct mycobacteria at a single amino acid resolution to accurately diagnose specific infections.[25] In the current study, we thus first identified semi-conserved peptide regions in ESX-1 and ESX-5 substrates from culture filtrate protein (CFP) samples of mycobacteria frequently detected in clinical isolates. We then analyzed their ability to serve as
species-specific biomarkers in an immunoprecipitation-based LC-MS (IP-MS) assay, where captured peptides were identified by parallel reaction monitoring (PRM), and their amino acid differences were used to identify their species origin. Diagnostic peptides analyzed in this approach identified target species with high accuracy (87.5%–100%) in early positive clinical MGIT cultures.

2 | METHODS

2.1 | BLAST analysis of Mtb CFP10 and EsxN protein sequence

Reviewed UniProt KB database entries for full-length Mtb CFP-10 (P9WNK5) and EsxN (P9WNJ3) protein sequence were searched against the UniProt KB bacterial proteome database using the UniProt The Basic Local Alignment Search Tool (BLAST) function to identify homologous proteins, using default search parameters (E-Threshold of 10, auto matrix selection, no filtering for low-complexity regions, and permitting gaps in the sequence alignments). To evaluate the number of gene duplication events in the 15 most common mycobacterial species that cause human disease,[26] exxB (CFP-10), and exxN gene sequences corresponding to protein hits for these species were evaluated against their corresponding genomes using the NCBI BlastN sequence alignment program with the setting “optimize for somewhat similar sequences”. For species that had sequence hits from unassembled whole-genome sequencing data, the GenBank protein ID corresponding to a hit was used to retrieve the sequence of the target peptide region protein. This information was gathered for CFP-10 from M. ulcerans (WP_012392185.1), M. scrofulaceum (ORB70459.1), and M. mucogenicum (WP_061000007.1). To confirm gene duplication events detected in several species, literature reviews were conducted to identify articles that examined gene duplications for the affected species: M. xenopi,[27] M. kansasii, M. szulgai, and M. marinum,[28] M. mucogenicum, [29] M. chelonae, and M. fortuitum.[30]

2.2 | Liquid culture of mycobacteria

Culture stocks for M. abscessus ATCC 19977, M. avium ATCC 25291, M. intracellulare ATCC 13950, and M. kansasii ATCC 12478 were purchased from ATCC, inoculated into 10 ml Middlebrook 7H9 broth supplemented with 10% ADC (5 mg/ml bovine serum albumin, 2 mg/ml dextrose, 3 µg/ml catalase final concentrations), 0.2% glycerol, and 0.05% Tween 80, and grown to saturation with 100 rpm rotation in a 37°C incubator. Due to the differential growth rate of these species, culture media was collected at different times post-inoculation for M. abscessus (2 days), M. avium (21 days), M. intracellulare (10 days), and M. kansasii (7 days). Saturated mycobacterial cultures were pelleted by centrifugation at 3,000 × g for 5 min, and supernatants were transferred into a sterile syringe and filtered through a 0.2 µm PES syringe filter, and the resulting CFP samples were stored at 4°C until use. M. tuberculosis CFP was obtained from BEI Resources (NR-14825). Mtb CFP obtained from BEI Resources (NR-14825) was derived from Glycerol-alanine-salts (GAS) media cultures, containing 2 mg/ml ammonium chloride, 1 mg/ml L-alanine, 0.3 mg/ml Bacto Casitone (Difco), 4 mg/ml dibasic potassium phosphate, 2 mg/ml citric acid, 50 µg/ml ferric ammonium citrate, 1.2 mg/ml magnesium chloride hexahydrate, 0.6 mg/ml potassium sulfate, 1.8 mM sodium hydroxide, and 1% glycerol. M. abscessus, M. avium, M. intracellulare, and M. kansasii CFP were provided by Dr. Shelley Haydel at Arizona State University’s Biodesign Institute.

2.3 | Protein filtration and digestion

Stored NTM CFP samples were passed through a 50 kDa filter unit (Amicon, Thermo) by 14,000 × g centrifugation at 4°C to reduce albumin present in the starting culture media, and analyzed for total protein concentration by BCA assay (Thermo Pierce). For NTM samples, 100 µg of this filtrate was reduced with 10 mM dithiothreitol (Sigma, USA) at 37 °C for one h, alkylated with 50 mM iodoacetamide (Sigma) at 25 °C for one h, and then digested with sequencing grade trypsin (Promega, USA) at a protein: enzyme ratio of 50:1 for 16 h at 37 °C. Mtb CFP obtained from BEI Resources protein additives was pre-concentrated,[31] but otherwise processed in the same manner as all other CFP samples.

2.4 | LC-MS/MS analysis of CFP tryptic digests

Tryptic peptide mixtures were separated with a linear gradient of 2%–37% buffer B (100% ACN and 0.1% formic acid) at a flow rate of 300 nl/min on an EASY-Spray C18 LC column (15 cm × 75 µm internal diameter, 3 µm particle size, Thermo Scientific). An UltiMate 3000 nanoLC system (Thermo Scientific) was online coupled to an Orbitrap Velos Pro (NTM samples) or Orbitrap Fusion Lumos instrument (Mtb sample) (Thermo Scientific). NTM and Mtb CFP samples were analyzed using different MS instruments and conditions to account for differences in complexity arising from the greater
protein contribution (5 mg/ml) by the media of the NTM versus \textit{Mtb} culture media, as indicated above. MS data were acquired in a data-dependent strategy selecting the fragmentation events based on the precursor abundance in the survey scan (mass-to-charge [m/z] 275–1850 for NTM samples and m/z 375–1700 for the \textit{Mtb} sample). The resolution of the survey scan was 120,000 at m/z 400.

For NTM samples, low-resolution MS/MS spectra were acquired in rapid collision-induced dissociation (CID) scan mode for the 15 most intense peaks from the survey scan. The normalized collision energy of 35 was used for fragmentation, with a dynamic exclusion of 40 s using an isolation window for MS/MS fragmentation set to 2 m/z. Resulting CID MS/MS spectra were searched against mycobacterial proteomes using Proteome Discoverer (version 2.4, Thermo Scientific) and the following parameters: trypic digestion with no more than two missed cleavage sites; precursor and product mass tolerances of 10 ppm and 0.6 Da respectively; cysteine carboxymethylation as a stable modification; and methionine oxidation as a dynamic modification. UniProt Proteome IDs for the databases of each mycobacteria isolate analyzed in this search, and their total number of protein sequence entries were: \textit{M. kansasii} strain ATCC12478 (UP0000017786; 5,843 entries), \textit{M. abscessus} strain ATCC19977 (UP000007137; 4,940 entries), \textit{M. intracellulare} strain ATCC 13950 (UP000008004; 5,131 entries), \textit{M. avium subsp. avium} strain H37Rv (UP0000019908; 6,021 entries) and \textit{M. intracellulare} strain H37Rv (UP0000001584; 3,993 entries).

For the \textit{Mtb} CFP sample, data-dependent MS/MS analysis was performed using a top-speed approach (cycle time of 3 s), normalized collision energy of 30 for Higher-energy C-trap dissociation (HCD), with dynamic exclusion set to 60 s, and the isolation window for MS/MS fragmentation set to 1.6 m/z. \textit{Mtb} HCD MS/MS spectra were searched against the TB proteome in UniProt using Proteome Discoverer (version 2.4; Thermo Scientific) and the following parameters: trypic digestion allowing no more than two missed cleavage sites; precursor and product mass tolerances of 10 ppm and 0.6 Da respectively; cysteine carboxymethylation as a stable modification; and methionine oxidation as a dynamic modification. After the database search, Percolator was used for peptide-spectrum match (PSM) validation with default settings. A decoy database was generated by reversing each protein sequence and a concatenated database was created from the non-decoy and the decoy databases were searched to evaluate the false discovery rate (FDR), calculated as the number of decoy hits divided by the number of target hits. Percolator analyses using a 1% FDR were sequentially employed to filter out high confidence matches at the PSM, peptide, and protein levels.

### 2.5 MGIT samples

This research project was reviewed and approved by the Institutional Review Board (IRB) at Tulane University prior to study initiation. Mycobacterial cultures were obtained from patients enrolled in IRB-approved protocols at the National Institutes of Health Clinical Center. Respiratory specimens were decontaminated using the N-acetyl-L-cysteine/sodium hydroxide method, concentrated by 15 min centrifugation at 3000 g, and pellets were suspended in 0.8 ml of sterile phosphate buffer saline (PBS). Microscopic examination for the presence of acid-fast bacilli (AFB) was performed by auramine-rodhamine staining (Becton Dickinson, Sparks, MD). All patient samples were inoculated into Bactec 960 MGIT tubes containing the manufacturer’s growth supplement (PANTA) and Middlebrook 7H11 agar (Remel, Lenexa, KS). Growth-positive MGIT samples were stained with auramine-rodhamine to detect AFB and/or contaminants and sub-cultured on solid media for species identification, and 3 ml was filtered sterilized by passage through a 0.22 µm filter and stored at -80°C for subsequent LC-MS/MS analysis. Mycobacteria isolated from positive MGIT cultures were identified by secAI gene sequencing directly from the culture and/or by MALDI-TOF MS after sub-culture on solid media. Multiplex PCR was employed to differentiate \textit{Mtb} from \textit{M. bovis}.

### 2.6 EsxN and CFP-10 peptide IP

Clinical MGIT sample aliquots (500 µl) were supplemented with 10 µl of 1 M Tris solution (adjusted to pH 8) and rotation-mixed at 37°C for 16 h with 5 µg of sequencing grade trypsin (1:600 enzyme/protein mass for a mean 6 mg protein/ml value). After digestion, samples were supplemented with 5 µl of 10% trifluoroacetic acid solution (adjusted to pH 7) and spiked with 40 µl of an internal standard (IS) peptide mixture (25 nM each) to permit accurate identification of the isolated target peptides. Custom rabbit polyclonal antibodies (GenScript, Nanjing, China) raised against \textit{Mtb} EsxN-p2 (AQAAASLEAEHQAIYVR) and \textit{Mtb} CFP-10-p2 (TQIDQVESTAGSLQGQWR) were suspended in PBS buffer (pH 7.4) separately, and 40 µg of antibody was bound to 3 mg of protein G Dynabeads (Thermo Fisher Scientific) in 400 µl of PBS buffer containing 0.2% (v/v) Tween-20 binding buffer, which was then washed twice with 200 µl of PBST binding buffer, and suspended in 50 µl of binding buffer. EsxN and CFP-10 peptide antibody beads (150 µg each) were simultaneously mixed with trypsin-digested CFP samples for 1 h at 25°C, then washed three times with 50 µl of binding buffer, and two times with 50 µl of LC grade water.
Development and analytical validation of targeted MS measurements

In this study, Tier 2 level PRM was employed to identify and quantify peptides generated by tryptic digestion of mycobacterial secreted proteins. To select the precursor ions of EsxN and CFP-10 peptide targets for PRM analysis, we retrieved eight protein sequences in April 2020 from the UniProt database for M. tuberculosis (Mtb), M. avium, M. kansasii, M. marinum, M. tuberculosis H37Rv (Mtb H37Rv), M. intracellulare (M. intracellulare), M. avium (M. avium), and M. kansasii ESAT-6-like protein (M. kansasii ESAT-6-like protein). Mass errors for these peptides were all within 0.1% of their theoretical molecular masses when analyzed by an AmaZon SL ion trap mass spectrometer (Bruker, MA). All these peptides were dissolved in 2% acetonitrile/0.1% trifluoroacetic acid to final a concentration of 50 µg/ml each to generate a multiplex IS peptide mixture, and a 1 µl aliquot was analyzed by LC-PRM-MS to obtain the retention times and MS/MS spectra for each of these peptides. A spectral library was built using the synthetic IS peptide for each targeted peptide, and these MS/spectra were searched against the eight target protein sequences and 236 common protein contaminants with Peaks Studio software (version 10.5). Identified peptide spectra were exported as mzXML files and imported into Skyline to build a spectral library. The dotp, rdotp, and peak area ratio values (endogenous/IS peptide) for target peptides detected in clinical MGIT samples were calculated in Skyline. EsxN-p2 fragment ions y1-y13 and b2-b3 and CFP-10-p2 fragment ions y1-y16 and b2-b3 were used to extract the ion chromatograms of each sample to achieve a full-length coverage to discriminate potential isomers. Endogenous (light) peptide fragment ions were required to co-elute in the same retention window as IS peptide (heavy) fragment peptide ions with a < 20 ppm mass error tolerance and to yield a dotp similarity score ≥0.9 upon Skyline analysis.

Identification of endogenous peptides was based on retention time and dotp, to avoid the potential loss of rdotp.
signal due to competition between high endogenous peptide level and the corresponding IS peptide. To evaluate the relative affinity of peptide antibodies for sequence variants of their target peptides, serial dilutions (L1-L6) of the IS solution containing 6.3, 3.1, 1.6, 0.8, 0.4, and 0.2 nM of each peptide were spiked into blank MGIT samples and analyzed by the standard IP-MS procedure, with the peak area of each peptide standard calculated by Skyline.

### 3 | RESULTS

#### 3.1 | Identification of conserved regions of CFPs as diagnostic peptides

LC-MS/MS analysis of equal amounts of trypsin digests of *M. tuberculosis*, *M. avium*, *M. intracellular*are, *M. kansasii*, and *M. abscessus* CFP samples that were ultrafiltered to deplete...
Protein sequences of CFP-10 and EsxN orthologs and/or paralogs detected in Mtb, M. avium, M. intracellulare, M. gordonae, and M. marinum were aligned to identify peptide targets suitable for species identification. Bioinformatic analysis of Mtb CFP-10 identified five tryptic peptides, two of which (CFP-10-p2 and -p4) distinguished CFP-10 expressed by Mtb and M. kansasii (Figure S1A). CFP-10-p2 was detected with the highest signal upon LC-MS/MS analysis of a trypsin digest of recombinant Mtb CFP-10 protein and was selected for further analysis (Figure S1C), while CFP-10-p2 ortholog sequences distinguished all four of the target species and one of the three M. gordonae strains.

Bioinformatic analysis identified five tryptic peptides that were conserved in EsxN orthologs/paralogs, which were identified for “Mtb, M. avium, M. intracellulare, M. gordonae, and M. kansasii”. Among these peptides, semi-conserved EsxN-p1, p2, and p5 sequences could distinguish EsxN orthologs/paralogs produced by these five species with variable discriminatory power to different numbers of species, groups of species, and strains (Figure S1B). EsxN-p2 produced the highest LC-MS/MS signal intensity of these three peptides upon analysis of a trypsin-digested recombinant Mtb EsxN protein sample and was selected for further evaluation (Figure S1C). M. avium EsxN-p2 peptide sequence was identical among all retrieved MAC subspecies and strains (Table S3), while Mtb EsxN-p2 sequence matched protein homologs detected in M. kansasii and M. marinum, and two different EsxN-p2 sequences were detected for different strains of M. gordonae (Figure 1D).

### 3.2 Detection efficiency for peptide variants following antibody capture

EsxN-p2 and CFP-10-p2 sequences selected as mass targets for MS-based mycobacterial species identification differed by one or more amino acids, although two different pairs of EsxN-p2 peptides had equivalent mass and were distinguished from each other by their different retention times and product ions (Figure 1D). Two EsxN-p2, one EsxO-p2, and three CFP-10-p2 sequences were synthesized with heavy isotope labels and subjected to MS/MS analysis to build a spectral library that could be used to identify target peptides specific for Mtb, M. avium, M. intracellulare, M. kansasii, and M. marinum in trypsin-digested MGIT cultures.

To exhibit the performance of peptide variant IP, we calculated the recovery rate as the percentage of total peptide fragment ion peak area before and after IP. In consistency with the above experiment result, 32.6 ± 7.5% and 32.0 ± 9.3% recovery rates were observed for Mtb and MAC EsxN-p2 respectively, whereas they decreased to 7.3 ± 1.1% for.
Mtb EsxO peptide variant (Figure S2), which was caused by antibody affinity variation. The CFP-10-p2 peptide variants, however, showed a quite similar recovery rate among three variants (41.2%–47.8%, Figure S3), supporting the feasibility of using our IP-MS assay in assessing the quantitative changes of each peptide variant in an unbiased way.

### 3.3 Method validation in MGIT samples

To evaluate the feasibility of the proposed assay IP-MS approach, MS detection criteria for EsxN-p2, EsxO-p2, and CFP-10-p2 target peptides were applied to analyze MGIT cultures from a cohort of 74 patients diagnosed with mycobacterial infections, including 52 patients diagnosed with MAC, Mtb complex, M. kansasii, or M. gordonae infections expected to express detectable variants of these target peptides.

Mtb and M. avium EsxN-p2 and Mtb EsxO-p2 target peptides and their heavy-isotope IS peptides were confidently identified in trypsin-digested MGIT cultures of patients infected with each of these species (Figure 2), with each pair of target and IS peptides exhibiting a distinct retention time.

Similar results were obtained when MGIT cultures from patients with active Mtb or M. kansasii infections were analyzed for their corresponding CFP-10-p2 target and IS peptide signals (Figure 3). This LC-MS/MS analysis also detected an extracted ion chromatogram signature similar to that of the M. marinum CFP-10-p2 sequence, but which yielded LC retention times and dotp values that differed from that characteristic of this peptide. Notably, the M. marinum and Mtb CFP-10-p2 peptide monoisotopic peaks differ by 14 Da (Figure S4), so the M+4 isotopic peak (+4) of the Mtb CFP-10-p2 IS peptide ([M+H]+ 2004.0+10) and the monoisotopic peak of M. marinum...
FIGURE 3 Spectral library of EsxN-p2 variants and liquid chromatography-tandem mass spectrometry (LC-MS) peak retention times. (A–C), MS/MS spectra of (A) *Mtb*, (B) *M. marinum*, and (C) *M. kansasii* CFP-10-p2 IS peptides with their annotated fragment ions. (D–F), Extracted ion chromatograms for stable-isotope-labeled IS (heavy) peptides and candidate pathogen-derived CFP-10-p2 peptides of in (D, E) *Mtb*-positive and (F) an *M. kansasii*-positive mycobacterial growth indicator tube (MGIT) samples, indicating the characteristic LC retention time for each candidate peptide (shaded region), where (E) indicates an interference peak determined to be the M+4 isotopic peak of the *Mtb* CFP-10-p2 IS peptide, as described in Figure S6.

CFP-10-p2 ([M+H]+ 2018.0) are captured by the same m/z screen. However, the stochastic distribution of the M+4 isotopic mass among *Mtb* CFP-10-p2 IS peptide fragments should reduce the detection of transition ions corresponding to shorter y-ion peptides and decrease its dotp score for *M. marinum* CFP-10-p2, as observed in this analysis, and these peptide ions and their fragment ions should be detected at different LC retention times.

MS peaks corresponding to the EsxN-p2 sequence shared by *Mtb*, MAC, and *M. kansasii* were detected only in MGIT samples obtained from patients infected with one or more of these pathogens (Figure 4). This shared EsxN-p2 signal was detected in all *Mtb* cases (100%; 22/22) and most *M. kansasii* cases (92.3%; 12/13), with all positive samples exhibiting robust dotp and peak area values (≥0.98 and >2×10^9, respectively), with the peak area of this shared EsxN-p2 peptide tending to be markedly higher in *M. kansasii* cases (median 3.62×10^9 vs. 4.93×10^8). Shared EsxN-p2 peptide signal was also sporadically detected in MGIT samples from patients diagnosed with MAC infections, likely indicating cases infected with particular subspecies or strains of *M. avium*, *M. intracellulare*, or *M. chimaera* that encode the shared EsxN-p2 peptide target sequence (Table S3).

*M. intracellulare* and *M. chimaera* are difficult to distinguish using established methods, [36,37] and four MGIT cultures were identified as *M. intracellulare/M. chimaera* positive after MALDI-MS Biotype analysis. IP-MS results also could not reliably distinguish *M. avium*, *M. intracellulare*, and *M. chimaera* MGIT culture samples, since strains or subspecies of these mycobacteria share one or more EsxN-p2 sequences (Table S4), and thus MGIT cultures with positive IP-MS MAC EsxN-p2 signal must be identified as resulting from MAC infections rather than specific MAC species, subspecies or strains.

Peaks matching MAC-specific EsxN-p2 peptide sequence were detected in all MGIT cultures from MAC cases (100%; 14/14), but not in cultures from individuals...
SAMPLE POSITIVITY FOR SPECIFIC CFP-10, ESXN, AND ESXO PEPTIDES IN MYCOBACTERIAL GROWTH INDICATOR TUBE (MGIT) SAMPLES OF 74 PATIENTS DIAGNOSED WITH MYCOBACTERIAL INFECTIONS. LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY (LC-MS/MS) DOTP AND PEAK AREA VALUES FOR CANDIDATE TARGET SIGNALS CALCULATED IN SKYLINE WERE JUDGED POSITIVE IF THEY EXCEED CUT-OFF VALUES FOR SIGNAL SIMILARITY (DOTP ≥ 0.9) AND INTENSITY (PEAK AREA ≥ 99TH PERCENTILE OF THE CORRESPONDING SIGNAL DETECTED IN 22 CULTURES POSITIVE FOR MYCOBACTERIA THAT DO NOT EXPRESS CFP-10, ESXN, AND ESXO). MKA = M. KANSASII. MMA = M. MARINUM. BLUE AND RED INDICATE NEGATIVE AND POSITIVE RESULTS, RESPECTIVELY.
with any other type of mycobacterial infection. Peaks corresponding to the \textit{Mtb} EsxO-p2 sequence were observed only in MGIT cultures from \textit{Mtb} cases. However, unlike the \textit{Mtb} EsxN-p2 signal that was detected in all these cases, only 13 of these 22 samples (59\%) had detectable \textit{Mtb}-specific EsxO-p2 signal, which was only \( \sim 7\% \) of the matching EsxN-p2 peak signal. None of the MGIT cultures from the negative control group (cases diagnosed with \textit{M. massilense}, or \textit{M. abscessus} or \textit{M. fortuitum} group infections), or from individuals diagnosed with \textit{M. gordonae} infections, revealed dotp or peak area values similar to those observed in patients with active \textit{Mtb}, \textit{M. kansasii}, or MAC infections (Table S4).

\textit{Mtb}-specific CFP-10-p2 peaks were detected only in MGIT cultures of patients with active \textit{Mtb} infections, and 95.5\% (21/22) of these cases had positive CFP-10-p2 signals (Figure 4). \textit{M. kansasii}-specific CFP-10-p2 peaks were also detected only in MGIT cultures from patients diagnosed with \textit{M. kansasii} infections, and 87.5\% (14/16) of these cases had positive CFP-10-p2 signals. Finally, \textit{M. kansasii}-specific CFP-10-p2, MAC-specific EsxN-p2, and shared EsxN-p2 peptide signal were detected in two MGIT cultures that grew both MAC and \textit{M. kansasii} (Figures 4 and 5).

Signal consistent with \textit{M. kansasii} CFP-10-p2 was also detected in three MGIT samples found to contain \textit{M. gordonae} by MALDI-MS Biotype assay results due to the equivalent mass of the CFP-10-p2 peptides produced by \textit{M. kansasii} and two \textit{M. gordonae} strains (a and d), despite two amino acid sequence difference (Figure 6A,B). However, the CFP-10-p2 peptides detected in the three \textit{M. gordonae} cultures revealed a 0.3-min LC retention time, poor similarity (dotp < 0.9), and five unique fragment ions (y8-y12) when compared to the \textit{M. kansasii} CFP-10-p2 peptide (Figure 6C,D) and were thus not considered to represent positive \textit{M. kansasii} CFP-10-p2 signal in this analysis. No signal matching the predicted CFP-10-p2 peak of \textit{M. marinum}, which primarily causes skin infections,[38] was detected in any MGIT sample, consistent with the lack of \textit{M. marinum} cases in this cohort.

Surprisingly, m/z targets with dotp values and RT times corresponding to \textit{Mtb} CFP-10-p2 were sporadically detected in MGIT samples containing mycobacteria that do not express CFP-10 or that express CFP-10 ortholog that cannot produce the target peptide. Peak areas in these samples were not greater than maximum values detected in samples with dotp values < 0.9, except when analyzed immediately after an \textit{Mtb} sample, suggesting the signal detected in these cases was due to an interfering peptide yielding a weak positive signal or a sample carryover effect.

To adjust for potential non-specific background or LC contamination events results from the 22 MGIT cultures of cases infected with rapidly growing NTM species (i.e., \textit{M. abscessus}, \textit{M. fortuitum}, and \textit{M. massilense}) that do not express any of the target peptides were used to specify the criteria for true positive peaks. Briefly, data above the 99th percentile of the extracted ion chromatogram peak area at the characteristic elution time for each target peptide in these 22 samples were selected as the threshold for a true positive signal. Log10 transformed values for these cut-off thresholds were 6.24 for \textit{Mtb}/MAC/\textit{M. kansasii} EsxN-p2 (m/z 797.42), 5.45 for \textit{Mtb} EsxO-p2 (m/z 803.44), 6.34 for MAC EsxN-p2 (m/z 804.43), 7.33 for \textit{Mtb} CFP-10-p2 (m/z 1002.5), 6.26 for \textit{M. kansasii} CFP-10 (m/z 1016.5) and 6.92 for \textit{M. marinum} CFP-10 (m/z 1009.5), while a dotp cut-off threshold of 0.9 was empirically set to reduce detection of non-specific background peaks. These two criteria disallowed 92\% (22/25) of the candidate peaks detected in non-\textit{Mtb} MGIT culture samples.

Thus, integrated criteria employing target peptide LC retention time, dotp, and peak area filtering criteria should permit accurate identification of all target peptides in clinical samples, even in the presence of high background signal from interfering peptide ions or peptide contaminants.

### 3.4 Distinguishing MTBC species by peptide variant combination

Identifying \textit{M. tuberculosis} complex (MTBC) species enables distinguishing between strict human and zoonotic TB and to trace source exposure during epidemiological studies. [39–41] It is difficult to use MALDI-TOF MS assays to differentiate MTBC species, including \textit{M. bovis}, \textit{M. bovis} BCG strain, and \textit{M. africanum} that often cause human diseases.[39,42] The feasibility of identifying MTBC species-distinguishable peptide variant markers through our IP-MS assay was validated using MGIT digest from the laboratory strains. Inherited lack of CFP-10-p2 is the signature of \textit{M. bovis} BCG strain, due to the deletion of RD1 region covering CFP-10 coding gene exsB in its genome.[43]

As a result, we easily distinguished BCG strains by this signature. The bioinformatic analysis predicts two EsxN paralog peptide variants among the involved MTBC species (Table S2). Our result showed that \textit{M. africanum} can be unambiguously differentiated from the rest of MTBC species by one of the two variants (Figure 7). \textit{M. bovis} shared most peptide variants with \textit{Mtb}, except for EsxO-p2. In our first validated cohort with 74 samples, we identified EsxO in 59\% (13/22) MTBC infected samples. Lacking EsxO in the rest 41\% of samples may be explained by either the selective expression of EsxO or the actual infection with \textit{M. bovis}. Further investigation of the species origin in these samples through sequencing is needed to elucidate the reason.
3.5 High-throughput MALDI-TOF MS as assay readout

To translate the proposed IP-MS assay into a clinical microbiologic laboratory, we tested the feasibility of employing a MALDI-TOF MS that is commonly equipped in such a laboratory as assay readout. To avoid the ambiguity of isomeric peptide variants such as those from *M. gordanae* and *M. kansasii*, we selected 21 MGIT samples including eight *Mtb*, three MAC, two *M. kansasii*, one *M. bovis* BCG, two *M. abscessus*, and 5 unrelated species as controls, and analyzed them by this approach. This IP-MS assay had an overall diagnostic sensitivity of 92.9% (13/14) with 100% specificity (7/7) (Table S5). This diagnostic sensitivity was reduced by the inclusion of a culture of a rare *M. bovis* BCG arising after vaccination since *M. bovis* BCG does not express EsxB and in the absence of EsxB expression, it is not possible to distinguish an infection arising from *M. kansasii* or an *Mtb* complex species using the shared EsxN-p2 variant.

3.6 4-h turnaround time setup for rapid diagnosis

Inspired by our previous work on developing a rapid diagnostic test for Ebola infectious disease,[44] we tested the 4-h sample preparation workflow in the multiplexed IP-MS assay of EsxN and CFP-10 peptide variants. We selected four representative MGIT samples infected with *Mtb* or *M. kansasii* or MAC, and shorten the time of trypsin digestion.
FIGURE 6  Differentiation of isomeric M. gordonae and M. kansasii. Culture filtrate protein (CFP) 10-p2 peptides (A, B), PRM-MS spectra for CFP-10-p2 peptides isolated from mycobacterial growth indicator tube (MGIT) cultures of patients infected with (A) M. kansasii and (B) M. gordonae. Monoisotopic masses of five unique fragment ions (colored; y8-y12) are indicated above their peaks and two amino acids that vary between these isomers are marked in red in their sequence. (C, D) Extracted ion chromatograms of fragment ions of the CFP-10-p2 peptides detected in (C) M. kansasii-positive and (D) M. gordonae-positive MGIT samples.

of MGIT samples from 16 to 1 h. It showed that the peptide variant targets were all identified correctly for their species origin (Figure S5–S10). All samples showed a relatively low abundance of peptide targets in the initial 16-h digestion method, as shown in Figure 4. The result proves that a 4-hr workflow, that is, 0.1-h pH adjustment, 1-h digestion, 1-h enrichment, 1-h washing/elution, and 0.5-h LC-MS/MS analysis allows mycobacterial species identification.

4 | DISCUSSION

Secreted proteins, particularly factors associated with virulence represent promising biomarkers for the rapid, species-specific diagnosis of infections from early-positive mycobacterial cultures inoculated with clinical samples. Candidate biomarker peptides analyzed in this study are derived from the ESX-1 and ESX-5 substates CFP-10 and EsxN, which play essential roles in mycobacterial growth and avoiding immune suppression, are detected at high abundance in CFP samples [22] and thus have potential utility as diagnostic biomarkers, as supported by our results.

Several NTM can produce infections resembling TB, but which respond to different drug regimens, and which may represent a significant percentage of suspected TB cases (4%–30%) [45,46] and chronic/multi-drug resistant TB cases (18%–27%) cases [47–49] Our IP-MS assay approach uses immunoprecipitation to capture and enrich a pair of semi-conserved peptides from CFP-10 and EsxN protein orthologs expressed by a number of slow-growing
mycobacteria responsible for most mycobacterial respiratory infections, and these peptides are then analyzed by LC-MS/MS to identify the mycobacteria responsible for their expression at the species or complex level. IP-MS analysis of clinical specimen MGIT cultures accurately distinguished patients with MAC infections from patients infected with \textit{Mtb}, \textit{M. kansasii}, or \textit{M. gordonae}. This has significant clinical implications, since effective treatment of MAC, \textit{Mtb}, and \textit{M. kansasii} infections require the use of different drug regimens,\cite{50,51} while \textit{M. gordonae}-positive cultures frequently represent MGIT culture contamination and this can thus affect the decision to treat a patient.\cite{14} Further, this IP-MS was able to identify both species in two MGIT cultures that grew both MAC and \textit{M. kansasii}. Such cultures can potentially confound species identifications in sequencing-, probe-, or MS-based methods if they produce overlapping target signals or signal fingerprints that obscure the identity of the species present.
Both immunoprecipitation and LC-MS/MS are essential for the performance of our described IP-MS assay. Specific antibodies that can bind sequence variants of CFP-10-p2 and/or EsxN-p2 sequences that are encoded by our target mycobacteria species of interest are required to selectively enrich these peptides to allow sensitive and specific detection against the high non-specific background produced by the protein constituents of the MGIT culture media, and other mycobacteria-derived proteins, particularly during the early growth phase when mycobacteria-derived proteins are present at very low concentration relative to the protein content contributed by the MGIT culture media. Specific antibodies produced for this proof-of-principle study were raised against full-length *Mtb* CFP-10-p2 and EsxN-p2 peptide sequence and yielded variable detection efficiency in an IP-MS assay when employed to detect equimolar amounts of species-specific synthetic peptide variants spiked into MGIT culture media. This variation appeared to arise from differences in antibody affinity rather than LC-MS/MS detection efficiency among these variants since LC-MS/MS analysis of an equimolar mixture of all CFP-10-p2 and EsxN-p2 IS peptides in PBS did not exhibit significant peak area differences (data not shown). The reduced affinity of the anti-CFP-10-p2 antibodies for the *M. kansasii* peptide did decrease its detection rate in early positive MGIT samples, although it might increase the amount of time required to detect positive biomarker results when attempting to detect positive cultures and identify mycobacterial species prior to MGIT culture positivity. Affinity differences for species-specific CFP-10-p2 and EsxN-p2 could, however, be attenuated by producing monoclonal antibodies to eight to eleven amino acid regions of CFP-10-p2 and EsxN-p2 sequence that contain single amino acid variations and selecting for clones that exhibit similar affinity for all target peptides.

LC-MS chromatography and spectra data both provide information that can distinguish species-specific isomers of target peptides (e.g., *M. gordonae* and *M. kansasii* CFP-10-p2) by the characteristic retention times of their peptide ions and co-eluting fragment ions. Species-specific target peptides are identified using pre-established criteria for the retention times of their individual peak ions and their resulting unique fragment ions, the similarity of these spectra to MS/MS reference libraries for the target peptides, and peak area thresholds that differentiate signal from background. Spectral libraries are also employed in commercial MALDI-TOF-MS-based diagnostic assays, but strain-specific differences and growth conditions can alter mycobacterial protein expression to influence spectral library data and subsequent species identifications that rely on the correspondence with species-specific protein expression signatures. Such mycobacterial phenotype differences do not affect IP-MS results, however, since this assay relies on the positive detection of distinct target peptide sequences from two actively secreted virulence factors to make species-specific identifications.

This study indicates that IP-MS assay can rapidly and specifically identify mycobacteria commonly detected in mycobacterial isolates from clinical specimens (*M. tuberculosis*, MAC, and *M. kansasii*), as well as one species frequently associated with MGIT contamination (*M. gordonae*), at the first sign of MGIT growth. Small proof-of-concept studies demonstrated that our target peptides were also efficiently detected when trypsin digestions were reduced to 1 h and when MALDI-TOF MS was used as the assay readout to further the reduced assay performance time and increase its throughput. However, further studies should evaluate whether this approach could be employed for target detection prior to MGIT positivity. Further studies should also investigate the potential of this approach to identify mycobacterial infections using plasma or serum as the analysis sample, since we have reported that a similar approach that analyzes a different CFP-10 peptide can detect all manifestations of active TB disease, including culture-negative pulmonary and extrapulmonary cases, in a diverse set of affected patient populations.[52,53]

**ACKNOWLEDGEMENTS**

The following reagents were obtained through BEI Resources, NIAID, NIH: Mycobacterium tuberculosis, Strain H37Rv, and CDC1551, Culture Filtrate Proteins, NR-14825, and NR-14826. The work was primarily supported by research funding provided by NIH [R01HD090927, R01AI122932, R01AI113725, and R21AI126361-01], the Department of Defense [W81XWH1910926], Arizona Biomedical Research Commission (ABRC) young investigator award, and the Weatherhead Presidential Endowment Fund.

**CONFLICT OF INTEREST**

T.H. and Christopher J. Lyon report other interests from NanoPin Technologies, Inc., outside the submitted work. In addition, T.H. and Qingbo Shu have a patent (“Compositions and methods of determining a level of infection in a subject”) licensed to NanoPin Technologies, Inc. The rest of the authors declare no competing interests.

**AUTHOR CONTRIBUTIONS**

Qingbo Shu, Jia Fan, Lingpeng Zhan, Xiangxing Kong, and Yifan He designed, performed, and analyzed the experiments. Meena Rajagopal, Suwatchareeporn Rotcheewaphan, and Adrian M. Zelazny provided the clinical isolates of MGIT samples and conducted mycobacterial identification using established methods, provided suggestions, and critically read the paper. Tony Hu and Wei Sha conceptualized and designed experiments, analyzed the
results. Qingbo Shu and Christopher J. Lyon contributed equally to the writing of the manuscript.

**ORCID**

Tony Hu [https://orcid.org/0000-0002-5166-4937](https://orcid.org/0000-0002-5166-4937)

**REFERENCES**

1. H. C. King, T. Khera-Butler, P. James, B. B. Oakley, G. Erenso, A. Aseffa, R. Knight, E. M. Wellington, O. Courtenay, *PLoS One* 2017, 12, e0173811.
2. E. Rallis, E. Koumantaki-Mathioudaki, *Expert Opin. Pharmacother.* 2007, 8, 2965.
3. W. Hoeftsloot, J. Van Ingen, C. Andrejak, K. Ângeby, R. Bauriaud, P. Bemer, N. Beylis, M. J. Boeree, J. Cacho, V. Chiêta, E. Rallis, E. Koumantaki-Mathioudaki, *Am. J. Respir. Crit. Care Med.* 2006, 173, 1097.
4. J. E. Richter, S. RüSch-Gerdes, D. Hillemann, E. Rallis, E. Koumantaki-Mathioudaki, *Am. J. Thorac. Dis.* 2001, 1769.
5. A. Shah, V. Bieren, *Front. Cell Infect. Microbiol.* 2016, 6, 49.
6. K. E. Holt, P. Mcdam, P. V. K. Thai, N. T. T. Thuong, D. T. M. Ha, N. N. Lan, H. N. Lan, N. T. Q. Nhu, H. T. Hai, V. T. N. Ha, G. Thwaites, D. J. Edwards, A. P. Nath, K. Pham, D. B. Ascher, J. Farrar, C. C. Khor, Y. Y. Teo, M. Inouye, M. Caws, S. J. Dunstan, *Nat. Genet.* 2018, 50, 849.
7. V. Srinivasan, V. T. N. Ha, D. N. Vinh, P. V. K. Thai, D. T. M. Ha, N. N. Lan, H. T. Hai, T. M. Walker, D. D. A. Thu, S. J. Dunstan, G. E. Thwaites, P. M. Ashton, M. Caws, N. T. T. Thuong, *Clin. Infect. Dis.* 2020, 71, e532.
8. W. Zhang, Q. Shu, Z. Zhao, J. Fan, C. J. Lyon, A. M. Zelazny, Ye Hu, *Clin. Proteome.* 2018, 15, 1.
9. D. E. Griffith, T. Aksamit, B. A. Brown-Elliott, A. Catanazzo, C. Daley, F. Gordin, S. M. Holland, R. Horsburgh, G. Huitt, P. F. Iademarco, M. Iseman, K. Olivier, S. Ruoss, C. F. Von Reyn, R. J. Wallace, K. Winthrop, *Am. J. Respir. Crit. Care Med.* 2007, 175, 367.
10. A. M. Abdallah, M. Rashid, S. A. Adroub, H. Elabdalouai, S. Ali, D. Van Soolingen, W. Bitter, A. Pain, *J. Bacteriol.* 2012, 194, 3282.
11. J. Van Ingen, R. De Zwaan, R. Dekhuijzen, M. Boeree, D. Van Soolingen, *J. Bacteriol.* 2009, 191, 5865.
12. P. R. K. Behra, B. M. F Pettersson, M. Ramesh, S. Dasgupta, L. A. Kirsebom, *Sci. Rep.* 2019, 9, 10259.
13. P. R. K. Behra, S. Das, B. M. F Pettersson, L. Shirreff, T. Ducote, K.-G. Jacobsson, D. G. Ennis, L. A. Kirsebom, *Sci. Rep.* 2019, 9, 4603.
14. K. M. Dobos, K. Swiderek, K. H. Khoo, P. J. Brennan, J. T. Belisle, *Infect. Immun.* 1995, 63, 2846.
15. P. G. Saleeb, S. K. Drake, P. R. Murray, A. M. Zelazny, *J. Clin. Microbiol.* 2011, 49, 1790.
16. C. S. Bakshi, D. H. Shah, R. Verma, R. K. Singh, M. Malik, *Vet. Microbiol.* 2005, 109, 217.
17. M. I. Gröschel, F. Sayes, R. Simeone, L. M. Majlessi, R. Brosch, *Emerg. Microbiol. Infect.* 2019, 4, 677.
18. S. Das, B. M. F Pettersson, P. R. K. Behra, A. Mallick, M. Cheramie, M. Ramesh, L. Shirreff, T. Ducote, S. Dasgupta, D. G. Ennis, L. A. Kirsebom, *Emerg. Microbiol. Infect.* 2019, 4, 677.
19. E. B. Sivasankar, E. M. Thomas, K. M. Morelli, D. J. Quinlan, *J. Clin. Microbiol.* 2019, 57, e00910.
20. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
21. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
22. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
23. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
24. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
25. J. T. O. Fall, E. C. Timme, L. M. Zou, N. H. L. Lan, N. T. T. Thuong, D. T. M. Ha, *J. Clin. Microbiol.* 2005, 43, 125.
39. A. O. Bajaj, S. Saraswat, J. E. A. Knuuttila, J. Freeke, J. B. Stielow, A. P. Barker, *Front. Cell Infect. Microbiol.* **2021**, *11*, 658880.

40. T. A. Halse, V. E. Escuyer, K. A. Musser, *J. Clin. Microbiol.* **2011**, 49, 2562.

41. P. Landolt, R. Stephan, S. Scherrer, *Sci. Rep.* **2019**, *9*, 1850.

42. J. A. O’Connor, et al., *J. Clin. Microbiol.* **2020**, *58*, e02077-19.

43. G. G. Mahairas, P. J. Sabo, M. J. Hickey, D. C. Singh, C. K. Stover, *J. Bacteriol.* **1996**, *178*, 1274.

44. Q. Shu, T. Kenny, J. Fan, C. J. Lyon, L. H. Cazes, T. Y. Hu, *PLoS Pathog.* **2021**, *17*, e1008039.

45. G. Aliyu, S. S. El-Kamary, A’Le Abimiku, C. Brown, K. Tracy, L. Hungerford, W. Blattner, *PLoS One* **2013**, *8*, e63170.

46. H. J. Jun, K. Jeon, S. W. Um, O. J. Kwon, N. Y. Lee, W. J. Koh, *Respir. Med.* **2009**, *103*, 1936.

47. A. H. Shahraki, P. Heidarieh, S. Z. Bostanabad, A. D. Khosravi, M. Hashemzadeh, S. Khandan, M. Biranvand, D. E. Schraufnagel, M. Mirsaeidi, *Eur. J. Intern. Med.* **2015**, *26*, 279.

48. G. Badoum, N. Saleri, M. S. Dembele, M. Ouedraogo, G. Pinsi, K. Boncoungou, V. Bonkoungou, E. Birba, P. Miotto, G. B. Migliori, D. M. Cirillo, A. Matteelli, *Eur. Respir. J.* **2011**, *37*, 1283.

49. M. Maiga, S. Siddiqui, S. DIALLO, B. DIARRA, B. Traoré, Y. R. SHEA, A. M. ZELAZNY, B. P. P. DEMBELE, D. GOITA, H. KASSAMBARA, A. S. HAMMOND, M. A. POLIS, A. TOUNKARA, *PLoS One* **2012**, *7*, e36902.

50. C. Gilpin, A. Korobitsyn, G. B. Migliori, M. C. Raviglione, K. Weyer, *Eur. Respir. J.* **2018**, *51*, 1800098.

51. C. L. Daley, J. M. Iaccarino, C. Lange, E. Cambau, R. J. Wallace, C. Andrejak, E. C. Böttger, J. Brozek, D. E. Griffith, L. Guglielmetti, G. A. Huitt, S. L. Knight, P. Leitman, T. K. Marras, K. N. Olivier, M. Santin, J. E. Stout, E. Tortoli, J. Van Ingen, D. Wagner, K. L. Winthrop, *Clin. Infect Dis* **2020**, *71*, 905.

52. C. Liu, Z. Zhao, J. Fan, C. J. Lyon, H.-J. Wu, D. Nedelkov, A. M. Zelazny, K. N. Olivier, L. H. Cazes, S. M. Holland, E. A. Graviss, Ye Hu, *Proc Natl Acad Sci USA* **2017**, *114*, 3969.

53. J. Fan, H. Zhang, D. T. Nguyen, C. J. Lyon, C. D. Mitchell, Z. Zhao, E. A. Graviss, Ye Hu, *BMC Med* **2017**, *15*, 188.

54. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsengberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaíno, *Nucleic Acids Res* **2019**, *47*, D442.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Q. Shu, M. Rajagopal, J. Fan, L. Zhan, X. Kong, Y. He, S. Rotcheewaphan, C. J. Lyon, W. Sha, A. M. Zelazny, T. Hu, *VIEW* **2022**, *3*, 20210019. [https://doi.org/10.1002/VIW.20210019](https://doi.org/10.1002/VIW.20210019)