Annals of Clinical Microbiology www.acm.or.kr

Review article

Challenges and advances in mycobacterial molecular typing

Minh Phuong Trinh, Min-Kyoung Shin
Department of Microbiology and Convergence Medical Science, College of Medicine, Gyeongsang National University, Jinju, Korea

Abstract

Mycobacterium tuberculosis (MTB) and nontuberculous mycobacteria (NTM) present distinct clinical and epidemiological challenges and thus require tailored genotyping approaches. MTB is a globally transmissible pathogen for which diagnostic and surveillance infrastructures are well defined, whereas NTM infections are environmentally acquired, taxonomically diverse, and increasingly prevalent among vulnerable populations. Molecular genotyping is indispensable for both pathogen groups, supporting outbreak investigation and drug resistance prediction for MTB and species-level identification and relapse-versus-reinfection distinction for NTM. In this review, the evolution of strategies for genotyping mycobacteria are outlined, and traditional techniques (e.g., spoligotyping and mycobacterial interspersed repetitive unit–variable number tandem repeat genotyping) and advanced methods (multilocus sequence typing and whole-genome sequencing) are compared. We highlight the divergent drivers of genotyping between MTB and NTM, examine key technical and interpretive challenges, and discuss how cross-learning between these two fields can accelerate innovation. Emerging technologies such as portable sequencing platforms, artificial intelligence-assisted analysis, and curated genomic databases are expanding access to high-resolution genotyping. However, significant gaps remain, particularly in standardizing NTM genomic analyses and integrating genotypic data into global surveillance systems. By exploring the intersections between MTB and NTM molecular epidemiology, a synergistic pathway for more precise, accessible, and effective mycobacterial genotyping is highlighted.

Keywords: Molecular typing, *Mycobacterium tuberculosis*, Nontuberculous mycobacteria

Check for updates

OPEN ACCESS

pISSN : 2288-0585
eISSN : 2288-6850

Ann Clin Microbiol 2025 September, 28(3):18
https://doi.org/10.5145/ACM.2025.28.3.7

Correspondence to
Min-Kyoung Shin
E-mail: mkshin@gnu.ac.kr

Received: July 22, 2025
Revised: August 22, 2025
Accepted: August 28, 2025

© 2025 Korean Society of Clinical Microbiology.

This is an Open Access article which is freely available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND) (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Background

Pulmonary diseases caused by infection with *Mycobacterium tuberculosis* (MTB) or nontuberculous mycobacteria (NTM) represent a significant global health challenge due to their diverse clinical manifestations and epidemiological profiles [1–3]. MTB is responsible for over 10 million new cases and 1.5 million deaths from tuberculosis (TB) annually, making it a major global health burden [3]. Reports of NTM infections in immunocompetent individuals and patients with chronic pulmonary diseases are increasing [3,4]. Although MTB and NTM both belong to the genus *Mycobacterium*, they differ markedly in their transmission dynamics, pathogenesis, host interactions, and clinical management [5] and thus have divergent

molecular genotyping requirements: MTB genotyping is primarily used for outbreak investigations and drug resistance surveillance, whereas NTM genotyping facilitates species identification and epidemiological analysis [6,7].

Historically, traditional phenotypic methods such as observations of growth rate, colony morphology, and pigment production as well as biochemical assays have been used for species identification but are often time consuming and lack discriminatory power, particularly for closely related mycobacterial taxa [1,2]. Although sequencing of conserved genetic markers such as the 16S rRNA, *hsp65*, and *rpoB* genes has markedly improved species-level identification, these methods remain insufficient for resolving subspecies-level variations and accurately discriminating among the more than 200 recognized *Mycobacterium* species [4,8]. Moreover, the resolving power of these gene sequencing methods is inadequate for genotyping applications that require precise strain-level differentiation. By contrast, more advanced tools such as multilocus sequence typing (MLST), variable number tandem repeat (VNTR) analysis, and whole-genome sequencing (WGS) enable high-resolution strain differentiation, transmission mapping, and resistance prediction [4]. Despite that MTB and NTM differ significantly in their genomic features, their shared genus-level classification allows some overlap in the molecular tools and strategies used for strain typing.

Objectives

In this review, we outline the distinct purposes and applications of molecular genotyping of MTB and NTM, compare the major techniques used for each pathogen, and explore how methodological insights from the genotyping of one can support and enhance that of the other.

Clinical and epidemiological differences between MTB and NTM

Epidemiological context: transmission and species diversity

Although MTB and NTM belong to the same genus, their transmission dynamics and ecological niches are fundamentally different. MTB is one of the leading causes of infectious deaths worldwide, with an estimated 10.8 million new cases and 1.25 million deaths reported in 2023. The highest TB burden is found in Southeast Asia (45%), Africa (24%), and the Western Pacific (17%) [9]. MTB is transmitted person-to-person via aerosolized droplets from patients with active lung disease and spreads rapidly in densely populated or vulnerable settings, such as among those with human immunodeficiency virus infection, individuals suffering malnutrition, and residents of healthcare facilities or correctional facilities [10]. In addition to causing primary disease, MTB can have a long asymptomatic latency period and it can reactivate years later. Such potential complicates its surveillance and requires genotyping tools to identify transmission routes, reactivation events, and the emergence of drug-resistant clones [11].

By contrast, NTM infections stem from environmental sources such as soil, water, and plumbing, with human-to-human transmission being rare [5]. Although the exact transmission route of NTM has not been

fully identified, most infections are sporadic and non-contagious. Therefore, the genotypic analysis of this pathogen is focused on species-level identification and environmental contamination source tracking rather than patient-to-patient transmission tracking. To date, more than 200 species of NTM have been reported and their genetic and phenotypic diversity is very large [1–3]. However, the main *Mycobacterium* species that cause human lung diseases, such as *Mycobacterium avium* complex (MAC), *M. kansasii*, and *M. abscessus*, are relatively limited [5].

Clinical characteristics: pathogenesis, drug response, and risk factors

MTB and NTM are distinct in their pathogenesis, drug responses, and major risk groups [5]. Although MTB primarily targets the lungs, it can invade extrapulmonary organs such as the lymph nodes, central nervous system, and bones in some patients [10]. Moreover, as described above, the fact that MTB can remain latent and asymptomatic for years after primary infection and later reactivate to cause active TB renders its epidemic surveillance and patient-to-patient transmission tracking difficult, and high-resolution genotyping is required to analyze its transmission route and reactivation events [11]. Isoniazid, rifampicin, pyrazinamide, and ethambutol are the first-line antibiotics used to treat drug-sensitive TB [12]. However, in patients with multidrug-resistant or extensively resistant disease, the treatment period is long and drug selection is limited. Recently, new classes of drugs such as bedaquiline and delamanid have been introduced to shorten the treatment period and improve therapeutic options for patients with drug-resistant disease [13]. However, long-term combination therapy, low treatment compliance, and drug misuse accelerate the occurrence of resistant strains; thus, the role of genotyping in predicting drug resistance is becoming important [13].

Clinically, NTM attacks manifest mainly as chronic pulmonary infections, particularly in older adults, patients with underlying lung diseases (e.g., bronchiectasis, chronic obstructive pulmonary disease, pneumoconiosis, previous history of TB, post-radiotherapy fibrosis, chronic pulmonary aspiration, and cystic fibrosis), and immunocompromised individuals [5]. The disease spectrum also includes lymphadenitis, skin and soft tissue infections, and disseminated diseases in patients with severe immunosuppression (e.g., patients with AIDS and transplant recipients) [5]. The primary objectives of MTB genotyping are to detect antibiotic resistance, track transmission, and investigate epidemics. According to Walker et al. [14], WGS can reveal genetic heterogeneity within assumed clusters, thereby improving epidemic classifications. For NTM, genotyping is used primarily to identify species and distinguish re-infection from relapse, thereby directly influencing the treatment duration and providing targeted environmental control strategies. For example, in a cohort study of patients with cystic fibrosis, WGS revealed that *M. abscessus* was transmitted from one patient to another, which resulted in revised methods being implemented to prevent infections [15]. These contrasting characteristics of MTB and NTM have shaped divergent pathways in strategies used for their molecular genotyping. A comparative summary is presented in Table 1.

Table 1. Divergent clinical and epidemiological drivers of genotyping for *Mycobacterium tuberculosis* and nontuberculous mycobacteria infections

	Mycobacterium tuberculosis	Nontuberculous mycobacteria
Infection source	Human-to-human transmission via aerosols	Environmental reservoirs: water, soil, biofilm, medical equipment
Transmission dynamics	High person-to-person spread, clustered outbreaks	Rare human-to-human transmission, pseudo-outbreaks often due to laboratory/medical contamination
Genotyping purpose	Transmission tracking, antimicrobial resistance, outbreak investigation Examples: A WGS study by Walker et al. [14] revealed that a presumed outbreak cluster in London identified by MIRU-VNTR included unrelated strains, prompting refined definitions for transmission.	Species identification, relapse vs. re-infection, source tracking Examples: A WGS study by Bryant et al. [15] identified direct patient-to-patient transmission of *M. abscessus* among patients with CF, revising the assumption that NTM are exclusively environmental.
Clinical manifestation	Pulmonary and extrapulmonary TB; granulomatous inflammation	Chronic pulmonary infection in bronchiectasis/CF, lymphadenitis, disseminated disease in immunocompromised patients
Resistance prediction	Correlates with known resistance mutations (*rpoB*, *katG*, *inhA*, *gyrA/B*)	More variable and species dependent, fewer validated resistance markers
Standardization	Well-developed platforms (SITVIT, ReSeqTB, TB Portals), WHO guidelines	Absence of centralized genomic databases, fragmented implementation of surveillance activities, limited WGS pipelines for routine use
Main genotyping tools	Spoligotyping, MIRU-VNTR, WGS, SNP-based clustering	*hsp65/rpoB* sequencing, MLST, WGS, gene markers
SNP threshold for relatedness	≤ 12 SNPs for recent transmission (standardized)	Undefined, varies by species (ranges < 10–500 SNPs), no gold standard cutoff
Implication of findings	Guides contact tracing, resistance treatment, public health response	Informs clinical management, environmental control, re-infection risk stratification

Abbreviations: WGS, whole-genome sequencing; MIRU-VNTR, mycobacterial interspersed repetitive unit–variable number tandem repeat; CF, cystic fibrosis; NTM, nontuberculous mycobacteria; TB, tuberculosis; WHO, World Health Organization; SNP, single nucleotide polymorphism; MLST, multilocus sequence typing.

Overview of techniques for the molecular genotyping of mycobacteria

Techniques for the molecular genotyping of MTB and NTM were developed on the basis of their different clinical and epidemiological characteristics (Table 1). Techniques for MTB focus on outbreak tracking and drug resistance surveillance, whereas those for NTM are aimed at species identification and environmental epidemiology analysis [1–3,16]. Accordingly, in this section, molecular genotyping strategies are divided into pattern-, gene-, and WGS-based categories on the basis of their technical principles and discriminatory power. Each method is compared and described in terms of its technical basis, clinical applicability, and relative advantages and limitations of use for genotyping MTB and NTM.

Pattern-based typing

Pulsed-field gel electrophoresis (PFGE) and restriction fragment length polymorphism (RFLP)

PFGE and RFLP were among the earliest molecular tools developed for differentiating mycobacterial strains. IS*6110*-RFLP exploits the variability in the number and position of the insertion sequence (IS) element IS*6110* in the MTB genome. This method provides high discriminatory power and stable fingerprint patterns and has historically enabled the precise tracking of transmission events and identification of

outbreak clusters [17–19]. However, RFLP-based methods for NTM are less standardized because of the high genomic diversity among species and the often-limited presence of IS elements. Thus, the methods require adaptations for generating distinct epidemiologically relevant banding patterns from clinical isolates. For example, IS*1245*-based RFLP was performed using restriction enzymes *BsaAI*, *PvuII*, and *NruI* to distinguish 17 *M. avium* isolates, which yielded polymorphic patterns but showed band clustering with *PvuII* and unexpectedly low-molecular-weight fragments, suggesting the presence of unidentified IS elements [20]. In another study performed to differentiate *M. avium* from *M. intracellulare* on the basis of the *groES* gene, polymerase chain reaction (PCR)-RFLP with *BamHI*, *BstNI*, and *HgaI* revealed three polymorphisms, with *HgaI* proving the most useful for resolving clinical isolates and complementing high-performance liquid chromatography [21]. Moreover, adaptations for species such as MAC have yielded some success in NTM strain typing [22]. However, despite their robustness, both PFGE and RFLP have significant limitations: they are labor intensive, time consuming, and culture dependent and require high-quality intact DNA. Moreover, the need for specialized software and gel-based interpretation restricts their routine use in most clinical laboratories and field settings [17].

Spoligotyping

Spoligotyping is a PCR-based method that targets the direct repeat region of the MTB genome, which contains conserved spacers interspersed with unique short sequences [23,24]. This method detects the presence or absence of 35–41 specific spacers and generates a digital pattern that is highly portable and suitable for database comparisons [23,24]. Remarkably, spoligotyping can be performed with as little as 10 fg of DNA (equivalent to DNA from 2–3 mycobacterial cells), making it useful for analyzing paucibacillary or archival specimens, including formalin-fixed paraffin-embedded tissues [23,24].

Although spoligotyping offers limited discriminatory power compared with other genotyping methods, it is particularly effective for identifying the lineage and global phylogeography of MTB. It has also been used to analyze select NTM species such as *M. intracellulare*, supporting clade-level classification and geographical distribution mapping studies. However, owing to its relatively low resolving power, spoligotyping alone is insufficient for high-resolution outbreak investigations and is best used in combination with more discriminatory techniques [25].

Mycobacterial interspersed repetitive unit-VNTR genotyping

The mycobacterial interspersed repetitive unit (MIRU)-VNTR method represents a significant advancement in genotyping owing to its reproducibility, scalability, and digital compatibility. The methodology is based on the examination of DNA segments that include tandem repetitive sequences, the number of copies of which differs between strains. This technique is based on PCR amplification, and the size of the amplified product is used to determine the number of repeats. According to a standard order, MIRU results are provided as designations of 15 or 24 characters, each of which represents the number of repeats at one of the loci [6]. MIRU-VNTR genotyping offers a digital output that is standardized and compatible with global databases for epidemiological surveillance [6].

The MIRU-VNTR method has become a widely adopted standard for genotyping MTB, particularly in

combination with spoligotyping for epidemiological surveillance. It has sufficient discriminatory power to delineate transmission clusters in most settings and is especially valuable when WGS is not feasible. However, its resolving power is lower than that of WGS, particularly for strains with low IS*6110* copy numbers [6]. Moreover, MIRU-VNTR genotyping has shown species-specific utility in differentiating NTM species. Ichikawa et al. [23] demonstrated that 16 MIRU-VNTR loci could reliably differentiate 50 genotypes among 74 *M. intracellulare* isolates, yielding a high Hunter–Gaston discriminatory index (0.988). However, the applicability of the method to NTM is limited by the genomic heterogeneity and variable repeat conservation among the species [23]. Consequently, NTM genotyping increasingly relies on gene-targeted sequencing techniques that offer higher taxonomic resolution and better reproducibility across diverse species [24–26].

Gene-based typing

NTM identification is commonly accomplished via gene-based typing, which targets conserved genes such as 16S rRNA, *hsp65*, *rpoB*, and *gyrB*. The sequencing of *hsp65* and *rpoB* provides superior resolution over that of 16S rRNA, especially for the MAC and *M. abscessus* complex subspecies [24–26].

Gene-based typing is used less commonly used for MTB strain differentiation owing to the clonal population structure and low genomic diversity of isolates [27]. Instead, targeted gene sequencing for diagnosing antimicrobial resistance is the main focus in the molecular diagnostics of MTB. For example, resistance-associated genes such as *gyrA/gyrB* for fluoroquinolone resistance, *katG* and *inhA* for isoniazid resistance, and *rpoB* for rifampicin resistance have been sequenced to find mutations that confer drug resistance [1,2]. These mutations are highly predictive of phenotypic resistance and form the molecular basis of several World Health Organization-endorsed diagnostic platforms, including GeneXpert MTB/RIF and line probe assays [28]. MLST analysis incorporating genes such as *argH* and *cya* enhances NTM classification, whereas it is rarely used for MTB because of its limited ability to resolve isolates at the strain level [25].

WGS-based typing

Currently, WGS represents the most comprehensive and high-resolution tool available for mycobacterial genotyping. It is superior to spoligotyping, MIRU-VNTR genotyping, and other techniques in determining genetic relatedness among MTB strains [29]. Two important WGS techniques are commonly applied: gene-by-gene typing, which extends MLST by detecting allelic variations in core or accessory genes [30]; and single nucleotide polymorphism (SNP) variant calling, which finds single nucleotide differences with a reference genome (e.g., *M. tuberculosis* H37Rv, GCF_000195955.2) and offers strong phylogenetic markers because SNP events are rare and homoplasy is low [31]. In the case of MTB, WGS allows for SNP-level discrimination between strains, providing unparalleled insights into the bacterial transmission networks, recent infection events, and in-host microevolution. Unlike traditional genotyping methods such as MIRU-VNTR that rely on repetitive loci and may overestimate relatedness, WGS offers precise phylogenetic resolution that allows for a more accurate delineation of transmission chains [14]. One study showed that

MIRU-VNTR genotyping overestimated TB cluster sizes in the UK Midlands, whereas WGS demonstrated that a large number of assumed epidemiologically connected patients were genetically unconnected, which led to improved outbreak definitions and contact tracing techniques [14]. In that UK study, which combined retrospective and longitudinal analyses, a mutation rate of approximately 0.5 SNPs/year was found in the genomic region analyzed. Samples obtained from patients living in close proximity to one another usually differed by fewer than four SNPs, but the largest difference was 10 SNPs, which were found in patients who developed recurrent TB over a 7–10-year span. A threshold of 12 SNPs was proposed to identify epidemiologically linked strains [14,32]. WGS also detects novel resistance mutations in genes, such as *rpoB*, *katG*, *inhA*, *gyrA*, and *eis*, which are often missed by conventional diagnostic techniques, thus supporting the integration of gene sequencing into routine TB surveillance in countries such as the UK and the Netherlands [33].

By resolving the issue of high genetic diversity among NTM strains, WGS enables novel species identification and strain-level discrimination [34]. Experiments to distinguish *M. marinum* from *M. ulcerans* have shown that WGS differentiates relapse from re-infection and clarifies virulence factors [34]. Revisions to infection control methods were prompted by a critical study by Bryant et al. [15], in which WGS confirmed the patient-to-patient transmission of *M. abscessus* in patients with cystic fibrosis. This finding highlighted the importance of high-resolution genome sequencing in comprehending the dynamics of NTM transmission and led to the reassessment of infection control procedures in cystic fibrosis clinics. Additionally, WGS can predict unknown NTM species, which facilitates the development of quick and efficient treatments for newly developing infections.

Bioinformatics tools such as TB-Profiler and Mykrobe help to streamline resistance and lineage studies for MTB but are less developed for NTM, for which standardized pipelines are needed [35]. As WGS becomes more accessible, its roles in enhancing the clinical management of MTB and NTM and the public health responses to both pathogens continue to grow. Table 2 presents a consolidated comparison of the key molecular techniques currently used for MTB and NTM genotyping, highlighting their respective principles, advantages, and limitations.

Challenges in mycobacterial molecular typing

Despite remarkable progress in the development of molecular genotyping techniques for mycobacteria, several key challenges continue to limit their widespread application and interpretive power in both clinical and epidemiological contexts. These limitations vary significantly between MTB and NTM owing to their differing biological characteristics and public health implications.

Table 2. Overview of molecular genotyping techniques for mycobacteria

Technique	Principle	Advantages	Limitations	MTB applications	NTM applications	Ref
Spoligotyping	PCR detection of presence/absence of unique spacer sequences in the direct repeat locus	Rapid, low DNA input (10 fg), standardized, suitable for degraded samples	Low discriminatory power, not suitable for outbreak confirmation	Lineage typing, strain tracking in low-resource settings	Limited, low resolution between NTM species	[22]
RFLP	RFLP analysis based on the insertion sequence IS*6110* (for MTB); for NTM, IS*1245*, IS*901*, IS*1311*, or *groES*-based RFLP variants	High discriminatory power, stable profiles (IS*6110* for MTB); for NTM, polymorphic patterns with *BsaAI*, *PvuII*, *NruI* (IS *1245*), *PvuII* (IS *901*/IS *1311*), or *BamHI* /*BstNI* /*HgaI* (*groES* P CR-RFLP)	Labor intensive, requires large DNA amounts, not suitable for low-copy IS*6110* strains (MTB); for NTM, band clustering with certain enzymes (Examples: *PvuII* in IS*1245*/IS*1311*), unexpected low-molecular-size fragments, limited standardization	Historical gold standard for outbreak studies	IS*1245*-based RFLP yields polymorphic patterns with *BsaAI* and *NruI* easier to detect, suggesting unidentified insertion elements; IS*901* RFLP identifies 6 profiles (E, F, G, M, Q, V) linked to flocks and transmission (Examples: pheasants/goshawks); IS*1311* RFLP with *PvuII* differentiates *M. avium* subspecies from human/animal origins, often combined with MIRU-VNTR; *groES*-based PCR-RFLP differentiates *M. avium* (25 isolates) and *M. intracellulare* (20 isolates) using *HgaI* polymorphism, complementing high-performance liquid chromatography	[20,21, 34,53, 54]
MIRU-VNTR	PCR-based detection of variable number of tandem repeats at multiple loci	High resolution, reproducible, database compatible, digital outputs	Less discriminatory than WGS, low-copy-number strains reduce resolution	Current global standard for MTB genotyping	Applied to select NTM species (Example: *M. intracellulare*)	[23]
16S rRNA sequencing	Amplification and sequencing of the conserved 16S ribosomal RNA gene	Broad-range detection, widely available, applicable to direct specimens	Cannot distinguish closely related species, low resolution (Example: *M. avium* vs. *M. intracellulare*)	Rarely used owing to poor discriminatory power	Common for genus-level identification, limited for subspecies resolution	[26]
hsp65, *rpoB*, *gyrB* sequencing	Gene-targeted sequencing of heat-shock protein, RNA polymerase, or gyrase genes	High species-level resolution, suitable for clinical ID, complements 16S rRNA	Still limited for some MAC/MABC species, not all targets validated	Limited, mainly used for species confirmation	Widely used for NTM species and subspecies discrimination, routine tool in clinical mycobacteriology labs	[25]
MLST	Sequencing of internal fragments of multiple housekeeping genes	Portable between laboratories, good for phylogenetics, population structure analysis	Low resolution for outbreak tracking, requires multiple loci sequencing	Rarely used	Increasingly used for taxonomy and evolutionary analysis	[30]
WGS	Sequencing of the entire genome	Highest resolution; detects SNPs, drug resistance, and phylogeny	Costly, bioinformatics heavy, requires high-quality DNA and reference databases	Drug resistance prediction, outbreak tracing, evolution	Drug resistance prediction, outbreak tracing, evolution,resolves relapse vs. re-infection, novel species ID, resistance markers	[14,15, 29,31]

Abbreviations: MTB, *Mycobacterium tuberculosis*; NTM, nontuberculous mycobacteria; Ref, reference; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; MIRU-VNTR, mycobacterial interspersed repetitive unit–variable number tandem repeat; ID, identification; MAC, *Mycobacterium avium* complex; MABC, *Mycobacterium abscessus* complex; WGS, whole-genome sequencing; SNP, single nucleotide polymorphism; MLST, multilocus sequence typing.

Low genetic diversity among MTB strains

One of the most significant limiting factors in the molecular genotyping of MTB is the low overall genetic diversity among its strains. Compared with many bacterial pathogens, MTB evolves slowly, with an estimated mutation rate of 0.3–0.5 SNPs per genome per year [36]. This genetic monomorphism presents difficulties in distinguishing among strains involved in recent transmissions as well as between those that are distantly related but still genetically similar. Consequently, conventional methods such as spoligotyping and MIRU-VNTR genotyping lack resolving power in identifying transmission chains in high-burden settings [14]. Despite that WGS offers much higher discriminatory power, it cannot distinguish the strains involved in short-term transmission events, especially in outbreak settings where clonal expansion dominates [14]. Furthermore, relying on a single molecular genotyping approach may be insufficient for a comprehensive epidemiological analysis. This highlights the need for integrated approaches that combine WGS, epidemiological metadata, and clinical contexts to trace transmissions accurately [36].

Differentiating relapse vs. re-infection in patients infected with nontuberculous mycobacteria

In contrast to MTB strains, NTM species such as *M. avium* and *M. abscessus* are characterized by higher genomic plasticity and diversity, posing a different set of challenges for their identification. A critical issue in clinical NTM management is differentiation between relapse (same strain reactivation) and re-infection (new strain acquisition) in patients with recurrent disease, especially chronic or recurrent pulmonary NTM infections [34,37]. WGS has emerged as a powerful tool for resolving this question, as it allows for high-resolution comparison of serial isolates. However, the absence of validated SNP thresholds for defining strain similarity in NTM remains a key barrier. In contrast to MTB, where a cutoff of 12 SNPs or less is widely accepted to suggest recent transmission [14], no such consensus exists for NTM species [14]. Studies on *M. avium* and *M. abscessus* have reported within-host SNP distances ranging from less than 10 to more than 500 SNPs, depending on the infection duration, anatomical sampling site, and patient immune status [14,37,38]. Recent studies have suggested that the intrastrain diversity in NTM may vary considerably depending on the species, host environment, and time between isolates. Additionally, the lack of species-specific reference databases, standardized analytical pipelines, and longitudinal genomic studies further impairs our ability to set meaningful benchmarks [34]. A step toward resolution for NTM species would involve large-scale, multicenter research efforts to define species-level mutation rates, within-host diversity norms, and stable SNP thresholds, analogous to the extensive molecular epidemiology studies that have been performed for MTB.

Difficulties in sample processing and WGS data interpretation

The technical complexities of sample processing and genome sequencing remain critical barriers, particularly in low-resource settings. Mycobacteria possess a thick lipid-rich cell wall that complicates DNA extraction and purification, requiring enzymatic pretreatment or mechanical lysis to achieve sufficient nucleic acid yield and quality [39]. Moreover, WGS requires high-coverage contamination-free sequences, which

may be difficult to obtain directly from clinical specimens such as sputum, especially in paucibacillary NTM infections. Although direct sputum sequencing is being explored for MTB [40], such approaches are still experimental for NTM, which frequently grow in low abundance and require culture enrichment. Additionally, these limitations are compounded by the high costs of sequencing platforms and the need for a bioinformatics infrastructure in places where MTB and NTM infection prevalence is the highest. The interpretation of WGS data also requires databases that may not be updated or compatible across regions [39].

Lack of global data synchronization and standardization

A major systemic challenge is the fragmentation and lack of synchronization of data in mycobacterial genomic databases. Although MTB genotyping has benefited from global initiatives, such as SITVIT2, ReSeqTB, and TB Portals, which provide curated and lineage-linked metadata, resistance prediction data, and epidemiological metadata that are interoperable across borders and laboratories, studies on NTM remain underserved by equivalent resources [41]. In many cases, sequencing data from NTM isolates are deposited in scattered repositories without regular updates and typically without accompanying standardized metadata such as patient demographics, clinical context, treatment history, or environmental sampling location, making cross-study comparisons difficult. Furthermore, differences in nomenclature, quality control, and sequencing methods across laboratories limit both data interoperability and the ability to identify global transmission networks and shared environmental sources [34,42].

The lack of data harmonization also weakens public health responses. For MTB, the detection of identical WGS profiles across countries has triggered cross-border outbreak investigations and interventions, as observed in the European multidrug-resistant TB cluster analysis by Walker et al.[14]. By contrast, no international surveillance mechanism exists to detect emerging pathogenic clones of NTM or to link infections in humans with environmental or nosocomial sources, despite increasing reports of global *M. chimaera* and *M. abscessus* outbreaks related to medical equipment contamination and healthcare settings [43].

Cross-learning opportunities and future innovations in mycobacterial molecular typing

Lessons from MTB genotyping to NTM genotyping

Advancements in the molecular genotyping of MTB have laid a robust foundation for enhancing the study of NTM. One significant contribution is the development of standardized pipelines such as ReSeqTB, a collaborative curated knowledge-based platform designed to standardize and aggregate global WGS data on MTB variants with phenotypic drug susceptibility testing and clinical data. The Unified variant pipeline within ReSeqTB facilitates consistent variant calling and lineage assignment, thereby promoting reproducibility and comparability across studies [41].

Artificial intelligence (AI) and machine learning have been increasingly applied to predict drug resistance phenotypes of MTB. For instance, machine learning algorithms have been developed to predict resistance to

first-line anti-MTB drugs on the basis of genomic data, thereby aiding in the rapid identification of multidrug-resistant MTB strains [44]. These AI-driven approaches can be adapted to predict resistance patterns in NTM, which is crucial given the diverse and often intrinsic resistance mechanisms among the species [45].

Low-input WGS techniques have been optimized to enable the sequencing of MTB in samples (e.g., sputum) with limited DNA quantities or poor quality. These methods demonstrate the feasibility of culture-free WGS, which is particularly beneficial in settings where culture facilities are limited or when rapid results are required. Adapting these low-input WGS protocols to NTM would overcome similar challenges and facilitate timely and accurate species identification and resistance profiling [40].

Contributions from NTM genotyping to MTB genotyping

Research on NTM has also provided valuable insights that can be used to enhance the molecular genotyping of MTB. Environmental surveillance approaches developed for NTM, which often inhabit diverse ecological niches, have highlighted the importance of environmental factors in mycobacterial transmission and infection [1–3]. For example, studies have identified geographical variations in the distribution of NTM species and environmental predictors of their presence, which provide useful data to inform public health strategies. Incorporating environmental surveillance into MTB control programs could improve our understanding of the transmission dynamics, particularly in regions where zoonotic or environmental reservoirs play a role [36].

NTM research has also emphasized the need for precise species-level differentiation owing to the clinical relevance of specific species. Advanced molecular techniques, including MLST and WGS, have been used to achieve accurate species identification [23,42]. Applying these high-resolution genotyping methods to MTB would enhance the detection of microevolutionary changes and transmission chains, particularly for outbreak investigations [46].

Furthermore, studies on biofilm formation by NTM have shed light on the adaptive mechanisms that contribute to their persistence and resistance. These findings are pertinent to MTB because biofilm-like growth is associated with drug tolerance and chronic infections [38]. Understanding the genetic and phenotypic adaptations related to biofilm formation may lead to the development of novel therapeutic strategies for targeting persistent mycobacterial populations [36].

Emerging technologies and global integration

The future of mycobacterial molecular genotyping is being reshaped by portable and scalable technologies that enable real-time, high-resolution genotyping at the point of care. Platforms such as MinION from Oxford Nanopore Technologies and iSeq from Illumina offer real-time, long-read sequencing capabilities in a compact format [35]. The portability and scalability of such platforms render them suitable for deployment in resource-limited settings and for the field-based surveillance of both MTB and NTM [35]. These devices have been successfully used for the rapid sequencing of MTB directly from clinical samples, enabling timely diagnosis and resistance detection in countries such as Madagascar and South Africa [35,46]. Votintseva et

al. [47] demonstrated the feasibility of generating actionable WGS results for MTB directly from clinical samples within 24 h, highlighting the potential of on-site sequencing in clinical workflows [47]. Moreover, advances in isothermal amplification, CRISPR-based diagnostics (e.g., SHERLOCK and DETECTR), and targeted amplicon sequencing may offer simpler point-of-care options suitable for use in resource-constrained areas [48,49].

Collaborative platforms such as GenomeTrakr and the European Nucleotide Archive facilitate the sharing of genomic data across institutions and countries. GenomeTrakr, a distributed network of laboratories coordinated by the US Food and Drug Administration, uses WGS for pathogen identification and source tracking, contributing to a global database that supports outbreak detection [41,50,51]. In addition to such platforms, several mycobacteria-specific genomic databases have emerged to support standardized research, clinical decision making, and global surveillance. ReSeqTB, a curated knowledge-based platform developed by the Critical Path to TB Drug Regimens Initiative, compiles standardized WGS data on MTB isolates worldwide, linking genomic variants to phenotypic drug resistance data [41]. This platform supports the development of genotypic resistance prediction tools and provides a common framework for comparing strains across studies and geographies. MycoBrowser, originally designed for the functional annotation of MTB H37Rv, has evolved into a comprehensive resource offering high-quality annotations of multiple mycobacterial taxa, including *M. leprae* and several NTM species. It enables an in-depth exploration of genes involved in drug resistance, cell wall synthesis, and virulence, thus guiding both diagnostics and therapeutic target discovery [50]. The Genome-wide *Mycobacterium tuberculosis* Variation (GMTV) database integrates mutation, resistance, and epidemiological data to track genomic variations across MTB lineages [51]. GMTV allows researchers to assess mutation hotspots and their correlation with drug resistance or fitness costs, thus enhancing our understanding of the evolutionary dynamics and adaptation of MTB strains. The repository TB Portals, which is supported by the US National Institutes of Health and Centers for Disease Control and Prevention, offers a patient-centric platform that integrates genomic, radiological, and clinical metadata from patients with multidrug-resistant and extensively drug-resistant MTB worldwide [41].

Despite these advances, equivalent infrastructures for NTM remain underdeveloped, impeding efforts to interpret genomic data and trace environmental or nosocomial sources. The recently launched NTM-Explorer platform represents a critical step toward consolidating genomic data and metadata on NTM isolates, fostering standardization, and supporting global strain tracking. However, curated and widely adopted databases analogous to ReSeqTB or TB Portals remain lacking for NTM. This gap is particularly concerning, given the taxonomic diversity, species-specific resistance mechanisms, and increasing healthcare-associated outbreaks of NTM, such as *M. chimaera* infections linked to heater-cooler units in cardiac surgery theatres [43].

Simultaneously, interdisciplinary collaboration among clinicians, microbiologists, epidemiologists, and bioinformaticians is essential for translating sequencing data into actionable insights. For example, interpreting whether a recurrent NTM infection reflects relapse or re-infection or predicting resistance patterns from novel variants requires integrated domain expertise. The international Comprehensive Resistance Prediction for Tuberculosis Consortium exemplifies how global-scale data sharing and coordinated analyses can accelerate the discovery of resistance mechanisms and improve diagnostic algorithms for MTB [52]. Extending such

collaborative models to the NTM domain will be key to advancing standardized interpretations and public health responses.

Ultimately, integrating portable diagnostic platforms with harmonized data infrastructures and interdisciplinary networks will be central to building a globally coordinated mycobacterial genomic surveillance system. Such achievement would serve precision medical care and improve real-time emergency responses to MTB and NTM outbreaks alike.

Conclusions

Molecular genotyping has become an essential tool for elucidating the transmission and resistance dynamics of MTB and NTM, which cause intractable chronic infections. However, the genotyping goals and methodologies for these two pathogens differ significantly owing to their different biological and clinical characteristics. The genotyping of MTB has been integrated into global surveillance systems through the use of standardized tools and high-resolution data integration. By contrast, despite the increasing clinical burden caused by NTM, particularly in older adults and immunocompromised populations, the genotyping of these species remains fragmented and unintegrated owing to a lack of consensus on species definitions, resistance factors, and transmission. As outlined in this review, lessons learned from MTB genotyping, for which advanced technologies such as integrated pipelines, curated databases, and AI-based interpretation have been applied, can inspire the maturation of NTM genotyping. Conversely, insights from NTM studies on environmental reservoirs, species complexity, and persistence are expected to inform clinical studies on MTB. The convergence of technology, bioinformatics, and interdisciplinary collaboration will be critical to not only bridging the knowledge gap between MTB and NTM but also establishing a truly integrated approach to the management of mycobacterial diseases.

Ethics statement

As this was not a human population study, neither approval by the Institutional Review Board nor informed consent was required.

Conflicts of interest

No potential conflicts of interest relevant to this article were reported.

Funding

This work was supported by the Gyeongsang National University Fund for Professors on Sabbatical Leave (2024), the Research Program funded by the Korea National Institute of Health (2025-ER2010-00), and the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education (2021R1I1A2045131), Republic of Korea.

Data availability

This review article does not involve the generation or analysis of new datasets. All data supporting the findings are derived from previously published studies, which are appropriately cited within the manuscript.

References

1. Turenne CY. Nontuberculous mycobacteria: insights on taxonomy and evolution. Infect Genet Evol 2019;72:159-68.
2. Forbes BA, Hall GS, Miller MB, Novak SM, Rowlinson MC, Salfinger M, et al. Practical guidance for clinical microbiology laboratories: mycobacteria. Clin Microbiol Rev 2018;31:e00038-17.
3. WHO. Global tuberculosis report 2023. Geneva: World Health Organization; 2023.
4. Zhang H, Tang M, Li D, Xu M, Ao Y, Lin L. Applications and advances in molecular diagnostics: revolutionizing non-tuberculous mycobacteria species and subspecies identification. Front Public Health 2024;12:1410672.
5. Griffith DE and Aksamit TR. Understanding nontuberculous mycobacterial lung disease: it's been a long time coming. F1000Res 2016;5:2797.
6. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 2006;44:4498-510.
7. Couvin D, Allaguy AS, Ez-Zari A, Jagielski T, Rastogi N. Molecular typing of *Mycobacterium tuberculosis*: a review of current methods, databases, softwares, and analytical tools. FEMS Microbiol Rev 2025;49:fuaf017.
8. Baldwin SL, Larsen SE, Ordway D, Cassell G, Coler RN. The complexities and challenges of preventing and treating nontuberculous mycobacterial diseases. PLoS Negl Trop Dis 2019;13:e0007083.
9. WHO. Global tuberculosis report 2024. Geneva: World Health Organization; 2024. p. 68.
10. Churchyard G, Kim P, Shah NS, Rustomjee R, Gandhi N, Mathema B, et al. What we know about tuberculosis transmission: an overview. J Infect Dis 2017;216:S629-35.
11. Auld SC, Kasmar AG, Dowdy DW, Mathema B, Gandhi NR, Churchyard GJ, et al. Research roadmap for tuberculosis transmission science: where do we go from here and how will we know when we're there? J Infect Dis 2017;216:S662-8.
12. Nahid P, Mase SR, Migliori GB, Sotgiu G, Bothamley GH, Brozek JL, et al. Treatment of drug-resistant tuberculosis an official ATS/CDC/ERS/IDSA clinical practice guideline. Am J Respir Crit Care Med 2019;200:E93–142.
13. Dheda K, Cox H, Esmail A, Wasserman S, Chang KC, Lange C. Recent controversies about MDR and XDR-TB: global implementation of the WHO shorter MDR-TB regimen and bedaquiline for all with MDR-TB? Respirology 2018;23:36-45.
14. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis 2013;13:137-46.
15. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. Lancet 2013;381:1551–60.

16. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J 2013;41:140-56.
17. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martín C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607-18.
18. Braden CR, Crawford JT, Schable BA. Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network. Emerg Infect Dis 2002;8:1210-5.
19. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J Infect Dis 1999;180:1238-44.
20. Garzelli C, Lari N, Nguon B, Cavallini M, Pistello M, Falcone G. Comparison of three restriction endonucleases in IS*1245*-based RFLP typing of *Mycobacterium avium*. J Med Microbiol 1997;46:933-9.
21. Aravindhan V, Sulochana S, Narayanan S, Paramasivam CN, Narayanan PR. Identification & differentiation *of Mycobacterium avium & M. intracellulare* by PCR- RFLP assay using the *groES* gene. Indian J Med Res 2007;126:575-9.
22. Molina-Moya B, Gomgnimbou MK, Spinasse L, Obasanya J, Oladimeji O, Dacombe R, et al. *Mycobacterium tuberculosis* complex genotypes circulating in Nigeria based on spoligotyping obtained from Ziehl-Neelsen stained slides extracted DNA. PLoS Negl Trop Dis 2018;12:e0006242.
23. Ichikawa K, Yagi T, Inagaki T, Moriyama M, Nakagawa T, Uchiya KI, et al. Molecular typing of *Mycobacterium intracellulare* using multilocus variable-number of tandem-repeat analysis: identification of loci and analysis of clinical isolates. Microbiology 2010;156(Pt 2):496-504.
24. Adékambi T, Colson P, Drancourt M. *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J Clin Microbiol 2003;41:5699–708.
25. Kim SH and Shin JH. Identification of nontuberculous mycobacteria using multilocous sequence analysis of 16S rRNA, *hsp65*, and *rpoB*. J Clin Lab Anal 2018;32:e22184.
26. Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA. A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One 2012;7:e30600.
27. Comas I and Gagneux S. The past and future of tuberculosis research. PLoS Pathog 2009;5:e1000600.
28. WHO. Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB). Geneva: World Health Organization; 2008.
29. Nikolayevskyy V, Niemann S, Anthony R, van Soolingen D, Tagliani E, Ködmön C, et al. Role and value of whole genome sequencing in studying tuberculosis transmission. Clin Microbiol Infect 2019;25:1377-82.
30. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 2013;11:728-36.
31. Stucki D and Gagneux S. Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. Tuberculosis 2013;93:30-9.
32. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, et al. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med 2013;10:e1001387.

33. 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;7:51.
34. Tortoli E, Meehan CJ, Grottola A, Fregni Serpini G, Fabio A, Trovato A, et al. Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. Infect Genet Evol 2019;75:103983.
35. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med 2019;11:41.
36. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol 2018;16: 202–13.
37. Lipworth S, Hough N, Leach L, Morgan M, Jeffery K, Andersson M, et al. Whole-genome sequencing for predicting clarithromycin resistance in *Mycobacterium abscessus*. Antimicrob Agents Chemother 2018;63:e01204-18.
38. Brown-Elliott BA, Nash KA, Wallace RJ Jr. Antimicrobial susceptibility testing, drug resistance mechanisms, and therapy of infections with nontuberculous mycobacteria. Clin Microbiol Rev 2012;25:545-82.
39. Diel R, Lipman M, Hoefsloot W. High mortality in patients with *Mycobacterium avium* complex lung disease: a systematic review. BMC Infect Dis 2018;18:206.
40. Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, et al. Direct whole-genome sequencing of sputum accurately identifies drug-resistant *Mycobacterium tuberculosis* faster than MGIT culture sequencing. J Clin Microbiol 2018;56:e00666-18.
41. Rosenthal A, Gabrielian A, Engle E, Hurt DE, Alexandru S, Crudu V, et al. The TB portals: an open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. J Clin Microbiol 2017;55:3267-82.
42. Tortoli E. Microbiological features and clinical relevance of new species of the genus *Mycobacterium*. Clin Microbiol Rev 2014;27:727-52.
43. van Ingen J, Kohl TA, Kranzer K, Hasse B, Keller PM, Katarzyna Szafrańska A, et al. Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. Lancet Infect Dis 2017;17:1033-41.
44. Deelder W, Christakoudi S, Phelan J, Benavente ED, Campino S, McNerney R, et al. Machine learning predicts accurately *Mycobacterium tuberculosis* drug resistance from whole genome sequencing data. Front Genet 2019;10:922.
45. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, et al. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics 2018;34:1666-71.
46. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. Nat Rev Microbiol 2019;17:533-45.
47. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 2017;55:1285-98.
48. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007;315:1709-12.
49. Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, et al. The new Xpert MTB/RIF ultra: improving detection of *Mycobacterium tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing. mBio 2017;8:e00812-17.

50. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis 2011;91:8-13.
51. Chernyaeva EN, Shulgina MV, Rotkevich MS, Dobrynin PV, Simonov SA, Shitikov EA, et al. Genome-wide *Mycobacterium tuberculosis* variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics 2014;15:308.
52. CRyPTIC consortium and the 100,000 genomes project; Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N Engl J Med 2018;379:1403-15.
53. Radomski N, Thibault VC, Karoui C, de Cruz K, Cochard T, Gutiérrez C, et al. Determination of genotypic diversity of *Mycobacterium avium* subspecies from human and animal origins by mycobacterial interspersed repetitive-unit-variable-number tandem-repeat and IS*1311* restriction fragment length polymorphism typing methods. J Clin Microbiol 2010;48:1026-34.
54. Moravkova M, Lamka J, Slany M, Pavlik I. Genetic IS*901* RFLP diversity among *Mycobacterium avium* subsp. *avium* isolates from four pheasant flocks. J Vet Sci 2013;14:99-102.