JKMS

Original Article
Infectious Diseases, Microbiology & Parasitology

Check for updates

Evaluation of Five User-Friendly Whole Genome Sequencing Software for *Mycobacterium tuberculosis* in Clinical Application

Namhee Kim [ID],[1,2*] **Kwang Hyuk Seok** [ID],[3*] **Soyoun Shin** [ID],[3,4] **Boram Kim** [ID],[2,5] **Hyunwoong Park** [ID],[1,2] **Eun Youn Roh** [ID],[1,2] **Jong Hyun Yoon** [ID],[1,2] and **Sue Shin** [ID] [1,2]

[1]Department of Laboratory Medicine, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Korea
[2]Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul, Korea
[3]Laboratory Medicine Center, The Korean Institute of Tuberculosis, Cheongju, Korea
[4]Seegene Medical Foundation, Daejeon, Korea
[5]Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

OPEN ACCESS

Address for Correspondence:
Sue Shin, MD, PhD
Department of Laboratory Medicine, SMG-SNU Boramae Medical Center, 20 Boramae-ro 5-gil, Dongjak-gu, Seoul 07061, Korea.
Email: jeannie@snu.ac.kr

*Namhee Kim and Kwang Hyuk Seok contributed equally to this work.

ORCID iDs
Namhee Kim [ID]
https://orcid.org/0000-0002-9851-1727
Kwang Hyuk Seok [ID]
https://orcid.org/0000-0003-3195-8775
Soyoun Shin [ID]
https://orcid.org/0000-0002-7097-4551
Boram Kim [ID]
https://orcid.org/0000-0001-5695-8902
Hyunwoong Park [ID]
https://orcid.org/0000-0001-7765-2259

ABSTRACT

Background: Whole genome sequencing (WGS) is an increasingly useful tool for tuberculosis (TB) diagnosis and disease management. In this study, we evaluated the utility of user-friendly WGS tools in reporting resistance profiles and identifying lineages of clinical TB isolates from South Korea.
Methods: Forty clinical samples from TB patients showing discrepancies between their rapid molecular and conventional drug susceptibility tests were used in this study. Among these clinical isolates, 37 strains were successfully evaluated via WGS software, using the GenTB, TB Profiler, PhyResSE, CASTB, and Mykrobe.
Results: More accurate and faster susceptibility results could be obtained with isoniazid than with rifampin. Using the phenotypic test as the gold standard, the isoniazid concordance rate between phenotypic drug susceptibility test (DST) and WGS (GenTB: 45.9%, TB profiler: 40.5%, PhyResSE: 40.5%, CASTB: 48.6%, and Mykrobe: 43.2%) was much higher than between phenotypic DST and rapid molecular genotypic DST (18.9%) among the 37 strains. In contrast, the rifampin concordance rate between phenotypic DST and WGS and that between phenotypic DST and rapid molecular genotypic DST was similar (81.1–89.2%). We also found novel mutations associated with INH in *katG* and *ahpC* gene region, not covered by the line probe assay. In addition, lineage analysis identified 81.1% of these samples as L2 East Asian lineage strains, and 18.9% as L4 Euro-American lineage strains.
Conclusion: WGS may play a pivotal role in TB diagnosis and the detection of drug resistance, genetic diversity, and transmission dynamics in the near future because of its accuracy, speed, and extensibility.

Keywords: *Mycobacterium tuberculosis*; Drug-Resistant Tuberculosis; Drug Susceptibility Test; Whole Genome Sequencing

Generated by xmlinkpress

JKMS

Eun Youn Roh
https://orcid.org/0000-0002-1787-7509
Jong Hyun Yoon
https://orcid.org/0000-0002-1252-4739
Sue Shin
https://orcid.org/0000-0003-4791-8671

Disclosure
The authors have no potential conflicts of
interest to disclose.

Data Availability Statement
The datasets used and/or analysed during
the current study are available from the
corresponding author on reasonable request.

Author Contributions
Conceptualization: Kim N, Seok KH, Park H,
Shin S. Data curation: Kim N, Seok KH. Formal
analysis: Kim N, Seok KH, Shin S. Funding
acquisition: Shin S. Investigation: Seok KH,
Shin S. Methodology: Kim N, Seok KH, Kim B.
Resources: Kim N, Roh EY, Yoon JH. Software:
Kim N, Kim B, Park H. Supervision: Shin S.
Validation: Seok KH, Shin S. Writing - original
draft: Kim N. Writing - review & editing: Roh
EY, Yoon JH, Shin S.

INTRODUCTION

Tuberculosis (TB) is a contagious disease caused by *Mycobacterium tuberculosis*, which
particularly affects the lungs and is easily spread through the air.[1,2] In response to active
TB eradication programs, the global incidence rate and mortality rate for TB have been
decreasing by 2% and 3%, respectively, every year.[3,4] However, TB remains one of the top 10
causes of death, taking the lead for a single infectious disease (pre-coronavirus disease era).[5]
Moreover, as Korea has ranked with the highest TB incidence rate among the Organization
for Economic Co-operation and Development countries, the Korean government established
the first (2013–2017) and second (2018–2022) National Strategic Plans for TB with the goal of
lowering the incidence rate to 40 per 100,000 by 2022.[6-8]

Multidrug-resistant (MDR)-TB is resistant to at least isoniazid (INH) and rifampin (RMP),
the backbone compounds of standard first-line TB treatments. Moreover, extensively
drug-resistant-TB is more resistant to drugs than MDR-TB is, including INH, RMP,
fluoroquinolones, and at least three injectable second-line drugs [amikacin (AMK),
kanamycin, or capreomycin].[1,2] Because drug-resistant TB is resistant to the most potent TB
drugs, the remaining treatment options are less effective, have more side effects, have longer
treatment periods, and are more expensive; hence, their treatment success rate is extremely
low,[9,10] representing one of the largest obstacles to TB eradication. Resistant TB infection
is caused by 1) direct transmission from other resistant TB patients and/or 2) selective
pressure owing to incorrect drug use (such as incorrect drug selection and poor compliance).
Therefore, the quick and accurate diagnosis of TB and drug resistance are key factors to
determine the outcome of the fight against TB.

Recently, the importance of next-generation sequencing (NGS) has been emphasised in
antimicrobial susceptibility testing, including in TB patients. This new technique can
overcome the limitations of speed and accuracy of existing tests such as rapid molecular
genotypic tests and conventional phenotypic tests. Thus, in 2018 the World Health
Organization (WHO) published new guidelines concerning 'The use of next-generation
sequencing technologies for the detection of mutations associated with drug resistance in
Mycobacterium tuberculosis complex: technical guide' to encourage the use of NGS.[2,11] NGS
analysis can be divided into targeted sequencing and whole-genome sequencing (WGS). NGS
based WGS is a method used to identify variants in the entire genome. This method is better
suited to finding rare and novel mutations and obtaining information on several targets at
once. For example, in addition to resistance, strain analysis is also possible, and it is very
helpful for epidemiological investigations[11-13]; therefore, several governmental public health
centres and large-scale medical centres in the United Kingdom, Italy, the United States, and
Taiwan have tried using NGS based WGS for TB analysis.[14-20]

This study was designed to assess the usefulness of WGS for TB treatment in the clinical
setting. We collected cases with INH- or RMP-related discrepancies between their rapid
molecular and phenotypic drug susceptibility tests (DSTs) and reviewed the limitations of these
two methods. In addition, we evaluated five user-friendly and free WGS analysis tools, GenTB,
TB profiler, PhyResSE, CASTB, and Mykrobe, which were specified in the WHO guideline,[11] and
whether these WGS tools could enhance the outcome of TB diagnosis and treatment.

METHODS

Sample selection

We selected stored residual samples from patients diagnosed with TB who visited the Boramae Medical Center between 2010 and 2018 and underwent a DST. These samples were then evaluated using two distinct DST methods—conventional phenotypic DST and rapid molecular genotypic DST—that were performed simultaneously with colonies cultured on solid media.

The conventional phenotypic DST used the absolute concentration method with a solid L-J medium. Briefly, *M. tuberculosis* samples were inoculated into medium containing various concentrations of antimicrobial compounds, and their susceptibility and resistance were determined using changes in the growth status of each inoculum. These assays used standard concentrations for each drug, including INH 0.2 μg/mL, RMP 40.0 μg/mL, ethambutol 2.0 μg/mL, levofloxacin 2.0 μg/mL, moxifloxacin 1.0 μg/mL, ofloxacin 4.0 μg/mL, cycloserine 30.0 μg/mL, streptomycin 10.0 μg/mL, AMK 30.0 μg/mL, kanamycin 30.0 μg/mL, capreomycin 40.0 μg/mL, ethionamide 40.0 μg/mL, *para*-aminosalicylic acid 1.0 μg/mL, and rifabutin 20.0 μg/mL. However, it should be noted that pyrazinamide susceptibility cannot be accurately assayed using this method; thus, it was determined by a pyrazinamidase activity test.

Rapid molecular genotypic DST was completed using a line probe assay,[21] GenoType MTBDR plus ver. 1.0 (Hain Lifescience, Nehren, Germany). This method checks for *katG-* and *inhA-*related INH resistance and *rpoB-*related RMP resistance; if the mutation-probe binds and/or the wild-type probe is missing, this assay will report a positive incidence for this type of resistance.

A total of 1,008 cases (907 patients; some were duplicates) were sent for both phenotypic and genotypic DSTs over the study period (2010 to 2018). Of these, 52 isolates showed discrepancies between the two DST methods, with 40 strains being placed at −80°C. The antibiotic susceptibility patterns for each of these strains are summarised in **Table 1**.

WGS

The stored TB strains were re-cultured on solid medium (3% Ogawa medium) for 2 months, and 37 strains underwent successful WGS analysis (three samples failed to return to viability following storage). WGS was performed using the Korean Institute of Tuberculosis protocol. Briefly, the *Mycobacterium* strains were mixed with glass beads and inactivated by heating before being subjected to nucleic acid extraction. The genomic DNA was then evaluated for quality and concentration before being used as a template for the production of a dual-index paired-end library using the Nextera DNA Flex Library Prep kit (Illumina, San Diego, CA, USA). Library quality control was then performed using the HS NGS fragment kit on a 5200 Fragment Analyzer System (Agilent Technologies, Santa Clara, CA, USA). Once the samples passed quality control, the library was normalised to an equal concentration, and sequencing was performed (600 cycles, paired-end, Miseq Reagent kit ver. 3) on a Miseq platform (Illumina). FASTQ data were then analysed using five free, user-friendly online tools

Table 1. Isoniazid and rifampin sensitivity and resistant patterns in the 40 study samples

Isoniazid				Rifampin			
Phenotype	Genotype	No.		Phenotype	Genotype	No.	
S	R	19	(16 RMP S & 3 RMP R)	S	R	6	(2 INH S & 4 INH R)
R	S	14	(8 RMP S & 6 RMP R)	R	S	1	(1 INH S & 0 INH R)

Phenotype = conventional phenotypic drug susceptibility test, Genotype = rapid molecular drug susceptibility test, S = susceptible, R = resistant, INH = isoniazid, RMP = rifampin.

[GenTB (https://gentb.hms.harvard.edu), TB profiler (https://tbdr.lshtm.ac.uk), PhyResSE (https://bioinf.fz-borstel.de/mchips/phyresse/), CASTB (http://castb.ri.ncgm.go.jp/CASTB/introduction.html), and Mykrobe (https://www.mykrobe.com)] mentioned in the WHO TB NGS technical guide.[11] A summary of the characteristics of each software is provided in **Supplementary Table 1**. Notably, when all 37 FASTQ paired (forward and reverse) files were uploaded simultaneously, the results were returned overnight.

Ethics statement
This study was reviewed and approved by the Institutional Review Board (IRB) at the Boramae Medical Center (IRB number 10-2019-1) with a waiver of consent.

RESULTS

Demographic and medical history data
The clinical and demographic data of each of the 40 enrolled samples are summarised in **Table 2**. The study cohort comprised 27 male patients (27/40, 67.5%) and 13 female patients

Table 2. Clinical information of the 40 study samples recorded in electronic medical record

Subject No.	Subject information			Treatment plan	Anti-tuberculosis drugs regimen	Treatment period	Outcomes
	Age	Sex	Previous TB history				
1	30	M	None	RR or MDR	EMB+PZA+MFX+CS+KM+PTH+PAS → PZA+MFX+CS+KM+PTH → PZA+MFX+CS+KM+PTH+LZD+DLM	1.5 yr → 2 yr → 2 yr	Failed
2	26	F	None	Other resistant	RMP+PZA+LFX+CS+KM	2 yr	Failed
3	67	M	None	RMP side effect	INH+EMB+LFX+CS+PAS	2 yr	Success
4	43	M	None	Standard treatment	INH+RMP+EMB+PZA	9 mon	Success
5	49	M	Unknown	Standard treatment	INH+RMP+EMB+PZA	1 mon (died during treatment)	Not evaluated
6	55	M	None	RR or MDR	EMB+PZA+LFX+SM+CS+PAS+AMG+CLR	5 yr	Failed
7	56	M	Unknown	HR	RMP+EMB+PZA+LFX+KM	2 mon (died during treatment)	Not evaluated
8	58	F	None	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
9	70	F	Unknown	Unknown	Unknown	Unknown	Not evaluated
10	63	M	None	Standard treatment	INH+RMP+EMB+PZA	10 mon	Success
11	89	F	None	Standard treatment	INH+RMP+EMB+PZA	5 mon (died during treatment)	Not evaluated
12	44	M	None	HR	RMP+EMB+PZA	6 mon (follow-up loss)	Not evaluated
13	59	F	None	HR	RMP+EMB+PZA	6 mon	Success
14	40	M	Yes (4 yr ago)	RR or MDR	EMB+PZA+LFX+PTH	2 wk (died during treatment)	Not evaluated
15	77	F	None	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
16	62	F	None	RR or MDR → HR	EMB+PZA+MFX+CS+KM+PAS → RMP+EMB+PZA+MFX	2 mon → 9 mon	Success
17	71	M	None	HR	RMP+EMB+PZA+MFX	8 mon	Success
18	74	M	None	RR or MDR	PZA+LFX+CS+KM+PTH	1 mon (died during treatment)	Not evaluated
19	52	M	Unknown	RR or MDR	PZA+LFX+KM+PTH+PAS	7 mon (transferred out)	Not evaluated
20	67	F	Yes (30 yr ago)	HR and RMP side effect	EMB+PZA+MFX+CS+SM → RMP+EMB+PZA+MFX+CS+AMK → PZA+ LFX+PAS+LZD+CLR	8 mon → 2 mon → 1 yr	Failed
21	29	F	None	RR or MDR	EMB+PZA+MFX+BQ+LZD+CS+AMG+PTH+PAS	2 yr	Failed
22	48	M	Yes (3 years ago)	HR	RMP+EMB+PZA+LFX+CS+AMK	3 mon (follow-up loss)	Not evaluated

(continued to the next page)

Table 2. (Continued) Clinical information of the 40 study samples recorded in electronic medical record

Subject No.	Subject information			Treatment plan	Anti-tuberculosis drugs regimen	Treatment period	Outcomes
	Age	Sex	Previous TB history				
23	79	M	None	HR	RMP+EMB+PZA+MFX	8 mon	Success
24	79	M	Unknown	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
25	51	M	Yes (4 yr ago)	Standard treatment	INH+RMP+EMB+PZA	1 mon (transferred out)	Not evaluated
26	59	M	Yes (7 yr ago)	RR or MDR	PZA+MXF+CS+AMK+PTH → INH+EMB+PZA+CS+KM+PAS+LZD	18 mon → 2 yr	Failed
27	54	F	None	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
28	45	F	Yes (40 & 2 yr ago)	RR or MDR	PZA+KM+PTH+PAS+AMC+LZD+DLM	2.5 yr	Failed
29	50	F	None	Standard treatment	INH+RMP+EMB+PZA	2 mon (transferred out)	Not evaluated
30	35	M	None	RR or MDR	INH+PZA+MFX+CS+KM+PTH+PAS	2 yr	Success
31	82	F	None	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
32	81	M	None	Standard treatment	INH+RMP+EMB+PZA	6 mon	Success
33	60	M	None	RR or MDR	EMB+PZA+MFX+CS+SM+PTH	1 mon (died during treatment)	Not evaluated
34	38	M	None	RR or MDR	PZA+MFX+CS+KM+PTH → INH+RMP+EMB+MFX	3 mon → 9 mon	Success
35	56	M	None	HR	RMP+EMB+PZA+MFX	7 mon	Success
36	39	M	Yes (10 yr ago)	RR or MDR	PZA+LFX+CS+KM+PTH	12 mon	Success
37	41	M	None	Standard treatment	INH+RMP+EMB+PZA	2 mon (follow-up loss)	Not evaluated
38	64	M	None	RR or MDR	PZA+LFX+CS+KM+PTH	2 yr	Success
39	68	M	None	Unknown	Unknown	1 mon (transferred out)	Not evaluated
40	34	M	None	Standard treatment	INH+RMP+EMB+PZA	4 mon (follow-up loss)	Not evaluated

TB = tuberculosis, RR = rifampin-monoresistant tuberculosis, MDR = multidrug-resistant tuberculosis, EMB = ethambutol, PZA = pyrazinamide, MFX = moxifloxacin, CS = cycloserine, KM = kanamycin, PTH = prothionamide, PAS = para-aminosalicylic acid, LZD = linezolid, DLM = delamanid, RMP = rifampin, LFX = levofloxacin, SM = streptomycin, AMG = aminoglycoside, CLR = clarithromycin, HR = isoniazid-monoresistant tuberculosis, Other resistant = resistant to the primary (standard treatment) drugs other than rifampin and isoniazid, INH = isoniazid, AMK = amikacin, BQ = bedaquiline.

(13/40, 32.5%); seven patients were aged under 40 years (7/40, 17.5%), 17 were between 40 and 59 years (17/40, 42.5%), 13 were between 60 and 79 years (13/40, 32.5%), and three patients were > 80 years old (3/40, 7.5%); and all of them (40/40, 100%) were Korean (East Asian). As for their past medical history, 28 patients (28/40, 70.0%) were newly diagnosed with TB, and seven (7/40, 17.5%) had a history of TB infection and treatment. Among the 28 newly diagnosed patients, nine (9/28, 32.1%) were treated with MDR, RMP-monoresistant, or RMP-side effect TB regimens, and among the seven re-treatment patients, five (5/7, 71.4%) were treated with MDR, RMP-monoresistant, or RMP-side effect TB regimens. In addition, the 14 patients who were treated with a standard treatment regimen (for sensitive TB) showed a 0% treatment failure rate, but the 24 patients who were infected with resistant TB showed a 29.2% (7/24) treatment failure rate. Thus, most of the re-treatment patients were infected with resistant TB and had a significantly high probability of treatment failure.

Conventional phenotypic DST and rapid molecular genotypic DST

The detailed resistance patterns including both first- and second-line drugs of each sample are described in **Table 3**. The discrepancies between the conventional phenotypic and rapid molecular genotypic DSTs are also indicated in **Table 3**, where the genotypic results are underlined.

Table 3. Conventional phenotypic and rapid molecular drug susceptibility patterns in the 40 study samples

Subject No.	Phenotypic DST															Genotypic DST
	First line therapy				Group A[a]			Group B[a]		Group C[a]					Etc.	LPA
	INH	RMP	EMB	PZA	FQ LFX	FQ MFX	FQ OFX	CS	SM	AMG AMK	AMG KM	AMG CPM	ETH	PAS	RBT	INH, RMP
1	R	R	S	S	S	S	S	S	R	S	S	S	S	S	S	INH S, RMP R
2	R	S	R	S	S	S	S	S	R	S	S	S	R	S	S	INH S, RMP S
3	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
4	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
5	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
6	S	R	R	S	R	R	R	S	R	R	R	R	R	S	R	INH R, RMP R
7	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
8	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
9	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S, RMP S
10	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
11	R	S	S	S	R	R	R	S	S	S	S	S	S	S	S	INH S, RMP S
12	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S, RMP S
13	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S, RMP S
14	R	R	S	S	S	S	S	S	S	S	S	S	S	S	R	INH S, RMP R
15	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
16	R	S	S	S	S	S	S	S	R	S	S	S	S	S	S	INH R, RMP R
17	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
18	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP R
19	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S, RMP R
20	R	S	S	S	S	S	S	S	R	S	S	S	S	S	S	INH R, RMP R
21	R	S	S	S	R	R	R	S	R	S	S	S	S	S	S	INH R, RMP R
22	R	S	S	S	S	S	S	S	R	S	S	S	S	S	S	INH S, RMP S
23	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
24	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
25	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
26	S	R	S	S	R	R	R	S	R	S	S	S	R	S	S	INH R, RMP R
27	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
28	R	R	S	S	R	R	R	S	S	S	S	S	S	S	R	INH S, RMP R
29	S	R	S	S	S	S	S	S	S	S	S	S	S	S	R	INH S, RMP S
30	S	R	R	S	S	S	S	S	S	S	S	S	S	S	R	INH R, RMP R
31	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
32	S	S	S	S	S	S	S	S	R	S	S	S	S	S	S	INH R, RMP S
33	R	R	R	S	S	S	S	S	S	S	S	S	S	S	R	INH S, RMP R
34	S	S	S	S	S	S	S	S	R	S	S	S	S	S	S	INH S, RMP R
35	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S
36	R	R	S	S	S	S	S	S	S	S	S	S	S	S	R	INH S, RMP R
37	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S
38	R	R	S	S	S	S	S	S	R	S	S	S	S	S	R	INH S, RMP S
39	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH S, RMP S
40	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	INH R, RMP S

DST = drug susceptibility test, LPA = line probe assay, INH = isoniazid, RMP = rifampin, EMB = ethambutol, PZA = pyrazinamide, FQ = fluoroquinolone, LFX = levofloxacin, MFX = moxifloxacin, OFX = ofloxacin, CS = cycloserine, SM = streptomycin, AMG = aminoglycoside, AMK = amikacin, KM = kanamycin, CPM = capreomycin, ETH = ethionamide, PAS = para-aminosalicylic acid, RBT = rifabutin.
[a]Classification of medication for multidrug-resistant tuberculosis: modified from World Health Organization consolidated guidelines on drug-resistant tuberculosis treatment (2019) [9].
Underline: The discrepancies between the phenotypic and genotypic DST are indicated, where the discrepant genotypic results are underlined.

WGS

Thirty-seven strains were successfully revived and subjected to WGS. The mean coverage depth was 112.06 ± 33.13 (48.24 ± 13.68–229.62 ± 40.4). The software analysis results are shown in **Table 4** and **Supplementary Table 2**.

Lineage

Of the 37 samples assessed by WGS, 30 (81.1%) were identified as the L2 East Asian lineage, with the Beijing genotype being the major determining feature, and seven (18.9%) were identified as belonging to the L4 Euro-American lineage (**Table 4**).

Table 4. Lineage and anti-tuberculosis drug resistant analysis using the five whole genome sequencing analysis tools

Subject No.	Lineage	Sub-lineage	Resistant mutations				
			GenTB (ver. 2.1)	TB profiler (ver. 3.0.1)	PhyResSE (ver. 1.0)	CASTB (ver. 1.5)	Mykrobe (ver. 0.9.0)
1	L2 East-Asian	2.2.2	INH, RMP, EMB, SM	INH, RMP, EMB, PZA, SM, ETH	INH, RMP, EMB, PZA, SM	INH, RMP, EMB, PZA, SM	INH, RMP, EMB, PZA, SM
2	L2 East-Asian	2.2.1	- -	- -	- -	- -	- -
3	L2 East-Asian	2.2.2	INH, ETH	INH, ETH	INH, ETH	INH -	INH -
4	L2 East-Asian	2.2.1	INH -	INH -	INH -	INH -	INH -
5	L2 East-Asian	2.2.2	INH, ETH	INH, ETH	INH, PZA, ETH	INH -	INH -
6	L2 East-Asian	2.2.1	INH, RMP, EMB, OFX, SM, AMG, ETH	INH, RMP, EMB, PZA, SM, AMG, ETH	INH, RMP, EMB, SM, AMG, ETH	INH, RMP, EMB, SM, AMK	INH, RMP, EMB, SM, AMG
7	L2 East-Asian	2.2.1	- -	- -	- -	- -	- -
8	L4 Euro-American	4	INH, ETH	INH, ETH	INH -	- -	INH -
9	L2 East-Asian	2.2.1	- ETH	- -	- -	- -	- -
10	L2 East-Asian	2.2.1	- -	- -	- -	- -	- -
11	L2 East-Asian	2.2.2	- -	- -	PZA -	- -	- -
12	L4 Euro-American	4.2.1	INH, ETH	INH, ETH	INH, ETH	INH -	INH -
13	L2 East-Asian	2.2.2	INH -	INH -	INH, PZA	INH -	INH -
14	L2 East-Asian	2.2.2	INH, RMP, EMB -	INH, RMP -	INH, RMP, EMB -	INH, RMP, EMB -	RMP -
15	L2 East-Asian	2.2.1	INH, ETH	INH, ETH	INH -	INH -	INH -
16	L2 East-Asian	2.2.1	INH, RMP, SM, AMK	INH, RMP, SM, PAS	INH, RMP, SM	INH, RMP -	INH, SM
17			Culture failure				
18	L2 East-Asian	2.2.1	INH, RMP -	INH, RMP -	INH, RMP -	INH, RMP -	INH, RMP -
19	L2 East-Asian	2.2.1	- -	- -	- -	- -	- -
20	L2 East-Asian	2.2.2	INH, RMP, EMB, SM	INH, RMP, EMB, SM, PAS	INH, RMP, PZA, SM	INH, RMP, SM	INH, EMB, SM
21	L2 East-Asian	2.2.1	INH, RMP, FQ, SM	INH, RMP, PZA, FQ, SM	INH, RMP, FQ, SM	INH, RMP, PZA, OFX, SM	PZA, SM
22	L2 East-Asian	2.2.2	INH, EMB -	INH, EMB -	INH, EMB -	INH, EMB -	EMB -
23	L2 East-Asian	2.2.2	INH, ETH	INH, ETH	INH, PZA, ETH	INH -	INH -
24	L2 East-Asian	2.2.2	INH, ETH	INH, ETH	INH, PZA, ETH	INH -	INH -
25	L2 East-Asian	2.2.2	INH, ETH	INH, ETH	INH, PZA, ETH	INH -	INH -
26	L2 East-Asian	2.2.1	INH, FQ, SM, ETH	INH, PZA, FQ, SM, ETH	INH, FQ, SM, ETH	INH, RMP, OFX, SM	INH, PZA, FQ, SM
27	L2 East-Asian	2.2.1	INH, OFX, ETH	INH, ETH	INH -	- -	INH -
28	L2 East-Asian	2.2.2	INH, RMP, EMB, PZA, FQ	INH, RMP, EMB, PZA, FQ	RMP, EMB, PZA, FQ	RMP, EMB, OFX	INH, RMP, EMB, PZA, FQ
29	L2 East-Asian	2.2.1	- -	- -	- -	- -	- -
30	L4 Euro-American	4.2.2	INH, RMP, EMB, PZA, ETH	INH, RMP, EMB, PZA, ETH	INH, RMP, EMB, PZA, ETH	INH, RMP, EMB, PZA -	INH, RMP, EMB, PZA -

Table 4. (Continued) Lineage and anti-tuberculosis drug resistant analysis using the five whole genome sequencing analysis tools

Subject No.	Lineage	Sub-lineage	Resistant mutations				
			GenTB (ver.2.1)	TB profiler (ver.3.0.1)	PhyResSE (ver.1.0)	CASTB (ver.1.5)	Mykrobe (ver.0.9.0)
31	L4 Euro-American	4	INH, -	INH, SM	INH -	INH -	INH -
32	L2 East-Asian	2.2.2	INH, SM, ETH	INH, SM, ETH	INH, PZA, SM, ETH	INH, SM	INH, SM
33	L2 East-Asian	2.2.1	INH, RMP, EMB, OFX	INH, RMP, EMB, FQ	INH, RMP, EMB -	RMP, EMB -	INH, RMP, EMB -
34	L2 East-Asian	2.2.2	RMP -	RMP -	RMP -	RMP -	RMP -
35	L4 Euro-American	4	INH, ETH	INH, ETH	INH, ETH	INH -	INH -
36	L4 Euro-American	4.2.1	INH, RMP, EMB, PZA -	INH, RMP, EMB -	RMP, EMB -	INH, RMP, EMB, PZA -	INH, RMP -
37	L4 Euro-American	4.2.1	- -	- -	- -	- -	INH -
38	L2 East-Asian	2.2.1	RMP, SM	RMP, SM	RMP, SM	INH, RMP, SM	INH, RMP, SM
39			Culture failure				
40			Culture failure				

INH = isoniazid, RMP = rifampin, EMB = ethambutol, SM = streptomycin, PZA = pyrazinamide, ETH = ethionamide, FQ = fluoroquinolone, OFX = ofloxacin, AMG = aminoglycoside, AMK = amikacin, PAS = para-aminosalicylic acid.

Drug susceptibility

The list of resistant drugs, as identified by WGS, is shown in **Table 4**, and a detailed list of the related variants is shown in **Supplementary Table 2**. These were organised using each online tool, GenTB, TB profiler, PhyResSE, CASTB, and Mykrobe. The concordance rates between the INH phenotypic DST and WGS were GenTB: 45.9% (17/37), TB profiler: 43.2% (16/37), PhyResSE: 40.5% (15/37), CASTB: 48.6% (18/37), and Mykrobe: 43.2% (16/37). The concordance rates between the RMP phenotypic DST and WGS were GenTB: 81.1% (30/37), TB profiler: 81.1% (30/37), PhyResSE: 81.1% (30/37), CASTB: 83.8% (31/37), and Mykrobe: 489.2% (33/37).

DISCUSSION

This study was performed to evaluate the value that WGS adds to the commonly used protocols for TB DSTs in Korea. Our data revealed that WGS allows for the collection of more diverse data than the current methods used for this type of evaluation.

First, WGS allowed us to evaluate both DST and lineage distribution simultaneously. In the present study, the lineage distribution of the TB strains was consistent with previously reported data from East Asia,[17,22] with the L2 East Asian lineage being dominant (70–90%) and the L4 lineage being less prevalent (~10%). Lineage analysis is useful for epidemiological purposes and can be used to confirm the chain of infection,[19] which is helpful for patient management, as specific lineages are known to be highly correlated with specific resistance characteristics. In particular, the Beijing lineage, which was predominant in our cohort, presents with a high mutation rate when infecting humans and has been frequently associated with drug resistance in various regions.[23-25]

Second, WGS can uncover novel and rare mutations that may be beyond the scope of the existing rapid molecular tests (line probe assays and commercial real-time polymerase

chain reaction). Previous studies reported that WGS is not inferior to these tests, even demonstrating more accurate results in some datasets.[26,27] In addition, our study clearly demonstrates the advantage of evaluating INH susceptibility/resistance using WGS as the concordance rate between the phenotypic DST and WGS (GenTB: 45.9% [17/37], TB profiler: 43.2% [16/37], PhyResSE: 40.5% [15/37], CASTB: 48.6% [18/37], and Mykrobe: 43.2% [16/37]) was much higher than that between the phenotypic DST and line probe assay (18.9% [7/37]). The reason for this large difference is likely associated with the fact that INH resistance-related mutations are scattered across a number of genes, including *katG*, *fabG1-inhA*, *oxyR-ahpC*, *ndh*, *kasA*, and *iniB*, meaning that commercial rapid molecular tests cannot cover them all. We also found novel mutations associated with INH in regions not covered by the line probe assay (*katG*: p.Q50P, p.W191R, and c.649_650insG & *ahpC*: c.-54C>T and c.-48G>A). Although the impact of these mutations remains ambiguous,[28] it is expected that their relationship with INH resistance can be revealed in future investigations.

For RMP, the concordance rate of the phenotypic DST and WGS, and phenotypic DST and line probe assays were very similar. This is likely because nearly all (95% or more) RMP-resistant strains harbour mutations within the 81 bp RMP-resistance-determining region (RRDR) in *rpoB*. Therefore, the line probe assay showed high sensitivity, despite only covering a short portion of this gene. However, it should be noted that approximately 5% of RMP-related mutations occur outside the RRDR sequence and may be missed when using rapid molecular tests.

Notably, we should be careful in interpreting the above INH and RMP concordance rates. This study was conducted only with about 40 strains showing discrepancies between their phenotypic DST and line probe assay. This sample selection bias significantly lowered the INH and RMP concordance rates. Thus, the above concordance rates were only used to explain the new mutations missing in the line probe assay, show detailed WGS information, and compare the five WGS software tools in the study. The fact that, for all 938 patients diagnosed with pulmonary TB in the Boramae Medical Center for 9 years, the INH concordance rate between the phenotypic DST and line probe assay was 95.6%, and RMP concordance rate between the phenotypic DST and line probe assay was 98.6%.

Finally, WGS is helpful when evaluating whether to use a drug that is not included in the commercial DST assays, such as quinolone or aminoglycoside. In addition, WGS examines the whole genome, allows for the storage of these data, and is at least 1 month faster than phenotypic DST. Hence, WGS is a promising tool for predicting resistance to new drugs, aiding in drug development, and managing TB in the future. It is clearly a useful tool for DST. In addition, since the five analysis tools evaluated in this study are free and very convenient to use, they can be easily accessed by clinicians and staff at general laboratories who are not highly trained bioinformatics experts. However, since each has different target antimicrobial agents and its own characteristic analysis mechanism, it is necessary to select the use after clearly understanding the strengths and weaknesses of each tool.

When we assessed the five tools with about 40 strains, there were no significant differences in the concordance of the drug-susceptible/resistant categories for these data from any of these online tools. Nevertheless, more importantly, each tool has specific its own mechanism such as bioinformatics pipeline, targeted drugs, relation genes, and upload system (**Supplementary Table 1**). Therefore, since CASTB and Mykrobe were the least up-to-date and provided the least information, we would recommend that they be used only

for auxiliary or confirmatory analyses. In contrast, GenTB was the most recently updated software and reported the most variants, including those that were not associated with antimicrobial resistance, making it a good screening tool but one that requires additional data trimming for clinical application. Thus, while GenTB demonstrates several significant advantages, the tool is not yet sophisticated enough on its own and should be supported by an additional evaluation methodology. Our experience suggests that this supporting tool should be PhyResSE, as it clearly displays the mapping results; the website would facilitate the evaluation and confirmation of the GenTB data.

While our data suggest that WGS has several advantages over current DST evaluations, the WGS technique has some limitations that must be overcome. It is very important to obtain high-quality whole-genome data for WGS. This can be complex in the case of Mycobacterium as these species produce a highly complex lipid cell layer that can interfere with genome extraction.[2,11] In addition, since live *M. tuberculosis* has a high biohazard level, additional protection steps are recommended, making it difficult to complete genome extractions in a typical laboratory environment. Therefore, a simple, efficient, high-quality TB genome extraction method using direct samples, such as sputum, is needed to facilitate the widespread adoption of these technologies. In addition, considering the high cost and labour-intensive aspects, it is difficult to manage TB WGS in a general hospital laboratory with current technology. It would be effective that TB WGS to be processed in the central reference laboratory by collecting dozens of priority samples (acquired from TB reactivated patients, patients having severe anti-TB drug side effects, patients failing to first-line anti-drug, and so on) from several general laboratories. With remarkable advances in science and medicine, the technology speed is accelerating and the cost is falling, therefore, it is worth trying to introduce clinical use beyond research in general laboratories in near future.

Validated, standardised, and stable analysis tools, and database expansion are necessary to reach the full potential of high-throughput sequencing-based analyses. In the fields of genetic diseases and cancer, several NGS databases have been established with a high degree of investment to tackle various research questions. However, no such established databases exist for microorganisms. The coronavirus disease pandemic has highlighted the importance of microbial genome analysis, emphasizing the need for a centralised sequence database for such organisms, with many countries, institutions, and companies working toward its establishment. Therefore, in the post-coronavirus disease era, database management for various microorganisms, including *M. tuberculosis*, should be systematised, substantially improving the potential of effective TB management.[29,30]

In summary, our data show that WGS is extremely useful for anti-TB DST. WGS identifies a much broader range of variants (especially for INH) compared with traditional rapid molecular techniques. In addition, WGS can confirm results more than 4 weeks earlier than phenotypic DST can, allowing for improved treatment planning and the prevention of TB transmission. Despite these advantages, additional research and development is needed to ensure that WGS can be adopted as a general laboratory diagnostic tool. However, considering the current speed of technological development and accumulated experience, we believe that these shortcomings will soon be overcome. In the near future, we expect that WGS will emerge as a commonplace technique that can be applied directly to culture-free specimens (such as sputum and bronchial wash) to determine the resistance of infectious agents to both first- and second-line anti-TB drugs within 1 week of specimen collection.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Overview of the five user-friendly and free next generation sequencing analysis tools for tuberculosis (assessment date: May 2021)

Click here to view

Supplementary Table 2
List of total variants associated with anti-tuberculosis drug using the whole genome sequencing analysis tools

Click here to view

REFERENCES

1. Centers for Disease Control and Prevention. Basic TB facts 2016. https://www.cdc.gov/tb/topic/basics/default.htm. Updated 2016. Accessed April 23, 2020.

2. Furin J, Cox H, Pai M. Tuberculosis. *Lancet* 2019;393(10181):1642-56.
   PUBMED | CROSSREF

3. World Health Organization. Tuberculosis, fact sheets 2020. https://www.who.int/news-room/fact-sheets/detail/tuberculosis. Updated 2020. Accessed July 2, 2021.

4. World Health Organization. Global tuberculosis report 2020. https://www.who.int/publications/i/item/9789240013131. Updated 2020. Accessed July 2, 2021.

5. World Health Organization. The top 10 causes of death 2020. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Updated 2020. Accessed July 2, 2021.

6. Kim JS, Shin JY, Kong IS. Characteristics of the notified tuberculosis - Republic of Korea, 2019. *Public Health Wkly Rep* 2020;13(14):832-50.

7. Korea Centers for Disease Control and Prevention. *Annual Report on the Notified Tuberculosis in Korea 2019*. Cheongju, Korea: Korea Centers for Disease Control and Prevention; 2020.

8. Go U, Park M, Kim UN, Lee S, Han S, Lee J, et al. Tuberculosis prevention and care in Korea: evolution of policy and practice. *J Clin Tuberc Other Mycobact Dis* 2018;11:28-36.
   PUBMED | CROSSREF

9. World Health Organization. *WHO Consolidated Guidelines on Drug-Resistant Tuberculosis Treatment 2019*. Geneva, Switzerland: WHO; 2019.

10. Joint Committee for the Revision of Korean Guidelines for Tuberculosis and Korea Center for Disease Control and Prevention. *Korean Guidelines for Tuberculosis*, 4th ed. 2020.

11. World Health Organization. *The Use of Next-Generation Sequencing Technologies for the Detection of Mutations Associated With Drug Resistance in Mycobacterium tuberculosis Complex: Technical Guide 2018*. Geneva, Switzerland: WHO; 2018.

12. Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: a retrospective observational study. *PLoS Med* 2016;13(10):e1002137.
   PUBMED | CROSSREF

13. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med* 2020;12(1):114.
   PUBMED | CROSSREF

14. Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P, et al. Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: a large-scale prospective assessment of performance against line probe assays and phenotyping. *J Clin Microbiol* 2018;56(2):e01480-17.
   PUBMED | CROSSREF

15. GOV.UK. England world leaders in the use of whole genome sequencing to diagnose TB. https://www.gov.uk/government/news/england-world-leaders-in-the-use-of-whole-genome-sequencing-to-diagnose-tb. Updated 2017. Accessed July 2, 2021.

16. Cabibbe AM, Walker TM, Niemann S, Cirillo DM. Whole genome sequencing of *Mycobacterium tuberculosis*. *Eur Respir J* 2018;52(5):1801163.
    PUBMED | CROSSREF

17. Genestet C, Hodille E, Berland JL, Ginevra C, Bryant JE, Ader F, et al. Whole-genome sequencing in drug susceptibility testing of *Mycobacterium tuberculosis* in routine practice in Lyon, France. *Int J Antimicrob Agents* 2020;55(4):105912.
    PUBMED | CROSSREF

18. Goig GA, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Navarro D, Borrás R, et al. Whole-genome sequencing of *Mycobacterium tuberculosis* directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study. *Lancet Microbe* 2020;1(4):e175-83.
    PUBMED | CROSSREF

19. Wu X, Gao R, Shen X, Guo Y, Yang J, Wu Z, et al. Use of whole-genome sequencing to predict Mycobacterium tuberculosis drug resistance in Shanghai, China. *Int J Infect Dis* 2020;96:48-53.
    PUBMED | CROSSREF

20. Vogel M, Utpatel C, Corbett C, Kohl TA, Iskakova A, Ahmedov S, et al. Implementation of whole genome sequencing for tuberculosis diagnostics in a low-middle income, high MDR-TB burden country. *Sci Rep* 2021;11(1):15333.
    PUBMED | CROSSREF

21. Nathavitharana RR, Cudahy PG, Schumacher SG, Steingart KR, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. *Eur Respir J* 2017;49(1):1601075.
    PUBMED | CROSSREF

22. Takii T, Seki K, Wakabayashi Y, Morishige Y, Sekizuka T, Yamashita A, et al. Whole-genome sequencing-based epidemiological analysis of anti-tuberculosis drug resistance genes in Japan in 2007: application of the genome research for Asian tuberculosis (GReAT) database. *Sci Rep* 2019;9(1):12823.
    PUBMED | CROSSREF

23. Cox HS, Kubica T, Doshetov D, Kebede Y, Rüsch-Gerdess S, Niemann S. The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. *Respir Res* 2005;6(1):134.
    PUBMED | CROSSREF

24. European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis. Beijing/W genotype Mycobacterium tuberculosis and drug resistance. *Emerg Infect Dis* 2006;12(5):736-43.
    PUBMED | CROSSREF

25. Hakamata M, Takihara H, Iwamoto T, Tamaru A, Hashimoto A, Tanaka T, et al. Higher genome mutation rates of Beijing lineage of Mycobacterium tuberculosis during human infection. *Sci Rep* 2020;10(1):17997.
    PUBMED | CROSSREF

26. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* 2015;7(1):51.
    PUBMED | CROSSREF

27. Nguyen TN, Anton-Le Berre V, Bañuls AL, Nguyen TV. Molecular diagnosis of drug-resistant tuberculosis; a literature review. *Front Microbiol* 2019;10:794.
    PUBMED | CROSSREF

28. World Health Organization. Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance 2021. https://www.who.int/publications/i/item/9789240028173. Updated 2021. Accessed October 17, 2021.

29. Ruhwald M, Carmona S, Pai M. Learning from COVID-19 to reimagine tuberculosis diagnosis. *Lancet Microbe* 2021;2(5):e169-70.
    PUBMED | CROSSREF

30. The Lancet Public Health. Renewing the fight to end tuberculosis. *Lancet Public Health* 2021;6(5):e260.
    PUBMED | CROSSREF