



mlstverse-web: advancing real-time identification of mycobacteria from sputum using targeted sequencing NALC-Seq

Yuki Matsumoto¹ · Kiyoharu Fukushima^{1,2,3} · Shunran Zhang¹ · Yuzy Fauzyah¹ · Daisuke Motooka^{1,5} · Haruko Saito² · June Yamauchi² · Tadayoshi Nitta² · Takuro Nii⁴ · Takanori Matsuki⁴ · Kazuyuki Tsujino⁴ · Keisuke Miki⁴ · Hiroshi Kida² · Shota Nakamura^{1,5,6}

Received: 27 February 2025 / Revised: 21 April 2025 / Accepted: 30 April 2025
© The Author(s) 2025

Abstract

The identification of mycobacteria, including *Mycobacterium tuberculosis* (MTB) and non-tuberculous mycobacteria (NTM), remains a critical challenge in clinical and public health settings due to their pathogenicity and increasing drug resistance. Traditional diagnostic methods, such as PCR and mass spectrometry, are limited by species detectability and prolonged culture requirements. To address these limitations, this study introduces a novel approach, named NALC-Seq, which integrates next-generation sequencing (NGS) and target capture sequencing for the direct and comprehensive identification of mycobacteria from sputum samples. A total of 54 patients were prospectively or retrospectively enrolled between April 6, 2021, and November 8, 2022, and sputum samples were subjected to next generation sequencing. The NALC-Seq methodology utilizes custom-designed RNA probes for targeted enrichment of mycobacterial DNA, coupled with species identification via the mlstverse-web system. This system integrates web, upload, head, and cloud nodes to streamline data management and analysis. Our evaluation demonstrated that NALC-Seq achieved high sensitivity (98.1%) and accuracy (83.3%) in detecting diverse mycobacterial species, including rare and drug-resistant subspecies. There were discrepancies with MGIT-Seq in nine samples. These samples exhibited low smear positivity rates, and the detection of environmental mycobacteria suggested potential contamination. Furthermore, real-time identification using the MinION device significantly reduced the turnaround time from 686 to 19 h. These findings highlight the potential of NALC-Seq and mlstverse-web as an integrated diagnostic solution for overcoming the limitations of current methods. The proposed approach offers a rapid, accurate, and comprehensive strategy for mycobacterial identification, with significant implications for clinical diagnostics and public health surveillance.

Keywords Mycobacteria · Species identification · Real-time diagnostics · Targeted capture sequencing · Drug resistance prediction

✉ Kiyoharu Fukushima
fukushima@imed3.med.osaka-u.ac.jp

✉ Shota Nakamura
nshotan@gen-info.osaka-u.ac.jp

¹ Department of Infection Metagenomics, Bioinformatics Center, Research Institute for Microbial Diseases (RIMD), Osaka University, Suita, Osaka, Japan

² Department of Clinical Laboratory, National Hospital Organization, Osaka Toneyama Medical Center, Toyonaka, Osaka, Japan

³ Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Osaka, Japan

⁴ Department of Respiratory Medicine, National Hospital Organization, Osaka Toneyama Medical Center, Toyonaka, Osaka, Japan

⁵ Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Osaka, Japan

⁶ Center for Infectious Disease Education and Research, Osaka University, Suita, Osaka, Japan

Background

Mycobacteria constitute a substantial genus comprising over 200 species, including *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis (TB), *Mycobacterium leprae*, the causative agent of leprosy, and non-tuberculous mycobacteria (NTM). TB ranks among the top three infectious diseases globally, with an estimated 10.6 million cases reported (Pai et al. 2023). These mycobacterial species exhibit pathogenic potential and frequently display resistance to multiple drugs, necessitating varied treatment strategies. Multidrug- or rifampicin-resistant MTB has afflicted approximately 450,000 individuals in 2021, primarily in developing countries (Vanino et al. 2023; Prasad et al. 2014). Conversely, in developed nations, the incidences of NTM diseases are increasing whereas the number of TB cases is declining (Falkinham 2016; Chin et al. 2020). NTM strains demonstrate a broad antimicrobial resistance, with significant variation among subspecies, notably the *Mycobacterium abscessus* complex, which exhibits high-level multidrug resistance (Saxena et al. 2021; Johansen et al. 2020; Shahraki et al. 2015). In contrast, *Mycobacterium avium* complex (MAC), one of the most frequently encountered NTM species, often exhibits multidrug resistance, posing a significant challenge in clinical treatment (Busatto et al. 2019). Misdiagnosis of NTM diseases has been reported in regions with high TB prevalence (Saptawati et al. 2019; Maiga et al. 2012). Infections caused by multidrug-resistant MTB and emerging NTM strains pose significant global health challenges (Adjemian et al. 2017; Prevots and Marras 2015; Namkoong et al. 2016).

Comprehensive identification of mycobacterial species is time-consuming due to the limited testing methods covering the full spectrum of mycobacterial diversity, often creating bottlenecks in treatment. While PCR-based methodologies are commonly employed for mycobacterial species identification, their accuracy is constrained within a limited species range (Avaniss-Aghajani et al. 1996; Hyeoung et al. 2000; Horne et al. 2019). Mass spectrometry offers comprehensive identification potential but is limited to species level accuracy and requires bacterial culture, extending the process by up to eight weeks (Alcolea-Medina et al. 2019; Claydon et al. 1996). Although next-generation sequencing (NGS) allows accurate identification at the subspecies level and facilitates phenotype estimation, it suffers from prolonged culture periods akin to conventional methods. Proof of concept studies utilizing NGS for rapid mycobacterial identification have reported same-day identification, particularly focusing on tuberculosis from sputum samples (Votintseva et al. 2017). Alternatively, highly sensitive identification methods have

been proposed using RNA probes targeting mycobacteria, enabling major mycobacterial identification from sputum and cultures using a Mycobacteria Growth Indicator Tube (MGIT) (Kambli et al. 2021; He et al. 2020; Mann et al. 2023). Despite the existence of some proof-of-concept studies, routine mycobacterial identification still relies on bacterial culture, which limits the practical application of direct sputum sequencing. Furthermore, the spectrum of mycobacterial species identifiable through direct sequencing remains limited in comparison to the over 200 recognized species within the mycobacterial genus.

In a previous study, we developed the core system of *mlstverse*, a method for comprehensive mycobacterial species identification using NGS (Matsumoto et al. 2019). We successfully reduced the identification period to approximately 14 d from MGIT samples (Fukushima et al. 2023). However, the widespread adoption of NGS-based identification methods faces barriers, including a demand for bioinformatics expertise. The skill set required for handling NGS data differs from that of clinical tests conducted by laboratory technicians, necessitating bioinformaticians. Although these issues are being addressed through the development of various bioinformatics tools (Yang et al. 2022; Hunt et al. 2019; Florensa et al. 2022; Feldgarden et al. 2019; Sheka et al. 2021), conventional analysis pipelines require the initiation of analysis after sequence completion, resulting in delays. Unlike short-read sequencing, the clinical application of Nanopore sequencing provides sequential output of reads and the potential for shorter turnaround times, making it promising for rapid diagnostics in infectious diseases (Lang 2022; Zheng et al. 2023). In this study, we adapted our identification method to the MinION device, enabling real-time sequence analysis and immediate result visualization through a web interface. We also established a target capture sequencing-based method designed for mycobacteria, facilitating genomic DNA enrichment in sputum and enabling direct sequencing without culture. Using this platform, we demonstrated comprehensive mycobacterial identification at the subspecies level within 24 h.

Methods

System overview of *mlstverse*-web

The *mlstverse*-web system comprises four key components (Fig. 1A): *Web* nodes, serving as the front end for analysis and result visualization; *Upload* nodes, operating on a host connected to the MinION sequencer and facilitating the transfer of raw signal outputs during sequencing to cloud storage; *Head* nodes, responsible for managing file exchanges and job management; and *Cloud* nodes, which

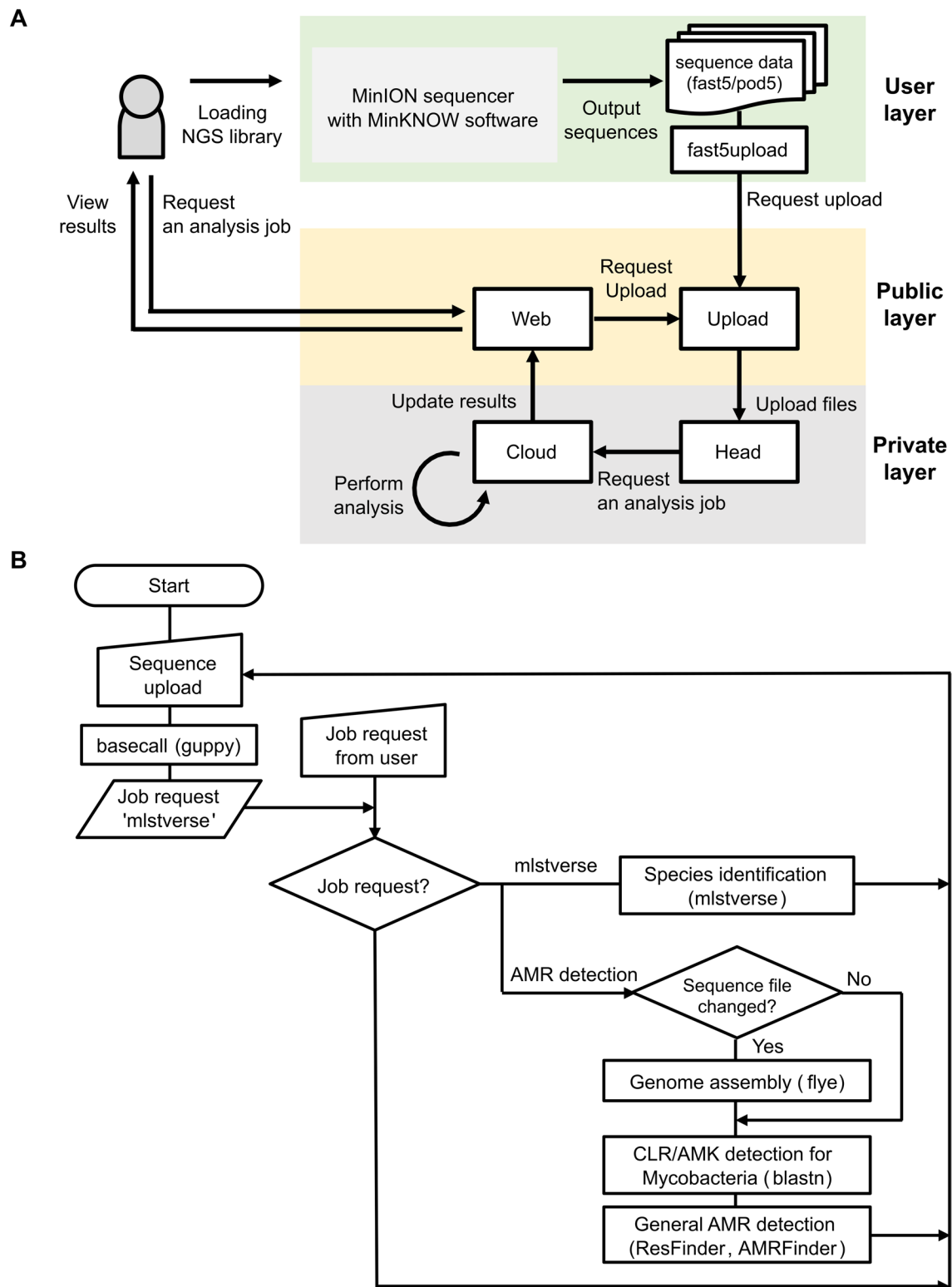


Fig. 1 System overview. **A** System components: The system comprises *Web*, *Upload*, *Head*, and *Cloud* nodes. **B** Analysis pipeline for species identification and antimicrobial resistance (AMR) detection

execute actual analysis processing in the cloud. Initially, upon commencing sequencing on the MinION device, the *Upload* node transfers the raw sequence files to cloud storage. The *Head* nodes monitor these file updates and, upon detection of an update, initiate a new analysis job through slurm, an open source job management system (Yoo et al. 2003). Once the sequencing files are uploaded, the analysis jobs are automatically executed on the cloud nodes. The analysis pipeline is illustrated in Fig. 1B. Given the prevalence of multidrug resistance in pulmonary mycobacterial infections, in addition to species identification, the system conducts drug resistance prediction through a web interface. This includes evaluating genetic markers such as *rrs*, *rhl*, and *erm(41)* to predict resistance to key drugs such as clarithromycin and amikacin. Moreover, assessments are conducted utilizing tools such as ResFinder and AMRFinder software (Fukushima et al. 2023; Florensa et al. 2022; Feldgarden

et al. 2019). Results are accessible via the *Web* node using a web browser, facilitating functional analyses, re-analyses, and results downloads. The *Web* node is publicly accessible at <https://mlstverse.org> for all users (Fig. 2).

Clinical sample collection, bacterial culture and MIC testing

A total of 54 patients were enrolled, including individuals previously diagnosed with relatively rare mycobacterial species. This cohort comprised 21 males and 23 females with an average age of 72.48 (± 13.63) years, as detailed in Supplementary Table 2. Sputum samples were prospectively collected between April 6, 2021, and May 27, 2021. Five of these samples were used for the evaluation of mycobacteria quantity via metagenomic shotgun sequencing, and 39 samples were used for NALC-Seq evaluation. In addition, 15

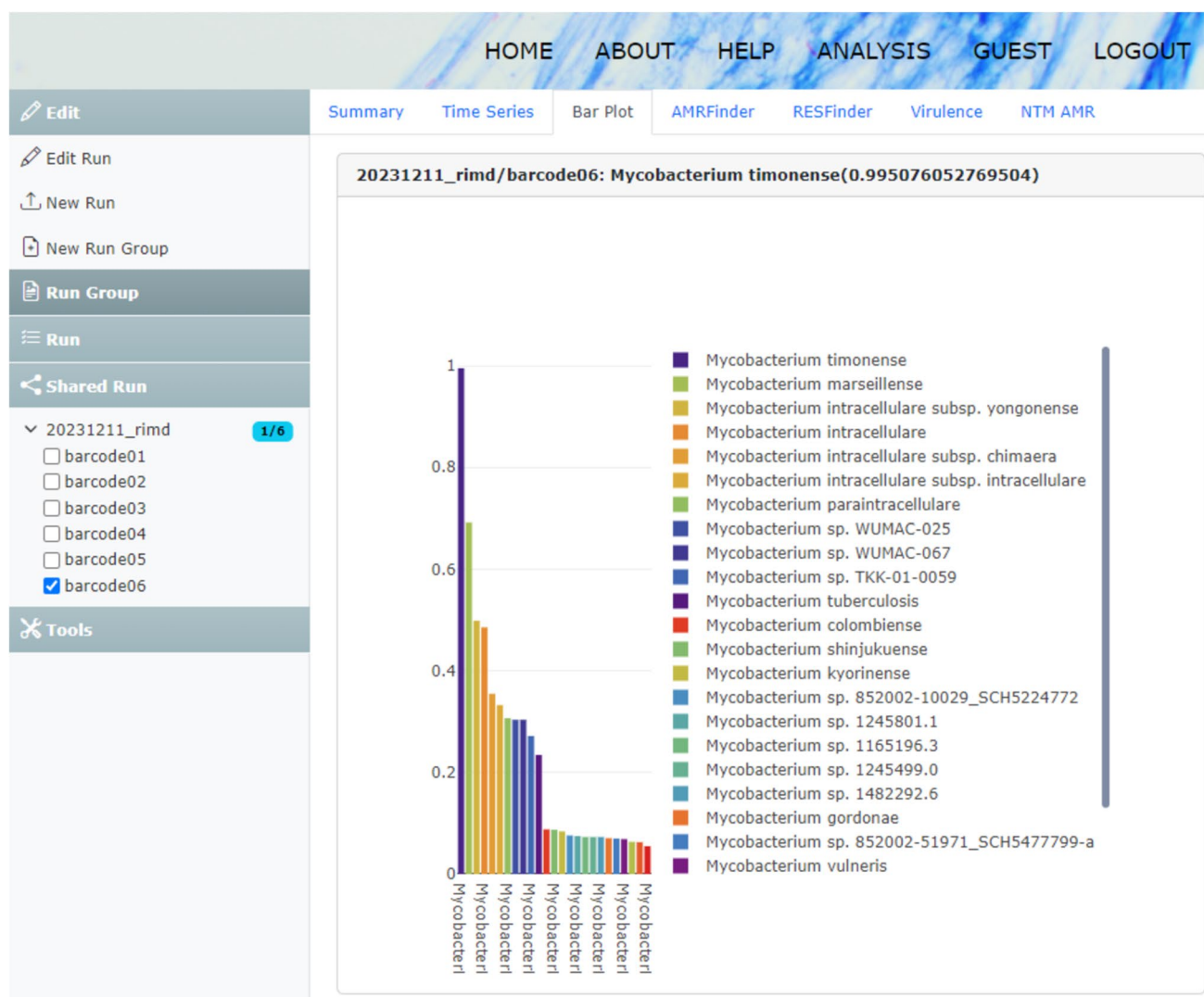


Fig. 2 Result overview on mlstverse-web

samples containing strains less frequently encountered in the prospective collection, such as MTB, rapidly growing mycobacteria (RGM), *Mycobacterium kansasii*, and *Mycobacterium lentiflavum*, were retrospectively selected from specimens collected between June 22, 2022, and November 8, 2022, and used for NALC-Seq evaluation. Sputum samples obtained from patients underwent preprocessing with the sputum agglutination procedure using Sputazyme (Kyokuto Pharmaceutical Industrial Co., Ltd., Tokyo, Japan) at room temperature for 15 min, followed by centrifugation and supernatant removal. The resulting pellet then underwent N-acetyl-L-cysteine-sodium hydroxide (NALC-NaOH) treatment using MycoPrep (Becton Dickinson, BD, USA) according to the manufacturer's protocol for 15 min, after which the supernatant was discarded. The processed material was used for conventional culture, MGIT culture, and DNA extraction. Conventional culture involved inoculation onto 2% Ogawa medium and incubation at 36 °C for up to 8 weeks. MGIT cultures were performed using the Bactec MGIT 960 system (BD) for up to 6 weeks. MIC testing for clarithromycin and amikacin was performed as previously described (Fukushima et al. 2023).

DNA extraction and library preparation

Genomic DNA was extracted from residual sputum and Mycobacteria Growth Indicator Tube (MGIT) cultures using the bead shaking method. Briefly, 250 µL of bacterial solution underwent mechanical disruption with 0.1 mm diameter glass beads (MN Bead Tubes Type B, cat# 740,812.50, Takara Bio, Shiga, Japan), using a Magnalyzer device for 45 s at a speed of 4500 rpm. Following mechanical disruption, the bead tubes were centrifuged at 13,000 rpm for 5 min, followed by heat treatment at 95 °C for 5 min using a block incubator. Subsequently, 100 µL of the supernatant was collected.

The mycobacterial gene panel was designed using SureDesign (<https://earray.chem.agilent.com/suredesign>) to cover the complete loci set of the MLST database of mlstverse. On average, 150 ng of extracted DNA was sheared using the Covaris M220 focused-ultrasonicator (Covaris, Woburn, MA, USA) to 600 bp. Sequence libraries comprising 54 samples were prepared using a custom SureSelect Low Input Target Enrichment System (Agilent Technologies Inc. Santa Clara, CA, USA), following the manufacturer's instructions, and sequenced using MGI DNBSEQ-G400 (MGI, Shenzhen, PRC) with paired-end 100 bp reads. The sequencing reads were analyzed and identified using mlstverse. For fifteen samples newly collected in this study, MGIT-Seq was performed and results were compared with the targeted capture sequencing. The sequencing reads were analyzed and identified using

the mlstverse. For the 15 samples newly collected in this study, MGIT-Seq was conducted, and the results were subsequently compared with those derived from targeted capture sequencing.

For real-time sequencing analysis, the library was prepared from the amplicon produced in the enrichment step of the 2nd PCR using the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies (ONT), Oxford, UK) with Native Barcoding Expansion 1–12 (EXP-NBD104, ONT) and sequenced using MinION with flow cell R9.4.1 (FLO-MIN106D, ONT).

Data analysis and pipeline

Raw output data obtained from MinION sequencing was basecalled using guppy software on high accuracy mode. Long reads (> 1000 bp) with an average quality score below Q10 were filtered, while short reads from MiSeq and DNBSeq-G400 had adapter sequences and regions below Q30 trimmed. For mlstverse analysis, long and short sequence reads were mapped to MLST loci sequence file using the minimap2 (Li 2018) and bwa-mem (Li 2013), respectively. The MLST score, ranging from 0 to 1 according to the proportion of detected loci, was calculated based on mapping results using mlstverse (Matsumoto et al. 2019). Species were identified based on the profile with the highest score, and profiles with scores below 0.1 were excluded. Genome assembly was conducted using the entire fastq file using flye (Kolmogorov et al. 2019). Clarithromycin and amikacin resistance were predicted by searching the sequences of the *rrs*, *rrl* and *erm(41)* genes using blastn (Camacho et al. 2009) and evaluated the following method described in Fukushima et al. (2023). Resistance mutations A2058G/A2059G (*rrl*) and A1408G (*rrs*) were evaluated. For *erm(41)*, gene presence, truncations, and the T28 C mutation were assessed. General drug resistance detection was conducted using ResFinder (Florensa et al. 2022) and AMRFinder (Feldgarden et al. 2019).

Metagenomic sequencing analysis of sputum mycobacteria

A sequencing library was prepared from the genomic DNA derived from the five sputum samples using Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The sequencing run was performed using the MiSeq Reagent Kit v2 (300 Cycles, Illumina). The taxonomic compositions were evaluated using Kraken2 (Wood et al. 2019) and species identification was performed using mlstverse (Matsumoto et al. 2019).

Evaluation of target capture sequencing for mycobacterial detection

Mycobacterium agri JCM 6377 was cultured in Middlebrook 7H9 medium with OADC Enrichment (BD-BBL) at 37 °C with 150 rpm on a shaker incubator. Genomic DNA was extracted from the bacterial pellet using Power Soil Pro according to the manufacturer's instructions. The input gDNA was prepared for each 10 ng of human gDNA, adding 1, 10, 100, 1000, or 10,000 pg of *M. agri* gDNA, respectively. Sequence libraries were prepared using a custom SureSelect Low Input Target Enrichment System (Agilent Technologies Inc. Santa Clara, CA, USA), according to the manufacturer's instructions, and sequenced using MGI DNBSEQ-G400 (MGI, Shenzhen, PRC) with paired-end 100 bp reads.

Benchmarking of captured sputum sequences

Input DNA for benchmarking sensitivity and specificity of capture sequencing was prepared by mixing the gDNA of humans, Microbial Community DNA Standard (ZymoBIOMICS, D6305), and *M. agri* in the ratio of 1:0:0 (Human), 0:1:0 (Mock), 0:0:1 (*M. agri*), 1:0:1 (A50_H50), 0:1:1 (A50_M50), and 0:9:1 (A10_M90) respectively. Sequencing libraries were prepared in the same manner as others using custom SureSelect libraries.

Real-time mycobacterial detection of clinical sputum samples

A sequencing library was prepared from the genomic DNA derived from the sputum sample #54 utilizing custom SureSelect libraries for real-time sequencing analysis. The sequencing output of MinION was automatically uploaded to the mlstverse web server by the fast5upload program (<https://github.com/ymatsumoto/fast5upload>). The fast5 files were basecalled using guppy basecaller. Species identification analysis was performed using mlstverse v0.2.1 and the mlstverse.Mycobacterium.db database. All analyses were automatically submitted from the head node via the job management system and executed on the cloud node.

Results and discussion

Presence of mycobacterium in sputum metagenome

Clinical samples such as sputum typically comprise over 50% of DNA derived from the human genome or oral bacteria (Votintseva et al. 2017; Doughty et al. 2014). Nicholls et al. demonstrated that achieving adequate sequencing depth is challenging when the amount of target DNA in the

metagenome is very low, even with the use of whole genome amplification (Nicholls et al. 2019). Prior to evaluating target capture sequencing, we assessed the sputum metagenome for the presence of mycobacteria using metagenomic shotgun sequencing. We sequenced five clinical sputum samples that tested positive for smear tests at a grade of 1+ or higher. Among these reads, those originating from the *Mycobacterium* genus ranged from 0.1% to 34%, with an average of 7.0% (Fig. 3A). Leveraging this sequencing data, we conducted an MLST analysis. Only one sample, which exhibited the highest proportion of mycobacterium (34%), was successfully identified as *Mycobacterium avium* subsp. *hominissuis*, with an MLST score of 0.83 (Supplementary Table 1). As noted in previous studies (Votintseva et al. 2017; Doughty et al. 2014), we observed a very low abundance of mycobacteria in sputum. These findings were consistent with prior research from other investigators (Votintseva et al. 2017; Doughty et al. 2014), underscoring the necessity for genomic enrichment to obtain sufficient DNA from sputum samples.

Assessment of target capture sequencing for mycobacterial detection

To assess the efficacy of RNA probes in detecting mycobacteria from sputum samples, we prepared simple DNA mixtures comprising target (*M. agri*) and non-target (human) components. Subsequently, we conducted target capture sequencing to evaluate the sensitivity of the RNA probes. The proportion of reads mapped from the captured DNA ranged from a maximum of 97% with a 10 ng (10,000 pg) target to a minimum of 3% with a 1 pg target (Table 1). Notably, even with a low mapped read ratio of 0.10 from a 1 pg target, the MLST score reached a value of 1.00 (Table 2). Moreover, probe concentration did not significantly impact capture efficiency, even up to a tenfold dilution (0.1X) of the manufacturer-suggested concentration.

Subsequent specificity assessment involved six conditions of DNA mixtures (Fig. 3B), comprising target (*M. agri*) and non-target (human or commercial bacterial mock). In conditions where the input DNA included *M. agri* (specifically, in conditions labeled as *M. agri*, A50_H50, A50_M50, and A10_M90), all 24 loci present within the *M. agri* profiles were successfully detected, achieving MLST scores of 1.00 (Fig. 3C). Conversely, no identification scores were obtained in conditions lacking *M. agri* (labeled as Human and Mock), despite enrichment and alignment of some sequences to the database (illustrated in Figs. 3D and E).

The observed increase in non-target sequences alongside a decrease in target DNA is likely attributable to residual DNA in the tube during the bead purification process, a phenomenon known as carryover (Jennings et al. 2017). The estimated amount of carryover DNA is approximately 10 pg,

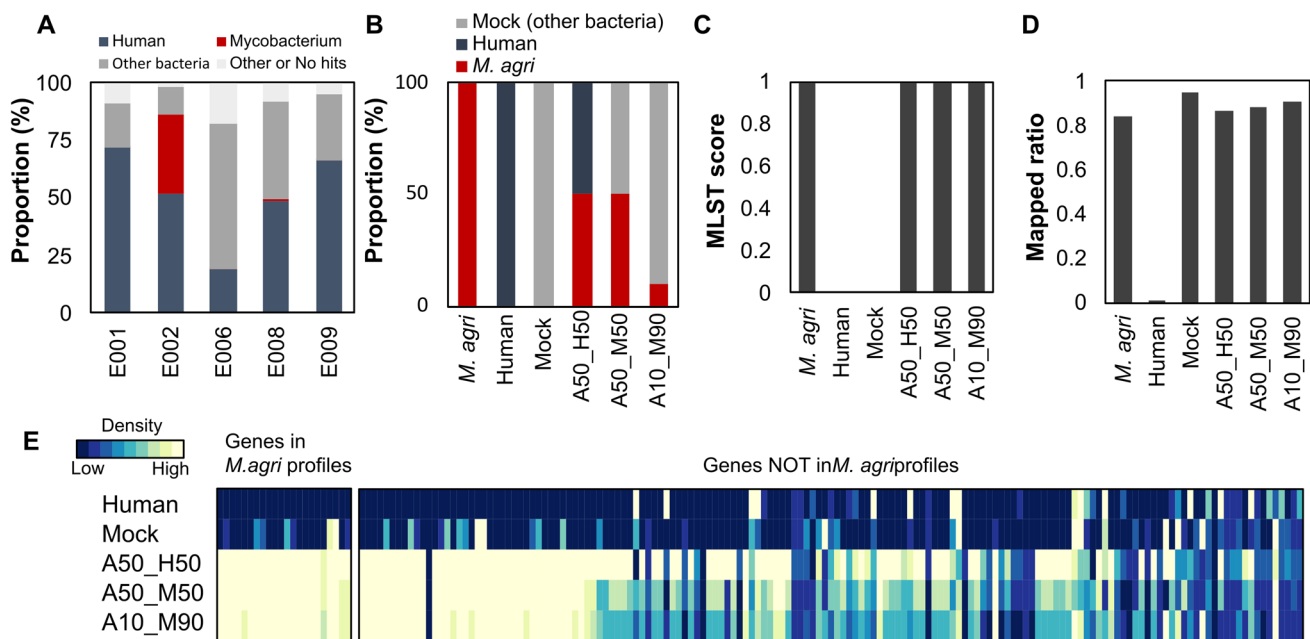


Fig. 3 Evaluation of sputum metagenome and NALC-Seq. **A.** Taxonomic compositions resulting from metagenomic sequencing. **B.** Artificial mock compositions for target capture sequencing. **C.** MLST

score identified as *M. agri*. **D.** Ratio of mapped reads in total reads. **E.** Density of reads mapped in each loci

Table 1 Mapped read ratio. Mapped read ratio in each condition when sequencing a total of 10 ng of mixed DNA comprising human and *M. agri* DNA. Each column of the table indicates the quantity of *M. agri* DNA

Amount of probes (X)	<i>M. agri</i> gDNA (pg)				
	1	10	100	1000	10,000
0.01	0.03	0.11	0.50	0.90	0.92
0.03	0.04	0.32	0.75	0.95	0.93
0.1	0.09	0.43	0.88	0.94	0.96
0.3	0.13	0.61	0.92	0.98	0.97
1	0.10	0.50	0.88	0.97	0.97

Table 2 MLST score using mlstverse

Amount of probes (X)	<i>M. agri</i> gDNA (pg)				
	1	10	100	1000	10,000
0.01	0.85	0.99	1.00	1.00	1.00
0.03	0.97	1.00	1.00	1.00	1.00
0.1	0.99	1.00	1.00	1.00	1.00
0.3	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00

as evidenced by the proportion of mapped sequences reaching 50% at a genomic DNA (gDNA) quantity of 10 pg. Thus, even from 1 pg of target within 10 ng of non-target DNA,

approximately 10% of target sequences can be obtained, enabling successful species identification. This detection threshold corresponds to a cell concentration of approximately 2.9×10^5 cells/mL, equivalent to a G2 on the Gaffky scale and a smear-positive of 1+ (Okada 1949). The designed RNA probe library facilitates highly sensitive mycobacterial detection, with an extremely low likelihood of false positives from non-mycobacterial DNA sources present in sputum samples. Subsequently, we coined this methodology NALC-Seq, after the NALC-NaOH treatment conducted prior to library preparation.

Comprehensive mycobacterial detection via target capture sequencing

We conducted the target capture sequencing, termed NALC-Seq, followed by identification analysis using mlstverse and compared the identification results with those of culture isolates. The identification results of culture isolates were as follows: twenty-one identified as *M. avium*, thirteen as *M. intracellulare*, five as *M. abscessus* subsp. *massiliense*, three as *M. abscessus* subsp. *abscessus/bolletii*, four as *M. tuberculosis*, four as *M. kansasii*, and one each as *Mycobacterium paragordoniae*, *Mycobacterium mucogenicum/phocaicum*, *Mycobacterium fortuitum* complex, and *M. lentiflavum* (Supplementary Table 3). Among these patients, 51 cases (98.1%) were identified as positive for either MTB or NTM.

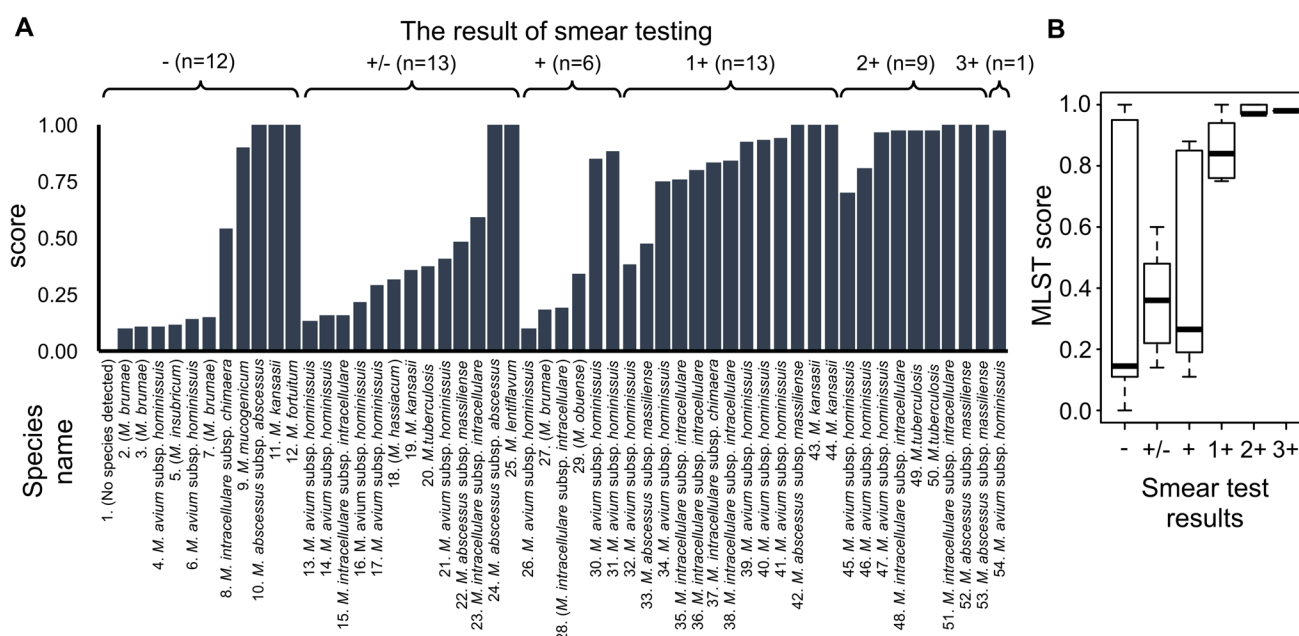


Fig. 4 Comprehensive species identification of mycobacteria from sputum samples using NALC-Seq. **A.** Top score in MLST analysis and identified species name. species names enclosed in parenthe-

ses represent inconsistent results with the MGIT-Seq. **B.** Boxplot of MLST score by Smear-Positive level

DNA extracted from 54 sputum samples underwent sequencing using NALC-Seq. Across all 54 samples, NALC-Seq demonstrated a sensitivity of 98.1%, yielding MLST scores in 53 out of the 54 samples (Fig. 4A and Table 3). Consistent identification results were observed in 45 out of 54 samples compared to culture isolates, resulting in an accuracy of 83.3% (Fukushima et al. 2023) (Supplementary Tables 3 and 4). Among the samples, nine exhibited discrepancies with MGIT-Seq results, being identified as *Mycobacterium brumae* and *Mycobacterium obuense* typically found in environmental sources such as water and soil, indicating possible environmental mycobacteria detection in the samples of lower smear positivity. In cases with strong smear positivity (1 + or higher), sensitivity and accuracy achieved 100.0%, with an average MLST score exceeding 0.75 (Figs. 4A and B).

Compared to drug susceptibility testing results for 49 isolates, discrepancies in resistance prediction were observed for 4 isolates with MGIT-Seq and 9 isolates with NALC-Seq; furthermore, there were discrepancies between MGIT-Seq and NALC-Seq results for 5 isolates. Among the discrepancies between MGIT-Seq and NALC-Seq, 2 isolates (#12, #15) were false negatives only with NALC-Seq. For the remaining 3 isolates (#4, #7, #23), false-positive clarithromycin resistance was detected by NALC-Seq from smear-negative samples. Notably, 2 of these 3 isolates with low smear grades exhibited low scores, suggesting potential false-positive detection due to sequences originating from pulmonary and oral bacteria.

The diagnostic validities of macrolide resistance prediction using NALC-Seq were as follows: sensitivity, 0.80; specificity, 0.912; positive predictive value, 0.80; and negative predictive value, 0.912. Regarding amikacin resistance, while resistance (MIC 64 ug/mL) was observed only in a single isolate (#46), the prediction based on sequencing data was negative, resulting in a discrepancy.

It is commonly acknowledged that the sensitivity of direct mycobacterial detection from sputum using PCR is approximately 80% (Noordhoek et al. 1994; Li et al. 2017; Anand and Biswas 2021; Sarro et al. 2021). Sarro et al. demonstrated the capability to detect MTB and common NTM, achieving a sensitivity of 83.3% and a specificity of 96.6% (Sarro et al. 2021) using multiplex PCR. He et al. conducted capture sequencing similar to our approach, achieving a sensitivity of 91.3% and an accuracy of 83.3% for 30 samples, including MTB and 32 common pathogenic NTM (He et al. 2020). These results indicate that NALC-Seq offers sensitivity and accuracy comparable to traditional direct identification methods from sputum. Additionally, the custom probes for NALC-Seq are designed based on the sequences of 186 mycobacterial species (Supplementary Table 5), significantly surpassing conventional methods (Matsumoto et al. 2019). The method successfully identified relatively rare species such as *M. abscessus*, *M. lentiflavum*, *M. mucogenicum*, *M. fortuitum*, and *M. kansasii* among the patients enrolled in this study. Based on these results, we established NALC-Seq as a method for comprehensive identification of a

Table 3 Comparison with conventional methods

sample#	smear	MIC tests		CLR resistance		NALC-Seq genotype	
		MIC	MIC	MGIT-Seq genotype		rrl erm(41)	
				rrl	erm(41)	rrl	erm(41)
1	-	S	0.125	S		S	
2	-	S	0.25	S		S	
3	-	S	0.25	S		S	
4	-	S	0.5	S		R	A2058G
5	-	S	0.25	S		S	
6	-	S	0.5	S		S	
7	-	S	0.25	S		R	A2058G
8	-	S	2	S		S	
9	-	S	0.5	S		S	
10	-	R	>64	R	Full	R	Full
11	-	S	0.25	S		S	
12	-	R	>64	S		S	Truncated
13	+/-	R	>32	R	Full	S	
14	+/-	S	0.25	S		S	
15	+/-	R	>32	R	A2058C	S	
16	+/-	S	0.5	S		S	
17	+/-	R	>32	R	A2058C	R	A2058C
18	+/-	Not tested		S		S	
19	+/-	S	0.125	S		S	
20	+/-	Not tested		S		S	
21	+/-	S	0.25	S		S	
22	+/-	S	0.125	S	Truncated	S	
23	+/-	S	0.125	S		R	A2058G
24	+/-	R	>64	R	Full	R	Full
25	+/-	S	0.25	S		S	
26	+	S	1	S		S	
27	+	Not tested		S		S	
28	+	S	4	S		S	
29	+	S	1	S		S	
30	+	R	>32	R	A2059C	R	A2059C
31	+	S	0.25	S		S	
32	1+	R	>32	S		S	
33	1+	S	0.25	S	Truncated	S	
34	1+	S	0.5	S		S	
35	1+	S	0.125	S		S	
36	1+	I	16	S		S	
37	1+	R	>32	R	A2059C	R	A2059C
38	1+	S	0.25	S		S	
39	1+	S	0.5	S		S	
40	1+	R	>32	S		S	
41	1+	R	>32	R	A2059G	R	A2059G
42	1+	S	0.25	S	Truncated	S	Truncated
43	1+	S	0.25	S		S	
44	1+	S	0.25	S		S	
45	2+	S	0.25	S		S	
46	2+	R	>32	R	A2058G	R	A2058G
47	2+	R	>32	R	A2058C	R	A2058C
48	2+	S	0.125	S		S	
49	2+	Not tested		S		S	
50	2+	Not tested		S		S	
51	2+	R	32	S		S	
52	2+	S	0.25	S	Truncated	S	Truncated
53	2+	S	0.25	S	Truncated	S	Truncated
54	3+	R	>32	R	A2059C	R	A2059C

diverse range of mycobacterial species using a standardized approach.

Real-time identification using NALC-Seq and mlstverse-web

To expedite species identification and achieve faster diagnosis, we demonstrated the integration of NALC-Seq with the MinION device along with a real-time analysis pipeline. We re-sequenced the genomic DNA of sample #54 (depicted in Fig. 4), previously identified as *M. avium* subsp. *hominissuis*, using the MinION and mlstverse-web system. Within 2 h of sequencing initiation, we obtained consistent identification results with the highest MLST score (Fig. 5A). The time series analysis of MLST scores (Fig. 5B) revealed that the score for *M. avium* subsp. *hominissuis* surpassed those of other species by over 50% within approximately 10 min and stabilized within 30 min. The total processing time for this real-time NALC-Seq analysis was 0.8 d (19 h), marking a significant reduction compared to conventional testing, which typically takes 28.5 d (686 h), and MGIT-Seq, which requires 17.3 d (414 h), and MGIT-Seq, which requires 17.3 d (414

h) (Fig. 5C) (Fukushima et al. 2023). NALC-Seq notably broadened the range of identifiable bacterial species, effectively overcoming limitations associated with existing capture sequencing and direct sequencing methods (Votintseva et al. 2017; Kambli et al. 2021; He et al. 2020; Mann et al. 2023). Additionally, using mlstverse-web system is not exclusive to NALC-Seq and can be seamlessly employed with just a MinION device and a laptop computer, eliminating the need for specialized bioinformatics expertise. This demonstration represents a substantial advancement over traditional methods, significantly reducing the time required for species identification and diagnosis from weeks to mere hours, through the seamless integration of NALC-Seq with the MinION platform and a real-time analysis pipeline.

The limitations of this study include the relatively small number of samples evaluated, the cost disadvantage compared to high-throughput short-read sequencers, the limited number of evaluated pretreatment methods, the limited types of detectable drug resistance, and the potential inability to detect species beyond the 186 mycobacterial species used to design the custom probes listed in Supplementary Table 5.

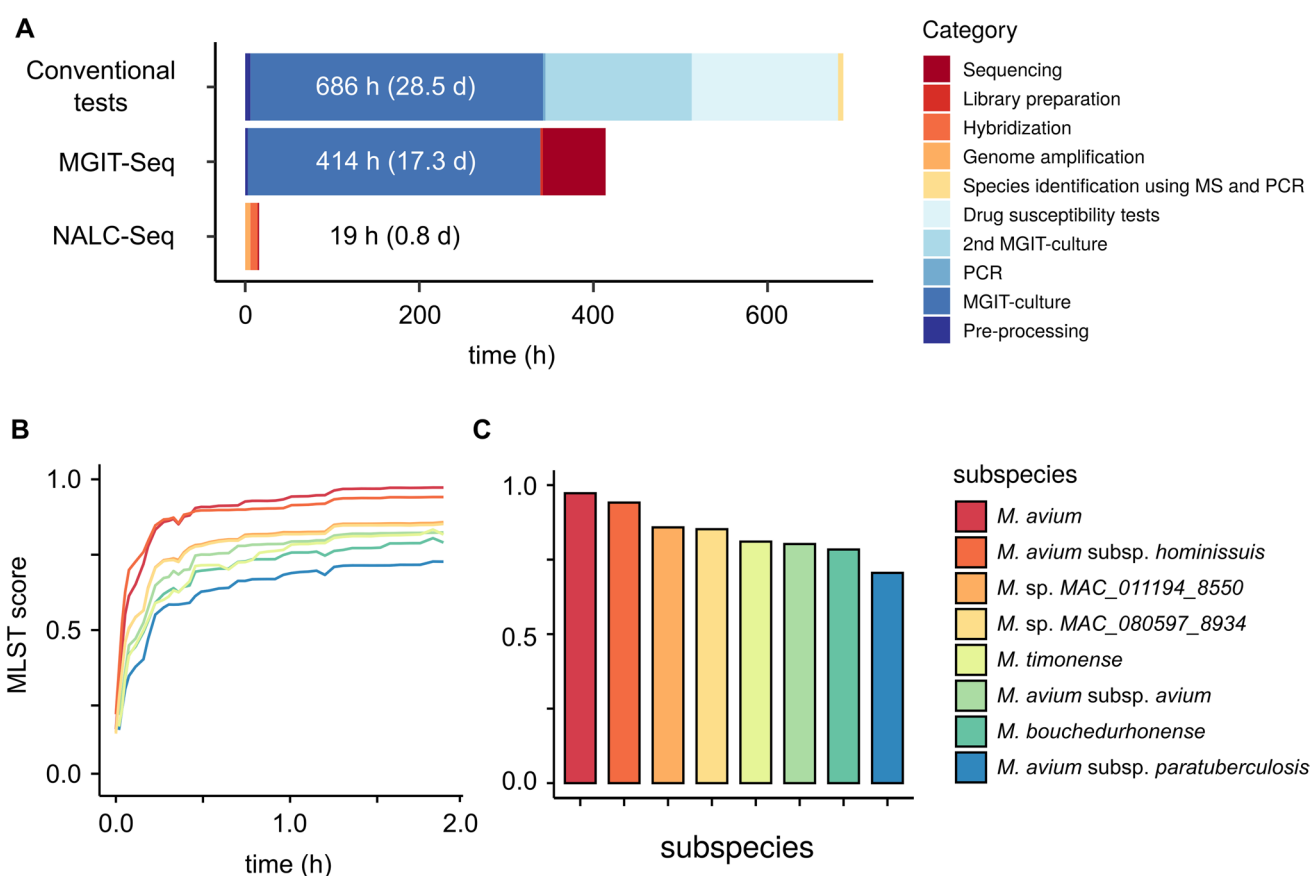


Fig. 5 Experimental flow and comparison of time required for NALC-Seq. **A.** Estimated duration in library preparation. **B.** Time course of MLST score. **C.** MLST score 2 h from sequencing start

Conclusion

In conclusion, the development of NALC-Seq represents a significant advancement in the diagnostics of mycobacterial infection. This study introduces a novel approach for the rapid and comprehensive identification of mycobacteria directly from sputum samples, effectively surpassing the limitations of traditional diagnostic methods. By integrating target capture sequencing with the mlstverse-web system and enabling real-time analysis on the MinION device, this research addresses the urgent need for expedited diagnostic methods amid escalating drug resistance and the global health burden associated with mycobacterial diseases. Our findings underscored the efficacy of NALC-Seq in identifying a wide range of mycobacterial species from smear-positive sputum samples, including rare and drug-resistant strains, with remarkable sensitivity and accuracy. This methodology not only streamlines treatment decisions by providing rapid and precise diagnostic information but also holds promise for significantly impacting public health through enhanced disease surveillance and control. Future research should prioritize further optimizing this technology to simplify its operation and broaden its applicability to a wide range of infectious agents, thus contributing to the global effort to combat infectious diseases.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10142-025-01609-x>.

Acknowledgements Not applicable

Authors' contributions YM, KF, and SN conceived and directed the project. KF, HK, HS, JY, TNit, TNii, TM, KT, and KM diagnosed NTM diseases and collected samples. YM and DM designed custom RNA probes. YF, DM, and YM conducted experiments. YM performed bioinformatic analyses. SZ and YM developed mlstverse-web system and real-time analysis pipelines. YM, KF, and SN drafted the manuscript. All authors contributed to various versions of the manuscript, providing comments and edits, and approved the final version.

Funding This work received support from Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers [JP19 K16643, JP20 KK0179, JP21 K08194, JP23 K07925] and Japan Agency for Medical Research and Development (AMED) under Grant Number [JP22ym0126809j0001, JP23ym0126815j0002].

Data availability The website of mlstverse is available at <https://mlstverse.org>. An automated uploader for raw sequencing data is publicly available on github (<https://github.com/yamatsumoto/fast5upload>). Sequencing data used in this study is deposited in National Center for Biotechnology Information with the accession PRJDB12894 and PRJNA1077745.

Declarations

Ethics approval and consent to participate The experimental protocol for data involving human participants followed the Ethical Guidelines of the Japan Ministries of Health and Labor for Medical and Health Research Involving Human Subjects. All the experiments were conducted in accordance with the principles outlined in the Declaration of

Helsinki. The ethics boards of Osaka University (31–8-3 and 583–3) and the National Hospital Organization, Osaka Toneyama Medical Center (TNH-R-2020020) approved this study. Written informed consent was obtained from all the participants.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adjemian J, Frankland TB, Daida YG, Honda JR, Olivier KN, Zelazny A et al (2017) Epidemiology of nontuberculous mycobacterial lung disease and tuberculosis, Hawaii, USA. *Emerg Infect Dis* 23:439–447
- Alcolea-Medina A, Fernandez MTC, Montiel N, García MPL, Sevilla CD, North N et al (2019) An improved simple method for the identification of Mycobacteria by MALDI-TOF MS (Matrix-Assisted Laser Desorption-Ionization mass spectrometry). *Sci Rep* 9:20216
- Anand AR, Biswas J (2021) TB or NTM: Can a new multiplex PCR assay be the answer? *EBioMedicine* 71:103552
- Avaniss-Aghajani E, Jones K, Holtzman A, Aronson T, Glover N, Boian M et al (1996) Molecular technique for rapid identification of mycobacteria. *J Clin Microbiol* 34:98–102
- Busatto C, Vianna JS, da Silva LV, Junior Ramis IB, da Silva PEA (2019) Mycobacterium avium: an overview. *Tuberculosis (Edinb)* 114:127–34
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- Chin KL, Sarmiento ME, Alvarez-Cabrera N, Norazmi MN, Acosta A (2020) Pulmonary non-tuberculous mycobacterial infections: current state and future management. *Eur J Clin Microbiol Infect Dis* 39:799–826
- Claydon MA, Davey SN, Edwards-Jones V, Gordon DB (1996) The rapid identification of intact microorganisms using mass spectrometry. *Nat Biotechnol* 14:1584–1586
- Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ (2014) Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* 2:e585
- Falkinham JO 3rd (2016) Current Epidemiologic Trends of the Non-tuberculous Mycobacteria (NTM). *Curr Environ Health Rep* 3:161–167
- Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I et al (2019) Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype

- correlations in a collection of isolates. *Antimicrob Agents Chemother* 63:e00483–e519
- Florensa AF, Kaas RS, Clausen PTL, Aytan-Aktug D, Aarestrup FM (2022) ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* 8:000748
- Fukushima K, Matsumoto Y, Matsuki T, Saito H, Motooka D, Komukai S et al (2023) MGIT-seq for the identification of nontuberculous mycobacteria and drug resistance: a prospective study. *J Clin Microbiol* 61:e0162622
- He Y, Gong Z, Zhao X, Zhang D, Zhang Z (2020) Comprehensive Determination of *Mycobacterium tuberculosis* and Nontuberculous *Mycobacteria* From Targeted Capture Sequencing. *Front Cell Infect Microbiol* 10:449
- Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D et al (2019) Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 6:CD009593
- Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB et al (2019) Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res* 4:191
- Hyeyoung L, Hee-Jung P, Sang-Nae C, Gill-Han B, Sang-Jae K (2000) Species identification of mycobacteria by PCR-restriction fragment length polymorphism of *terpoB* Gene. *J Clin Microbiol* 38:2966–2971
- Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J et al (2017) Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the association for molecular pathology and college of american pathologists. *J Mol Diagn* 19:341–365
- Johansen MD, Herrmann J-L, Kremer L (2020) Non-tuberculous mycobacteria and the rise of *Mycobacterium abscessus*. *Nat Rev Microbiol* 18:392–407
- Kambli P, Ajbani K, Kazi M, Sadani M, Naik S, Shetty A et al (2021) Targeted next generation sequencing directly from sputum for comprehensive genetic information on drug resistant *Mycobacterium tuberculosis*. *Tuberculosis* 127:102051
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546
- Lang J (2022) NanoCoV19: An analytical pipeline for rapid detection of severe acute respiratory syndrome coronavirus 2. *Front Genet* 13:1008792
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]*
- Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100
- Li S, Liu B, Peng M, Chen M, Yin W, Tang H et al (2017) Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. *PLoS ONE* 12:e0180725
- Maiga M, Siddiqui S, Diallo S, Diarra B, Traoré B, Shea YR et al (2012) Failure to recognize nontuberculous mycobacteria leads to misdiagnosis of chronic pulmonary tuberculosis. *PLoS ONE* 7:e36902
- Mann BC, Jacobson KR, Ghebrekristos Y, Warren RM, Farhat MR (2023) Assessment and validation of enrichment and target capture approaches to improve *Mycobacterium tuberculosis* WGS from direct patient samples. *J Clin Microbiol* 61:e0038223
- Matsumoto Y, Kinjo T, Motooka D, Nabeya D, Jung N, Uechi K et al (2019) Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles. *Emerg Microbes Infect* 8:1043–1053
- Namkoong H, Kurashima A, Morimoto K, Hoshino Y, Hasegawa N, Ato M et al (2016) Epidemiology of pulmonary nontuberculous mycobacterial disease. *Japan Emerg Infect Dis* 22:1116–1117
- Nicholls SM, Quick JC, Tang S, Loman NJ (2019) Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* 8:giz043
- Noordhoek GT, Kolk AH, Bjune G, Catty D, Dale JW, Fine PE et al (1994) Sensitivity and specificity of PCR for detection of *Mycobacterium tuberculosis*: a blind comparison study among seven laboratories. *J Clin Microbiol* 32:277–284
- Okada Y (1949) The Examination of Tubercle Bacilli in Sputum. *Tohoku J Exp Med* 50:389–398
- Pai M, Dewan PK, Swaminathan S (2023) Transforming tuberculosis diagnosis. *Nat Microbiol* 8:756–759
- Prasad R, Gupta N, Singh M (2014) Multidrug resistant tuberculosis: trends and control. *Indian J Chest Dis Allied Sci* 56:237–246
- Prevots DR, Marras TK (2015) Epidemiology of human pulmonary infection with nontuberculous mycobacteria: a review. *Clin Chest Med* 36:13–34
- Saptawati L, Kusumo H, Suryawati B (2019) Prevalence of Non-tuberculous *Mycobacteria* (NTM) in Surakarta, Indonesia: Higher Than Expected. *KnE Life Sciences* 4:132–8
- Sarro YDS, Butzler MA, Sanogo F, Kodio O, Tolofoudie M, Goumane MS et al (2021) Development and clinical evaluation of a new multiplex PCR assay for a simultaneous diagnosis of tuberculous and nontuberculous mycobacteria. *EBioMedicine* 70:103527
- Saxena S, Spaink HP, Forn-Cuní G (2021) Drug resistance in nontuberculous mycobacteria: mechanisms and models. *Biology* 10:96
- Shahraki AH, Heidarieh P, Bostanabad SZ, Khosravi AD, Hashemzadeh M, Khandan S et al (2015) “Multidrug-resistant tuberculosis” may be nontuberculous mycobacteria. *Eur J Intern Med* 26:279–284
- Sheka D, Alabi N, Gordon PMK (2021) Oxford nanopore sequencing in clinical microbiology and infection diagnostics. *Brief Bioinform* 22:403
- Vanino E, Granozzi B, Akkerman OW, Munoz-Torrico M, Palmieri F, Seaworth B et al (2023) Update of drug-resistant tuberculosis treatment guidelines: a turning point. *Int J Infect Dis* 130(Suppl 1):S12–S15
- Votintseva AA, Bradley P, Pankhurst L, Del Ojo EC, Loose M, Nilgiriwala K et al (2017) Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J Clin Microbiol* 55:1285–1298
- Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257
- Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G et al (2022) SAM-TB: a whole genome sequencing data analysis website for detection of *Mycobacterium tuberculosis* drug resistance and transmission. *Brief Bioinform* 23:bbac030.
- Yoo AB, Jette MA, Grondona M (2003) SLURM: Simple Linux Utility for Resource Management. In: *Job Scheduling Strategies for Parallel Processing*. Springer Berlin Heidelberg; p. 44–60
- Zheng P, Zhou C, Ding Y, Liu B, Lu L, Zhu F et al (2023) Nanopore sequencing technology and its applications. *MedComm* 4:e316

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.