Enfermedades Infecciosas y Microbiologia Clinica 44 (2026) 503038

Contents lists available at ScienceDirect

Enfermedades Infecciosas y Microbiologia Clinica

journal homepage: www.elsevier.es/eimc

Brief report

Identification of lineages and sublineages of the *Mycobacterium tuberculosis* complex using advanced molecular strategies and WGS in Seville, Spain (2015–2022)

Identificación de linajes y sublinajes del complejo Mycobacterium tuberculosis *mediante estrategias moleculares avanzadas y secuenciación en Sevilla, España (2015–2022)*

María Aznar[a], Mariana G. López[b], María José Torres[c], Eduardo Briones[d], Juan Francisco Medina[e], Verónica González-Galán[f,*]

[a] *Servicio de Microbiología y Parasitología clínica, UCEIMP, Hospital Universitario Virgen del Rocío. Instituto de Biomedicina de Sevilla (IBiS), Grupo de Investigación de Microbiología clínica y Molecular, Línea Micobacterias, Seville, Spain*
[b] *Unidad de Genómica de la Tuberculosis, Instituto de Biomedicina de Valencia (IBV, CSIC), Valencia, Spain*
[c] *Universidad de Sevilla. Instituto de Biomedicina de Sevilla (IBiS), Grupo de Investigación de Microbiología clínica y Molecular, Línea Micobacterias, Seville, Spain*
[d] *Distrito sanitario Sevilla. CIBERESP: Centro de Investigación en Red Epidemiología y Salud Pública, Seville, Spain*
[e] *Unidad Médico-Quirúrgica de Enfermedades Respiratorias, IBiS, Hospital Universitario Virgen del Rocío/Universidad de Sevilla, Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Seville, Spain*
[f] *Servicio de Microbiología y Parasitología Clínica, UCEIMP, Hospital Universitario Virgen del Rocío. Instituto de Biomedicina de Sevilla (IBiS), Grupo de Investigación de Microbiología Clínica y Molecular, Línea Micobacterias, Grupo GEIM-SEIMC y PII-TB& MNT, SEPAR, Seville, Spain*

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Lineages
Whole genome sequencing

Palabras clave:
Mycobacterium tuberculosis
Linajes
Secuenciación masiva

ABSTRACT

Introduction: Tuberculosis (TB) in the Virgen del Rocío university hospital health area, included in the Seville Health District, Spain, exceeds the rate of 17 cases/100,000 inhabitants. Within the *Mycobacterium tuberculosis* complex (MTBC), *M. tuberculosis sensu stricto* comprises seven phylogenetic lineages (L1–L4 and L7–L9). The objective of this work is to know the circulating lineages in the area.
Methods: Molecular typing and sequencing were performed to identify circulating lineages from 2015 to 2022.
Results: From 2015 to 2019, the following were identified: L4 (95.95%, n = 355), L2 (1.62%, n = 6), and L6 (1.08%, n = 4). From 2020 to 2022: L4 94.62% (n = 123); L1 1.5% (n = 2); L3 1.5% (n = 2) and L2 0.76% (n = 1).
Conclusions: The emergence of different lineages in our area and their recognition are key to contributing to disease control. These findings demonstrate the need to continue surveillance and apply sequencing to understand and control the dynamics of TB transmission.

RESUMEN

Introducción: La tuberculosis (TB) en el área sanitaria hospital universitario Virgen del Rocío, incluido en el Distrito sanitario Sevilla, España, supera la tasa de 17 casos/100.000 habitantes. Dentro del complejo *Mycobacterium tuberculosis* (MTBC), *M. tuberculosis sensu stricto* comprende siete linajes filogenéticos (L1–L4 y L7–L9). El objetivo de este trabajo es conocer los linajes circulantes del área.
Métodos: Se realizó tipificación molecular y secuenciación para identificar los linajes circulantes de 2015 a 2022.
Resultados: De 2015 a 2019 se identificaron: L4 (95,95%, n = 355), L2 (1,62%, n = 6) y L6 (1,08%, n = 4). De 2020 to 2022: L4 94.62% (n = 123); L1 1,5% (n = 2); L3 1,5% (n = 2) y L2 0,76% (n = 1).
Conclusiones: La aparición de linajes diferentes en nuestra área y reconocerlos es clave para contribuir al control de la enfermedad. Estos hallazgos evidencian la necesidad de continuar con la vigilancia y aplicar la secuenciación para controlar la dinámica de la transmisión de TB.

DOI of original article: https://doi.org/10.1016/j.eimc.2025.503038
* Corresponding author.
E-mail address: veronica.gonzalez.galan.sspa@juntadeandalucia.es (V. González-Galán).

Received 8 June 2025; Accepted 11 September 2025
Available online 2 January 2026
0213-005X/© 2025 Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Introduction and objectives

Tuberculosis (TB) remains the human infectious disease that causes the greatest number of deaths worldwide each year. The latest WHO report[1] states that in 2023 there were an estimated 10.8 million new cases and 1.25 million TB-related deaths. In our health area, Hospital Universitario Virgen del Rocío in the Seville Health District, the rate of TB exceeds 17 cases/100,000 population.[2]

The genotype of *Mycobacterium tuberculosis* complex (MTBC) strains plays an increasingly important role in understanding the epidemiology of tuberculosis.

Spoligotyping (spacer oligonucleotide typing) is a classic technique which analyses the presence or absence of 43 DR sequences; this provides low resolution, which is useful solely for classifying strains into major lineages. However, its discriminative ability is limited and can lead to errors, particularly in the form of homoplasy or convergent evolution, i.e. strains from different lineages showing similar patterns. For this reason, although it is accessible and inexpensive, its reliability for transmission studies or more detailed subtyping is poor. In contrast, WGS provides much higher genomic resolution, enabling accurate identification not only of the lineage to which a strain belongs, but also of its sublineages and geographic variants.

Some lineages have a wide geographical distribution, such as lineage 4 (L4), which is the most predominant globally. Meanwhile, other lineages are restricted to certain specific geographical areas, such as: L1 and L3 in regions around the Indian Ocean; L5 and L6 in West Africa (https://www.nature.com/articles/s41579-025-01159-w).

The aim of this work is to determine the population structure of MTBC strains circulating in the health area covered by our centre by means of complementary genotyping and next-generation sequencing methods.

Material and methods

This was a prospective observational study conducted in Seville (Spain) over seven years. Initially, molecular typing was performed by applying 43-spacer spoligotyping following the standard protocol; single nucleotide polymorphism (SNP) detection was performed by Taqman real-time PCR according to the protocols of Stucki et al., 2012. In the period 2015–2019, 462 MTBC strains were typed from clinical samples from a molecular epidemiology study on tuberculosis at Hospital Universitario Virgen del Rocío, Seville, Spain. From 2020, next-generation sequencing was used to analyse 130 strains from clinical samples using Illumina technology. Reads were analysed with the pipeline developed at the Tuberculosis Genomics Unit (https://gitlab.com/tbgenomicsunit/ThePipeline). Briefly, reads are cleaned and trimmed for quality (fastp), taxonomically filtered to remove contaminations (kraken); mapped (bwa) using the MTBC ancestor genome as a reference (http://tgu.ibv.csic.es/?page_id=1794) and variant calling of polymorphisms is performed (Varscan and GATK), both of single nucleotide (SNP) and insertions/deletions (InDels). Lineages are assigned using marker SNPs (https://pubmed.ncbi.nlm.nih.gov/25176035/), which in turn enable the identification of mixed infections.

Results

500 (84.46%) of the 592 strains available were typed. The quality of the DNA obtained in the first period did not allow results to be obtained using the spoligotyping technique in 92 strains. There were no strains typed by both techniques. Table 1 shows the molecular typing results from the first study period where six lineages were identified from those described in the literature, with L4 being the most common (95.95%, n = 355) followed by L2 (1.62%, n = 6) and L6 (1.08%, n = 4).

The results of the second period using next-generation sequencing were: L4 94.62% (n = 123); L1 1.5% (n = 2); L3 1.5% (n = 2); and L2 0.76% (n = 1). In addition, two samples (1.5%) were identified with strains of animal lineage, 1 *Mycobacterium bovis* and 1 *Mycobacterium caprae*. The L1 lineage and *M. caprae* appear in 2020, the L3 lineage in 2021 and *M. bovis* in 2022.

Table 1
MTBC Lineage 2015–2019 (370 strains).

MTBC lineage	Large sequence polymorphism (LSP)	N (%)
1	Indo-OCEANIC	1 (0.27)
2	East Asian	6 (1.62)
3	East African-Indian	2 (0.54)
4	Euro-American	355 (95.95)
5	*M. Africanum* West African 1	2 (0.54)
6	*M. Africanum* West African 2	4 (1.08)

The results of the sublineages are shown in Fig. 1, the most common being L4.1.2 (31.82%, n = 42), L4.3.2 (21.21%, n = 28) and L4.10 (19.70%, n = 26). Sublineages were obtained for the 130 strains sequenced using Illumina technology. No mixed infections were identified.

Discussion

For almost 30 years, molecular genotyping tools have been used to characterise strains through the presence/absence of spacer-sequences-spoligotyping. Whole genome sequencing (WGS) of *M. tuberculosis* strains has been a major breakthrough both for the diagnosis of resistance and for the identification of lineages and sublineages, and also enables the quantification of clustering using the genetic distance between strains as a measure to estimate transmission. Next-generation sequencing, as well as having greater discriminatory power, provides us with and allows us to access all the genomic information with a single determination, where we can obtain all the information simultaneously. In summary, while spoligotyping can be useful as an initial approach or in resource-limited contexts, whole genome sequencing is now established as the reference tool in molecular epidemiology, thanks to its high resolution, reliability and informative value. Although more costly and technically demanding, WGS offers a much more comprehensive approach to TB control and surveillance. Apart from phylogenetic information, WGS also makes it possible to detect mutations associated with antimicrobial resistance and analyse the genomic evolution of strains. We carried out this analysis and it is the subject of a separate publication currently being drafted. Due to the nature of the mutations found, the research team deemed it appropriate to include these findings in an article which is currently in preparation.

Throughout the period studied, in our area, the most common lineage is L4, also known as Euro-American, regardless of the technique used (95.95% in the first period of the study and 94.7% in the second). This is in line with data published by other authors, with Euro-American being the most prevalent and widespread lineage in Europe, America and Africa.[3]

Other lineages were described as restricted to specific geographical areas in Asia and Africa. According to Benjamin N. Koleske et al., these patterns are changing, and these lineages are also expanding to other parts of the world.[4] In the first period we detected 4 (1.08%) strains from the L6 lineage, 2 (0.54%) from L5, 355 (95.95%) from L4, 2 (0.54%) from L3, 6 (1.62%) from L2 and 1 (0.27%) from L1 (Table 1). However, in the second period we found no strains of the L6 and L5 lineages, similar numbers of L3 and L1 and a decrease in L2. This may be due to the smaller number of strains studied (130 strains) together with a shorter time period (2020–2022) or because those strains of the L6 and L2 lineages were part of intra-familial outbreaks, although we do not have that information. In the second period, sequencing techniques continued to detect lineage 1, 2 and 3 strains with similar frequencies (Fig. 1), demonstrating the robustness of the techniques used, both being valid

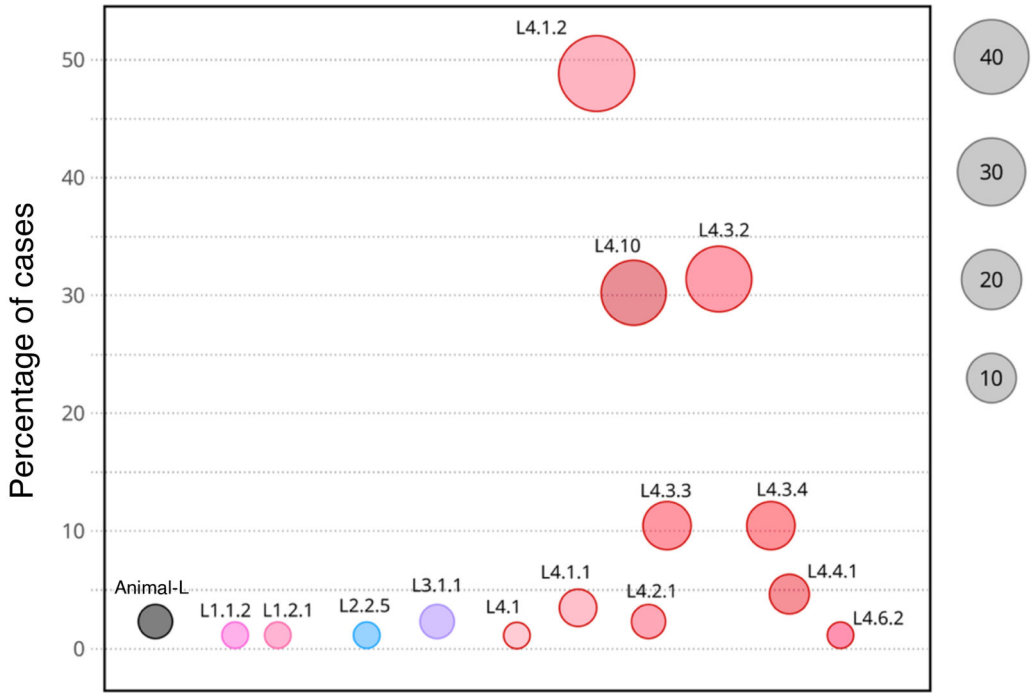

Fig. 1. *M. tuberculosis* sublineages. 2020–2022 (130 strains).

and reliable. Surveillance of these lineages is particularly important, as some of them are associated with increased virulence due to their high capacity to acquire resistance, such as lineage 2-*Beijing*.[4]

In addition, modern lineages are divided into sublineages which also correlate with human geographical distribution. According to the classification suggested by Stucki et al., we are talking about generalist and specialist sublineages.[5] Specialist sublineages are restricted geographically while generalist sublineages are geographically widespread. In our study the most common sublineages were L4.1.2 (Harlem), L4.3.2 (LAM) and L4.10 (PGG3) classified as generalists, which corresponds with the literature regarding their distribution. The L4.10 sublineage (one of the most common in our study) may be associated with a faster rate of acquisition of resistance mutations in West Africa.[6]

The results obtained show the diversity of MTBC circulating in Seville. Over the years of study, we have found strains belonging to six different lineages. Lineage 4 is the most common (until December 2022). However, the presence in our area of specific lineages which were initially limited to geographical areas such as the African and Asian continents, shows that in addition to transmission patterns, the introduction of new strains contributes to maintaining the TB epidemic in our region and is an important factor to be monitored.

The phylogeographic diversity of MTBC is attributed to ecological factors and recent demographic and epidemic events. Different MTBC lineages show variations in their ability to infect, which may indicate adaptation to specific human populations. In India, the Delhi and Beijing lineages are the most predominant, marking the country as an endemic focus for these lineages.[7] Beijing lineage strains have caused significant outbreaks of TB worldwide, sometimes with multiple drug resistance.[8] Specific associations have also been identified in different populations. For example, in Saudi Arabia, a direct link has been established between the *M. bovis* and Delhi lineages with lymph node tuberculosis and gastrointestinal tuberculosis.[9] In Tanzania, lineage 2 has been shown to be associated with being female and cases of retreatment.[10] In Uruguay, a marked increase in Haarlem lineage has been found among intensive care unit (ICU) patients and in the incidence of impaired liver function.[8] This is evidence that MTBC lineages influence the phenotypes of the disease and its epidemiological consequences. In our series L3 is a cluster in transmission composed of two strains and we lack epidemiological information linking them beyond sequencing, while lineage 1 strains are not in transmission. However, it should be noted that *M. bovis and M. caprae strains belong to a transmission cluster in the autonomous region of Andalusia referred to in the publication by* Martinez-Lirola et al.[11]

We found strains belonging to six different lineages, which shows that changes in the population structure of tuberculosis are taking place in our area. These findings demonstrate the need to continue surveillance and to apply molecular epidemiology in our setting in order to understand and control the dynamics of tuberculosis transmission.

Sources of funding

SEPAR GRANT Implementation of whole genome sequencing of *M. tuberculosis* for a multidisciplinary approach to tuberculosis. The main researcher is Verónica González Galán and the collaborators are: Juan Francisco Medina Gallardo, Carmen Calero Acuña, Diana Lucila Ramos Apolaya, Rafael Luque Márquez, Iñaki Comas Espadas, María José Torres Sánchez, Eduardo Briones de la Blanca. Grant reference number. SEPAR 943/2019.

New strategy, based on specific PCR applied to the *M. tuberculosis* complex for active surveillance of tuberculosis transmission in the population of the area covered by Hospital Universitario Virgen del Rocío de Sevilla, PS-0116-2016, Ministry of Health of the Autonomous Government of Andalusia.

Declaration of competing interest

Authors declare that there are no competing interests.

References

1. World Health Organization. *Global tuberculosis report 2024*. Geneva: WHO; 2024.
2. Tuberculosis en Andalucía. Informe de los años 2022-2023. Monografía 2025. Vol. 30, no 02. Available from: https://repositoriosalud.es/entities/publication/076e4545-b338-4116-8bb4-56f36ab9626a/full.
3. Koleske BN, Jacobs WR, Bishai WR. The *Mycobacterium tuberculosis* genome at 25 years: lessons and lingering questions. *J Clin Invest*. 2023;133:e173156.
4. Bainomugisa A, Duarte T, Lavu E, et al. A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage

strain identifies novel variation in repetitive PE/PPE gene regions. *Microb Genom*. 2018;4:e000188.
5. Stucki D, Brites D, Jeljeli L, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet*. 2016;48:1535–1543.
6. Senghore M, Diarra B, Gehre F, et al. Evolution of *Mycobacterium tuberculosis* complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. *Sci Rep*. 2020;10:327.
7. Gutierrez MC, Ahmed N, Willery E, et al. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg Infect Dis*. 2006;12:1367–1374.
8. Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg Infect Dis*. 2002;8:843–849.
9. Varghese B, Shajan SE, Sridharan J, et al. Impact of *Mycobacterium tuberculosis* complex lineages as a determinant of disease phenotypes from an immigrant rich moderate tuberculosis burden country. *Respir Res*. 2018;19:259.
10. Rutaihwa LK, Menardo F, Stucki D, Gagneux S, Borrell S. Insights into the genetic diversity of *Mycobacterium tuberculosis* in Tanzania. *PLoS One*. 2019;14:e0206334.
11. Martínez-Lirola M, Herranz M, Buenestado Serrano S, et al. A One Health approach revealed the long-term role of *Mycobacterium caprae* as the hidden cause of human tuberculosis in a region of Spain, 2003 to 2022. *Euro Surveill*. 2023;28:2200852, http://dx.doi.org/10.2807/1560-7917.ES.2023.28.12.2200852.