UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA

TESIS DOCTORAL

Nuevas metodologías para el diagnóstico de la tuberculosis bovina

New methodologies for the diagnosis of bovine tuberculosis

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Víctor Lorente Leal

DIRECTORAS

Lucía de Juan Ferré y Beatriz Romero Martínez

Madrid

Financiación y agradecimientos

La realización de esta Tesis Doctoral ha sido posible gracias al programa de financiación de la Universidad Complutense de Madrid–Santander Universidades, mediante un contrato predoctoral de Personal Investigador en Formación (CT17/17 - CT18/17), una ayuda para estancias breves en España y en el extranjero (EB25/20), así como una ayuda para estancias breves en el extranjero "Short Term Mission" (2018) y una ayuda de movilidad para congresos de la asociación Med-Vet-Net (2022).

La realización de las investigaciones que componen la presente Tesis Doctoral ha sido posible gracias a la financiación obtenida por parte de la Subdirección General de Sanidad e Higiene Animal y Trazabilidad, Dirección General de la Producción Agraria. Ministerio de Agricultura, Pesca y Alimentación (MAPA) y el Laboratorio de Referencia Europeo para la tuberculosis bovina. Parte del contenido de esta Tesis Doctoral ha sido financiada a través del proyecto de Tecnologías Avanzadas en Vigilancia Sanitaria (TAVS) de la Comunidad de Madrid.

A su vez, la realización de esta Tesis Doctoral ha sido posible gracias a la colaboración con las siguientes autoridades veterinarias regionales: Dirección General de Producción Agropecuaria e Infraestructuras Agrarias, Consejería de Agricultura y Ganadería (Junta de Castilla y León); Subdirección General de Producción Agroalimentaria, Dirección General de Agricultura, Ganadería y Alimentación (Comunidad de Madrid); Sección de Coordinación de Programas de Sanidad Animal, Departamento de Agricultura, Ganadería y Medio Ambiente (Gobierno de Aragón); Dirección General de Agricultura y Ganadería, Consejería de Agricultura Agua y Desarrollo Rural (Junta de Castilla-La Mancha); Servicio de Producción y Sanidad Animal, Consejería de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica (Generalitat Valenciana). Finalmente, agradecer la colaboración del Laboratorio Central de Sanidad Animal de Granada.


UNIVERSIDAD
COMPLUTENSE
MADRID


Santander
Universidades


GOBIERNO
DE ESPAÑA
MINISTERIO
DE AGRICULTURA, PESCA
Y ALIMENTACIÓN


MED • VET • NET
ASSOCIATION


EU-RL
Bovine
Tuberculosis
European Union Reference Laboratory


Comunidad de Madrid

Doctorado con mención internacional

La presente Tesis Doctoral cumple con los requisitos exigidos por la Universidad Complutense de Madrid para obtener la mención de Doctor Internacional:

1. Realización de una estancia mínima de tres meses en una institución de enseñanza superior o centro de investigación fuera de España:
    i. Primera estancia:
        - Centro receptor: Animal & Plant Health Agency, Weybridge, United Kingdom.
        - Investigador principal: Dr. Richard Ellis.
        - Duración: 1,5 meses (23/02/2019-07/04/2019).
    ii. Segunda estancia:
        - Centro receptor: University College Dublin, Dublin, Irlanda
        - Investigador principal: Dr. Stephen V. Gordon.
        - Duración: 1,8 meses (26/09/2020-21/11/2020).
2. El contenido de la Tesis Doctoral, con excepción del resumen, ha sido redactado en una de las lenguas habituales para la comunicación científica en su campo de conocimiento, distinta a cualquiera de las lenguas oficiales en España (inglés).
3. La Tesis Doctoral ha sido evaluada por dos expertos pertenecientes a alguna institución de educación superior o instituto de investigación no español.
4. El Tribunal evaluador de la tesis está compuesto por, al menos, un experto perteneciente a alguna institución de educación superior o centro de investigación no español.

Animal & Plant Health Agency

UCD DUBLIN

"There is nothing like looking, if you want to find something.

You certainly usually find something, if you look,

but it is not always quite the something you were after."

The Hobbit - J. R. R. Tolkien

Agradecimientos

Escribir y defender una Tesis Doctoral es un camino largo y tedioso en el que uno se centra tanto en su objetivo, que muchas veces pararse a disfrutar del paisaje parece algo secundario. Aunque a veces las fotos panorámicas puedan parecer las más increíbles, son aquellas que se toman con la gente que te ha acompañado durante todo el recorrido las que realmente merecen la pena. Ahora, echando la vista hacia atrás, puedo ver con alegría que en la mayoría de estas fotografías no cabe ni un alfiler. Estoy convencido de que, si no hubiese sido por todas y cada una de estas personas, llegar hasta este punto no habría sido posible. Es por ello que me gustaría emplear esta sección para agradecerle a todas esas personas que han formado parte de mi viaje a pesar de las dificultades.

En primer lugar, me gustaría agradecerles a mis directoras de tesis, Lucía y Bea, su enorme apoyo y esfuerzo durante todos estos años. Ha sido un verdadero placer trabajar con vosotras y me siento muy afortunado de haber podido hacerlo siempre entre risas y con buenas palabras. Me gustaría agradeceros especialmente que depositaseis en mi vuestra confianza hace cinco largos años, cuando llamé a VISAVET preguntando por un trabajo en el Servicio de Anatomía Patológica y, por cosas de la vida, acabé iniciando este apasionante viaje en el mundo de las micobacterias. Lucía, no puedo dejar de agradecerte que siempre me hayas apoyado en mis propuestas e ideas, aportando siempre una visión distinta y constructiva, y ayudándome a ver más allá de mis investigaciones. Bea, me gustaría agradecerte tu apoyo y paciencia durante estos cinco largos años; tu puerta ha estado siempre abierta para recibirme con una sonrisa, a pesar de lo ocupada que pudieses estar en ese momento, y tu visión positiva me ha ayudado a ver el lado bueno de las cosas (esto, para alguien tan "cenizo" como yo, no es asunto baladí). Manu, mi "director" honorífico, muchas gracias por enseñarme desde el principio a pensar con rigor y de manera crítica, por iniciarme en el maravilloso (y críptico) mundo de las PCR a tiempo real (sin ti, no creo que hubiese aprendido ni la mitad de lo que se), y por aguantar mis interminables preguntas, dudas y debates con una sonrisa en los labios.

Me gustaría agradecer a todo el equipo del Servicio de Micobacterias del Centro VISAVET (Tati, Nuria, Daniela, Laura Sánchez, Alex y Fran) todo su apoyo y ayuda durante este tiempo. Habéis sido una parte fundamental de este trabajo y siempre me habéis echado un cable en mis momentos de necesidad,

incluso cuando el tiempo no acompañaba. ¡Muchísimas gracias por todo! Fran, me gustaría agradecerte especialmente los esfuerzos que has realizado por facilitar la incorporación de mis proyectos dentro del flujo de trabajo del servicio y por tratar de darme una solución práctica a cualquiera de mis dudas o problemas.

Me gustaría mencionar a mis compañeros de travesía "visavetiana" que tantas alegrías (y viernes festivos) me habéis brindado. Sin vosotros, este camino no habría sido tan divertido ni especial; me habéis hecho sentir uno más desde el principio y ayudado a disfrutar más allá del aspecto académico. A los que llegasteis al principio del viaje; Marisauer, Alvaro Roy, Pater (gracias por recogerme de aquella cesta de mimbre), Eli (gracias por darme mi segundo trabajo como "chico del cable" de SEVISEQ), Irene Tomé, José Antonio, Chesko, Houses, Gabi, Sonia y Pikachu; muchas gracias por vuestra amistad, por iniciarme en los FNS y abrirme las puertas del "Carmelo's". Agradecimientos especiales para Mazas, por hacerme reír prácticamente todos los días y ayudarme a deshacerme de todo tipo de dulces. A Alvarito, muchas gracias por tu alegría y positivismo contagioso de por las mañanas. A Arantxa, mi compi de despacho congelador, ¡muchas gracias por ser siempre tan amable y cercana! A mis compañeros de viaje que llegaron un poco más tarde, pero que no por ello son menos especiales; Clara NED, Pedro, Irene Aldea (y tu amor incondicional por los Journal Club), Vicente, Jairo, Paloma, Javi, Javier Ortega (por preocuparte siempre por mí y por ofrecerte a ayudarme en lo que hiciese falta), a Alberto (por tu energía y espontaneidad); Laura, Claudia y Carmen (por ayudarme a retomar la normalidad tras el COVID sin ser yo nada de eso); a Blanca, Néstor y Agustín; ¡gracias a todos por acompañarme estos años! A Laura T., muchísimas gracias por tu ánimo constante en esta recta final de la tesis (te debo unas patatas revolconas). A mis súper compañeras de oficina; Coral, gracias por estar siempre ahí y por nuestras conversaciones "frikis", y Clara, gracias por tu confianza y por tu apoyo todos estos años. ¡Ambas habéis sido una parte muy importante de este camino y espero que siga siendo así! Finalmente, Pilar, me gustaría agradecer toda tu ayuda estos años (y tus códigos de R); has sido una compañera de trabajo genial y me has ayudado en los peores momentos, especialmente en la recta final que ha sido la escritura de esta tesis.

No puedo olvidarme del resto de miembros de VISAVET: Nerea, Irene, Marta, Fátima, Antonio, Joaquín, Lucas, Teresa, Sergio, Cris, Susana, Leydis (por enseñarme a comer Portugal), Almu, Carmen Bárcenas, Peter, David, Laura D. y Lorena. No importa lo duro que haya podido llegar a ser el camino, siempre

habéis sabido sacarme una sonrisa de alguna manera, y eso es algo de lo que me siento muy afortunado. Me gustaría además agradecerle a Marga su amabilidad todo este tiempo y a los chicos SIC (Carlos y Javi) por vuestras bromas y por permitirme hacer mis locuras informáticas. Finalmente, gracias Julio† por toda tu inestimable ayuda en temas estadísticos (y de filetes) y por tu sentido del humor†[1]. Es difícil encontrar a gente así, y aunque espero tener la misma suerte que he tenido estos cinco años en mis próximos pasos, me conformo con llevarme un montón de buenos recuerdos allá donde vaya.

Me gustaría también agradecer de manera especial a las dos personas que me han acompañado durante buena parte de mi experiencia como docente en la Facultad de Veterinaria. José Luis, gracias por hacerme sentir parte del equipo y permitirme aportar a tus clases, he disfrutado mucho todos estos años y he aprendido muchísimo contigo y los alumnos. Carmen, no tengo palabras para agradecer tu amabilidad y energía durante las clases que hemos dado juntos; muchísimas gracias por confiar siempre en mí y por darme todo tu apoyo.

También ha habido un buen grupo de personas que, más allá de la universidad, han sido un pilar fundamental durante estos años. Valeriano (a.k.a. Sr. Pimpón), muchas gracias por apoyarme en los momentos más difíciles, saber sacarme una sonrisa siempre y ayudarme a distraerme del trabajo cuando más lo necesitaba. Drosio, muchas gracias por apoyarme desde el principio de los tiempos, cuando ambos éramos dos estudiantes perdidos en Holanda y nos atiborrábamos a comida china como gremlins, y por recordarme aquellas cosas que me gustan de la vida (ya tu sah). A Torra, muchas gracias por tu sentido del humor y por preocuparte por mí y mi "depósito". ¿Creo que podemos decir que, esta vez, no he jugado tan mal? Y a ti, Paqui, muchas gracias por tus ánimos y alegría todo este tiempo (y por controlar al gremlin). A mis amigos inseparables del colegio (Alberto, Jorge y Aitor); porque habéis estado ahí desde mis comienzos y habéis sido una constante muy importante en mi vida. A ti, Aitor, me gustaría agradecerte especialmente tu apoyo durante la "Era Marquinote" y en los momentos más duros de la recta final. Finalmente, me gustaría dar las gracias a Criswi, mi "hermana menor"; gracias por estar ahí siempre, pase lo que pase.

---

[1] Ahí tienes tu crucecita.

I cannot leave without mention all those persons that have accompanied me during my travels around the world. I would like to thank Richard Ellis for helping me in my first steps with whole genome sequencing and making me feel so welcome at the Animal & Plant Health Agency. Susie and Sammy, I can't forget how well you treated me during my stay in the UK; you really made me feel at home and I will always remember you for this (and the crumpets)! Damien, thank you for all your support and help during my stay in Ireland, I learnt a great deal from you! And Stephen, thank you for making me feel part of the team in Ireland and for all of your help and hospitality. Finally, but not least, I would like to thank Jasmin and Nele, my two musketeers; without you and your support I wouldn't be where I am today. I am really proud of being your friend and I am very happy that we ended up fulfilling our objectives together, despite the distance.

A mi familia, me gustaría agradeceros todo el apoyo que me habéis ofrecido durante estos cinco largos años. Creo que no hay doctorando que no haya tenido que escuchar la pregunta "¿Y cuánto te queda?" al menos diez mil veces, pero con vosotros esto siempre ha sido una celebración. Paloma, Pepe, Loli, Lines, Paloma Jr., David, Mili, Efi, Carmen, Antonio, Antonio Sr., Rosa, Gema y Roberto ¡muchas gracias por vuestro apoyo, por darme siempre ánimos y por confiar en mí! A mi hermano, gracias por preocuparte siempre por mí y por mi futuro y por tratar siempre de aconsejarme de la mejor manera (a pesar de mi testarudez).

A Irene, que apareciste en mi vida al principio de este proyecto y me has acompañado desde entonces; no puedo haber encontrado a una compañera mejor. Has sido el apoyo más fuerte e importante de todos, ayudándome a seguir adelante cuando no me quedaban fuerzas y confiar en mí cuando yo no sabía hacerlo. Espero poder devolverte todo lo que has hecho por mí estos años, y hacerte sentir tan feliz como me he sentido yo por tenerte a mi lado. Espero que podamos seguir viajando juntos, allá donde haga falta (a ser posible, que sea fresquito).

Finalmente, no podría acabar esta sección sin darle las gracias a las dos personas más importantes de mi vida. A mis padres, porque lo habéis dado todo por mí y por mi hermano siempre, porque si no hubiese sido por vosotros y vuestro sacrificio no habría ido construyendo los cimientos del camino que me ha llevado hasta aquí. Creo que no hay ni habrá palabras suficientes para agradecer todo lo que habéis hecho por mí, y por ello este trabajo es tanto mío como vuestro.

Foreword

This doctoral thesis is a compendium of four research studies that have been published in indexed scientific journals. These studies have derived from the accomplishment of three research objectives focused on the development and evaluation of novel methodologies for the diagnosis of bovine tuberculosis.

The above mentioned studies reflect the work carried out by the PhD candidate during his matriculation in the Doctoral Programme in Veterinary Science and are in accordance with the rules and requirements of the Doctoral Commission of the Universidad Complutense of Madrid and the Department of Animal Health of the Faculty of Veterinary Medicine of the same university. In the first place, the suitability of the publication of this thesis in a publication format is sustained by the thematic cohesion of the included studies and the fact that the PhD candidate was first author in all of the publications. Furthermore, the scientific articles included herein have all been published in journals within the first quartile of the Journal Citation Reports (JCR). Finally, the publication of these scientific articles reflect the extensive skill set that the PhD candidate has acquired during his studies in relation to the highly diverse methodologies involved, as well as the analyses required for the scientific interpretation of the results.

The scientific articles included herein are:

1)	Authors	Lorente-Leal, V., Liandris E., Castellanos E., Bezos J., Domínguez L., de Juan, L. & Romero, B.
	Title	Validation of a Real-Time PCR for the Detection of *Mycobacterium tuberculosis* Complex Members in Bovine Tissue Samples.
	Publication date	04 of March of 2019
	Journal reference	Frontiers in Veterinary Science 6, article 61
	DOI	10.3389/fvets.2019.00061
	Position in JCR index	19/141 (Q1)
	Index factor	2.245
	Category	Veterinary Sciences
2)	Authors	Lorente-Leal, V., Liandris E., Pacciarini, M., Botelho, A., Kenny, K., Loyo, B., Fernández, R., Bezos J., Domínguez L., de Juan, L. & Romero, B.
	Title	Direct PCR on Tissue Samples to Detect *Mycobacterium tuberculosis* Complex: an Alternative to the Bacteriological Culture
	Publication date	21st of January 2021
	Journal reference	Journal of Clinical Microbiology
	DOI	10.1128/JCM.01404-20
	Position in JCR index	22/137 (Q1)
	Index factor	5.948
	Category	Microbiology
3)	Authors	Lorente-Leal V, Farrell D, Romero B, Alvarez J, de Juan L, Gordon SV.
	Title	Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization.
	Publication date	25th of November 2021
	Journal reference	Frontiers in Veterinary Science
	DOI	10.3389/fvets.2021.780018
	Position in JCR index	9/146 (Q1)
	Index factor	3.412
	Category	Veterinary Sciences

4)	Authors	Lorente-Leal V, Liandris E, Bezos J, Pérez-Sancho M, Romero B, de Juan L.
	Title	MALDI-TOF Mass Spectrometry as a Rapid Screening Alternative for Non-tuberculous Mycobacterial Species Identification in the Veterinary Laboratory.
	Publication date	28th of January 2022
	Journal reference	Frontiers in Veterinary Science
	DOI	10.3389/fvets.2022.827702
	Position in JCR index	9/146 (Q1)
	Index factor	3.412
	Category	Veterinary Sciences

Index

Summary

Bovine tuberculosis (bTB) is a chronic infectious disease of bovids that is caused by members of the *Mycobacterium tuberculosis* complex (MTBC), mainly *M. bovis* and *M. caprae*. Due to the negative impact of this infection in animal health and production, as well as to its zoonotic potential, bTB is subjected to surveillance and eradication programmes in different parts of the world. In the European Union, eradication of bTB is governed by the 'Animal Health Law' (AHL; Regulation 2016/429) and related legislation. In Spain, eradication of bTB is transposed from EU legislation through the Royal Decree 2611/1996 and the National Eradication Programme for bTB.

Eradication of bTB in the EU is based on a 'test and slaughter' strategy by which infection is detected through the use of official *ante-mortem* diagnostic methods, and eliminated through the culling of infected animals. Positive animals to official diagnostic tests may be subjected to additional tests, such as microbiology, in order to confirm the infection. The suspicion or confirmation of the infection in a herd entails the suspension or withdrawal of its OTF status, which results in a series of restrictions that can greatly affect the competitiveness of the herd.

Although eradication programmes have been in place in the EU for decades and have contributed to the significant decrease of the prevalence of bTB, this infection remains a problem in many regions. This is not only due to the complex

epidemiology of this infection, in which several domestic and wildlife reservoirs are involved, but also as a result of limitations in currently used diagnostic and characterisation tools, including the limited sensitivity (Se) and increased incubation times of microbiological culture, the limited resolution of conventional molecular characterisation methods used in outbreak investigations and the diagnostic interference caused by the presence of non-tuberculous mycobacteria (NTM) in *ante-mortem* diagnostic methods.

The main goal of this doctoral thesis was to develop novel diagnostic and characterisation tools that can be integrated into bTB eradication programmes and can contribute in enhancing the efficiency of the eradication efforts. This was achieved through three objectives: 1) the development of sensitive and rapid diagnostic methods for the detection of bTB; 2) the development of high-resolution molecular genetic methods for epidemiological investigations; and 3) the development of rapid screening tools for the identification of NTM species.

The AHL allows the use of other methods for the confirmation of bTB, in addition to microbiological culture, including the detection of MTBC-specific nucleic acids through molecular genetic methods, such as real-time PCR (qPCR). Therefore, the first objective was focused on the development and validation of two qPCR protocols targeting MTBC-specific genetic targets, the *mpb70* (study 1) and IS*6110* insertion sequence (study 2), respectively, for the direct detection of MTBC DNA in bovine tissue samples (direct qPCR). The diagnostic performance of these qPCRs was compared against microbiological culture in a panel of 200 and 985 fresh bovine tissue samples, respectively. The two qPCRs showed a high diagnostic performance, with Se and specificity (Sp) values of 88.41% and 92.37% for the *mpb70* PCR and 96.45% and 93.66% for the IS*6110* qPCR, respectively. The Se increased when MTBC DNA was confirmed in culture-negative/PCR-positive samples with additional molecular methods, showing a higher or comparable Se over culture. However, both qPCR protocols presented the advantage of being significantly faster than culture, requiring at most three days to obtain results in comparison with the weeks required by culture. Finally, a limited cross-reactivity was detected for the IS*6110* qPCR against a '*M. avium* subsp. *hominissuis*' strain and whole genome sequence (WGS) analysis revealed the presence of an IS*6110*-like element with 83% similarity to the IS*6110* element. Despite this cross-reactivity, no unspecific amplification was detected in any of the bovine tissue samples analysed in study 2, including 100 samples from 78 bTB-free herds. Overall, the results from studies 1 and 2 reflect the diagnostic potential of both qPCR protocols as an

alternative method to culture, with an increased sensitivity, speed and throughput capacity, as well as decreased costs.

Molecular characterisation techniques used for the genotyping of MTBC members analyse a set of molecular markers that comprehend a small fraction of the microorganism's genome. Although their use has allowed to unravel the genetic structure and relatedness of MTBC members, with important applications in large-scale molecular epidemiology studies, their reduced resolution limits their use in more localised studies, including herd breakdown investigations. The second objective of this thesis was focused on the evaluation of WGS, a high-resolution and high-throughput technique increasingly used to study bTB around the world, as a molecular characterisation method for bTB epidemiological studies. This objective was pursued through two different studies focusing on the comparison of performance of bTB variant calling methods (study 3) and their implementation in the assessment of within-herd genetic variation of the most frequent *M. bovis* genotype in Spain (SB0121) in chronically-infected herds (study 4).

Study 3 involved four different variant calling methods for *M. bovis* WGS data (vSNP, SNiPgenie, BovTB and MTBseq) and evaluated their performance, agreement and phylogenetic outputs in a panel of 47 artificial *M. bovis* genomes generated from a high bTB prevalence region in the Republic of Ireland (RoI). Different filter files were used to mask regions of the genome of MTBC members that are difficult to analyse. The agreement and performance between the pipelines was high, with precision and recall rates over 99.8% when the same filter files were used. Furthermore, the inferred phylogenetic trees were highly similar to the ones observed in the simulation and varied with the application of different filters, although small variations were observed in the positioning of highly related taxa due to the presence of polytomies. These results indicate that available variant calling methods show a high performance and are suitable for epidemiological studies of bTB.

Study 4 evaluated 70 *M. bovis* SB0121 isolates obtained from 22 herds, in six Spanish provinces, with persistent bTB infections in order to assess the use of WGS to establish the cause or persistence. The analysis revealed an overall high genetic diversity, with three major genetic clades. A low within-herd genetic diversity was detected in a group of herds, indicating the circulation of highly similar strains within the herds or their surroundings. Highly diverging strains were identified in other herds, indicating the presence of different sources of infection, such as contact with other herds or wildlife reservoirs. These findings

reflected a complex epidemiological scenario in which both highly similar and distantly related strains circulated in the herds and their surroundings. The analysis of additional samples from different wildlife species revealed the possible frequent between-species transmission of *M. bovis* from cattle and sympatric wildlife in one of the studied regions (Ciudad Real), which could play a major role in maintaining a diverse pool of *M. bovis* strains in certain environments. Finally, WGS revealed the presence of genetically similar strains between different herds, which suggested a shared infection source and possible epidemiological links. In conclusion, the results of this study highlight the potential of WGS in rapidly assessing within-herd diversity as a mean to detect possible bTB recurrence or re-introduction events from common or distant sources of infection, greatly enhancing outbreaks investigations in chronically infected herds and, ultimately, aiding in eradicating this infection.

The isolation of NTM from animals diagnosed with bTB significantly affects eradication efforts and the trust of stakeholders in eradication measures. Identification of NTM isolates is usually achieved through the use of Sanger sequencing, which presents a series of limitations due to the taxonomic variations in the genus and the genetic similarity of certain mycobacterial groups. The third objective of this thesis evaluated the use of matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF MS) as a rapid screening method for the identification of NTM isolates obtained from animal clinical samples (study 5). A total of 75 NTM isolates obtained from 20 animal species were identified using both MALDI-TOF MS and Sanger sequencing. Both methodologies identified a similar proportion of NTM isolates; 88% (n=66) of the isolates were reliably identified by both methods (n = 57) or could be identified by neither (n = 9). Nevertheless, MALDI-TOF MS achieved identification in a significantly reduced time in comparison to Sanger sequencing. The presence of unidentified isolates probably indicated the presence of unknown NTM species and a high diversity of NTM species in animals. Discordances were identified for both techniques and were related to low quality identifications or possible misidentifications due to the absence of spectral profiles from certain NTM species in current commercial databases, which are mainly based on clinical isolates from human medicine. Overall, the results of study 5 support the use of MALDI-TOF MS as a rapid and high-throughput method for the identification of NTM.

The studies included in this thesis highlight the potential of recent advances in molecular and proteomic methods in the study of bTB and other mycobacteriosis in animals. The implementation of the techniques and methods described herein will greatly improve the sensitivity, throughput capacity and cost-effectiveness of surveillance and eradication programmes in the EU.

Resumen

La tuberculosis bovina (TBb) es una enfermedad infecciosa crónica producida por distintos miembros del complejo *Mycobacterium tuberculosis* (CMTB). Debido al efecto negativo que tiene esta infección en la sanidad y producción animal, además de por su potencial zoonósico, la TBb está sometida a programas de vigilancia y erradicación en distintas partes del mundo. En la Unión Europea (UE), la erradicación de la bTB está enmarcada dentro de la 'Ley de Sanidad Animal' (AHL por sus siglas en inglés; Regulación 2016/429) y su legislación relacionada. En España, la erradicación de la bTB ha sido traspuesta de la legislación europea a través del Real Decreto 2611/1996 y el Programa Nacional de Erradicación de la TBb.

La erradicación de la bTB en la UE se basa en una estrategia de 'diagnóstico y sacrificio' mediante la cual la infección es detectada a través del uso de técnicas diagnósticas *ante-mortem* oficiales, y eliminada mediante el sacrificio de los animales infectados. Los animales que son positivos a estas pruebas pueden ser sometidos a pruebas adicionales, como el cultivo microbiológico, con el objetivo de confirmar la infección. La sospecha o confirmación de la infección en un rebaño implica la suspensión o la retirada de su calificación 'oficialmente libre de tuberculosis' (OTF por sus siglas en inglés), lo cual resulta en la implementación de una serie de restricciones que afectan gravemente a la competitividad del rebaño.

Aunque los programas de erradicación llevan establecidos en la UE desde hace décadas y estos han contribuido a la reducción significativa de la prevalencia de la bTB, esta infección sigue siendo un problema importante en muchas regiones. Esto no sólo es debido a su compleja epidemiología, en la cual diversas especies domésticas y silvestres están involucradas, sino también por las limitaciones de los métodos diagnósticos y de caracterización actuales; entre ellas, la limitada sensibilidad (Se) y los largos tiempos de incubación de las técnicas de confirmación convencionales (i.e. cultivo microbiológico); la limitada resolución y capacidad de procesado de las técnicas convencionales de caracterización molecular empleadas en el estudio de brotes; y la interferencia diagnóstica producida por la presencia de micobacterias no tuberculosas (NTM por sus siglas en inglés) en las técnicas diagnósticas *ante-mortem*.

El objetivo principal de esta tesis doctoral ha sido el desarrollo de nuevas metodologías con el fin de mejorar los programas de erradicación actuales y futuros. Esto se ha conseguido a través de tres objetivos, subdivididos a su vez en cinco estudios distintos, enfocados en 1) el desarrollo de métodos diagnósticos rápidos y sensibles para la confirmación de la bTB; 2) el desarrollo de herramientas de caracterización genética de alta resolución para estudios epidemiológicos; y 3) el desarrollo de herramientas de cribado rápido para la identificación de NTM en muestras clínicas.

La implementación de la AHL ha permitido el uso de otros métodos de confirmación de la bTB además del cultivo microbiológico, como pueden ser la detección de ácidos nucleicos específicos del CMTB a través de métodos genéticos moleculares como la PCR a tiempo real (qPCR). Por lo tanto, el primer objetivo se centró en el desarrollo y validación de dos protocolos de qPCR dirigidos a la detección de dos dianas genéticas específicas del CMTB, el gen *mpb70* (estudio 1) y la secuencia de inserción IS*6110* (estudio 2), respectivamente, para la detección directa de ADN específico del CMTB en muestras de tejido bovinas (qPCR directa). La capacidad diagnóstica de estas qPCRs fue comparada frente al cultivo microbiológico en un panel de 200 y 985 muestras de tejido bovino, respectivamente. Las dos qPCRs mostraron una elevada capacidad diagnóstica, con valores de sensibilidad (Se) y especificidad (Sp) diagnósticas del 88,41% y del 92,37% para la PCR dirigida al *mpb70* y 96,45% y 93,66% para la PCR dirigida a la IS*6110*, respectivamente. La confirmación de la presencia de ADN del MTBC en 85,36% de las muestras cultivo-negativas y PCR-positivas del estudio 2 indicaron una mayor sensibilidad de la qPCR dirigida a la IS*6110*, mientras que la confirmación en tan sólo el 50%

de este tipo de muestras en el estudio 1 indicaron una capacidad comparable a la del cultivo. Sin embargo, ambas técnicas tienen además la ventaja de ser significativamente más rápidas que el cultivo microbiológico, requiriendo como mucho tres días para la obtención de resultados, frente a las semanas que requiere el cultivo. Finalmente, se detectó reactividad cruzada de la qPCR dirigida a la IS*6110* frente a una cepa de '*M. avium* subsp. *hominissuis*', y un análisis de secuenciación masiva (WGS por sus siglas en inglés) reveló la presencia de una secuencia genética con un 83% de identidad frente a la secuencia IS*6110*. A pesar de esta reactividad cruzada, no se detectó ninguna amplificación inespecífica en ninguno de las muestras de tejidos analizadas en el estudio 2, incluyendo 100 muestras procedentes de 78 rebaños calificados como OTF. En conclusión, los resultados de los estudios 1 y 2 reflejan el potencial diagnóstico de ambos protocolos de qPCR como técnicas alternativas al cultivo, con una elevada sensibilidad, velocidad y capacidad de análisis, además de costes reducidos.

Las técnicas de caracterización molecular empleadas para el genotipado de los miembros del CMTB analizan un set de marcadores moleculares que corresponden a una pequeña fracción del genoma del microorganismo. Aunque su uso ha permitido revelar la estructura genética y de parentesco entre los miembros del CMTB, con importantes aplicaciones en estudios de epidemiología molecular a gran escala, su limitada reducida resolución limita su uso en estudios más localizados, como pueden ser la investigación de brotes. El segundo objetivo de esta tesis se centró en la evaluación de la WGS, una técnica de elevada resolución y capacidad cada vez más empleada en el estudio de la bTB en todo el mundo, como una técnica de caracterización molecular para estudios epidemiológicos. Este objetivo se desarrolló a través de sendos estudios enfocados en la comparación de la capacidad de los métodos de análisis de variantes disponibles para bTB (estudio 3) y su implementación en el estudio de la variabilidad intra-rebaño del genotipo más frecuente de *M. bovis* en España (SB0121) en rebaños con infecciones crónicas (estudio 4).

El estudio 3 se centró en cuatro herramientas de análisis de variantes para el estudio de datos genómicos de *M. bovis* (vSNP, SNiPgenie, BovTB y MTBseq) y evaluó su rendimiento, concordancia y conclusiones filogenéticas en un panel de 47 genomas artificiales de *M. bovis* generados a partir de un set de datos procedente de una región de alta prevalencia de bTB en la República de Irlanda. Se emplearon distintos archivos de filtrado para enmascarar aquellas regiones del genoma del MTBC que son difíciles de analizar. La concordancia

y el rendimiento de estas herramientas fue elevada, con unos valores de precisión y Se por encima del 99,8% cuando se emplearon los mismos archivos de filtrado para cada herramienta. Aunque los árboles inferidos a partir de los resultados de dichas herramientas variaron con la aplicación de los distintos filtros, estos fueron muy similares entre sí, salvo por pequeñas variaciones observadas en el posicionamiento de taxones altamente relacionados debido a la presencia de politomías. Estos resultados indican que las herramientas de análisis de variantes disponibles son adecuadas para su uso en estudios epidemiológicos de bTB.

En el estudio 4 se evaluaron 70 aislados de *M. bovis* SB0121 obtenidos de 22 rebaños, procedentes de seis provincias españolas, con TBb persistente con el objetivo de evaluar el uso la WGS para determinar las causas de su cronicidad. El análisis reveló una elevada diversidad genética de este genotipo, con tres clados genéticos principales. Se detectó una baja diversidad genética intra-rebaño en un grupo de rebaños, lo cual indica la circulación de aislados similares en estos rebaños o sus alrededores. Se identificó una elevada variabilidad en otros rebaños, indicando la presencia de distintas fuentes de infección, como otros rebaños o reservorios silvestres. Estos resultados reflejaron un escenario epidemiológico muy complejo en el que tanto cepas similares como menos relacionadas circulaban en los rebaños y sus alrededores. El análisis de muestras adicionales procedentes de distintas especies cinegéticas reveló la transmisión inter-especie de *M. bovis* entre bovino y fauna silvestre simpátrica en una de las regiones estudiadas (Ciudad Real), lo cual podría ser un reservorio de *M. bovis* con una gran diversidad genética. Finalmente, el estudio genómico reveló la presencia de aislados genéticamente similares entre distintos rebaños, lo cual sugiere la presencia de una fuente común de infección y posibles vínculos epidemiológicos. En conclusión, los resultados de este estudio subrayan el potencial de la WGS para evaluar la diversidad intra-rebaño como una forma de detectar posibles fenómenos de recurrencia o reintroducciones de TBb, mejorando así las investigaciones de brotes en rebaños con infecciones crónicas y, de manera última, ayudando a la erradicación de esta enfermedad infecciosa.

El aislamiento de NTM en animales diagnosticados con bTB afecta de manera significativa a los esfuerzos realizados para la erradicación de esta infección, así como a la confianza de los distintos sectores involucrados hacia los métodos diagnósticos empleados en los programas de erradicación. La identificación de aislados de NTM suele realizarse mediante la secuenciación Sanger, la cual presenta una serie de limitaciones debido al gran número de variaciones

taxonómicas del género *Mycobacterium* y la elevada similitud genética de determinados grupos micobacterianos. El tercer objetivo de esta tesis doctoral se centró en la evaluación de la espectrometría de masas (EM) por desorción/ionización láser asistida por una matriz con detección de masas por tiempo de vuelo (MALDI-TOF por sus siglas en inglés) como una herramienta rápida de cribado con la cual identificar aislados de NTM obtenidos de muestras clínicas animales (estudio 5). La capacidad de identificación de la EM MALDI-TOF Bruker BioTyper se evaluó en un panel de 75 aislados de NTM obtenidos de 20 especies animales, y fue comparada frente a la identificación por secuenciación Sanger. Ambas metodologías identificaron una proporción similar de aislados de NTM; el 88% (n = 66) de los aislados fueron identificados de manera fiable por ambos métodos (n = 57) o no pudieron ser identificados por ninguno (n = 9). Sin embargo, la EM MALDI-TOF obtuvo las identificaciones en un tiempo mucho menor que la secuenciación Sanger. La presencia de aislados no identificados probablemente indica la presencia de especies de NTM desconocidas y una elevada diversidad de especies NTM en animales. Se detectaron algunas discordancias entre ambas técnicas que estaban relacionadas con la presencia de identificaciones de baja calidad o posibles identificaciones erróneas debido a la ausencia de perfiles proteicos para determinadas especies de NTM en las bases de datos comerciales, que han sido construidas generalmente a partir de aislados clínicos procedentes de la medicina humana. En general, los resultados del estudio 5 refuerzan el uso de la EM MALDI-TOF como un método rápido y de alta capacidad para la identificación de NTM.

Los estudios incluidos en esta tesis doctoral subrayan el potencial de los recientes avances en métodos moleculares y proteómicos para el estudio de la bTB y otras micobacteriosis animales. La implementación de las técnicas y métodos descritos en esta tesis servirán para mejorar la sensibilidad, capacidad y relación costo-rendimiento de los programas actuales y futuros de vigilancia y erradicación de la bTB.

1. Introduction

1.1. Tuberculosis

1.1.1. Animal and bovine tuberculosis

Tuberculosis (TB) is a chronic infectious disease that affects a wide range of mammals, including humans, and is caused by a closely-related group of mycobacteria defined as the *Mycobacterium tuberculosis* complex (MTBC). Traditionally, the members of the MTBC have been considered as separate species, but due to their high genetic similarity they are currently considered variants of *M. tuberculosis* (Riojas et al., 2018). Although the modern taxonomy of the MTBC will be included in this thesis, the infrasubspecific term "variant" ('var.') will be avoided due to the extended use of the traditional nomenclature in veterinary literature.

The MTBC includes two human-adapted variants, *M. tuberculosis sensu stricto* and *M. africanum,* and nine animal-adapted variants and strains: *M. bovis, M. caprae, M. microti, M. pinnipedii, M. orygis, M. suricattae, M. mungi,* the dassie bacillus and the chimpanzee bacillus. Although the inclusion of *M. canettii* in the MTBC is controversial, it is important to mention it here as it is considered to be related to the most recent common ancestor (MRCA) of the MTBC (Supply *et al.*, 2013; Supply & Brosch, 2017).

Tuberculosis in animals, or animal TB (aTB), is mainly restricted to non-human mammals and can affect an extremely diverse array of species, including bovids (e.g. cattle, buffaloes, bisons, goats or antilopes), swine (e.g. domestic pigs and wild boar), cervids (e.g. deer or moose), pinnipeds (e.g. seals), mustelids (e.g. badgers), felines and canines (both domestic and wild), elephants, rhinoceros, small rodents (e.g. voles) and non-human primates.

Infections in animals are mainly caused by the animal-adapted MTBC members, since animal infections with the human-adapted members, such as *M. tuberculosis* or *M. africanum,* are rare (Romero *et al.*, 2011; Ameni *et al.*, 2013; Paudel & Sreevatsan, 2020). Transmission of TB between animals is considered to occur mainly via the respiratory route, although other routes such as ingestion or contamination of wounds have also been described, especially in other MTBC members like *M. mungi* (Alexander *et al.*, 2016). A more detailed description of the epidemiology and transmission of aTB is given in section 1.3.

While certain MTBC members show a reduced host range, such as *M. mungi,* which causes TB in mongoose, others can infect a wider range of animal species, such as *M. bovis*. In fact, *M. bovis* is regarded as the hallmark of animal-adapted MTBC variants and is majorly responsible for the large number of animal species affected by aTB (Malone & Gordon, 2017). As a generalist pathogen, *M. bovis* has been shown to infect an extremely diverse range of wild and domestic animals, including cattle, goats, sheep, wild boar, deer or badgers, among others.

More specifically, *M. bovis* is an extremely relevant pathogen of cattle, which led to its recognition as the main causal agent of bovine TB (bTB). Furthermore, due to the economic importance of cattle around the world and the zoonotic potential of bTB, *M. bovis* has become the most studied animal-adapted member of the MTBC (OIE, 2018).

Originally, bTB was considered a disease of bovines caused by *M. bovis*. However, the increased isolation of this pathogen in other bovids and the finding that other members of the MTBC, such as *M. caprae* (Aranaz *et al.*, 2003; Rodriguez *et al.*, 2011) or *M. orygis* (van Ingen *et al.*, 2012), are also significant pathogens in this group of animals have recently led to the redefinition of the concept of bTB in certain parts of the world, such as the EU (Bovine Tuberculosis Subgroup of the Task Force, 2013).

Throughout this thesis, the term aTB will be used to describe TB caused by all members of the MTBC in all mammalian species, excluding humans, whereas

bTB will specifically be used to address the infection of bovids by MTBC members, mainly by *M. bovis* and *M. caprae*.

1.1.2. History of TB, bTB and impact of zoonotic TB

Bovine TB is an important disease of cattle which cannot be understood without considering the history of humankind and human TB, as this disease has probably accompanied human populations since their early migrations from the African continent and the first domestication of cattle (Hershkovitz *et al.*, 2008; Brites & Gagneux, 2017; Brites *et al.*, 2018). Human TB has been a major cause of mortality throughout history; descriptions of the disease and its clinical forms have populated many historical civilizations, including ancient Greece and the Roman empire, and the causes of *phtisis,* or consumption, have been many and varied (Karamanou & Androutsos, 2012; Cambau & Drancourt, 2014). Despite its long history, it was not until 1882 that the causal agent of TB, the "tuberculous bacillus", was isolated by Robert Koch. With it, the establishment of "Koch's postulates" in 1884 revolutionised our understanding of infectious diseases until today.

By that time TB was also known as a disease of mammals, as bovines and other mammalian species were affected by it, with a symptomatology and pathology similar to that seen in humans (Atkins, 2016). Certainly, TB in bovines was already considered an important veterinary disease with great economic, agricultural and hygienic consequences. While the name *M. tuberculosis* was used to describe the human pathogen (Lessel, 1960), *M. bovis* was established as the causal agent of TB in bovines (Karlsson & Lessel, 1970). Since then, an increasing number of variants have been described in the MTBC, all of which present varying levels of host adaptation and clinical significance despite their limited genetic diversity (Brites *et al.*, 2018).

Nowadays, human TB is considered the leading cause of death from a single infectious agent in the world, with current estimations suggesting that approximately one quarter of the world's population is infected, 10 million people develop the disease each year, whereas 1.5 million die as a result of TB (WHO, 2020). In humans, TB is mainly a respiratory disease that affects the lungs (pulmonary TB), since it is usually transmitted via the airborne route. However, the infection may be transmitted through other routes, such as the oral route, and to other tissues and organs, such as the gastrointestinal tract, the central nervous system or the skeletal system (extrapulmonary TB).

Importantly, the *M. tuberculosis* variants responsible for TB in animals can also infect and cause significant disease in humans (zoonotic TB or zTB). The predominant causal agent of zTB is considered to be *M. bovis,* due to its global expansion, wide host range and prevalence. In 2019, the WHO estimated that 140,000 new cases of zTB were caused by *M. bovis* and 11,400 people died of this disease (WHO, 2020). Incidence of the disease is low but well-documented in high income countries (HICs) (Majoor *et al.*, 2011; Davidson *et al.*, 2017) and is mainly observed in lower-middle-income countries (LMICs) (Muller *et al.*, 2013); especially in the African continent (68,900 cases) and south-eastern Asia (43,400 cases) (WHO, 2020). However, these values probably represent an underestimation of the true incidence of the disease due to many factors, but most importantly as a result of a lack of routine disease surveillance in LMICs, and the inability of standard TB diagnostic protocols to differentiate between *M. tuberculosis* and *M. bovis* (Olea-Popelka *et al.*, 2017). Such differentiation is only possible with the use of genetic molecular methods, which are expensive and not always available in these settings. Furthermore, zTB generally produces the extrapulmonary form of TB (Majoor *et al.*, 2011; Davidson *et al.*, 2017), which limits the isolation of *M. bovis* through standard procedures such as sputum collection.

The use of *M. bovis* as a proxy of zTB has historically led to the underestimation of other possible causal agents of zTB, such as *M. caprae* or *M. orygis*. Only recently has the EU started to stratify zTB cases as caused by *M. bovis* or *M. caprae,* revealing that the latter represented 2.25% of the total (n = 89) of zTB notified cases in the EU in 2019 (ECDC, 2021). The available information strongly suggests that human *M. caprae* infections may be restricted to the European continent and the surrounding Mediterranean region (Tar *et al.*, 2009; Bayraktar & Togay, 2011; Lahlou *et al.*, 2012). Isolations outside of this region have only been reported twice in scientific literature (USA and Australia), and were either found to be related to a European migrant (Sintchenko *et al.*, 2006) or their origin could not be ascertained (Hlavsa *et al.*, 2008). Furthermore, a molecular epidemiological study in southern Asia has recently proposed that *M. orygis* may be the major cause of zTB in the Indian subcontinent, since it was the only animal-adapted strain isolated (n = 0.7%, 7) in a longitudinal study carried out in India (Duffy *et al.*, 2020).

Zoonotic TB is mainly transmitted by the ingestion of contaminated dairy products, including milk and cheese (Harris *et al.*, 2007), or raw meat from infected animals. The significance of the oral route of infection could partly explain the association of extrapulmonary TB in patients with zTB (Durr *et al.*,

2013). The implementation of milk pasteurisation, appropriate cooking and abattoir inspections have been crucial in order to reduce the foodborne transmission of zTB (WHO *et al.*, 2017). In fact, ingestion of unpasteurized milk and dairy products was the main cause of paediatric lymphadenitis in HICs before the implementation of pasteurization of milk and is still considered the main risk factor for zTB around the world (Thoen *et al.*, 2006). Direct contact with infected animals is also a recognised mode of transmission and is a common risk factor for animal-related professions such as livestock farmers (especially in pastoral communities), veterinarians, abattoir workers or hunters (Vayr *et al.*, 2018). Therefore, the control of the infection in the animal population is also critical in order to prevent the transmission of zTB (WHO *et al.*, 2017).

Transmission of zTB is governed by multiple sociodemographic and economic factors that make its control a complicated task to achieve unless a multidisciplinary approach is promoted. The consumption of raw dairy and meat products is still a common practice in many cultures of the world and food hygiene and security measures can be hard to implement without the appropriate economical and anthropological measures. The presence of TB in domestic and wild animal species will, therefore, always pose a risk of infection for humans unless the disease is controlled in the animal population, especially in those cases in which eradication programmes are not economically or environmentally feasible (Arnot & Michel, 2020). Furthermore, although rare, the possibility of anthroponotic infections of *M. tuberculosis* and other MTBC members to other animal species, such as cattle, has been described (Romero *et al.*, 2011). As a result, zTB cannot be fully addressed until human and animal health are both taken into account through a One Health perspective (WHO *et al.*, 2017).

1.2. Aetiology of tuberculosis in animals

1.2.1. Genus *Mycobacterium*

1.2.1.1. Common features and taxonomy

The genus *Mycobacterium* is a member of the *Mycobacteriaceae* family, included within the phylum *Actinobacteria,* class *Actinobacteria*, order *Actinomycetales*, suborder *Corynebacterineae* (Goodfellow *et al.*, 2012). Mycobacteria form an extensive group of bacteria; the current List of Prokaryotic Names with Standing Nomenclature includes more than 200 species (Parte *et al.*, 2020). The genus includes obligate and opportunistic pathogens and saprophytes, isolated from multiple sources including animals and environmental samples.

Mycobacteria are Gram-positive, non-motile, bacilli which are usually obligate aerobes but that in many cases can resist micro-aerobic conditions. They are asporogenous and, while most grow on relatively simple substrates, others require the addition of specific supplements for their growth, such as mycobactin for *M. avium* subspecies *paratuberculosis* (MAP) (Goodfellow *et al.*, 2012).

Colony morphology varies within the genus, and has been associated to pathogenicity in certain groups, as is the case with the rugose morphology of the MTBC (Boritsch *et al.*, 2016a). The composition of their cell wall differs from other bacteria due to the presence of mycolic acids. These are 2-alkyl, 3-hydroxy long-chain fatty acids that can be found associated with the peptidoglycan components of the bacterial cell wall or unbound to the cell envelope (Marrakchi *et al.*, 2014). Mycolic acids are responsible for the acid-alcohol resistance (acid-fastness) of mycobacteria, an important property shared with *Nocardia* by which bacilli resist decolouration by acids routinely used in staining procedures, and that is amply exploited in the diagnosis of TB (Reynolds *et al.*, 2009).

Classification of mycobacteria was initially based on the comparison of phenotypic characteristics, mainly growth rate (rapid *vs.* slow growing mycobacteria) and pigmentation (Runyon, 1959). Rapid and slow growing mycobacteria (RGM and SGM, respectively) tend to generate colonies before and after seven days, respectively. Interestingly, the most important pathogens within the genus *Mycobacterium* are slow growing, whereas most saprophytes are part of the rapid growing group. The classification based on pigmentation

was originally established for SGM and comprises three groups: photochromogens (pigmentation develops in the presence of light), scotochromogens (pigmentation develops in both darkness and light) and non-photochromogens (weak pigmentation if any) (Runyon, 1959). Cultivability has also been used to classify mycobacteria; non-cultivable or difficult to cultivate mycobacteria include *M. leprae* and *M. lepraemurium*, whereas cultivable mycobacteria include the rest of species.

Mycobacteria are also frequently classified based on their capacity to produce TB; tuberculous mycobacteria correspond to the members of the MTBC, whereas non-tuberculous or atypical mycobacteria (NTM), or Mycobacteria other than TB (MOTT), would include the rest of mycobacterial species, with the exception of *M. leprae*. Non-tuberculous mycobacteria are ubiquitous in the environment and can also be important primary and secondary pathogens of several animal species; such as MAP in cattle or *M. marinum* in fish (Biet & Boschiroli, 2014; Puk *et al.*, 2018). In addition, NTM have received increased attention due to their relevance as opportunistic pathogens in humans (Biet *et al.*, 2005), as well as their capacity of interfering with bTB diagnostic methods (see sections 1.4 and 1.6).

1.2.1.2. Genetic features of mycobacteria and the MTBC

Apart from their lipid-rich cell envelope, mycobacteria are commonly known for their high genomic G/C content (> 60%) (Fedrizzi *et al.*, 2017). Despite being a homogenous group within the *Actinomycetales,* genome size varies considerably within the genus, ranging between 3 to 8 Mb (Tortoli *et al.*, 2017).

Genomic diversity is also heterogeneous across the genus, varying within the different mycobacterial groups; while certain complexes, such as the *M. terrae* complex, present a high genetic diversity (figure 1), others, such as the MTBC, are highly clonal microorganisms and share more than 99.9% of their genome at the nucleotide level (Tortoli *et al.*, 2017). Furthermore, pan-genome analyses of mycobacterial genomes have revealed a highly diverse gene content with a great variety of functions (Fedrizzi *et al.*, 2017). Some of these genes are conserved among different species and complexes, whereas others appear to be specific to certain species. However, the function of most of the genes identified in mycobacterial genomes is poorly understood (Fedrizzi *et al.*, 2017).

Based on genetic similarity, mycobacteria can be divided into different monophyletic and paraphyletic groups, commonly known as complexes (Koh, 2017). Some of these complexes comprise important pathogens of animals and

humans, such as the MTBC, or the *Mycobacterium avium* complex (MAC). Recently, the use of Whole Genome Sequencing (WGS) has allowed for the proposal of a more comprehensive classification of mycobacteria into five emended genera; *Mycobacterium, Mycobacteroides, Mycolicibacter, Mycolicibacterium* and *Mycolicibacillus* (Gupta *et al.*, 2018). However, the implementation of such nomenclature remains controversial (Tortoli *et al.*, 2019; Armstrong & Parrish, 2021; Meehan *et al.*, 2021).

Traditionally, genomic plasticity in mycobacteria has not been attributed to Lateral Gene Transfer (LGT) events, partly as a result of their rigid and complex cell wall (Gray & Derbyshire, 2018). However, the discovery of distributive conjugative transfer (DCT) in *M. smegmatis* and *M. canettii* (Gray *et al.*, 2013; Boritsch *et al.*, 2016b), and the detection of episomal genetic elements in *M. avium* or *M. abscessus* (Matsumoto *et al.*, 2014; Uchiya *et al.*, 2015), suggest that this may be more frequent than previously anticipated, at least in certain mycobacterial species. Nevertheless, the overall impact of LGT in genomic evolution of mycobacteria has been proposed to be lower than previously expected (Fedrizzi *et al.*, 2017).

As previously mentioned, TB is caused by a closely-related group of acid-fast bacilli (AFB), the MTBC. The main members of the MTBC share more than 99.9% of their genome composition, but nevertheless present different host-adaptation. Genomic variation differs between members of the complex, being more pronounced between members of the human-adapted strains than between the animal-adapted strains (Malone & Gordon, 2017).

The development of molecular genetic methods, including WGS led to the detection of a series of genetic features that are shared between the different members of the MTBC. These genetic elements have been crucial to understand the evolution of this group of pathogens, as well as to develop diagnostic and molecular typing methods to be used in the control and eradication of TB in humans and animals. The different genetic features are described in this section, whereas the molecular genetic methods used for their analysis will be detailed in sections 1.4 and 1.5 of this thesis.

Figure 1. Phylogenetic tree of the genus *Mycobacterium* constructed from WGS data. Source: Tortoli *et al*., 2017. With permission from Infection, Genetics and Evolution.

Regions of difference (RD)

Comparative genomic studies of MTBC members allowed the identification of genomic regions that are deleted in some variants and conserved in others (Mahairas *et al.*, 1996). These regions were termed Regions of Difference (RDs) or Large Sequence Polymorphisms (LSPs) and can affect up to 10 kbp. Many of these deletions are conserved in different strain groups and lineages and were used to define the currently accepted evolutionary scenario of the MTBC by which strains evolve through clonal genome reduction (Brosch *et al.*, 2002). In this scenario, the MTBC is divided into different lineages which correspond to *M. tuberculosis sensu stricto* (L1 – L4), *M. africanum* (L5 and L6) and the animal-adapted MTBC members, such as *M. bovis* or *M. orygis* (Comas *et al.*, 2013). Consequently, all MTBC members with exception of *M. tuberculosis* L1 – L4 share the RD9 deletion, whereas *M. africanum* L6 and the animal-adapted members share RD7, RD8 and RD10. In addition, *M. bovis* presents the RD4 deletion (figure 2).

Despite the specificity of several RDs, other genomic deletions converge in different variants. For example, RD3 and RD11, two regions that contain prophage sequences (see section below), have been found to diverge between tuberculous bacilli strains (Brosch *et al.*, 2002). Interestingly, *M. microti, M. mungi, M. suricattae, M. bovis* BCG and the dassie bacillus all present different deletions in the RD1. In the case of *M. bovis* BCG, the RD1 deletion has been associated with its attenuation and is suspected to be partially responsible for the attenuated phenotype of *M. microti,* which shows a significant overlap to the RD1 deletion of *M. bovis* BCG (Lewis *et al.*, 2003; Brodin *et al.*, 2004a). Another important example of RD convergence is RD900, which is conserved in *M. orygis* but varies between *M. bovis* and *M. caprae* strains (Brites *et al.*, 2018). Additionally, although the RD4 is specific of *M. bovis*, certain strains of *M. caprae* have been shown to have a partial deletion in this region (Domogalla *et al.*, 2013).

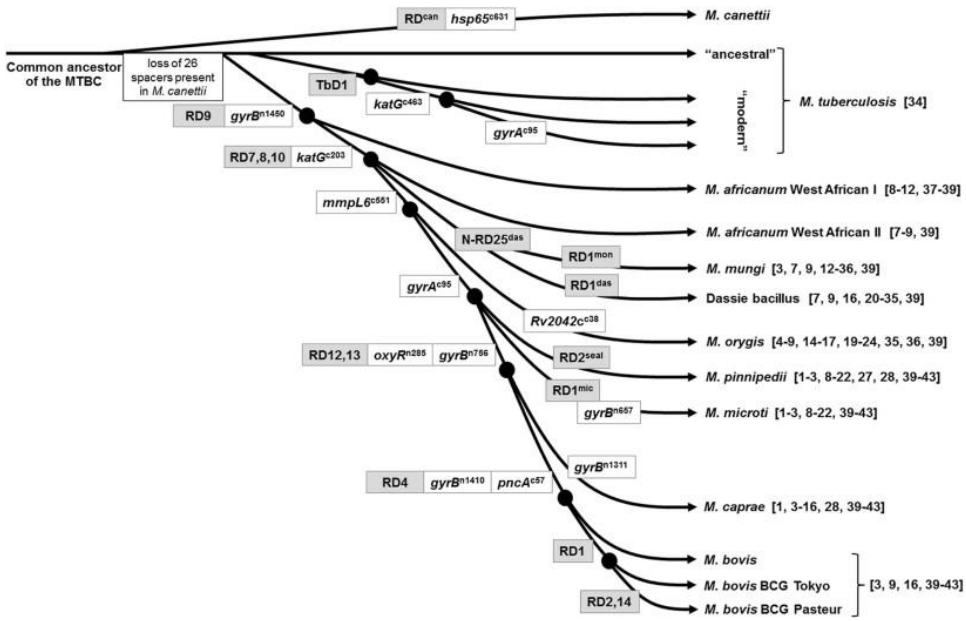


Figure 2. Phylogeny of the MTBC based on the presence/absence of regions of difference (RDs) and Single Nucleotide Polymorphisms (SNPs). Square brackets correspond to Direct Repeat (DR) spacer deletions. Source: Rodríguez-Campos *et al*., 2014. With permission from Research in Veterinary Science.

Insertion sequences

Insertion sequences (IS) are small transposable elements that encode for no other function but that which is involved in their own mobility (Mahillon & Chandler, 1998). In comparison to other transposable elements, such as retrotransposons, these sequences are relatively small and compact (< 2.5 kb). In the majority of IS, a transposase is confined within two terminal inverted repeats, termed left and right inverted repeats, respectively. In addition, the transposase may be encoded by one or two open reading frames (ORF).

Insertion sequences are classified into different families according to sequence homology, ORF organisation or target sites, and can be found in different bacterial and archaeal genomes (Mahillon & Chandler, 1998). Elements from different families have been described in mycobacteria, such as the IS*3*, IS*256* or the IS*110* families. Some of these IS may be specific to a group of mycobacteria; the IS*1081* and IS*6110* elements, representative members of mycobacterial IS*256* and IS*3* families, respectively, are considered specific of the MTBC, whereas the IS*1311* and IS*1245* are specific of the MAC. In turn, some

IS are specific to one mycobacterial species or subspecies, such as the IS*900* or IS*901* in MAP or *M. avium* subsp. *avium*, respectively (Turenne *et al.*, 2007). However, homologous sequences have been identified for IS previously considered specific of different mycobacteria, such as the IS*6110*-like element identified in *M. smegmatis* or the IS*1245* element in *M. kansasii* (Coros *et al.*, 2008; da Silva Rabello *et al.*, 2010).

Of all of the IS elements present in MTBC genomes (Gordon *et al.*, 1999b), the IS*6110* and IS*1081* elements are probably the most relevant, since they are extensively used as molecular markers for the detection and/or characterisation of tuberculous bacilli. Of special importance is the IS*6110* element, previously known as IS*986* in *M. tuberculosis* and IS*987* in *M. bovis* BCG (Hermans *et al.*, 1990; Thierry *et al.*, 1990; Hermans *et al.*, 1991). This mobile element is distributed across the genome of MTBC members and its location and copy number has been found to differ both between MTBC variants and strains from the same variant (Yeh *et al.*, 1998). For example, *M. tuberculosis* and *M. orygis* are known to harbour several copies of the IS*6110* element (Gordon *et al.*, 1999b; van Ingen *et al.*, 2012), whereas *M. bovis* usually contains a single copy, or even no copies (Steensels *et al.*, 2013), which could be related to a reduced transposition frequency in this variant (van Soolingen *et al.*, 1994; Gonzalo-Asensio *et al.*, 2018) and/or an over-representation of *M. bovis* genotypes that usually contain a single copy, such as the European 1 (Eu1) complex, in current datasets (Berg *et al.*, 2011; Smith, 2012). The IS*1081* element, in turn, appears to be present in a stable number of copies (n = 6) in MTBC genomes, with limited differences in its distribution and therefore transposition frequency between variants (van Soolingen *et al.*, 1992).

Due to their intrinsic capacity to alter the target genome and lead to deleterious events, ISs can play a critical role in the evolution of microorganisms, especially in the case of the MTBC in which a limited genetic diversity and LGT events are observed for compensation (McEvoy *et al.*, 2007). Transposition of ISs usually occur at random positions, but certain regions of the genome present a higher frequency of transposition than others. For example, the IS*6110* element is frequently found within the Direct Repeat (DR) locus, which is considered an integration hot spot in the genome of MTBC members (Hermans *et al.*, 1991). Additional targets for IS*6110* integration are multiple gene families, including *pe/ppe* genes and genes of unknown function (Yesilkaya *et al.*, 2005). In addition to gene interruption, IS*6110* transposition could lead to altered gene expression by acting as a mobile promoter, or could lead to deletion of

genomic regions as a result of recombination between two adjacent copies (Soto *et al.*, 2004).

Prophage sequences and the REP13E12 element

Prophages are bacteriophage genomes that have integrated into bacterial chromosomes or exist as an episomal element (Canchaya *et al.*, 2003). Several prophages and prophage-like sequences have been identified in mycobacterial genomes and classified into clusters according to sequence similarity and functional capacities (Fan *et al.*, 2014; Glickman *et al.*, 2020). In the case of MTBC genomes, two prophages have been identified; the first one, phiRv1, is present in *M. tuberculosis* and *M. bovis,* whereas the second, phiRv2 has only been found in the former.

These elements probably do not encode for an active prophage, but phiRv1 has been shown to be able to excise and integrate *in vitro*, which could explain the finding of a nearly identical phiRv1 copy at two different locations in two *M. tuberculosis* type strains, H37Rv and CDC 1551 (Bibb & Hatfull, 2002). Furthermore, the regions containing the two prophages (RD3 and RD11) have been found to diverge between tuberculous bacilli strains and are, therefore, not considered suitable molecular markers (Brosch *et al.*, 2002). Although RD3, containing phiRv1, is absent in *M. bovis* BCG, this region is not suspected to affect virulence since its absence has also been described in many clinical isolates of virulent *M. bovis* (Mahairas *et al.*, 1996).

The phiRv1 element in *M. tuberculosis* H37Rv is integrated within a copy of REP13E12, a degenerate repetitive element that is present in seven copies in the genome. REP13E12 elements vary in sequence length and structure, containing one or two ORFs in an overlapping or truncated fashion (Gordon *et al.*, 1999b). In addition, they are suspected to contain potential chromosomal attachment sites that promote prophage integration (Gordon *et al.*, 1999b), which is further supported by the fact that the phiRv1 copy of *M. tuberculosis* CDC 1551 is integrated in a different copy of the REP13E12 element (Bibb & Hatfull, 2002). The REP13E12 elements are considered to be specific to the MTBC, since they have also been detected in other variants, such as *M. bovis, M. africanum* or *M. microti,* among others (Gordon *et al.*, 1999b).

Variable number of tandem repeats (VNTRs)

Variable number of tandem repeats are a type of repetitive element found in genomes of eukaryotes and prokaryotes. In the classification of repetitive sequences, VNTRs are considered minisatellites, which correspond to genomic loci in which a short repetitive sequence is replicated in a tandem fashion, leading to the appearance of different alleles containing a variable number of repeats.

Minisatellites were originally described in eukaryotes in the late 1960s, but were not identified in MTBC members until the availability of an increasing number of genomic sequences from *M. tuberculosis* H37Rv in the 1990s (Goyal *et al.*, 1994; Frothingham, 1995; Supply *et al.*, 1997; Frothingham & Meeker-O'connell, 1998). Due to their variable nature, VNTRs rapidly became a popular genetic marker for the characterisation of MTBC members, as will be discussed later (see section 1.5.3.2).

Since their discovery, several groups of VNTR loci have been described, albeit with different nomenclatures; these include the exact tandem repeats or ETRs (Frothingham & Meeker-O'connell, 1998), Queen's University Belfast (QUB) VNTRs (Roring *et al.*, 2002; Skuce *et al.*, 2002), the mycobacterial interspersed repetitive units (MIRUs) (Supply *et al.*, 1997; Supply *et al.*, 2000), the major polymorphic tandem repeats or MPTRs (Frothingham, 1995) or the Mtb-v or *Mycobacterium tuberculosis* variation loci (Mtb-v) (Spurgiesz *et al.*, 2003). In some cases, certain VNTR loci have received more than one name; for example, the ETR-D corresponds to MIRU-4, ETR-F to Mtb-v 20 and MIRU-23 to QUB-2531c (Roring *et al.*, 2002; Supply *et al.*, 2006). In order to standardise the plethora of available VNTR nomenclatures, a system was established by which VNTR names were based on the first four digits of their position in the *M. tuberculosis* H37Rv genome (Smittipat & Palittapongarnpim, 2000). Under this convention, for example, the ETR-D locus is called VNTR 0580. Nevertheless, the original aliases are still frequently used in scientific literature.

The different VNTR loci in MTBC genomes usually vary in repeat length and number. For instance, ETRs are usually 53-79 base pairs (bp) long, and can contain up to seven repeats according to the locus (Frothingham & Meeker-O'connell, 1998). On the other hand, MIRU repeats can span between 40 to 100 bp long and can be divided into three major types: type I sequences are usually up to 77 bp long, whereas type II and III usually contain a gap sequence of 24 or 15 bp, respectively (Supply *et al.*, 1997; Supply *et al.*, 2000). Additional features can be observed that differentiate between the different VNTRs; for

example, MIRUs are usually present in intergenic regions and can, in many cases, contain small ORFs, whose initiation and stop codons can overlap with surrounding genes (Supply *et al.*, 1997). In other cases, VNTR loci can be contained within ORFs, usually in *pe/ppe* genes, such as several QUBs (e.g. QUB-11a) and the MPTR (Frothingham & Meeker-O'connell, 1998; Skuce *et al.*, 2002).

With exception of VNTRs that are part of protein-encoding genes, such as the MPTR in *pe/ppe* genes, little is known about the real function of these repetitive sequences. It has been proposed that certain VNTRs could have a role in the regulation of gene expression when they are upstream of translational start sites (Tantivitayakul *et al.*, 2010).

Proline-glutamate (PE) and proline-proline-glutamate (PPE) families of proteins

The *pe/ppe* genes are the most abundant genes in the genomes of MTBC members, comprising almost 10% of their coding capacity (Cole *et al.*, 1998). These genes tend to be clustered within the genomes and usually contain multiple copies of GC-rich repetitive sequences, such as the polymorphic GC-rich repetitive sequences (PGRSs) or the MPTRs. As previously mentioned, these repetitive sequences were already described as abundant repetitive sequences in MTBC genomes and were considered important molecular markers for genotyping of tuberculous bacilli isolates before the description of *pe/ppe* genes (Hermans *et al.*, 1992; van Soolingen *et al.*, 1994; Poulet & Cole, 1995).

The MPTR elements are structurally similar to the DR locus and consist of a 10 bp repeat with a consensus sequence of 5'GCCGGTGTTG3' (Hermans *et al.*, 1992). In contrast, PGRSs are composed of a 9 bp repeat (5'CGGCGGCAA3') that is believed to have originated from the expansion of a CGG or CRR triplet, where R is a purine nucleotide (Poulet & Cole, 1995). Homologous sequences have been identified in other mycobacteria, such as *M. gordonae* or *M. gastrii,* and sequence diversification appears to have been higher for MPTR repeats than for PGRS repeats (Hermans *et al.*, 1992; Poulet & Cole, 1995). At the amino acid level, PE/PPE proteins present a relatively conserved N-terminal domain containing Pro-Pro and Pro-Glu motifs (Gey van Pittius *et al.*, 2006; Sampson, 2011). The more diverse C-terminal domains, in which PGRS or MPTR repeats are found, are comprised of polypeptide motifs composed of Gly and Ala. According to these features, PE/PPE proteins are subdivided into different

families, such as PE_PGRS, PPE_MPTR or PPE_SVP subfamilies (Gey van Pittius *et al.*, 2006).

Homologous genes have been found in NTMs, such as *M. marinum* or *M. smegmatis,* and genetic studies have identified at least five coevolving *pe/ppe* gene subfamilies in the *Mycobacterium* genus, which agree with their structural classification (Gey van Pittius *et al.*, 2006). Interestingly, the *pe/ppe* genes found in RGM are associated with the *esx-1* gene cluster, which is also frequently observed in slow growing mycobacteria (see below). In the latter case, additional *pe/ppe* genes appear to have expanded from an ancestral *pe/ppe* and early-secreted antigenic target 6 (ESAT-6) cluster by *esx* gene cluster duplication, followed by subduplications of the *pe/ppe* genes (Gey van Pittius *et al.*, 2006). As a result, there is a great variety of *pe/ppe* genes in MTBC genomes, which in turn are considered to be the most variable genetic elements in tuberculous bacilli. However, this feature mostly holds true for *pe_pgrs* and *ppe_mptr* genes, with many other *pe/ppe* genes being relatively conserved (Gey van Pittius *et al.*, 2006).

Polymorphisms in *pe/ppe* genes usually arise from intramolecular recombination or DNA polymerase slippage, which in turn lead to SNPs, frameshifts or in-frame indels (McEvoy *et al.*, 2012). Several *pe/ppe* genes have been proposed to be hotspots for recombination events or the insertion of transposable elements (McEvoy *et al.*, 2009a; 2009b), although these appear to be more frequent in *ppe* genes than in *pe* genes (McEvoy *et al.*, 2012). Specific *pe/ppe* gene repertoires have been identified in different *M. tuberculosis* lineages (Phelan *et al.*, 2016). Despite their apparent variability, the rate of evolution of *pe/ppe* genes between closely related isolates appears to be much more limited than originally hypothesized (McEvoy *et al.*, 2012). Polymorphisms in *pe/ppe* genes have also been described in *M. bovis*, both by deletion and recombination events (Patane *et al.*, 2017; Lasserre *et al.*, 2018; Reis & Cunha, 2021), although the extent of the latter phenomenon in this variant remains controversial (Reis & Cunha, 2021).

Despite the extensive literature related to PE/PPE proteins, their true function remains unknown in the majority of cases (De Maio *et al.*, 2020). Due to their highly polymorphic nature, they are suspected to be involved in antigenic variation, which is usually supported by their frequent association with mycobacterial secretion systems and extracellular location. However, this theory is controversial and experimental studies have revealed that these proteins may be involved in distinct roles in MTBC biology, such as cellular

adhesion, host pathway activation, virulence or iron uptake, which in turn may vary under certain environmental conditions (Gey van Pittius *et al.*, 2006; De Maio *et al.*, 2020).

The direct repeat (DR) locus

The DR locus was accidentally discovered in the early 1990s during the characterisation of the IS*986* element in *M. bovis* BCG and corresponds to the clustered regularly interspersed short palindromic repeat (CRISPR) locus of MTBC members (Hermans *et al.*, 1991). Although CRISPR loci and their associated proteins (i.e. CRISPR-associated or Cas proteins) have been shown to mediate an "adaptive immunity" response against foreign DNA in prokaryotes (Barrangou *et al.*, 2007), the function of this locus in the MTBC is unknown.

This region of the genome is composed of a series of conserved 36 bp-long DRs, interspersed with non-repetitive spacer sequences between 34 and 41 bp-long (Kamerbeek *et al.*, 1997). The combination of a DR and its spacer sequence is termed direct variable repeat or DVR (Groenen *et al.*, 1993). Recently, genomic analysis of the DR locus in different MTBC members has unravelled its surrounding structure (Refregier *et al.*, 2020), revealing that it is usually preceded by a cluster of *cas* genes and is flanked by two external DRs (DR0), which in turn are bordered by specific sequences of 48 bp and 148 bp, respectively. A total of 69 spacers have been described in the MTBC, which increase to 94 for *M. canettii* (van Embden *et al.*, 2000; van der Zanden *et al.*, 2002). As previously mentioned, many MTBC strains harbour one IS*6110* element within the DR locus (Hermans *et al.*, 1991), but strains with 0 to 3 copies of the IS*6110* have also been described (Hermans *et al.*, 1991; Groenen *et al.*, 1993; Refregier *et al.*, 2020). It is unclear if the presence of IS*6110* elements in the DR region depends on a preferential integration or a decreased excision frequency in this locus.

Microevolution of the DR region has been amply described in the MTBC and can occur both by deletion and expansion (Warren *et al.*, 2002; Refregier *et al.*, 2020). Different variants of the MTBC present characteristic DVR patterns, which has resulted in the use of this locus as a molecular marker in epidemiological studies (see section 1.5.4.2). However, similar DVR patterns can arise as a result of different deletion events (convergent evolution), which can have important implications in the use of this region as a phylogenetic marker.

Expansion of the DR locus usually occurs through insertion of the IS*6110* element or duplication of DVRs (van Embden *et al.*, 2000; Warren *et al.*, 2002; Refregier *et al.*, 2020). Duplications have been proposed to occur in tandem and involve a single or adjacent DVRs, or they can appear in what has been termed as rearrangement duplications, by which duplicates of a specific DVR appear at a different position with respect to their original location (Refregier *et al.*, 2020). Deletions occur mainly through recombination between two adjacent DRs but can also affect a larger number of DVRs. In addition, the IS*6110* element is considered a major cause of large deleterious recombination events (Roychowdhury *et al.*, 2015), which also extends to the DR locus; when two IS*6110* elements are inserted in this region, homologous recombination may occur and result in the deletion of the DVRs located in between the two copies. Nevertheless, the extent to which this phenomenon affects *M. bovis* genomes could be limited, due to the reduced number of IS*6110* elements usually present in this variant.

ESAT-6 secretion systems

During the 1980s and 1990s, several studies focused on the analysis of *M. tuberculosis* culture supernatants in order to detect possible immunogenic proteins (Andersen *et al.*, 1991; Andersen *et al.*, 1992; Brodin *et al.*, 2004b). During these studies, a highly antigenic protein, ESAT-6, was identified in the lower molecular fraction of the culture filtrate proteins (CFPs) (Sorensen *et al.*, 1995).

The study of the genome of *M. tuberculosis* revealed that the ESAT-6 encoding gene is located in RD1 (Gordon *et al.*, 1999a), which explains why this antigen is not secreted by *M. bovis* BCG. Furthermore, the discovery of ESAT-6-like proteins, although with a wide variation in sequence identity, led to the definition of the ESAT-6 family of proteins, which additionally includes the CFP-10 and CFP-7 proteins (Brodin *et al.*, 2004b). In contrast to ESAT-6, these proteins contained a signalling peptide, which indicated the existence of a specific secretory mechanism in tuberculous bacilli. Several ESAT-6 family proteins can be found encoded across the genomes of MTBC members, and some of these are located within RDs, such as RD5 and RD8 (Gordon *et al.*, 1999a).

Further evaluation of the genome revealed the localisation of several ESAT-6 family encoding genes within clusters of conserved genes (*esx* gene clusters) (Bottai *et al.*, 2014). Further characterisation of these genes led to the description of the ESAT-6 secretory system (ESX), also called the type VII secretion system. This is a multiprotein complex composed of ESX-conserved

proteins, secretion-associated proteins and PE/PPE proteins. Several ESX paralogs have been identified, of which ESX-4 is considered to be the ancestral locus, since it is found in other SGM, such as *M. avium,* as well as in RGM, such as *M. smegmatis* (Gey Van Pittius *et al.*, 2001).

1.2.2. General description of MTBC members

1.2.2.1. *Mycobacterium tuberculosis*

Also known as 'Koch's bacillus' due to its original isolation by Robert Koch in 1882 and its establishment as the causal agent of human TB (Koch, 1882). It is responsible for human TB in the majority of cases, and its isolation in other animal species is uncommon. *Mycobacterium tuberculosis* is especially adapted to human hosts and has been found to be attenuated in different animal species, especially in cattle (Whelan *et al.*, 2010a). This phenotype was already noted in 1889 as a differential feature from the bovine tuberculous bacilli (Smith, 1898).

Most of the animal infections are considered to have an anthroponotic origin, since they mainly occur among animals with a close contact with humans, such as cattle and other production animals (Chen *et al.*, 2012; Kassa *et al.*, 2012; Ameni *et al.*, 2013; Cadmus *et al.*, 2016; Palaniyandi *et al.*, 2019), as well as companion animals like dogs and, very rarely, cats (Aranaz *et al.*, 1996b; Erwin *et al.*, 2004; Parsons *et al.*, 2012). Wildlife in captivity has also been shown to be infected with *M. tuberculosis* due to contacts with infected caretakers (Adeogun *et al.*, 2016); especially non-human primates and elephants (Montali *et al.*, 2001; Paudel & Sreevatsan, 2020).

Sustained animal infection with *M. tuberculosis* is rare in free ranging animals, in which most species are considered dead-end hosts (O'Reilly & Daborn, 1995). However, transmission has been identified from experimentally infected dogs to other animals, and small outbreaks of TB have been recorded in zoos or among elephant caretakers (Montali *et al.*, 2001; Stephens *et al.*, 2013).

1.2.2.2. *Mycobacterium africanum*

Mycobacterium africanum was first described in the year 1968 by Castets and colleagues (Castets *et al.*, 1968). Based on the phenotypic diversity of *M. africanum* isolates, this variant was historically classified into two lineages (David *et al.*, 1978); *M. africanum* type I (West African clade), now subdivided into West African 1 and 2 (genomic lineages 5 and 6, respectively), and *M. africanum* type II (East African clade), now revisited as *M. tuberculosis*

"Uganda" type through genomic analyses (Mostowy *et al.*, 2004; de Jong *et al.*, 2010; Gagneux, 2018).

Infection with *M. africanum* is almost exclusively restricted to the African continent, especially west Africa, as infections in other countries have been sporadic and mostly restricted to west African migrants (de Jong *et al.*, 2010). Despite its close proximity to certain animal-adapted variants, such as *M. mungi* or *M. suricattae* (Brites *et al.*, 2018), no animal reservoir has been identified for *M. africanum* so far (Yeboah-Manu *et al.*, 2017). In a similar manner to *M. tuberculosis*, isolation of *M. africanum* in animals has sporadically been documented. Infected cattle have been identified in Ghana (Asante-Poku *et al.*, 2014) or Nigeria (Cadmus *et al.*, 2006). Contaminated milk and dairy products have also been found in Nigeria (Cadmus *et al.*, 2010; Agada *et al.*, 2014). *Mycobacterium africanum* has also been isolated from other domestic species, such as pigs (Alfredsen & Saxegaard, 1992) and goats (van den Heever, 1984), and in wildlife in captivity, such as hyraxes or monkeys (Thorel, 1980; Gudan *et al.*, 2008). Importantly, genetic similarities between *M. orygis* and *M. africanum* may have led to the misclassification of *M. africanum* infections in different animal species (Rahim *et al.*, 2007; Dawson *et al.*, 2012; van Ingen *et al.*, 2012; Rahim *et al.*, 2017).

1.2.2.3. ***Mycobacterium bovis***

Shortly after the discovery of the human tuberculous bacillus, Theobald Smith proposed its differentiation from the bovine tuberculous bacillus, isolated from cattle, due to its differential pathogenesis in animals (Smith, 1898). Despite its early discovery, it was not until the 1970s that the bovine tuberculous bacillus was officially catalogued as a separate species, *Mycobacterium bovis* (Karlsson & Lessel, 1970). An important phenotypic difference of *M. bovis* with *M. tuberculosis* and other MTBC members is its natural resistance to one of the first-line drugs employed against TB, pyrazinamide (Olea-Popelka *et al.*, 2017).

Until the discovery of more recent MTBC variants, *M. bovis* was traditionally considered the unique aetiological agent of bTB and zTB. Although other variants have been included in this list, *M. bovis* is without doubt the most extended MTBC member in non-human mammals. Its high prevalence in cattle has had an enormous economic impact around the world, mainly due to reduction in herd yields, but also due to the tremendous effort and cost of control and eradication programmes.

Mycobacterium bovis was initially isolated from cattle, which is considered to be its natural reservoir. However, and unlike other MTBC members for which there seems to be a reduced number of maintenance hosts, *M. bovis* is a generalist pathogen that is able to maintain an infectious cycle in a diverse array of domestic and wild animals. These maintenance hosts vary according to the geographic region and include, among many others, Eurasian badgers (Corner *et al.*, 2011), wild boar (Naranjo *et al.*, 2008), red deer or brush tail possums (Nugent *et al.*, 2015a).

Other domestic and wildlife species have been shown to be infected with *M. bovis*, although they are usually considered spillover hosts (Mukundan *et al.*, 2015). Among domestic species, *M. bovis* has been isolated from goats (Crawshaw *et al.*, 2008), sheep (Munoz Mendoza *et al.*, 2012), pigs (Muwonge *et al.*, 2012; Bailey *et al.*, 2013) or South American camelids (Twomey *et al.*, 2007; García-Bocanegra *et al.*, 2010), among others. In the case of wildlife species, *M. bovis* has also been isolated from a wide array of animals, including wood bison (Miller & Sweeney, 2013), wildebeest or impala (Hlokwe *et al.*, 2014)], roe deer (Delahay *et al.*, 2007), feral pigs (Nugent *et al.*, 2015b), foxes (Michelet *et al.*, 2018a) or even large felines, such as lions (Viljoen *et al.*, 2015). Some of the abovementioned species may play a significant role as maintenance hosts in certain environments or conditions, such as feral pigs in New Zealand or domestic goats in Spain (Bezos *et al.*, 2012b; Zanardi *et al.*, 2013; Nugent *et al.*, 2015b).

1.2.2.4. ***Mycobacterium bovis* BCG**

The attenuated strain of *M. bovis* was fortuitously obtained by Léon Charles Albert Calmette and Jean-Marie Camille Guérin after a series of passages on potato slice medium supplemented with ox bile and glycerol (Sakula, 1983). It took 13 years of additional passages until enough attenuation was observed in different animal models for it to be used in humans for the first time in 1921 (Sakula, 1983). One hundred years after its first human use, this strain of *M. bovis* is still the only licensed TB vaccine in humans.

It is estimated that more than 4 billion doses of BCG vaccines have been administered worldwide since 1921 (Tran *et al.*, 2014), most of these being achieved with the use of only three strains: BCG-Denmark, BCG-Japan and BCG-Bulgaria (Ritz *et al.*, 2008). Despite its extensive use in humans, its use in domestic animals has been limited, since protection against infection is incomplete and vaccination may interfere with current diagnostic techniques (Buddle *et al.*, 2018). In contrast, vaccination with *M. bovis* BCG has been

proposed as an effective method to control the infection in wildlife populations (Tompkins *et al.*, 2009).

The extensive study of *M. bovis* BCG strains has allowed the identification of some of the molecular mechanisms responsible for their attenuation, which can aid in the development of new and better vaccines, as well as tests for differentiating infected from vaccinated animals (DIVA). One such modification is the deletion of RD1, which is shared among all of the known BCG strains and is considered one of its most important attenuation mechanisms (Lewis *et al.*, 2003).

1.2.2.5. ***Mycobacterium caprae***

This variant was first described in 1999 after its isolation from a goat with disseminated TB (Aranaz *et al.*, 1999). With the exception of a general susceptibility to pyrazinamide and a weak hydrolysis of Tween 80 after ten days, *M. caprae* is phenotypically indistinguishable from *M. bovis*. Therefore, it was not until the implementation of molecular genetic methods that differentiation of *M. caprae* from *M. bovis* could be achieved (Aranaz *et al.*, 2003).

Mycobacterium caprae is considered, along with *M. bovis*, one of the main causal agents of caprine TB. Goats are considered the main reservoir of *M. caprae* in Spain, although during the years 2004-2009, this variant was detected in 0.85-6.67% of bovine herds diagnosed with bTB (Rodriguez *et al.*, 2011), indicating that it can also be a relevant pathogen in cattle, especially in those regions where interspecies interactions are possible. However, the fact that the majority of *M. caprae*-infected cattle were raised in farms without any contacts with goats suggests that other means of infection are possible (Rodriguez *et al.*, 2011). Certainly, wild boar has also been found to be infected with *M. caprae* in Spain (Rodriguez *et al.*, 2011), and genetic similarities between *M. caprae* isolates from goats and wild boar indicate that the infection can be maintained in wild ecosystems, similarly to *M. bovis*.

The frequent isolation of *M. caprae* in goats appears to be limited to Spain, since *M. caprae* infections are usually described in cattle and red deer in other regions of continental Europe (Prodinger *et al.*, 2005; Amato *et al.*, 2019; Ciaravino *et al.*, 2021; European Food Safety *et al.*, 2021), including regions surrounding the Alps, such as Switzerland, south Germany, northern Italy and Austria (Prodinger *et al.*, 2002; Domogalla *et al.*, 2013; Ghielmetti *et al.*, 2017; Magnani *et al.*, 2020); and countries in eastern Europe, such as Bulgaria (Valcheva *et al.*, 2020; Valcheva *et al.*, 2022) and Poland (Krajewska-Wedzina

et al., 2018). The prevalence of *M. caprae* in goats in Spain could be a result of host-specific factors or due to the increased farming of this species in this country. However, this could also be a result of an increased surveillance in this species, since goats have been included in the Spanish eradication programme under certain situations, such as when a cattle herd has epidemiological links or shares premises with goats (MAPA, 2022). Certainly, the isolation of *M. caprae* from a goat farmer in Greece, one of the most important goat farming countries in the EU, could indicate that this variant may also be present in goats in this country (Papaventsis *et al.*, 2021).

Cattle infections with *M. caprae* have not been described in Ireland and the United Kingdom. Outside Europe, infection of cattle with *M. caprae* has been rarely described and has been restricted to countries in the Mediterranean basin, such as Algeria and Tunisia (Sahraoui *et al.*, 2009; Lamine-Khemiri *et al.*, 2014). However, isolation of *M. caprae* in reindeer and sheep from northern China suggests that this variant may be found in other continents (Zeng *et al.*, 2013).

In addition to caprines and bovines, *M. caprae* has been isolated from other domestic species, including sheep (Munoz Mendoza *et al.*, 2012; Vidal *et al.*, 2018), free-ranging pigs (Amato *et al.*, 2017) or rabbits (Sevilla *et al.*, 2020). Infection due to *M. caprae* has also been recorded in several animal species held in captivity, such as camels (Erler *et al.*, 2004; Pate *et al.*, 2006; Infantes-Lorenzo *et al.*, 2020), European bison (Pate *et al.*, 2006; Didkowska *et al.*, 2021), Bornean elephants and a Siberian tiger (Lantos *et al.*, 2003). Finally, in addition to wild boar, *M. caprae* has been isolated from several wildlife species, such as roe deer (Orlowska *et al.*, 2020), foxes (Steinparzer *et al.*, 2020) or grey wolves (Orlowska *et al.*, 2020).

1.2.2.6. ***Mycobacterium microti***

Mycobacterium microti was first identified in 1937 and was considered the causative agent of TB in small rodents, or vole TB (Wells, 1937). Until the implementation of genotypic methods, *M. microti* could not be appropriately differentiated from *M. africanum* and *M. bovis* due to its varying phenotype (Levy-Frebault & Portaels, 1992). Nowadays, genetic analyses have related this variant with *M. pinnipedii* (Brites *et al.*, 2018).

The reservoir host of *M. microti* is considered to be the field vole, with a prevalence that ranges between 7 to 31% that appears to vary according to the season, place and method of capture (Wells & Robb-Smith, 1946;

Cavanagh *et al.*, 2002; Burthe *et al.*, 2008). Although *M. microti* has been isolated all year round, prevalence seems to follow a temporal trend, being highest during spring and early summer (Wells & Robb-Smith, 1946; Burthe *et al.*, 2008). An interesting feature of *M. microti* infections in voles is the presence of skin lesions, suggestive of transmission through direct contact (Kipar *et al.*, 2014), and an extremely slow growth rate, requiring a minimum of 8 weeks for its primary isolation in most cases (van Soolingen *et al.*, 1998; Boniotti *et al.*, 2014). This could lead to an underestimation of its true prevalence and, therefore, molecular genetic methods have become particularly useful to detect this pathogen (Pérez de Val *et al.*, 2019).

Detection of *M. microti* in humans is rare, but has been shown to affect both immunocompetent and immunocompromised patients, leading to severe forms of disease and death in some cases (van Soolingen *et al.*, 1998; Niemann *et al.*, 2000; Emmanuel *et al.*, 2007; Panteix *et al.*, 2010; Crawshaw *et al.*, 2014). Apart from humans and small rodents, *M. microti* has been isolated from a variety of hosts, albeit sporadically. Infections have been reported in farm animals, like pigs, goats or cattle(Jahans *et al.*, 2004; Taylor *et al.*, 2006; Michelet *et al.*, 2016; Michelet *et al.*, 2017b), as well as in companion animals, like cats and dogs (Deforges *et al.*, 2004; Rüfenacht *et al.*, 2011; Michelet *et al.*, 2015b). Importantly, *M. microti* is the main causal agent of feline TB, which is considered an emerging disease in cats (Peterhans *et al.*, 2020). Infections have also been detected in free living wildlife, such as wild boars, badgers or deer (Emmanuel *et al.*, 2007; Boniotti *et al.*, 2014; Michelet *et al.*, 2015a; Chiari *et al.*, 2016; de Val *et al.*, 2019; Ghielmetti *et al.*, 2021b), as well as in captive or domestic animals like meerkats or new word camelids, respectively (Oevermann *et al.*, 2004; Zanolari *et al.*, 2009; Palgrave *et al.*, 2012). The increasing detection of *M. microti* in European wildlife suggests that its presence could be more extended than anticipated (Pérez *et al.*, 2002; Tagliapietra *et al.*, 2021). Furthermore, the detection of *M. microti* in five skin test-positive cattle in France highlights the risk for possible diagnostic interference in bTB eradication campaigns and should be assessed in the future, especially in officially tuberculosis free (OTF) regions (Michelet *et al.*, 2015a; Michelet *et al.*, 2020).

1.2.2.7. ***Mycobacterium pinnipedii***

Between 1986 and 1995, a series of MTBC isolates from captive and wild pinnipeds prompted the description of the 'seal bacillus', formally defined as *M. pinnipedii* at the beginning of the twenty-first century (Cousins *et al.*, 2003).

In a similar manner to *M. caprae*, the main phenotypical difference of *M. pinnipedii* from *M. bovis* is its susceptibility to pyrazinamide. Several isolates have also been found to be weakly resistant to thiophen-2-carboxylic acid hydrazide, and give weak or positive reactions to the niacin accumulation test.

The host range of *M. pinnipedii* seems to be more restricted to that observed for *M. bovis, M. caprae* or *M. microti*. This variant has been isolated from pinnipeds around the world, especially in captivity, but also from free-living specimens in the southern hemisphere (Cousins *et al.*, 2003; Silva-Pereira *et al.*, 2019). However, the isolation of *M. pinnipedii* in Hector's and bottlenose dolphins indicates that other marine mammals are also susceptible to infection (Roe *et al.*, 2019). Interestingly, disease pathology in pinnipeds is highly similar to that observed in other mammalian species, and the predominant affection of the respiratory tract strongly suggests the inhalatory route as the main transmission mechanism in these animals (Moser *et al.*, 2008; Kriz *et al.*, 2011).

Terrestrial mammals, such as tapirs, camels or cattle have also been shown to be infected with *M. pinnipedii*, although an epidemiological link to seals or sea lions is either demonstrated or suspected in all cases (Moser *et al.*, 2008; Loeffler *et al.*, 2014). The lack of multiple TB cases in the same cattle herd (Loeffler *et al.*, 2014), as well as the occurrence of infections in animals in captivity suggest that *M. pinnipedii* does not transmit efficiently between terrestrial hosts.

Mycobacterium pinnipedii has rarely been isolated in humans, and infections have mainly been reported in personnel from zoos and marine parks (Thompson *et al.*, 1993; Kiers *et al.*, 2008; Macedo *et al.*, 2020). Recently, an infection with *M. pinnipedii* has been reported in a 79-year old female patient with no apparent epidemiological link to pinnipeds or zoos, suggesting that other yet unknown infectious origins might be possible (Zmak *et al.*, 2019). Interestingly, *M. pinnipedii*-related genomes have been recovered from three 1000-year old skeletons from Peru, supporting that the introduction of TB into the New World probably occurred before European contact (Bos *et al.*, 2014). Initially, pinniped-to-human transmission was suspected to have been the cause of the spread of these strains in the coastal populations. However, recent data have revealed the presence of ancient *M. pinnipedii* genomes in inland

regions of Peru and Colombia, which could suggest its dissemination beyond the coast through human-to-human transmission or the presence of multiple animal-to-human transmission events (Vågene *et al.*, 2022).

1.2.2.8. ***Mycobacterium orygis***

Formerly named the 'Oryx bacillus' due to its first isolation from two East African oryxes in 1976 (Lomme *et al.*, 1976), *M. orygis* was formally described as a new member the MTBC only recently (van Ingen *et al.*, 2012). Since its original description, *M. orygis* has been isolated from several animal species, including large artiodactyls [oryxes (van Soolingen *et al.*, 1994), gazelles (van Soolingen *et al.*, 1994; Smith *et al.*, 2006b), deer (Smith *et al.*, 2006b; Thapa *et al.*, 2015), antelopes (Smith *et al.*, 2006b; Thapa *et al.*, 2015), waterbucks (Smith *et al.*, 2006b), camels (Kinne *et al.*, 2006; Wernery *et al.*, 2007), African buffalos (Gey van Pittius *et al.*, 2012) and cattle (Dawson *et al.*, 2012; Rahim *et al.*, 2017; Refaya *et al.*, 2019)], greater one-horned rhinoceros (Thapa *et al.*, 2016) and rhesus monkeys (Rahim *et al.*, 2017).

Most human infections have been described in, or have epidemiological links with, South-east Asia (Dawson *et al.*, 2012; van Ingen *et al.*, 2012; Marcos *et al.*, 2017), where *M. orygis* is considered to be endemic. Furthermore, the increased identification of *M. orygis* in the Indian subcontinent suggests that this variant may be the leading cause of zoonotic TB instead of *M. bovis* in this region (Duffy *et al.*, 2020).

1.2.2.9. ***Mycobacterium mungi***

Mycobacterium mungi was initially recovered from an outbreak of TB that affected banded mongooses at the northern border of Chobe National Park, Botswana, in 1999 (Alexander *et al.*, 2002), and was identified as *M. tuberculosis* due to its phenotypical and genetic similarity with this species. The reoccurrence and spread of the epizootics between 2000 – 2010 led to the description of *M. mungi* as a novel MTBC species (Alexander *et al.*, 2010). Since it's description, *M. mungi* has only been described in this region of Africa, where outbreaks are continuously being reported and monitored, with as many as 93% (14/15) of the tracked mongoose troops in the region being infected as of 2018 (Alexander *et al.*, 2018). Prevalence within these troops ranges from 3 to 17% of specimens, with high mortality among affected individuals in a relatively short period of time (2 – 3 months since onset of symptoms) when compared to other members of the MTBC (Alexander *et al.*, 2010).

The clinical picture of *M. mungi* infection seems to differ from other MTBC infections (Alexander *et al.*, 2010; 2016; 2018). Although the affection of the respiratory tract is usually present, this mainly occurs in severe disseminated cases. An additional feature of TB by *M. mungi* is the presence of lesions in the nasal cavity and skin, and a high number of lesions in liver, spleen and anal glands/sacs (Alexander *et al.*, 2018).

Attempts to culture the microorganisms have been limited to date (Alexander *et al.*, 2010; Alexander *et al.*, 2016), but molecular approaches through real-time PCR (qPCR) have putatively identified high bacterial DNA concentrations in the anal gland and its secretions, the nasal planum, as well as in the bladder, indicating that transmission of this variant may take advantage of the social behaviour of mongooses, which involves scent inspection, autogrooming or allogrooming in scent marked areas or scent marking with anal gland secretions in injuries (Alexander *et al.*, 2016).

1.2.2.10. ***Mycobacterium suricattae***

Being first isolated in 2013 from TB-symptomatic meerkats in South Africa, *M. suricattae* is the latest addition to the MTBC (Parsons *et al.*, 2013). Preliminary genetic studies indicated that this variant is closely related to the dassie bacillus (see below), although relatedness with the chimpanzee bacillus (see below) and *M. mungi* has also been indicated (Dippenaar *et al.*, 2015; Brites *et al.*, 2018). Similar to *M. mungi* in mongooses, *M. suricattae* is highly pathogenic for meerkats.

There is no current evidence for the prevalence of *M. suricattae* in wild meerkats, but previous lack of awareness may have led to its erroneous classification as *M. tuberculosis* or *M. bovis* infections (Parsons *et al.*, 2013). Furthermore, infections in humans or in other animal species have been reported to date.

1.2.2.11. ***Mycobacterium canettii***

This variant of tuberculous mycobacteria was first described by Georges Canetti in the 1960s (Supply & Brosch, 2017). Since its discovery, less than 100 strains of *M. canettii* have been isolated, of which the majority were obtained from patients that had contact with the Horn of Africa.

Mycobacterium canettii is frequently considered a member of the MTBC, although this inclusion is not without controversy due to its significant phenotypic and genetic differences (Supply & Brosch, 2017). Firstly, whereas

the classical colony morphology of MTBC variants is rugose, *M. canettii* presents a smooth phenotype, which has granted its nomenclature as the smooth tubercle bacillus (Boritsch *et al.*, 2016a). Secondly, virulence of *M. canettii* is significantly lower when compared to other MTBC members (Blouin *et al.*, 2014; Supply & Brosch, 2017). In addition, while the genomic diversity of MTBC members has been shown to be very limited, *M. canettii* strains present an extremely high genetic diversity, which is partly related to their larger and recombinative genomes, and LGT capacity (Boritsch *et al.*, 2016a; Boritsch *et al.*, 2016b). In fact, a WGS study in a selection of *M. canettii* strains revealed a higher genomic diversity than the one observed for all of MTBC members around the world, which is interesting considering its limited geographical extent (Supply *et al.*, 2013).

Outbreaks of *M. canettii* have been described in literature (Blouin *et al.*, 2014). However, no animal nor environmental reservoir has been described for this variant to date.

1.2.2.12. The dassie bacillus

The dassie bacillus was first isolated in 1954 from an epithelioid granuloma obtained from the lung of a Cape hyrax, commonly referred to as the dassie in South Africa (Wagner *et al.*, 1958). Morphological and growth requirements suggested that this pathogen was related to *M. microti* (Smith, 1960), and initial experimental infections revealed that the dassie bacillus was attenuated in guinea pigs, mice, voles and rabbits (Wagner *et al.*, 1958; Smith, 1965; Cousins *et al.*, 1994). The current prevalence of infection with the dassie bacillus is unknown, since infections in free-living animals have only been reported at the time of its initial discovery and in 2006 (Parsons *et al.*, 2008). No humans or other animal species have been found to be infected by the dassie bacillus; although an infection in a meerkat has been reported, this may have been caused by *M. suricattae* (Mostowy *et al.*, 2004).

1.2.2.13. The chimpanzee bacillus

The chimpanzee bacillus was isolated from an elderly chimpanzee female in Côte d'Ivoire (Coscolla *et al.*, 2013). The animal had died as a consequence of a leopard attack and granulomatous lesions were observed in the liver, spleen and mesenteric LNs. Although no other lesions were found in the body, the inhalatory route of infection could not be discarded because most of the lungs had been consumed by the leopard. Isolation of typical MTBC colonies that were further characterised using molecular methods, including WGS, identified

it as a strain closely related to *M. africanum, M. mungi* and the dassie bacillus. No other descriptions of the chimpanzee bacillus have been registered to date.

1.2.3. **Evolution of the MTBC**

The definition of the MTBC as a highly clonal group of bacteria was primarily based on the high congruence of phylogenetic trees obtained from different chromosomal markers, mainly SNPs and RDs (Smith *et al.*, 2006a). With the exception of highly polymorphic loci, the almost complete absence of recombination in MTBC genomes entails that polymorphisms cannot be restored to their previous form and are therefore accumulated in a directional manner (Brosch *et al.*, 2002; Smith *et al.*, 2006a), which has two important implications: 1) phylogenies of MTBC members are highly congruent, which means that a group of strains sharing a set of phylogenetically informative markers most probably are related to each other; and 2) a strong chromosomal linkage disequilibrium indicates that the MRCA probably contained these markers.

As a result of these discoveries, an evolutionary model was proposed in which the MTBC was defined as a series of clades of host-adapted ecotypes that share certain phylogenetically informative polymorphisms, including RDs, SNPs and spacer deletions (figure 2), which evolved from hypothetical ancestral strains (Brosch *et al.*, 2002; Smith *et al.*, 2006b). For example, the RD4 deletion and the SNPs in the *pncA* and *oxyR* genes in *M. bovis* suggest its descent from an RD4-containing ancestor, which in turn evolved from an ancestral strain shared with *M. caprae*.

Clonal microorganisms are highly susceptible to reductions in diversity caused by population bottlenecks and selective sweeps (Smith *et al.*, 2006a). Therefore, it is believed that the sets of phylogenetic markers of the different lineage ancestors appeared in populations that overcame severe reduction in size, possibly during their adaptation to new host species. Then, through clonal expansion, these clones outlived the diversity present in each of the preancestral populations and expanded as a result of selection or population sampling as they invaded new geographical regions (Smith *et al.*, 2006a). In the context of bTB, several processes have been hypothesised to be responsible for the current diversity of *M. bovis* genomes, such as selection and effective population reduction driven by test and slaughter regimes, geographical subdivisions or the availability of different hosts, such as wildlife species (Smith *et al.*, 2006a).

The high resolution and capacity of WGS methods have allowed the scientific community to expand the phylogenetic relationships identified through traditional genetic methods, increasing our knowledge on the evolution of the MTBC as well as the population structure of its different members.

One of the most striking findings of WGS analyses of tuberculous bacilli was the discovery of a strong recombinative signal in the genomes of *M. canettii* (Supply *et al.*, 2013), which contrasts with the highly clonal structure of the rest of the tubercle bacilli. It is generally considered that the MTBC evolved from a pool of *M. canetti*-like mycobacteria which lost its recombinative capacity during the evolution to become a highly successful pathogen (Supply *et al.*, 2013; Supply & Brosch, 2017). However, the drivers and determinants of this evolutionary process remain unknown.

An additional milestone in the understanding of MTBC evolution was achieved with the availability of the full genomic sequences of *M. tuberculosis* and *M. bovis* (Gagneux, 2018). Before the availability of these sequences, it was hypothesised that human TB originated from animals during the Neolithic period, based on the broad host range of *M. bovis*. However, the finding that the genome of *M. bovis* is smaller than *M. tuberculosis*, as well as the detection of RDs which are lost in a specific fashion between the different members of the MTBC could support an inverse scenario in which humans transmitted TB to animals.

Whole genome analyses of MTBC genomes divided the human-adapted strains into seven distinct genetic lineages and the animal-adapted variants into one single lineage (Comas *et al.*, 2013). The analysis of a larger number of genomic sequences from different animal-adapted variants improved the resolution of this lineage, revealing that this group of mycobacteria could be further divided into four different clades (A_1 – A_4) (Brites *et al.*, 2018). While clades A_1, A_2 and A_4 include MTBC ecotypes that affect multiple animal species, such as *M. bovis*, clade A_3 includes only *M. orygis* (figure 3). Interestingly, A_1 includes the most divergent variants from *M. bovis*, such as *M. mungi, M. suricattae,* the chimpanzee bacillus and the dassie bacillus. In addition, these genomes are more closely related to the human-adapted variants, being in turn monophyletic with *M. africanum* L6 and paraphyletic with the rest of the animal-adapted variants.

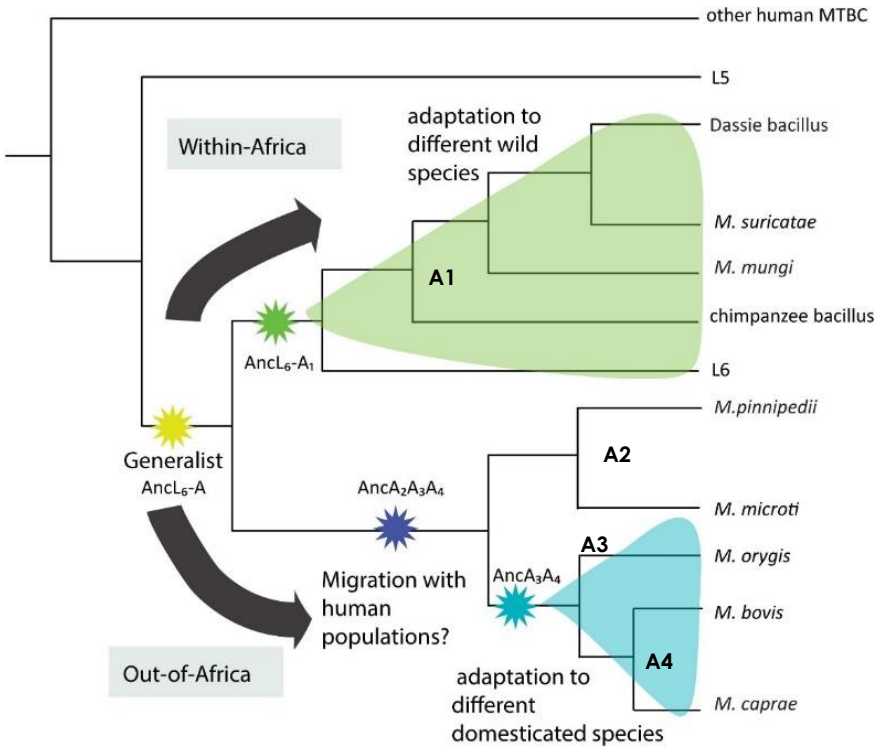


Figure 3. Evolutionary model for the animal-adapted lineage of the MTBC. Anc = ancestral strains of the animal-adapted lineages. A = animal-adapted lineage. Modified from: Brites *et al.*, 2018. with permission from Frontiers in Microbiology.

The fact that clade A_1 is monophyletic with *M. africanum* L6, which causes up to 50% of TB cases in Africa (Kim *et al.*, 2002), has been used to propose that this variant may have originated from an unknown animal reservoir. However, no such reservoir has been identified so far and, therefore, there is a possibility that *M. africanum* may indeed be human-adapted. Brites *et al*. proposed an evolutionary model in which the ancestor of L6 and the animal-adapted strains was probably a generalist pathogen that, through a differentiation event, led to the appearance of two ancestral populations: Ancestral (Anc) L6-A_1 and Anc$A_2A_3A_4$ (figure 3). The former population could have adapted to humans in the case of L6 and to different wild species in the case of A_1. The fact that the extant maintenance hosts of the members of clades A_3 and A_4 are domesticated livestock suggests that the AncA_3A_4 may have accompanied human populations during their migration out of Africa (Brites *et al.*, 2018). Considering the host range and distribution of *M. orygis*, Brites *et al*. hypothesised that clades A_3 and A_4 may reflect the two independent ancestral auroch domestication events that are known to have occurred in history in the

Fertile Crescent (*Bos taurus*) and the Indus Valley (*Bos indicus*) (Brites *et al.*, 2018), leading to the distribution of *M. bovis* in western countries and *M. orygis* in Southeast Asia (Duffy *et al.*, 2020).

1.3. Epidemiological aspects of bTB

1.3.1. Distribution of bTB

1.3.1.1. Global distribution

Bovine TB is a widely distributed disease (figure 4), which was present in 82 out of the 188 countries that reported their epidemiological situation between the years 2017 and 2018 (Murai *et al.*, 2019). This widespread distribution is thought to have occurred as a result of the success of *M. bovis* in infecting cattle, coupled with the transport of these animals around the world during human migrations.

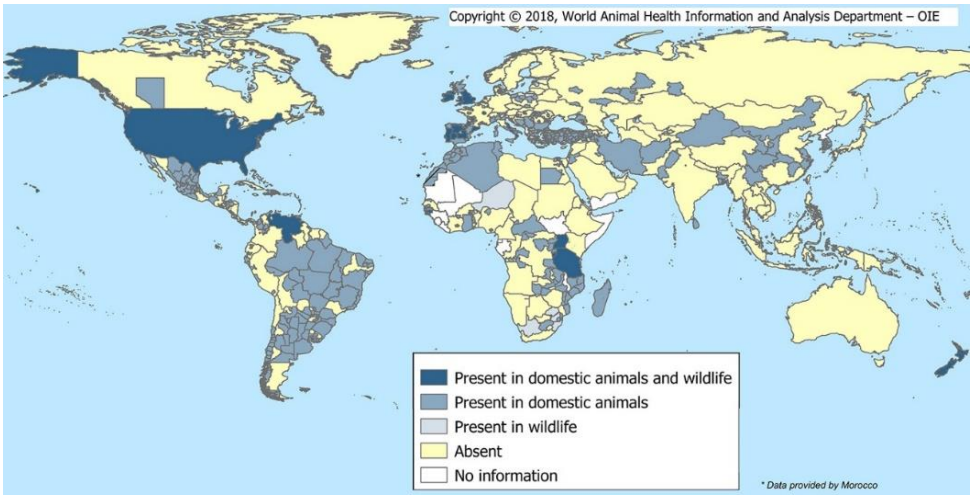


Figure 4. Global distribution of bTB in the year 2018. Figure obtained from the OIE Bulletin 'Panorama 2019 – 1'.

1.3.1.2. Situation of bTB in the EU

There has been a great improvement in the status of bTB in the EU since the implementation of the first eradication programmes in the 20th century, with 17 OTF countries and an overall herd prevalence of 0.4% in 2020 (ECDC, 2021). Despite the overall low prevalence, the situation of bTB in Europe is heterogeneously distributed across its regions, with prevalence ranging from 0-4.7%, being highest in the RoI (ECDC, 2021). Among non-Member States and pre-accession countries, herd prevalence in 2020 ranged between 0-10.7%,

with the maximum prevalence being recorded in the United Kingdom (ECDC, 2021). These values reflect the enormous variety of breeding systems and environmental conditions in Europe, that give rise to different epidemiological conditions that could affect the prevalence of this infectious disease (Reviriego Gordejo & Vermeersch, 2006).

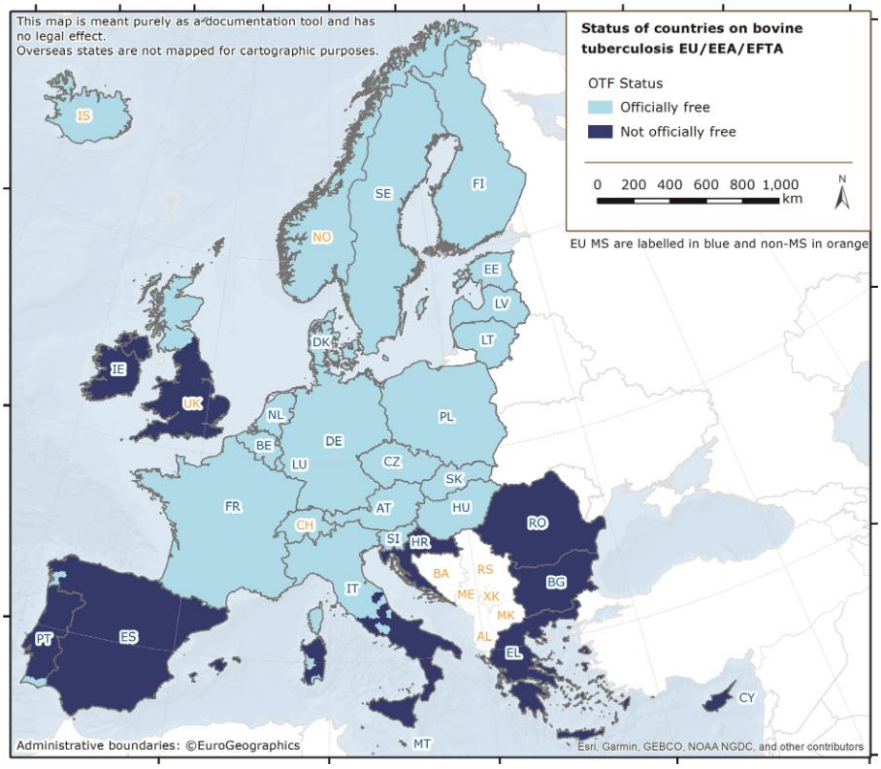


Figure 5. Bovine TB status of EU Member States and non-Member States during the year 2020. Image obtained from the EFSA Journal 2021.

Non-OTF countries are mainly restricted to the periphery of the EU and include Bulgaria, Croatia, Cyprus, Greece, Ireland, Italy, Malta, Portugal, Romania and Spain (figure 5) (ECDC, 2021). Three of these Member States declared OTF regions in 2020: Portugal, Spain and Italy.

1.3.1.3. **Situation of bTB in Spain**

Since 2016, herd prevalence in Spain has decreased considerably from 2.87% to 1.61% in 2020 (MAPA, 2022). Despite this decreasing trend, the prevalence of bTB is heterogeneously distributed across Spain; most of the infections occur in the southwestern regions of the country (figure 6), and the highest herd prevalence was described in Castile-La Mancha (4.37%), Extremadura (3,24%) and Andalusia (2.70%). Despite this situation, the majority (96.93%) of cattle herds in Spain obtained the bTB-free status (T3) in 2020, and three autonomous regions have become OTF (Galicia, Asturias and the Canary Islands). Furthermore, one additional Autonomous Region and 10 provinces have initiated the process to obtain the OTF status.

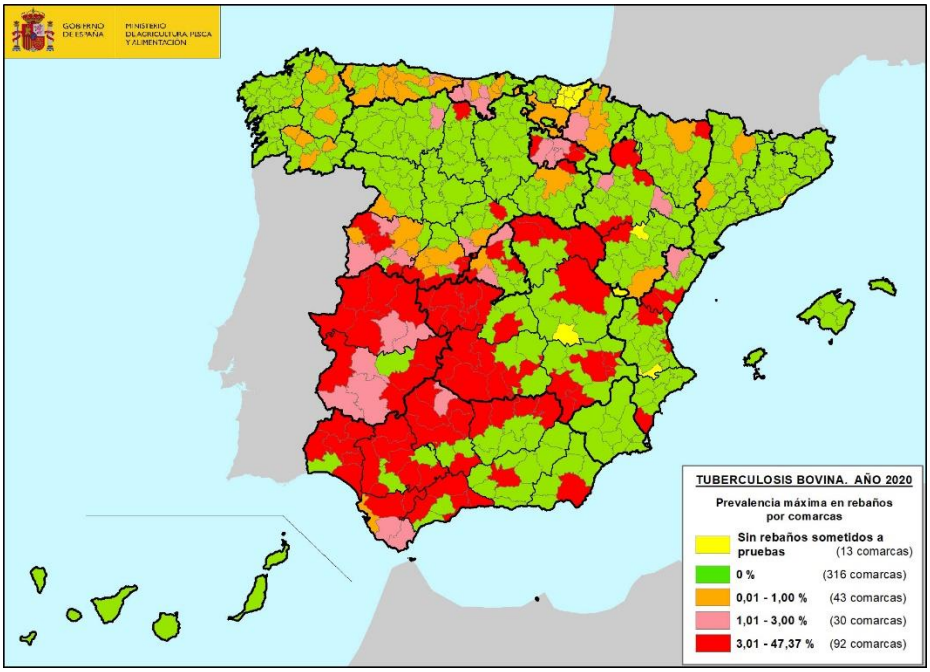


Figure 6. Distribution of bTB herd prevalence in Spain during the year 2020. Image obtained from the Spanish eradication programme for bTB (2022).

1.3.2. Transmission routes

Inhalation is considered the main transmission route of bTB, as tuberculous lesions are frequently observed in the respiratory tract and LNs from the head and thorax of infected animals (Corner, 1994; Cassidy, 2006) (see section 1.6.2.10). Intensive husbandry settings or other scenarios in which a large number of animals are in close contact with each other are considered a risk factor for airborne transmission of bTB (Ameni *et al.*, 2007; Menin *et al.*, 2013). However, the detection of lesions in other LNs, especially the mesenteric and hepatic LNs, suggest that the ingestion of contaminated food and water, or direct contact with contaminated fomites, may also be an important route of infection, especially in less intensive settings, such as communal pastures (Ameni *et al.*, 2007; Liebana *et al.*, 2008; Ameni *et al.*, 2011). In addition, the infection can spread to the supramammary LNs, leading to the contamination of milk and therefore posing a high risk of infection for calves and humans that ingest it (Bruning-Fann *et al.*, 2001).

1.3.3. Reservoirs of bTB and their significance

As previously mentioned, the major maintenance host of *M. bovis* has traditionally been considered to be cattle. However, other domestic and wildlife species can also act as such in certain epidemiological contexts. This is reflected in the fact that 35.4% of countries that reported bTB infections have also reported TB in wildlife (Murai *et al.*, 2019).

Continuous contact between infected cattle and other animals poses a significant risk for spillover events and/or the establishment of a new infectious cycle. This not only can have a negative ecological effect if wild endangered species are affected, but can seriously hamper eradication efforts due to the appearance of new sources of infection that are difficult to control (Arnot & Michel, 2020). Certainly, wildlife reservoirs play a role in the maintenance of bTB in different countries, such as the Eurasian badgers in the United Kingdom (UK) and the Republic of Ireland (RoI) (Corner *et al.*, 2011), white-tailed deer in North America (Miller & Sweeney, 2013; Salvador *et al.*, 2019), African buffaloes and greater Kudus in South Africa (Arnot & Michel, 2020), or brush tail possums in New Zealand (Nugent *et al.*, 2015a).

In Spain, wild boars and red deer are considered two of the most important maintenance hosts of bTB, especially in southcentral and southwestern regions (Naranjo *et al.*, 2008). In these areas, a series of factors concur that may facilitate the transmission of TB between wildlife and cattle; 1) the density of

wild ungulates is high and is distributed both in natural reserves and hunting estates (Gortazar *et al.*, 2008; Delibes-Mateos *et al.*, 2009); 2) TB is endemic in wild ungulates in these regions (Vicente *et al.*, 2013); and 3) there is a high proportion of extensive farms in which cattle can easily come into contact with infected wildlife.

Once an infectious cycle becomes established in wildlife, the control of TB requires the implementation of additional control measures in the animal species involved. These measures may include the active surveillance of wildlife populations using different diagnostic methods (Thomas *et al.*, 2021), the control of TB reservoir populations through hunting (Donnelly *et al.*, 2007; Garcia-Jimenez *et al.*, 2013), the implementation of risk mitigation protocols in cattle farms (Martinez-Guijosa *et al.*, 2021), or vaccination of wildlife species (Tompkins *et al.*, 2009). In support of the relevance of wildlife in the maintenance of bTB, the Spanish government recently established a legal framework for the surveillance and control of aTB in wildlife (Royal Decree 138/2020).

1.4. Diagnosis of bTB

There is a great variety of diagnostic techniques available for bTB, but only certain methods are officially recommended for routine and/or ancillary use in control and eradication campaigns, either by national governments or international organisations, such as the OIE and the EU. More details on the surveillance and eradication of bTB will be presented in section 1.6.

Diagnostic methods for bTB can be divided into two major groups according to the detection of the host's response towards the pathogen (indirect methods) or the pathogen itself (direct methods) (table 1). These methods, in turn, can be classified as *ante-mortem* or *post-mortem* whether they are performed before or after the death of the animal, respectively. Indirect methods include immunological tests, such as the tuberculin skin test, the interferon gamma release assay (IGRA) or serological tests, and standard histopathological analyses that identify the host's response against the pathogen (e.g. tuberculous lesions). In contrast, direct methods include microbiological culture, molecular methods, such as PCR, or the detection of bacilli through acid-fast staining or immunohistochemistry.

Type	Application	Method
Indirect	*Ante-mortem*	Tuberculin skin test
		IGRA
		Serology
	Post-mortem	Gross pathology & histopathology (host response)
Direct	*Post-mortem*	Culture isolation
		Molecular methods (PCR)
		Histopathology (pathogen detection)

Table 1. Diagnostic methods used for the diagnosis of bTB.

The ability of a diagnostic test to adequately detect the infection in an animal or a herd (diagnostic performance) depends on several factors, such as the epidemiological situation of the region under study, the stage of infection and bacterial load in the animal, as well as its immune status. An important factor in bTB diagnostic tests is the lack of a "golden standard" that can be used in order to compare and establish other techniques. This role has usually been attributed to microbiological culture, despite its technical limitations (see below). As a result, there is a great variability in scientific literature regarding the performance of bTB diagnostic methods and, therefore, extrapolations to different epidemiological scenarios should be carried out with caution. The increased application of Bayesian statistics to calculate the performance of bTB diagnostic methods without the need of a reference standard has become a promising alternative (Courcoul *et al.*, 2014). However, the use of Bayesian models still requires the selection of appropriate prior values for Se and Sp (Alvarez *et al.*, 2012), which make any test comparison data highly valuable.

1.4.1. *Ante-mortem* diagnosis of bTB

1.4.1.1. Clinical diagnosis of bTB

The usual clinical signs of TB in animals include weakness, anorexia and emaciation, fluctuating fever, dyspnoea and cough, low grade pneumonia, diarrhoea, lymphadenitis and, ultimately, death (Cousins, 2001; OIE, 2018). Although these symptoms may be suggestive of TB, they are not pathognomonic and usually appear in advanced stages of the disease (figure 7). Importantly, infected animals are able to transmit the infection before they exhibit obvious clinical signs and therefore pose a significant risk for other animals and humans. Due to the continuous surveillance of bTB, clinical signs in bovines are not frequently observed, although they may be detected in other animal species in which surveillance is not established, such as wildlife animals.

1.4.1.2. Intradermal tuberculin tests

The intradermal tuberculin (IT) test, or skin test, is one of the oldest bTB diagnostic techniques, yet it has resulted to be the most useful and significant method in its control and eradication (Good *et al.*, 2018). Its origins are tied to the search for a treatment or prophylaxis for TB in 1890, when Robert Koch created the first tuberculin (Koch's old tuberculin or KOT). Despite its failure as a treatment for TB, the immune reaction that was induced by KOT led to its use as a diagnostic method in the first eradication programme that was established in Denmark in 1890 by Bernard Bang. The initial success of this procedure led to its extended use in other countries and even to its use in humans by Charles Mantoux in 1908.

Principle of the technique

The basis for this test relies on the sensitisation of TB-infected animals towards tuberculins or Protein Purified Derivatives (PPDs). When these are inoculated in the skin of the animal, a delayed type IV hypersensitivity reaction leads to a local inflammatory response that can be observed and measured (Monaghan *et al.*, 1994; de la Rua-Domenech *et al.*, 2006). Animals older than six weeks are tested, with a minimum interval of at least 42 days between tests in order to avoid desensitisations due to repeated Inoculations (Cousins & Roberts, 2001; de la Rua-Domenech *et al.*, 2006). An animal is considered a reactor according to a matrix of skin-fold thickness measurements and clinical signs in the site of injection, including diffuse or extensive oedema, exudation, necrosis, pain or inflammation.

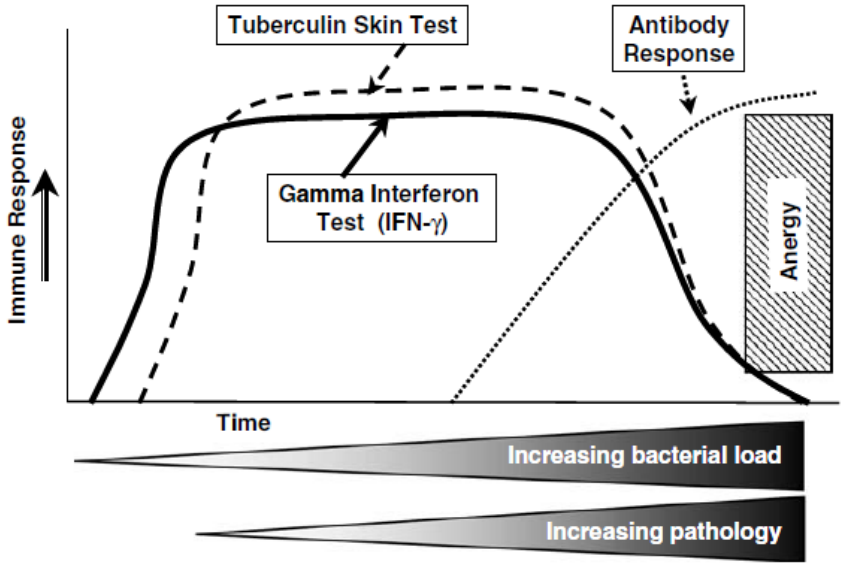


Figure 7. Evolution of immune responses towards bTB infection measured by *ante-mortem* indirect methods. Source: de la Rua-Domenech et al., 2006, with permission from Research in Veterinary Science.

Protein Purified Derivatives are obtained from an inactivated culture of mycobacteria, which is then purified using filtration and precipitation methods to obtain a high protein content extract. Although KOT was obtained from pure cultures of *M. tuberculosis,* nowadays PPDs used in bovines are derived from *M. bovis* AN5 (bovine PPD or bPPD) or *M. avium* D4ER (avian PPDs or aPPDs). Avian PPDs are used to detect cross-reactions in IT tests, since some antigens in bPPDs are shared with other mycobacteria (Infantes-Lorenzo *et al.*, 2017). As a result, skin tests can be classified into single and comparative intradermal tuberculin tests (SIT and CIT tests, respectively) whether bPPD is used alone or in combination with aPPDs, respectively (Bezos *et al.*, 2014). In the case of CIT tests, a reaction towards the aPPD suggests infection or co-infection with *M. avium* or related species, and positivity is established based on the difference between the skin-fold thickness and/or the presence of clinical signs observed for both PPDs.

In Europe, tuberculins are usually injected in the neck of the animal and readouts are carried out after 72 hours. However, injection of PPDs in other body parts is possible, such as in the caudal fold of the tail, which is commonly used in the USA, Canada or New Zealand (Palmer & Waters, 2011; Harrington *et al.*, 2014; Livingstone *et al.*, 2015).

Both SIT and CIT tests can be used for the diagnosis of bTB. For example, SIT tests are the method of choice in the eradication programmes in Spain, whereas CITs are employed in Ireland or the UK, where exposure to NTMs in cattle is high (Allen *et al.*, 2018). The selection of one or the other is mainly dictated by bTB prevalence, but also due to the possible exposure to NTMs (Monaghan *et al.*, 1994).

The OIE has established a set of thresholds to be used when reading out skin tests, although countries may adjust these on their surveillance and eradication programmes based on bTB prevalence or other epidemiological factors. In SIT tests, animals that present an increase of more than four mm in skin-fold thickness are considered positive, whereas a 2-4 mm increase is considered an inconclusive reaction and a limited swelling inferior to two mm is considered negative (OIE, 2018). In any case, animals with clinical signs in the site of injection, such as exudation or necrosis, are also considered reactors. In CIT tests, the difference between the skin-fold thickness between aPPD and bPPDs must be more than four mm for a reaction. A difference between 2-4 mm is interpreted as an inconclusive reaction, and a difference less than two mm is reported as a negative reaction. Alternatively, inconclusive reactors may be considered positive under certain prevalence scenarios or regions (severe interpretation), and are re-tested after at least 42 days (de la Rua-Domenech *et al.*, 2006).

Diagnostic performance of skin tests

The diagnostic performance of skin tests has been extensively evaluated in the scientific literature (de la Rua-Domenech *et al.*, 2006; Bezos *et al.*, 2011a). Experimental comparisons with microbiological culture, considered as the reference method for the confirmation of bTB infections, have shown a varying diagnostic sensitivity (Se) and specificity (Sp) between both techniques. The SIT is considered more sensitive than the CIT test, with Se values ranging between 59-100% and 68.6-95.5%, respectively (de la Rua-Domenech *et al.*, 2006; Nuñez-Garcia *et al.*, 2018). On the other hand, the SIT test is less specific than the CIT test, with Sp ranging between 75.5-99.4% and 99-100%, respectively. More recent studies employing Bayesian modelling have been used to establish similar diagnostic Se and Sp values for SITs and CIT tests (Alvarez *et al.*, 2012; Praud *et al.*, 2015; Lahuerta-Marin *et al.*, 2018).

Skin tests are considered a good test for the detection of bTB infections at the herd level, but their performance is greatly affected by several factors (de la Rua-Domenech *et al.*, 2006). These include the prevalence of bTB in the region under study (Monaghan *et al.*, 1994); differences in interpretation criteria (e.g. severe interpretation for inconclusive reactors); the sensitisation, vaccination or infection with NTM (Hope *et al.*, 2005; de la Rua-Domenech *et al.*, 2006; Garrido *et al.*, 2013; Seva *et al.*, 2014; Allen *et al.*, 2018) or the immune impairment of the host as a result of other infections, early post-partum or generalised TB (anergy), among others (de la Rua-Domenech *et al.*, 2006).

Interference due to infection with NTM is an important limitation of skin tests and the main reason for the use of CIT tests when no bTB is suspected (Garrido et al., 2013). However, their lower Se means that certain bTB-infected animals may be missed during testing, which entails a higher risk of persistence of the infection in the herd. In addition to the use of CIT tests, there have been several efforts to find and implement alternative MTBC-specific antigens to the PPDs in order to increase the Sp of skin tests, such as ESAT-6, CFP-10 or the Rv3614c (Mb3645c) protein (Bezos *et al.*, 2014). Although peptide cocktails usually present a lower Se when compared to PPDs, certain combinations have been shown to have a comparable performance to that of the skin test (Casal *et al.*, 2012; Jones *et al.*, 2012; Xin *et al.*, 2013; Srinivasan *et al.*, 2019; Srinivasan *et al.*, 2020). Furthermore, the use of specific peptide cocktails could be used as DIVA diagnostic test in the future (Srinivasan *et al.*, 2019; Srinivasan *et al.*, 2020).

1.4.1.3. Interferon Gamma Release Assays (IGRAs)

Principle of the technique

In 1990, Rothel and colleagues devised an *in vitro* method, the IGRA, that detected the secretion of IFN-γ by bTB-sensitised lymphocytes collected from the blood of animals suspected to be infected with bTB (Rothel *et al.*, 1990). During the procedure, extracted blood is incubated for 18-24 hours with bPPDs or aPPDs, after which plasma is collected for the detection of IFN-γ using a sandwich Enzyme-Linked Immunosorbent Assay (ELISA). The optical density (OD) is compared to that obtained from unstimulated blood and/or blood stimulated with aPPD. If the OD is above a certain threshold (defined by the manufacturer and/or official legislation), the reaction is considered positive.

Diagnostic performance of IGRAs

In a similar manner to skin tests, diagnostic performance varies in scientific literature according to the studied population as well as due to variations in reagents and the interpretation criteria or the reference method used for comparison (de la Rua-Domenech *et al.*, 2006). When compared to other methods, diagnostic Se and Sp values for IGRA tests range from 67-100% and 70-99%, respectively (de la Rua-Domenech *et al.*, 2006; Nuñez-Garcia *et al.*, 2018). Bayesian estimations using latent class analysis (LCA) indicate a higher Se of IGRAs when compared to skin tests (49-83% *vs.* 57-76%), although at the expense of a decrease in Sp (88-97.9% *vs.* 99.3-100%) (Clegg *et al.*, 2011; Alvarez *et al.*, 2012; de la Cruz *et al.*, 2018).

Several factors may affect the performance of IGRAs. Due to the use of PPDs for lymphocyte stimulation, cross-reactivity can also occur due to infection or vaccination with NTM (de la Rua-Domenech *et al.*, 2006; Alvarez *et al.*, 2008; Alvarez *et al.*, 2009b; Bezos *et al.*, 2012a; Roupie *et al.*, 2018). Although other antigens, such as ESAT-6 or CFP-10, have shown to have better diagnostic Sp values than PPDs, they need to be formulated in combination (e.g. ESAT-6/CFP-10) in order to achieve comparable sensitivity values (Bezos *et al.*, 2014; Lahuerta-Marin *et al.*, 2018).

Sample processing time is also an important factor that can affect the performance of IGRAs; blood samples need to be appropriately collected and transported, and need to be processed within the first eight hours, limiting the use of this technique if laboratories are far away from the sampling source (Rothel *et al.*, 1992). This effect is thought to occur due to a depletion of cellular viability and a loss of T cell stimulation, which leads to a decrease in IFN-γ production and in OD signals. Although several studies have shown that a delay of up to 24 hours in the processing of blood samples does not have a significant effect in test performance as long as blood is transported and stored in appropriate conditions (Ryan *et al.*, 2000; Gormley *et al.*, 2004; Coad *et al.*, 2007; Schiller *et al.*, 2009), contrary results have also been reported (Bezos *et al.*, 2011b). In contrast, a reduction of IGRA signal strength was observed when using different peptide cocktails, such as ESAT-6 or CFP-10 (Coad *et al.*, 2007).

In a similar manner to IT tests, the immune status of the animal can also have an important effect on IGRA performance; it may decrease in anergic, stressed, immunosuppressed or senescent animals (de la Rua-Domenech *et al.*, 2006). Furthermore, calves younger than six months have been shown to enact non-specific IFN-γ reactions, possibly due to the production of this cytokine by

innate NK cells (Olsen *et al.*, 2005), and therefore IGRAs are not recommended in this group of animals.

Skin tests and IGRAs detect cell-mediated immune (CMI) responses, although they appear to target detect different sub-populations of infected animals, particularly animals in early stages of infection (figure 7) (Pollock *et al.*, 2005). The application of IGRA is especially useful in combination with skin tests, leading to an increased sensitivity when used on infected herds (Gormley *et al.*, 2006; Alvarez *et al.*, 2012; Anon., 2013).

1.4.1.4. **Serological tests**

Skin tests and IGRAs are dependent on the host's CMI response and may therefore not be able to detect animals with an impaired immunity, such as anergic animals. Humoral responses are thought to play a secondary role in bTB infections as they mainly develop in later stages of the disease, especially during disseminated TB (Pollock *et al.*, 2005; Welsh *et al.*, 2005). Therefore, detection of antibodies could help to remove unreactive animals that could pose a risk for the rest of the herd.

Serological tests measure the humoral immune response in bTB infections by detecting antibodies against specific MTBC antigens, such as MPB70, MPB83, ESAT-6 or the P22 complex, which may be formulated alone or in combination (Bezos *et al.*, 2014; Casal *et al.*, 2017). They exhibit a Sp between 67.9-99.4% and a varying Se, ranging from 7.4%-96% depending on the studied population, antigens analysed or test kits used (Whelan *et al.*, 2010b; Waters *et al.*, 2011; Casal *et al.*, 2014; McCallan *et al.*, 2017; Infantes-Lorenzo *et al.*, 2019). Interestingly, performance has been found to be improved when testing takes place several days after skin tests are applied, possibly as a result of a boosting effect, with Se values ranging between 66.7-88.5% (Palmer *et al.*, 2006; Casal *et al.*, 2014; McCallan *et al.*, 2017).

The most extended serological tests are based on ELISA technologies. However, only the IDEXX *M. bovis* Antigen test kit is currently recommended by the OIE for its use as a supplementary method to detect *M. bovis* infections. Other serological platforms exist, such as lateral flow assays (Bermúdez *et al.*, 2012), or fluorescence polarisation assays (Jolley *et al.*, 2007), but are not currently used for routine diagnosis of bTB.

1.4.2. *Post-mortem* diagnosis of bTB

1.4.2.1. Pathological diagnosis of bTB

General pathology of bTB

The persistent nature of MTBC members makes bTB a progressive disease in which different infection stages frequently coexist within the host (Palmer *et al.*, 2007). The pathology of the disease is a highly complex process and the interplay of pathogen and host factors, such as strain virulence, host susceptibility and immune status, usually define its outcome.

During the initial stages of infection, tuberculous bacilli are phagocytosed by macrophages at the site of entry, usually alveolar spaces or mucous membranes (Domingo *et al.*, 2014). Once the initial infection has been established, infected macrophages generate new infectious foci in draining LNs as they migrate from the primary entry point, forming the primary complex (Domingo *et al.*, 2014). The primary complex may be incomplete or complete whether lesions are present only at LNs or at both the organ of entry and draining LNs, respectively.

These locations can develop into granulomatous lesions, called tuberculoid granulomas or tubercles, in which the homing of macrophages and T helper cells and the consequent inflammatory response lead to the formation of a fibrous tissue that encapsulates the infection. The presence of a strong CMI response in certain animals may prevent the progression of granulomas and their expansion to other organs. However, the immune response is usually unable to clear the infection and granulomatous lesions evolve in size and morphology (Wangoo *et al.*, 2005), with different levels of immune cell infiltration, caseation, mineralisation and necrosis. These lesions usually extend from the primary complex to other tissues in what is known as post-primary or chronic TB.

As the immune response of the host fails to control the infection, lesions progress and disseminate to other regions and organs of the body in what is known as generalised TB (Neill *et al.*, 2001; Domingo *et al.*, 2014). Cavitation is considered to be infrequent in bovines and mainly occur in late stages of the disease (Cassidy, 2006; Helke *et al.*, 2006; Ozturk-Gurgen *et al.*, 2020), although it has also been identified early in the infection (Palmer *et al.*, 2007).

It is generally accepted that most human beings infected by *M. tuberculosis* will manage to control the infection through the initial innate immune response,

either by full clearance of the pathogen or control of its replication (latent tuberculosis) (Getahun *et al.*, 2015). Although this is a classic feature of human TB infections, latent infection has not been clearly demonstrated in cattle (Alvarez *et al.*, 2009a; Sabio *et al.*, 2020). Evidence is mainly based on the isolation of tubercle bacilli from animals without visible lesions, although thorough examination of organs has revealed that apparently healthy tissues may indeed contain small tuberculous lesions (McIlroy *et al.*, 1986). Due to the risk of persistently infected animals in maintaining the infectious cycle, there is an increased interest in identifying these animals, especially in advanced stages of eradication programmes in which most infected animals are detected at early infection stages and may not present visible lesions.

Granulomatous lesions are not limited to bTB and may be caused by a diverse array of pathogens, including fungi (e.g. *Rhizomucor*), parasites (e.g. *Echinococcus*) and other bacteria (e.g. *Rhodococcus, Corynebacterium, Nocardia* and NTM) (Liebana *et al.*, 2008; Michelet *et al.*, 2018b). Therefore, the detection of gross lesions cannot be used to confirm bTB and tissue samples need to be collected for analysis with complementary *post-mortem* methods, such as histology, immunohistochemistry, microbiological culture or PCR.

Histopathology

Histopathology is a simple and inexpensive method by which tissue samples are fixated in formalin and analysed with Hematoxylin-Eosin and other stains (e.g. Zielh-Neelsen staining). Hematoxylin-Eosin staining allows the detection of the host's response against the pathogen through the detection of microscopic granulomatous lesions, which can be divided into four types (types I – IV) according to morphological criteria (Wangoo *et al.*, 2005). Ziehl-Neelsen staining is then used to differentially stain acid-fast bacteria, such as the members of the MTBC. However, acid-fast staining is not considered a specific method, since other mycobacteria, as well other *Actinomycetales*, such as *Nocardia*, are also acid-alcohol resistant.

Diagnostic Se and Sp for histopathology range between 72.1-93% and 92.3-95.2%, respectively, when employed on reactor animals and compared to microbiological culture (Varello *et al.*, 2008; Lopes *et al.*, 2020), with Se being highest among samples with visible lesions (VLs) (Varello *et al.*, 2008). Bayesian modelling has shown similar diagnostic performance, with diagnostic Se and Sp values ranging between 78.4-93.6% and 83.3-98.9%, respectively (Courcoul *et al.*, 2014; Pucken *et al.*, 2017; Larenas-Muñoz *et al.*, 2022).

1.4.2.2. **Microbiological culture**

Microbiological culture includes a diverse array of microbiological methods for the primary isolation and growth of MTBC members obtained from different types of samples, such as tissues or milk (Franco *et al.*, 2013). Once microbial growth is detected, presence of MTBC members needs to be confirmed by ancillary methods, such as acid-fast staining, biochemical assays or genetic methods (Gormley *et al.*, 2014).

Culture is considered the standard method for the confirmation of bTB infections, and is commonly used to assess the performance of other diagnostic methods. In addition, the isolation of MTBC is a key step for molecular characterisation techniques which are used in molecular epidemiological studies.

The overall performance of microbiological culture in isolating *M. bovis* and other members of the MTBC can be significantly affected by several factors related to the collection of samples, their processing and the culture media used.

Sample collection and pre-processing

Sample collection should ideally be carried out in aseptic conditions in order to avoid cross-contamination with environmental microorganisms, and transport should take place at an appropriate temperature in order to reduce sample degradation and growth of other microorganisms present in the sample (Gormley *et al.*, 2014). Samples should be stored at 4-6 ºC if they are going to be processed within 24-48 hours, or should be frozen if they will be processed at a later time (Gormley *et al.*, 2014).

Tissues are the most frequently analysed samples and, in order to release the bacteria from the complex tissue matrix, samples are homogenised in a liquid solution, usually phosphate saline buffer. Different tissue samples may be pooled and homogenised together.

Decontamination of samples

Due to the fastidious nature of MTBC members, their growth can easily be masked by contaminating microorganisms, which can originate from the host's microbiota, inappropriate sample handling or preservation, as well as from other coexisting infections. As a result, samples obtained for culture require a decontamination step by which they are incubated with aggressive chemical reagents.

The most commonly used decontaminants in veterinary samples are hexadecylpyridinium or cetidylpyridinium chloride (HPC or CPC), sodium hydroxide (NaOH), benzylkonium chloride (BC), oxalic acid (OA) and sulphuric acid (H_2SO_4) (de Lisle & Havill, 1985; Corner & Trajstman, 1988; Corner *et al.*, 1995; Corner *et al.*, 2012; Gormley *et al.*, 2014). Mycobacteria are partially resistant to decontamination due to their robust cell wall, and grow on inoculated culture media after this procedure. Survival of *M. bovis* in tissue homogenates has been found to vary across decontamination reagents, ranging between 15.1-70.6% when using recommended concentrations (Corner *et al.*, 1995). However, the amount of viable bacilli after decontamination not only depends on the type and amount of decontaminant used, but also on the amount of processed tissue, the bacterial load in the sample and the time of exposure (Corner *et al.*, 1995; Ambrosio *et al.*, 2008; Corner *et al.*, 2012; Medeiros *et al.*, 2012). Therefore, decontamination conditions need to be finely tuned in order to achieve the best recovery and decontamination rates (Corner *et al.*, 1995).

An important factor that defines the selection of decontaminants is the type of media that will be used subsequently for isolation, since not all media platforms are compatible with all decontaminants. For example, although HPC is a commonly used and very efficient decontaminant, it is not recommended for liquid Mycobacterial Growth Indicator Tubes (MGIT) (Siddiqui & Rüsh-Gerdes, 2006). In the latter case, NaOH or OA are recommended as decontamination reagents instead.

With exception of detergent-based decontaminants (i.e. HPC or BC), decontamination procedures require a subsequent neutralisation step, usually by addition of an acid solution in alkali-based decontaminants (e.g. orthophosphoric acid for NaOH) (Corner *et al.*, 1995), or an alkali solution in acid-based decontaminants (e.g. NaOH for H_2SO_4) (de Azevedo Issa *et al.*, 2017). After neutralisation, homogenates are centrifuged and pellets are inoculated on different culture media and incubated at 37 ºC.

Types of culture media, incubation times and nutritional requirements

There is a wide variety of culture media, both solid and liquid, that can be employed for isolation of MTBC members (table 2). Performance of isolation and incubation times vary between platforms and therefore laboratories need to evaluate the choice of culture media in order to guarantee the best results.

Solid media can be classified into agar based media, such as Middlebrook 7H11 agar, or egg based media, such as Löwenstein-Jensen (LJ) (table 2). Many solid media, such as LJ, are able to inhibit the growth of most microorganisms other than mycobacteria (MOTM) due to the presence of malachite green, a dye with antimicrobial properties that gives these media their characteristic green hue. Furthermore, the growth of MOTM can also be limited with the supplementation of antibiotics, such as polymixin B, amphotericin B, carbenicillin and trimethoprim (e.g. Stonebrink media).

Liquid or broth-based media can be classified into two different groups according to the presence or absence of an indicator that reveals when bacterial growth occurs. In the BACTEC MGIT 960 systems, growth is detected through a fluorescent signal that is emitted as a result of oxygen consumption. Previous to this method, there was the BACTEC 460 platform, no longer in use, which was based on the detection of a radioactive signal emitted when microorganisms present in the media degraded a C^{14}-labelled substrate (Siddiqui & Rüsh-Gerdes, 2006). Finally, the VersaTrek Myco platform measures changes in pressure inside tubes as a result of gas consumption/production. Similar to solid media, specific combinations of antibiotics can also be added to liquid cultures, such as PANTA in MGIT, which contains polymyxin B, amphotericin B, nalidixic acid, trimethroprim and aziocillin.

It is important to note that not all culture media support growth of all members of the MTBC; due to a point mutation in the *pykA* gene, *M. bovis, M. africanum* and *M. microti* cannot metabolise glycerol and therefore are unable to grow on media that contain this compound as the only source of carbon, such as Löwenstein-Jensen (Keating *et al.*, 2005). In order for these variants to grow, supplements need to be added to the medium, such as sodium pyruvate or aspartic acid. Furthermore, to allow the growth of MTBC members, certain media require the addition oleic albumin dextrose catalase (OADC), such as Middlebrook 7H9.

Media platform	Component basis	Example
Solid	Agar based	Middlebrook 7H10/7H11
		Blood agar B83
	Egg based	Löwenstein-Jensen
		Coletsos
		Stonebrink
Liquid	Indicator	BACTEC MGIT 960
		BACTEC 460
		VersaTrek Myco
	Non-indicator	Middlebrook 7H9

Table 2. Microbiological culture media used for the isolation of MTBC members.

Among solid media, growth of *M. bovis* is considered to be faster on agar-based media, but egg-based media presents a higher sensitivity (Corner *et al.*, 2012). The suggested minimum incubation time for solid media is 8 weeks, but incubation up to 12 weeks is recommended (OIE, 2018).

Broth-based platforms allow for higher throughput of samples and a more rapid time to detection, since sample readout is automated and MTBC members tend to grow better and faster on these media (Gormley *et al.*, 2014). When compared to solid media, the BACTEC MGIT 960 platform achieves higher *M. bovis* recovery rates to those observed in solid media (81.9-94.6% *vs.* 74.4-80.6%), with average detection times of approximately 2-3 weeks and 3-6 weeks, respectively (Hines *et al.*, 2006; Robbe-Austerman *et al.*, 2013). Despite the reduced time to detection, the recommended incubation times for liquid media range between 6-8 weeks for maximum sensitivity.

Contamination rates appear to be higher on liquid media than on solid media (Robbe-Austerman *et al.*, 2013; Yates *et al.*, 2017), although contradicting results have been observed in literature (Hines *et al.*, 2006). These differences may be related to other methodological factors, such as the type and time of exposure to the decontaminant, or the antibiotic and concentrations used (Gormley *et al.*, 2014).

Culture performance and limitations of culture

Performance of culture is usually measured through the recovery rate, that is, the number of suspicious or high risk samples (i.e. obtained from reactor animals and/or tissues with VLs) that are positive to *M. bovis* or other MTBC members. This varies significantly between studies, ranging from 19–100% depending on different factors, such as the prevalence of the disease in the sample population, the pathological status of the analysed samples, the decontamination procedure or the culture media used (Hines *et al.*, 2006; Parra *et al.*, 2008a; Costa *et al.*, 2013; Araújo *et al.*, 2014a; Araújo *et al.*, 2014b; Yates *et al.*, 2017; Lopes *et al.*, 2020; Pozo *et al.*, 2021).

The pathology of the tissue samples submitted for microbiological analysis is critical when determining the success rate of culture, since samples from reactor animals with VLs are more likely to contain tuberculous bacilli. Indeed, when employed on high bTB risk samples, recovery rates can be high, ranging between 62.5–100% (Parra *et al.*, 2008a; Araújo *et al.*, 2014a; Araújo *et al.*, 2014b; de Azevedo Issa *et al.*, 2017; Pozo *et al.*, 2021).

On the other hand, the recovery rate of culture falls to 19.5-32.2% when employed on reactor animals with non-visible lesions (NVLs) (Parra *et al.*, 2008a; Araújo *et al.*, 2014a; Araújo *et al.*, 2014b), probably due to the difficulty of obtaining an appropriate sample from apparently healthy tissues and the negative effect of decontamination in cellular viability, especially in paucibacillary samples with low bacterial loads. This highlights the importance of the visual inspection and collection of the majority of target samples (LNs) at the abattoir (see section 1.6.2.3), and to finely tune the decontamination protocol. If the concentration of the decontaminant and contact time is excessive, recovery of MTBC members will be low, whereas if decontamination is insufficient, other microorganisms will grow instead (Corner *et al.*, 1995; Gormley *et al.*, 2014). Different decontaminants may have different mechanisms of action depending on the metabolic status of the tuberculous bacilli in the sample (Corner *et al.*, 2012), and may therefore be better suited for these type of samples (Medeiros *et al.*, 2012).

Due to the previously mentioned limitations of *ante-mortem* tests and slaughterhouse surveillance, not all samples sent for microbiological analysis will contain tuberculous bacilli or enough loads to survive decontamination. Therefore, the limited Se of slaughterhouse surveillance and microbiological culture may lead to an underestimation of the true prevalence of infection (de la Rua-Domenech *et al.*, 2006). When used as a standard method, the lower

performance of microbiological culture will negatively impact the diagnostic performance of any comparative method. As a result, there is an increased interest in the use of Bayesian modelling to evaluate the diagnostic performance of diagnostic methods for bTB without the need of using any comparative technique, such as culture. By using this methodology diagnostic Se and Sp values for culture have been established between 78.1-88.9% and 99.1-100%, respectively (Courcoul *et al.*, 2014; Pucken *et al.*, 2017).

1.4.2.3. Molecular genetic methods for MTBC detection

Molecular genetic methods can be used for multiple purposes, such as detection and/or identification of MTBC members or their molecular characterisation for epidemiological and evolutionary purposes. The introduction of these methods revolutionised TB diagnostics and allowed great advances in the knowledge of the evolution and epidemiology of human and animal TB, including the description of novel MTBC members. Molecular genetic methods mainly used for the confirmation of infection will be detailed in this section, whereas other genetic methods employed in identification or molecular characterisation will be detailed in section 1.5.

The development of nucleic acid amplification techniques through Polymerase Chain Reaction (PCR) allowed for a diverse array of rapid, sensitive and specific techniques which were not only suitable for the confirmation of the infection and characterisation of MTBC members, but also paved the way for their direct detection on different matrices.

Polymerase chain reaction (PCR)

Polymerase chain reaction is a widely used technique in molecular biology due to its versatility and high analytical capacity. Analytical sensitivity, or Limit of Detection (LOD), of this method is high, since it is able to detect only a few copies of the target nucleic acids in the reaction mixture (Costa *et al.*, 2014). If appropriate DNA targets and oligonucleotides are selected, analytical specificity of PCR is also very high and amplification of unrelated targets is highly unlikely.

End-point PCR

End-point PCR is a molecular technique in which the amplified product is detected at the end of the PCR reaction through the use of agarose gel electrophoresis and DNA intercalating dyes. Although PCR was invented and implemented in the mid-1980s, it was not until the early 1990s that this technique was used in bTB diagnostics (Cousins *et al.*, 1991; Costa *et al.*, 2014). Initial

studies described the selection of MTBC-specific targets, such as the *mpb70* gene, and the use of PCR for MTBC identification in DNA samples obtained from pure cultures (Cousins *et al.*, 1991). Due to the analytical capability of PCR and its rapid detection time when compared to culture (few hours *vs.* weeks), there was a significant interest in implementing PCR as a direct detection method in fresh or formalin-fixed tissue samples (direct PCR) instead of microbiological culture (Liébana *et al.*, 1995; Wards *et al.*, 1995; Miller *et al.*, 1997; Coetsier *et al.*, 2000; Zanini *et al.*, 2001). Since then, numerous studies have been published regarding the performance of direct PCR on fresh tissues, with diagnostic Se and Sp values ranging between 63–100% and 50–100%, respectively (Liébana *et al.*, 1995; Wards *et al.*, 1995; Taylor *et al.*, 2001; Zanini *et al.*, 2001; Taylor *et al.*, 2007b; Stewart *et al.*, 2013; Sanchez-Carvajal *et al.*, 2021). In the case of formalin-fixed tissues, diagnostic sensitivity has been shown to be high (88–93%) in samples with high loads of acid-fast bacilli (Miller *et al.*, 1997; Coetsier *et al.*, 2000), but reduced when used on paucibacillary samples or tissue samples with negative culture results, probably as a result of molecular changes in the fixated sample, such as protein coagulation or DNA degradation (Coetsier *et al.*, 2000; Miller *et al.*, 2002).

In addition to its high analytical Se, end-point PCR presents multiplexing capacity, which allows the use of different primer combinations for the detection of different genetic targets. In bTB diagnostics, multiple copy genetic targets, such as the IS*6110* or IS*1081* elements, are frequently used, since they offer higher sensitivity than single copy targets, such as the *mpb70* gene (Wards *et al.*, 1995; Taylor *et al.*, 2007b). However, other genetic targets, such as the *HupB* gene or a deletion in the *mce-3* operon (Mishra *et al.*, 2005; Mittal *et al.*, 2014), have also been used.

Real-time PCR (qPCR)

The use of real-time PCRs was introduced in the field of bTB diagnosis during the early 2000s and has rapidly been adopted in many different veterinary laboratories around the world. Real-time PCR technologies are based on the real-time detection of fluorescence emitted during the amplification of DNA. Fluorescence is emitted either by DNA-intercalating fluorophores, such as SYBR-Green, or fluorescent oligonucleotide probes. The latter present a higher specificity, since intercalating fluorophores bind to any double-stranded DNA molecule irrespective of its origin, whereas probes are based on sequence complementarity.

There is a wide array of probes that may be used in qPCR applications, but the most extended are hydrolysis probes. A fluorophore is attached to one extreme of the oligonucleotide, whereas a quencher is attached to the other. When the fluorophore is excited by the laser in the thermal cycler, fluorescence is inhibited by the transfer of energy to the quencher through fluorescence resonance energy transfer. When amplification of DNA takes place, the probe binds to the amplifying product and is excised by the 3' exonuclease activity of the DNA polymerase, therefore releasing the fluorophore into the solution and allowing the emission of fluorescence.

Target amplification is determined according to the cycle threshold (Ct), which corresponds to the reaction cycle in which fluorescence surpasses a certain threshold, which corresponds to a significant increase in fluorescence with respect to the baseline or background signal.

Under optimised conditions, amplification of DNA is considered proportional to the emitted fluorescence and, due to this, real-time PCRs are commonly known as quantitative PCRs (qPCRs). However, not all real-time PCRs are quantitative, since only the use of a standard curve (absolute quantification) or a reference target (relative quantification) can accurately establish the amount of copies of the genetic target originally present in the sample. In the rest of applications, qPCRs can only indicate a relative quantification with respect to the rest of the samples and the amplification controls, based on the comparison of Ct values. Nevertheless, due to the extended use of this term, qPCR will be used to refer to both qualitative and quantitative real-time PCRs in this thesis.

Real-time PCRs offer several advantages over end-point PCRs. Firstly, due to the small amplicon size [60–200 base pairs (bp)] and the real-time detection of fluorescence, reaction times are reduced. Secondly, sample post-processing is not required, which limits the risk of cross-contamination due to the management of highly-concentrated DNA amplicons. Thirdly, due to the diversity of fluorophores and quenchers, qPCRs can be easily multiplexed to detect different targets simultaneously. Finally, qPCRs can be used to quantify the amount of template DNA, which in turn can be correlated with the bacterial load in the sample. However, in order to obtain a robust quantification, extra optimisation steps are required to guarantee the maximum efficiency of amplification, since small variations in efficiency between runs can significantly alter quantification results.

Several genetic targets have been used for real-time detection of MTBC members in tissue samples (direct qPCR), such as the IS*6110* and IS*1081* elements, the TbD1 region, or the *DevR* or *rv2807* genes (Parra *et al.*, 2008a; Thacker *et al.*, 2011; Costa *et al.*, 2013; Araújo *et al.*, 2014a; Araújo *et al.*, 2014b; Courcoul *et al.*, 2014; Sevilla *et al.*, 2015; Michelet *et al.*, 2018b). In a similar manner to end-point direct PCR, diagnostic performance of direct qPCR varied between these studies, with diagnostic Se and Sp values ranging between 66.7–100% and 88.7–100%, respectively (Parra *et al.*, 2008a; Thacker *et al.*, 2011; Costa *et al.*, 2013; Araújo *et al.*, 2014a; Araújo *et al.*, 2014b; Courcoul *et al.*, 2014).

Factors affecting PCR performance

The variation in diagnostic performance of PCR (end-point and qPCR) between studies can be related to several factors, such as limitations in the reference method used for comparison or in the DNA extraction procedure. The number of samples used and their pathological status, can also affect performance results.

The higher analytical Se of PCR in comparison to culture and its independence of cellular viability implies that certain tissue samples may test positive to the PCR but remain negative to culture, especially in samples with low bacterial loads that may not survive decontamination. As a result, the apparent decreased Sp of PCR with respect to culture may correspond in many cases to the reduced sensitivity of the latter. However, cross-reactivity of primer pairs with different NTMs have been described for different genetic targets in both end-point PCR and qPCR, such as the IS*6110* element (da Silva Rabello *et al.*, 2010; Thacker *et al.*, 2011; Rabello *et al.*, 2012).

Similar to microbiological culture, Bayesian modelling has been used to assess the diagnostic performance of qPCR; diagnostic Se and Sp values were 87.7% and 97.0%, respectively (Courcoul *et al.*, 2014). Although Sp values were slightly higher for microbiological culture (Sp = 99.1%), Se was significantly lower (Se = 78.1%).

The lower diagnostic Se reported in certain studies reflects the complexity of extracting mycobacterial DNA of sufficient quantity and quality for PCR detection. Mycobacterial DNA is enclosed within mycobacterial cells with a highly resistant cell wall, which are usually confined inside the lysosomes of macrophages which, in turn, are found within a granuloma or embedded in the tissue matrix (Radomski *et al.*, 2013). Furthermore, granulomas may appear casseated, calcified or filled with pus in advanced lesion stages, which further

complicates the extraction and amplification of DNA from this matrix due to the presence of PCR inhibitors (Stewart *et al.*, 2013).

Common PCR inhibitors include excess concentrations of host DNA or organic compounds, such as urea, collagen or bile salts. As a result, inhibition controls (ICs) are frequently used in order to assess the presence or absence of inhibitors in the DNA sample. Internal controls can be endogenous or exogenous whether the control is contained within the sample or added during or posteriorly to the extraction. In turn, ICs can be classified into homologous or heterologous whether they are detected by the same or a different primer set than the one used for the primary target, respectively.

The complexity behind mycobacterial DNA extraction from fresh tissues has led to a wide array of available protocols over the years. Although commercial kits are available, they are usually modified to be adapted to the specific requirements of granulomatous lesions (Radomski *et al.*, 2013; Fell *et al.*, 2016). Extraction methods often involve the use of enzymes (e.g. proteinase K or lysozyme), surfactants (e.g. SDS or Tween 20), mechanical lysis (e.g. bead beating or grinding) or a combination thereof, and purification of DNA is then achieved either by phenol:chloroform:isoamyl alcohol (PCI) extraction or through column-based methods (van Soolingen *et al.*, 1991; Zanini *et al.*, 2001; Radomski *et al.*, 2013; Leth *et al.*, 2017). The combination of enzymatic and mechanical lysis methods, followed by the use of column-based methods for DNA purification, is nowadays frequently employed in many laboratories around the world, offering good results and eliminating the use of the highly toxic PCI (Radomski *et al.*, 2013).

In addition to the improvements in DNA extraction efficiency, other strategies have been followed to obtain more concentrated and pure mycobacterial DNA. A popular approach is the use of sequence capture platforms, which use biotinylated probes to hybridise MTBC DNA in samples after the mechanical and biochemical lysis of tissues, followed by capture of the biotinylated complexes through streptavidin-coupled magnetic beads (Taylor *et al.*, 2001; Taylor *et al.*, 2007b; Parra *et al.*, 2008a; Fell *et al.*, 2016). In addition to concentrating DNA, sequence capture methods allow for a better removal of PCR inhibitors. In turn, immunomagnetic separation has also been used for concentration of *M. bovis* bacilli in tissue samples, although with limited sensitivity (Stewart *et al.*, 2013).

The use of larger amounts of tissue for DNA extraction could potentially increase sensitivity of PCR detection, especially in paucibacillary samples. However, an excessive amount of tissue can be detrimental to DNA extraction either by an increase of inhibitors, especially host DNA, or by a decreased efficiency of purification methods (Radomski *et al.*, 2013). However, the Matrix Lysis procedure has recently been shown to achieve good PCR results when using tissue quantities of up to 10 grams (Leth *et al.*, 2017).

Finally, the reamplification of DNA through nested PCR has been shown to achieve a higher sensitivity than a single amplification round (Wards *et al.*, 1995; Taylor *et al.*, 2007b). Although good diagnostic performance has been achieved with this methodology (Costa *et al.*, 2013), its use entails an increased risk of cross-contamination due to the processing of concentrated amplified products and should be weighed against its possible advantages.

Other nucleotide amplification assays

Loop-mediated isothermal amplification (LAMP)

Loop-mediated isothermal amplification is a sensitive and inexpensive method that is based in auto-cycling strand displacement for the DNA amplification of the target sequence using a combination of four to six primer sequences and a DNA polymerase with a high strand displacement (Notomi *et al.*, 2000). Due to the lack of denaturalisation steps, LAMPs require the use of stable temperatures (60 – 65 °C) and can be carried out in standard heat blocks or water baths, which allows its use in low income settings. Positive amplification is detected either through visual inspection of turbidity, or by the addition of SYBR-Green to the test tube. Tomita and colleagues described a method based in the addition of calcein and manganous ion to detect DNA amplification without the need of post-processing (Tomita *et al.*, 2008). Despite its promising applications, LAMP has only been used to identify MTBC members, such as *M. tuberculosis sensu stricto* or *M. bovis*, in DNA obtained from culture isolates (Kapalamula *et al.*, 2021), or in human clinical samples, such as sputum, blood or pleural fluid (Zhu *et al.*, 2009).

Digital amplification methods (dPCR and dLAMP)

Digital amplification methods are based on the partitioning of amplification reactions into hundreds to millions of separate reactions by the use of microwells, capillaries, oil emulsions or arrays (Kuypers & Jerome, 2017). Since the conceptualisation of digital PCR (dPCR) in the 1990s, other platforms have been devised, such as droplet dPCR and digital LAMP (dLAMP). Droplet dPCR

is considered a highly sensitive and specific method, with very promising applications in the diagnosis of human TB (Nyaruaba *et al.*, 2019). Its implementation in the detection of MTBC members in bovine samples has been limited (Larenas-Muñoz *et al.*, 2022), but could be especially useful in paucibacillary samples. Digital LAMP, on the other hand, has been used to detect *M. bovis* DNA in spiked milk samples with a very high analytical sensitivity (14 Colony Forming Units/mL) and robustness (Tao *et al.*, 2020).

1.5. Identification and characterisation of mycobacteria

In the context of bTB, the identification of MTBC members is important in order to be able to differentiate possible origins of infection. However, the identification of mycobacterial species is not limited to the MTBC; NTM are also isolated in some animal species and could interfere with the different diagnostic methods. As previously mentioned, several NTM species or other microorganisms can cause granulomatous lesions or lead to unspecific immunological reactions to the skin test or IGRA. As a result, it is critical for animal health authorities to be able to identify these species in order to detect false positive results, especially in low prevalence settings in which the isolation of MTBC members is rare and control strategies are limited to slaughterhouse surveillance. In addition, the molecular characterisation of MTBC isolates is crucial in order to establish the genetic relatedness between strains, which in turn can be used in epidemiological investigations.

The first two parts of this section will describe different phenotypic and genotypic methods used for the identification of mycobacteria in general, including the MTBC. The third part will focus specifically on genotypic methods for the molecular characterisation of MTBC members, mainly *M. bovis*.

1.5.1. Phenotypic methods for the identification of mycobacteria

1.5.1.1. Biochemical assays

Before the advent of genetic molecular methods, mycobacterial species were identified based on their phenotypic properties, such as their growth rate and pigmentation (i.e. the Runyon classification), or through several biochemical assays. These assays were mainly used to identify the most relevant mycobacterial species, such as the MTBC, *M. avium, M. scrofulaceum* or *M. kansasii*, among others (Bhalla *et al.*, 2018). Commonly used biochemical assays include the niacin accumulation test, the nitrate reduction test, the Tween 80 hydrolysis test or the tellurite reduction test. Some of these methods

are more specific than others; while the niacin reduction test can be used with all mycobacteria, the tellurite reduction test can be used to identify members of the MAC (Shinnick & Good, 1995).

Although these tests can be useful to identify certain mycobacteria, they can be cumbersome to carry out due to the extensive taxonomy of this bacterial group, the fastidious growth of some of its members and the need to evaluate a large panel of phenotypic features for an appropriate identification (Tortoli, 2010). In addition, similar to other bacterial species, phenotypic variation can occur.

1.5.1.2. **Chromatographic methods**

A common phenotypic approach to identify mycobacteria is the analysis of their mycolic acids through chromatographic methods, such as thin-layer chromatography (TLC) or high-performance liquid chromatography (HPLC) (Butler & Guthertz, 2001). This technique analyses the lipid fractions of pure mycobacterial cultures obtained after their saponification and posterior processing and separation through chromatographic methods, obtaining a series of characteristic patterns or chromatograms.

Although mycolic acids are present in other actinobacteria, TLC revealed the presence of at least 7 different types of mycobacterial mycolic acids (Luquin *et al.*, 1991). Some of these mycolic acids are specific to certain mycobacterial species, while others are shared between all mycobacteria; for example, while the ω-1-methoxy-mycolates (type VII mycolic acids) are specific of certain RGM, such as *M. fortuitum, M. peregrinum* or *M. porcinum,* all mycobacteria contain α-mycolic acids (type I mycolic acids) (Luquin *et al.*, 1991; Tortoli, 2003). As a result of the limited number of mycolic acid types and the shared patterns between mycobacteria, TLC has not proved very useful for the chemotaxonomic characterisation of mycobacterial species (Tortoli, 2003).

In contrast, HPLC allows for a much higher resolution when compared with TLC, since each species is characterised by a particular chromatographic pattern. Representative chromatograms have been obtained for many mycobacterial species; however, minor variations in the heights of certain peaks have been identified among isolates of the same species (Butler & Guthertz, 2001). Libraries of chromatograms of different mycobacterial species have been created to be used as a reference. Nevertheless, certain species of mycobacteria have identical or very similar chromatograms, hampering their appropriate identification through HPLC (Tortoli, 2003).

In comparison to biochemical assays, mycolic acid analysis offers a unified procedure that can give results in a reduced amount of time. Despite its apparent simplicity, mycolic acid analyses require costly instrumentation and trained personnel that can effectively interpret the chromatographic patterns obtained through this technique.

1.5.1.3. **Mass spectrometry**

The application of mass spectrometry (MS) for the identification of bacterial organisms was proposed during the 1970s (Anhalt & Fenselau, 2002), but it was not until the 1990s that this methodology was first used on bacterial isolates without any pre-treatment (Claydon *et al.*, 1996; Holland *et al.*, 1996).

Identification of bacterial isolates using MS is based on the use of "soft" ionisation techniques, such as matrix-assisted laser desorption/ionisation — time of flight (MALDI-TOF) MS, which allows the analysis of large non-volatile molecules, such as proteins (Croxatto *et al.*, 2012). In this method, the analyte, usually a portion of biomass obtained from a bacterial culture or a protein extract, is co-crystallised with an acidic matrix, such as a-cyano-4-hydroxycinnamic acid (HCCA), on a metal target plate. The sample-matrix is then ionised using a laser, resulting in protein ions of different sizes that are desorbed from the matrix. These ions are then accelerated by a potential difference through an acceleration grid until they collide against a detector. The time of flight of these ions depends on their mass to charge (m/z) ratio; since MALDI-TOF mostly generates single charged ions, the m/z ratio mainly reflects the mass of the ion. Spectral profiles are then represented as peaks that correspond to the frequency at which ions of a specific m/z ratio, or mass, collide against the detector. The majority of the peaks detected through MALDI-TOF MS correspond to conserved proteins with house-keeping functions, such as ribosomal proteins (Croxatto *et al.*, 2012; Alcaide *et al.*, 2018).

Similar to HPLC, the principle behind microbial identification through MALDI-TOF MS is based on the generation of characteristic patterns or spectral profiles which can be unique for a certain species or bacterial group. The obtained spectral profiles are then compared against a database of previously defined spectral profiles and an identification is achieved based on the similarities between them.

Due to its speed, sensitivity and relative simplicity, since intact cells obtained directly from cultures can be used in many cases, MALDI-TOF MS advanced the field of clinical microbiology by allowing a rapid identification of many

important human and veterinary fungal and bacterial pathogens (Clark *et al.*, 2013; Randall *et al.*, 2015). Several commercial platforms are available, such as the Bruker BioTyper or the Vitek-MS systems. These platforms contain their own standardised spectral databases and recommend specific extraction protocols. An advantage of these systems is the possibility of expanding their databases to include spectra from species or strains not included in the commercial versions. In the context of mycobacterial identification, the Bruker BioTyper is the most frequently used MALDI-TOF MS platform (Alcaide *et al.*, 2018).

The implementation of MALDI-TOF MS in the identification of mycobacteria faced a series of limitations (Alcaide *et al.*, 2018). Due to the biosafety requirements of certain mycobacterial species (e.g. MTBC), isolates require a previous inactivation in order to be analysed outside biosafety level 3 (BSL-3) facilities (Clark *et al.*, 2013). The complex cell wall composition and aggregation capacity of mycobacteria greatly reduce the quality of the obtained spectra. Therefore, a great amount of research was invested in establishing the effect of culture conditions on MALDI-TOF MS performance and developing more efficient and standardised extraction protocols for mycobacteria. These publications addressed the inactivation of mycobacterial isolates through ethanol or heat treatment, the dispersal of cellular aggregates via grinding, the use of different solvent compositions or the improvement in the penetration of extraction reagents by way of different lysis procedures, including mechanical lysis (bead-beating), sonication or freeze-thawing (El Khechine *et al.*, 2009; Saleeb *et al.*, 2011; Balada-Llasat *et al.*, 2013; Machen *et al.*, 2013; Rodriguez-Temporal *et al.*, 2018).

During the posterior validation of MALDI-TOF MS in mycobacterial identification, an overall increase in performance was frequently described when commercial databases were expanded with spectra obtained from their private mycobacterial collections (Machen *et al.*, 2013; Mather *et al.*, 2014). This not only underscores the importance of achieving an appropriate representation of mycobacterial strains with good quality spectra in the available databases, but also the need to establish standardised extraction protocols, since variations in culture and extraction procedures with respect to those used for the development of the database could lead to a reduced performance (Clark *et al.*, 2013; Machen *et al.*, 2013). In this aspect, the continuous updates of spectra libraries has significantly improved the identification of mycobacteria (Rodriguez-Sanchez *et al.*, 2016; Rodriguez-Temporal *et al.*, 2017; Alcaide *et al.*, 2018).

Commercial MALDI-TOF MS systems use confidence scores to categorise the similarities between the obtained spectra with those stored in the database. According to the Bruker and Vitek systems, a score above 2.0 or 90%, respectively, can be considered a high-confidence identification to the species level. A lower score between 1.7 and 2.0 can be considered a low-confidence identification to the genus level. Nevertheless, studies have revealed a high level of agreement between genotyping and MALDI-TOF MS identifications at lower confidence scores than those recommended by the manufacturer, similar to what has been reported for other Gram positive microorganisms, as well as yeasts (Buchan & Ledeboer, 2013; McElvania Tekippe *et al.*, 2013; Mather *et al.*, 2014; Rodriguez-Sanchez *et al.*, 2016). Consequently, several publications have recommended a different cut-off value to that established by the manufacturer. A cut-off value of 1.8 could be used to establish a high-confidence identification in the Bruker BioTyper, whereas a value between 1.6 and 1.8 could be used to establish a low-confidence identification (Alcaide *et al.*, 2018). In the case of the Vitek-MS platforms, score values between 80 – 90% can also be considered as high-confidence identifications, followed by 60 – 80% as low-confidence and less than 60% as non-reliable (Martiny *et al.*, 2012).

In addition to its speed and accuracy, MALDI-TOF MS has become an important microbial identification method due to its reduced running costs; although the initial investment is high, due to the cost of equipment and commercial databases, the high throughput of this technique added to the inexpensive consumables needed make it a highly cost-effective method (Saleeb *et al.*, 2011). However, MALDI-TOF MS still faces several limitations when used for mycobacterial species identification. Firstly, MALDI-TOF MS is not able to effectively differentiate between closely related mycobacterial species or groups of interest, including the members of the MTBC or *M. intracellulare* and *M. chimaera*. Several studies have evaluated the possibility of improving its resolution by analysing a large variety of spectra from different strains of the MTBC and MAC, indicating that with an appropriate database it could be possible to differentiate these species in the future (Pranada *et al.*, 2017). Secondly, less prevalent species are not well represented in commercial databases (Costa-Alcalde *et al.*, 2019). This issue is further accentuated when considering that most databases mainly contain spectra from human isolates, with limited representation of environmental and animal sources.

Although the majority of studies related to mycobacterial identification through MALDI-TOF MS have been carried out in the human clinical setting (Alcaide *et al.*, 2018), this methodology could also be an interesting alternative in the veterinary microbiology laboratory. The Bruker BioTyper has been recently used to identify more than 100 NTM isolates, belonging to 24 different mycobacterial species, from wild boars in Switzerland (Ghielmetti *et al.*, 2021a). The high agreement between MALDI-TOF MS and Sanger sequencing of housekeeping genes in this study indicates that proteomic analysis could be an interesting alternative to genotyping methods for the rapid identification of NTM in the veterinary setting. Furthermore, with appropriate main spectra prophile (MSP) sampling, closely related veterinary relevant mycobacterial species could be differentiated in the future, including the MTBC or *M. avium* subspecies (Ravva *et al.*, 2017; Ricchi *et al.*, 2017; Bacanelli *et al.*, 2019).

1.5.2. Genetic molecular methods for the identification of mycobacteria

The advances in the analyses and detection of genetic sequences have revolutionised the ability of microbiology laboratories to identify bacterial isolates. In the case of mycobacteria, the most popular approaches are based on PCR, DNA probe hybridisation or Sanger sequencing of housekeeping genes (Forbes *et al.*, 2018). Although WGS has been used to evaluate the phylogeny of mycobacteria (Tortoli *et al.*, 2017; Gupta *et al.*, 2018), it has not found yet a proper use in the routine identification of NTM isolates due to its increased cost and informatics requirements.

1.5.2.1. PCR

In addition to detecting mycobacteria directly from tissue samples, PCR can also be used to identify mycobacterial isolates. Different molecular targets are available for this purpose, such as complex-specific markers, including the IS*6110* and IS*1081* elements in the MTBC (Liebana *et al.*, 1996), or the *mpb70* gene (Wilton & Cousins, 1992; Liebana *et al.*, 1996); or species-specific markers such as the IS*1311* element for *M. avium,* the IS*900* for MAP (Turenne *et al.*, 2007; Shin *et al.*, 2010) or different RDs and SNPs present in different MTBC members (see section 1.5.3).

1.5.2.2. Genetic probes

Two major types of probe assays are used to identify mycobacterial species; direct DNA probe assays and direct line probe assays (LPAs). The first category encompasses assays that use probes that bind directly to the extracted genomic DNA, whereas the second includes those that require a preliminary DNA amplification step, usually a PCR. Commercial kits are available for both types of assays; for example, AccuProbe is a direct DNA probe assay for the detection of the MTBC, the MAC, *M. avium, M. intracellulare, M. gordonae* or *M. kansasii* (Richter et al., 1999), while the INNO-LiPA Mycobacteria is a LPA that can detect the genus *Mycobacterium* as well as 16 different mycobacterial species (Tortoli *et al.*, 2003).

Despite the accuracy of probe-based assays in the identification of mycobacteria, these methods face a series of limitations. Firstly, the number of species that can be identified is limited when compared to the extensive taxonomy of the genus *Mycobacterium*. Secondly, small polymorphisms in the genetic sequences targeted by the probes can prevent hybridisation. Thirdly, a high similarity between mycobacterial species can lead to cross-reactions (Lumb *et al.*, 1993; Tortoli *et al.*, 2010). Nevertheless, genetic probes are still commonly used for the identification of different mycobacterial species (Forbes *et al.*, 2018).

1.5.2.3. Sanger sequencing

Due to its increased resolution and versatility, Sanger sequencing of conserved genetic regions is considered the reference method for the identification of mycobacterial species (Tortoli, 2010). The most frequently used target is the 16S rRNA gene (Tortoli, 2010), although other genes, such as the *hsp65* and the *rpoB* genes (Telenti *et al.*, 1993; Adekambi *et al.*, 2003; Turenne *et al.*, 2006), are frequently used as complementary targets to increase the resolution in those cases in which mycobacterial species share identical 16S rRNA sequences. While the *rpoB* gene is usually employed in the identification of RGM, the *hsp65* gene is used for the SGM.

The region of the 16S rRNA usually selected for sequencing analyses is a fragment that includes the first 500 bp of the 5' extreme, which are considered more variable, although in certain cases the whole gene may be analysed (Tortoli, 2010). In the case of the *rpoB*, a fragment of 764 bp located towards the 3' end of the gene is used (Adekambi *et al.*, 2006). For the *hsp65* gene, two different fragments are usually analysed; the first one, commonly referred to as

the "Telenti fragment", corresponds to 439 bp in the 5' region and is used to analyse SGM in general (Telenti *et al.*, 1993), whereas the second targets most of the remaining part of the gene (1,059 kbp) and is used for the identification of *M. avium* subspecies as well as *M. intracellulare* (Turenne *et al.*, 2006).

In addition to the previous genetic targets, other genes can also be used for the identification of NTM species, such as *recA* (Blackwood *et al.*, 2000), *sodA* (Zolg & Philippi-Schulz, 1994) or *gyrB* (Kasai *et al.*, 2000). These targets have also been used for the description of novel NTM species (Adekambi *et al.*, 2006; Guerin-Faublee *et al.*, 2013; Konjek *et al.*, 2016). The 16S–23S rRNA internal transcribed spacer (ITS) is an important genetic target that warrants further mention since it is highly polymorphic among mycobacteria and has been used to separate closely related SGM species, such as members of the MAC (Roth *et al.*, 1998). Furthermore, most of the currently accepted MAC members were identified based on their different ITS sequences (Tortoli *et al.*, 2004; Murcia *et al.*, 2006; van Ingen *et al.*, 2009).

Once the genetic targets are amplified and sequenced, the output sequence files need to be processed and compared against a database of annotated sequences in order to obtain an adequate identification. Manually curated databases for the 16S rRNA (RIDOM) and *hsp65* genes of many of the known mycobacterial species have been developed (Harmsen *et al.*, 2003; Dai *et al.*, 2011), but they have been discontinued and are no longer accessible. Therefore, microbiological laboratories may create their own databases or access non-curated databases (Tortoli, 2010), such as the international nucleotide sequence database collaboration, which includes the DNA Data Bank of Japan, the European Bioinformatics Institute and the National Centre for Biotechnology Information (NCBI) (Tortoli, 2010). The latter databases are updated daily and may contain a large amount of erroneous sequences, making them difficult to use in standardised procedures.

Identification is then based on the similarities between the sequenced isolates and the sequences deposited in the database. At least 99% identity has been suggested as a threshold for an identification to the species level using the 16S rRNA and *hsp65* genes (Tortoli, 2010; Dai *et al.*, 2011), although a lower percentage (e.g. 98%) may be accepted for complementary targets, such as the *rpoB* gene (Adekambi *et al.*, 2003). However, due to the increased number of mycobacterial species without standing nomenclature, the high similarity between certain mycobacterial species and the limited number of high quality sequences for less prevalent NTM species, the identification process is not as

straightforward as it could initially seem and requires experienced personnel acquainted with the extensive taxonomy of the genus.

1.5.3. **Molecular characterisation of MTBC members**

Despite the availability of different molecular tools for the identification of mycobacteria, identification and characterisation of MTBC members has traditionally been a complex endeavour due to their similar morphology (with the exception of *M. canettii*) or their highly similar biochemical properties. In addition, their clonal genomes have limited the use of more refined genetic methods, such as Sanger sequencing, since the targets used for identification are usually identical between MTBC members.

Due to these difficulties, the field has relied on the development of highly specific molecular methods. Due to the great public health importance of *M. tuberculosis*, the vast majority of strain characterisation methods were developed and extrapolated from this species to other MTBC members. Genetic characterisation of MTBC variants and strains has been critical to unravel their geographic distribution and evolution, as well as to understand the complex epidemiology of the group. In addition, many of the new variants have been described and characterised only recently, highlighting the limitations that the field has encountered for decades.

1.5.3.1. **First genotyping methods of MTBC strains**

In the context of bTB, the first genetic methods for the characterisation of animal-adapted MTBC members were developed during the 1980s and 1990s, and therefore most studies targeted *M. bovis*. These methods were based on the processing and analysis of whole genome DNA, such as restriction endonuclease analysis (REA) and Pulse-field gel electrophoresis (PFGE) (Gormley *et al.*, 2014).

Restriction endonuclease analysis consists in the digestion of pure chromosomal DNA with restriction enzymes and the visualisation of the obtained patterns through agarose gel electrophoresis. These patterns are then compared between strains and similarities are used to establish epidemiological relationships. This technique was initially useful to identify common sources of *M. bovis* infection from different hosts, such as cattle, possums, or cats (Collins *et al.*, 1994), and was used to establish new variants of tuberculous bacilli, such as *M. pinnipedii* in seals (Cousins *et al.*, 1993; Thompson *et al.*, 1993). However, its implementation has not been continued due to the laboriousness of the

technique, as well as to difficulties in interpreting the obtained patterns due to the presence of a large number of fragments (Gormley *et al.*, 2014).

Pulse field gel electrophoresis was implemented as a means to increase the resolution of the large number of restriction fragments generated by REA (Feizabadi *et al.*, 1996). It has been used to differentiate between *M. bovis* strains from different regions, as well as infections with more than one strain of *M. bovis* in cattle. Furthermore, PFGE pin-pointed the divergence of isolates of tuberculous bacilli from seals and cattle (Feizabadi *et al.*, 1996). Despite its utility, PFGE has not been routinely implemented due to its laboriousness and the development of improved molecular methods.

1.5.3.2. **Partial genome genotyping methods**

The development of DNA probing and genetic amplification techniques has led to the appearance of a wide range of techniques that are based on the analysis of different genomic regions. These methods were much easier to interpret than previous ones and therefore became widely used around the world. Their higher resolution when compared to previous methods greatly improved our knowledge on the genetic structure of the MTBC and its epidemiology (Gormley *et al.*, 2014).

Restriction fragment length polymorphisms (RFLPs)

In a similar manner to REA, RFLP is based on the digestion of chromosomal DNA with restriction endonucleases (van Embden *et al.*, 1993). The selection of the genetic target is crucial for the performance of the technique, since it determines the types of enzymes and probes used as well as the level of divergence between isolates (Durr *et al.*, 2000). Repetitive elements are usually the target of choice for the development of DNA probes due to their higher polymorphic potential (Kremer *et al.*, 1999). The most common repetitive elements used in RFLP are the IS*6110* transposon, the PGRS and the DR region (Hermans *et al.*, 1991; van Soolingen *et al.*, 1991; Ross *et al.*, 1992). After the DNA is fragmented, it is separated using gel electrophoresis and is transferred to a nitro-cellulose or nylon membrane through Southern blotting. The transferred DNA is then labelled using a specific DNA probe targeting the genetic target.

The IS*6110* has been commonly used in the study of *M. tuberculosis* in human TB, since this element is present in a large number of copies and therefore shows a high discriminatory power (Kremer *et al.*, 1999). In the case of *M. bovis* typing, the resolution achieved by this technique is usually limited due to the fact that most isolates tested contain a single copy of the IS*6110* element (Perumaalla *et*

al., 1996; Romano *et al.*, 1996). As a result, when studied through RFLP, these isolates frequently show only one type of banding pattern and, therefore, relatedness cannot be assessed (Skuce *et al.*, 1996). Nevertheless, exceptions to this situation have been described in literature; for example, studies in Spain have described isolates with a higher number of copies (2-13 copies) and patterns (Liébana *et al.*, 1997; Aranaz *et al.*, 1998). In addition, other animal-adapted members, such as *M. caprae* or *M. pinnipedii* are also known to carry more than one copy of the IS*6110* element and present characteristic RFLP patterns (Cousins *et al.*, 1993; Aranaz *et al.*, 1999; Zumarraga *et al.*, 1999a). Due to the limitations in the discriminatory power of IS*6110* when investigating strains with low copy numbers, this target is not routinely used for bTB epidemiological investigations.

Another MTBC-specific mobile element, the IS*1081* transposon, has also been targeted for RFLP analysis. Although this element is usually present in six copies in the genome of several MTBC isolates, including *M. bovis*, it shows very limited discrimination and is therefore not used in genotyping studies (Aranaz *et al.*, 1998; Kremer *et al.*, 1999). However, IS*1081*-RFLP has been shown to be a reliable method to differentiate between *M. bovis* and *M. bovis* BCG (van Soolingen *et al.*, 1992).

As previously mentioned, PGRS are abundant in MTBC genomes and have been shown to produce a high number of RFLP types in *M. bovis* isolates with low IS*6110* copy numbers (Cousins *et al.*, 1998). As a result, this target has frequently been preferred in RFLP analyses of *M. bovis* (Aranaz *et al.*, 1998; Cousins *et al.*, 1998; Costello *et al.*, 1999). However, in certain geographic regions, such as Northern Ireland or Spain, as well as in *M. caprae* infections, the variation between PGRS types has been shown to be similar to that observed for IS*6110* (Skuce *et al.*, 1994; Aranaz *et al.*, 1998). Furthermore, the complex fingerprints generated through PGRS-RFLP make this method difficult to interpret, especially at fragment sizes below 2.0 kb, in which resolution is reduced (Skuce *et al.*, 1996; Cousins *et al.*, 1998).

The high polymorphism of the DR region has also made it an interesting target for MTBC isolate typing through RFLP (Kremer *et al.*, 1999). Direct repeat RFLP has a higher discriminatory power than IS*6110*-RFLP in *M. bovis* strains with a single copy of the IS*6110*-element or in *M. caprae* isolates (Perumaalla *et al.*, 1996; Romano *et al.*, 1996; Aranaz *et al.*, 1998). However, this target usually shows a reduced resolution when compared to PGRS due to the limited

distribution of the DR region within the genome (Cousins *et al.*, 1998; Costello *et al.*, 1999).

Despite the apparent advantages of DR and PGRS over IS*6110* in RFLP typing, the combination of these targets achieves a higher discrimination than when used alone (Skuce *et al.*, 1996). However, these methods are highly laborious, time consuming and require large amounts of genomic DNA, usually one or two µg (van Embden *et al.*, 1993). Furthermore, data interpretation and standardisation can be complex when a high number of fragments or types are analysed, hampering interlaboratory comparisons.

Direct variable repeat spacer oligonucleotide typing (DVR-spoligotyping)

Spoligotyping is a PCR-based method in which the DR region is amplified using a biotinylated specific flanking primer. The amplicons are then hybridised to a membrane containing specific oligonucleotides for the different spacer sequences. Membranes are incubated with streptavidin-peroxidase conjugate and developed on an X-ray film using a chemiluminescent reagent (Kamerbeek *et al.*, 1997). In addition to this membrane-based platform, DVR-spoligotyping has also been adapted to the microbead Luminex system (Cowan *et al.*, 2004). Presence or absence of the different spacers in the membrane is then recorded in binary format and used to generate a specific pattern or spoligotype. Since the evolution of the DR region is considered unidirectional and to mainly occur as a result of subsequent deletion events (Romero *et al.*, 2008), spoligotype patterns showing the same or cumulative deletions are usually considered related.

Several sets of spacer oligonucleotides have been described for spoligotyping, but the most commonly used set contains 43 spacers (first generation membrane), which were originally described in *M. tuberculosis* and *M. bovis* BCG (Kamerbeek *et al.*, 1997). Not all of the spacers in the DR region are useful for genotyping of MTBC variants; for example, spacers located downstream of the IS*6110* element, which is usually located between genomic spacers 34 and 35 (membrane spacers 24 – 25), are frequently shared between *M. bovis* isolates and therefore offer a low discriminatory capacity (Caimi *et al.*, 2001). A second generation membrane containing additional spacers was developed in order to increase the discriminatory power of spoligotyping in other MTBC members (van Embden *et al.*, 2000; van der Zanden *et al.*, 2002). While the new spacers achieved a higher discrimination for *M. bovis, M. microti* and *M. canettii* strains (van der Zanden *et al.*, 2002; Javed *et al.*, 2007), it did not

improve the differentiation of *M. caprae* strains, which appear to have a limited spoligotype diversity.

Spoligotyping also allows the identification of MTBC variants, since the absence of a combined set of spacers is highly specific for certain MTBC members. For example, spacers 3, 9, 16 and 39–43 are typically absent in *M. bovis* (van der Zanden *et al.*, 2002), whereas the deletion of spacers 1, 3–16, 28 and 39–43 are usually features of *M. caprae* (Aranaz *et al.*, 1999; Rodriguez *et al.*, 2011).

Spoligotyping offers a reduced discriminatory power when compared to IS*6110* and PGRS RFLP in high IS*6110* copy strains, such as *M. tuberculosis* strains. On the other hand, when a low number of IS*6110* copies is present, such as in many *M. bovis* strains, spoligotyping has been shown to have a higher discriminatory power than IS*6110* RFLP (Aranaz *et al.*, 1996a; Zumarraga *et al.*, 1999b). As a result, spoligotyping is usually preferred as the reference genotyping method in bTB. In addition, spoligotyping is a relatively simple and rapid technique, which requires a reduced amount of DNA in comparison to RFLP, significantly reducing the amount of culture and incubation times needed (Kamerbeek *et al.*, 1997; Hewinson *et al.*, 2006). Finally, data interpretation is very simple, although it greatly depends on the quality of the membrane, and can easily be stored and exchanged between laboratories (Kremer *et al.*, 1999). This has allowed for the creation of a global spoligotype database for *M. bovis* (www.mbovis.org), which at the time this thesis was written holds more than 2,200 different spoligotypes.

Spoligotyping has been used for decades in epidemiological studies of TB in cattle and other animal species (Aranaz *et al.*, 1998; Gortazar *et al.*, 2005), and has been used to point out possible links between infected cattle and wildlife reservoirs (Romero *et al.*, 2008). In addition, the generalised application of spoligotyping as a genotyping method for *M. bovis* greatly improved our knowledge regarding the overall population structure of this pathogen in different regions of the world. Importantly, the genetic diversity of *M. bovis* has been usually associated with the diversity of spoligotype patterns observed in a geographic region and has been seen to vary significantly in certain countries. For example, the spoligotype diversity in the UK and the Republic of Ireland (RoI) is reduced when compared to other countries, such as Spain or France (Costello *et al.*, 1999; Haddad *et al.*, 2001; Smith *et al.*, 2003; Rodriguez-Campos *et al.*, 2013).

Many countries contain a dominant spoligotype or share spoligotypes with other countries (Hewinson *et al.*, 2006), which is useful for epidemiological studies and can be used for international source validation, such as through transboundary movements of animals across trade routes between countries. For example, Spain has three main spoligotypes (SB0121, SB0134 and SB0339) and shares its most frequent spoligotype (SB0121) with Portugal (Duarte *et al.*, 2008; Rodriguez *et al.*, 2010). In addition, Tunisia and France, which have strong commercial ties, share their three most frequent spoligotypes: SB0120, SB0121 and SB0134 (Djemal *et al.*, 2017). Spoligotype patterns can also be geographically localised within countries to certain regions, indicating a local-scale distribution and the existence of shared sources of infection (Smith *et al.*, 2003; Rodriguez *et al.*, 2010; Egbe *et al.*, 2017). Certain spoligotype patterns have also been associated with domestic or wildlife animals and could be used in epidemiological investigations to suggest possible origins of infection (Rodriguez *et al.*, 2010). Furthermore, strains belonging to specific spoligotypes could also be associated with the development and severity of tuberculous lesions in cattle (Gomez-Buendia *et al.*, 2021).

Further analyses of dominant spoligotype patterns in different regions led to the identification of clonal populations, named clonal complexes (Smith, 2012), with shared chromosomal deletions (RDs), SNPs and spoligotype patterns. Four major *M. bovis* clonal complexes have been described to date: African 1 (Af1) (Muller *et al.*, 2009), African 2 (Af2) (Berg *et al.*, 2011), European 1 (Eu1) and European 2 (Eu2) (Smith *et al.*, 2011; Rodriguez-Campos *et al.*, 2012b).

The Af1 complex is mainly present in West-Central African countries, such as Nigeria, Mali, Chad or Cameroon, and is characterised by the RDAf1 and the spacer 30 deletions (Muller *et al.*, 2009). In turn, the Af2 complex, which is dominant in eastern Africa, has the RDAf2 and spacer 3 to 7 deletions as molecular features (Berg *et al.*, 2011). The Eu1 complex is a clonal complex that is dominant in the UK and the RoI, but is also globally distributed, being found in South Africa, the USA, Central America and several South American countries, such as Argentina or Chile (Smith *et al.*, 2011). European 1 is characterised by an 806 bp long deletion within the *treY* gene, named RDEu1 or RD17, and a deletion in spacer 11. The Eu2 complex is a clonal population of *M. bovis* that is mostly present in the Iberian Peninsula (Rodriguez-Campos *et al.*, 2012b), but is also frequent in other countries like Brazil (Rodrigues *et al.*, 2021), and is characterised by a deletion of spacer 21 and a SNP in the *guaA* gene. All strains from the described clonal complexes share specific spoligotype signatures, which correspond to the spacers deleted in the pattern

of their MRCA and from which the rest of spoligotypes observed in the complex were derived from (Smith, 2012). For example, the spoligotype signature of Af1 is SB0944, whereas the spoligotype signature of Eu2 is SB0121.

Additional spoligotype patterns outside of the major clonal complexes have been identified in certain regions, suggesting that novel clones or minor clonal complexes may exist (Smith, 2012). For example, up to 40% of the strains isolated in Mali showed a distinct spoligotype signature to that expected for the Af1 complex (Muller *et al.*, 2009). In addition, strains from Madagascar lack a set of different spacers that suggest the presence of a different clonal complex, provisionally named African 3 (Af3) in that region (Rasolofo Razanamparany *et al.*, 2006). Another group of *M. bovis* strains containing a BCG-like spoligotype patterns, such as SB0134 or SB0120, have been frequently described in France, Italy, as well as in Iran, Algeria, Ethiopia and Zambia (Haddad *et al.*, 2001; Boniotti *et al.*, 2009; Smith, 2012; Almaw *et al.*, 2021). Although this group of strains have traditionally been called BCG-like strains, they have recently been proposed as the Eu3 clonal complex (Rodriguez-Campos *et al.*, 2014; Branger *et al.*, 2020; Almaw *et al.*, 2021).

Regardless of its advantages and extended use, spoligotyping presents a series of limitations that could affect its performance under different scenarios. The overall discriminatory power of spoligotyping can be significantly affected by the genetic diversity of MTBC strains in the region under study (Gormley *et al.*, 2014). Due to this variation in diversity, spoligotyping may not be a suitable epidemiological tool in all settings. Spoligotyping presents a better resolution in countries with a higher genetic diversity, such as Spain (Rodriguez *et al.*, 2010), Portugal (Duarte *et al.*, 2008), France (Haddad *et al.*, 2001) or Italy (Boniotti *et al.*, 2009) in mainland Europe, Cameroon in Africa (Egbe *et al.*, 2017), Mexico in North America or Brazil in South America (Gutierrez Reyes *et al.*, 2012; Zumarraga *et al.*, 2013). On the other hand, several countries around the world present a relatively low diversity of spoligotypes, such as the UK and the RoI in Europe (Costello *et al.*, 1999; Smith *et al.*, 2006a) or Australia in Oceania (Aranaz *et al.*, 1998). In these cases, spoligotyping may not achieve the best resolution possible. In addition, due to the recombinative nature of deletions in the DR region, similar DVR deletions can also appear through convergent evolution (i.e. homoplasies), leading to the same spoligotype pattern on phylogenetically unrelated strains (Smith *et al.*, 2006a; Comas *et al.*, 2009). Furthermore, the obtained spoligotype is a result of the cumulative amplification of spacers and therefore, it is not suitable to identify mixed infections (Romero *et al.*, 2008). In

order to overcome some of these limitations, additional ancillary molecular methods, such as MIRU-VNTR, can be used.

Mycobacterial interspersed repetitive units variable number of tandem repeats (MIRU-VNTR)

Contrarily to spoligotyping, which only targets the DR region, this method is based on PCR amplification of a set of MIRU-VNTR loci using specific flanking primers and the subsequent evaluation of the length of the amplicons through capillary or gel electrophoresis in order to assess the number of repeats. Relatedness is then established according to the number of copies in each set of MIRUs.

Alleles with a high allelic diversity are usually selected for MIRU-VNTR analyses. However, not all of them are appropriate for epidemiological studies for several reasons, such as an excessive amount of variation, difficulties in reproducibility or interpretation of results, or amplification failure through PCR (Supply *et al.*, 2006). Several sets of MIRU-VNTR loci have been studied in *M. tuberculosis* (Supply *et al.*, 2000; Le Fleche *et al.*, 2002; Savine *et al.*, 2002; Kremer *et al.*, 2005; Kam *et al.*, 2006), and two standardised sets of 12 and 24 MIRUs have been selected and are commonly used in epidemiological studies (Supply *et al.*, 2006). In addition, a small and an intermediate set of nine and 15 MIRUs with high allelic diversity have been shown to also achieve high resolution (Supply *et al.*, 2006).

Several panels of VNTR loci have been assessed for the typing of *M. bovis* strains around the world with very good results (Roring *et al.*, 2002; Skuce *et al.*, 2002; Allix *et al.*, 2006; Boniotti *et al.*, 2009; Duarte *et al.*, 2010). Although a small number of studies have specifically dealt with the use of MIRU-VNTR for *M. caprae* typing (Boniotti *et al.*, 2009; Rodriguez *et al.*, 2011; Lamine-Khemiri *et al.*, 2014), this technique appears to be a good alternative with high discriminatory capacity (Prodinger *et al.*, 2005).

A minimum standardised set of six VNTR loci (QUB 3232, ETR-A, ETR-B, ETR-D, QUB 11a, QUB 11b, QUB 26 and ETR-D) was defined by a European consortia and is currently proposed by the European Union Reference Laboratory (EU-RL) for bTB (VENoMYC, 2006). However, Member States use slightly different sets, comprising 6-13 loci. This is a result of the apparent discordance between the discriminatory capacity of several VNTR loci according to the region being studied (Hauer *et al.*, 2016). For instance, QUB 11a has frequently been shown to have a high discriminatory power in several countries, such as Portugal or the

RoI (Duarte *et al.*, 2010; McLernon *et al.*, 2010), but in other regions of the world, such as Italy or Tunisia, discrimination has been shown to be moderate to low, respectively (Boniotti *et al.*, 2009; Lamine-Khemiri *et al.*, 2014). This discordance could be related to the differences in the genetic diversity of the bacterial population, which in turn could be related to the genotypes present in a given region (Rodriguez-Campos *et al.*, 2013; Hauer *et al.*, 2016). Furthermore, some of the most variable loci, such as QUB 3232, usually show amplification problems and are therefore not frequently used (Martinez *et al.*, 2008; Boniotti *et al.*, 2009; Bolado-Martinez *et al.*, 2015).

MIRU-VNTR usually shows a high discriminatory power when compared to other traditionally used typing techniques. The overall discriminatory capacity of spoligotyping is reduced when compared to MIRU-VNTR typing (Roring *et al.*, 2002; Skuce *et al.*, 2002; Duarte *et al.*, 2010). In general, there is a good agreement between spoligotyping and MIRU-VNTR, with most of the VNTR patterns being confined to a certain spoligotype (McLernon *et al.*, 2010). This makes MIRU-VNTR an interesting method to study highly prevalent genotypes, such as the SB0121 in Spain (Rodriguez-Campos *et al.*, 2013). Nevertheless, the detection of identical VNTR genotypes in different spoligotypes has been recorded (Duarte *et al.*, 2010; McLernon *et al.*, 2010). This is frequently found in spoligotypes with only one spacer difference and could be due to microevolution of the DR region (Rodriguez *et al.*, 2011).

The discriminatory capacity of MIRU-VNTR has also been shown to be higher than IS*6110* RFLP typing (Allix *et al.*, 2006), but is reduced when compared to RFLP typing using a combination of probes (IS*6110*, PGRS and RD) (McLernon *et al.*, 2010). Nevertheless, MIRU-VNTR is usually preferred over RFLP as a typing method due to its speed and simplicity. Furthermore, MIRU-VNTR typing can easily identify mixed infections or microevolution within an animal by the detection of more than one allele for one or several VNTR loci (Allix *et al.*, 2006; Romero *et al.*, 2008).

The application of MIRU-VNTR to population surveys allowed for a better resolution of clonal strains of *M. bovis* circulating in different regions. For example, the use of ETR loci in Western Africa revealed that the most common spoligotype (SB0944) could be differentiated into 16 genotypes, some of which were more frequent in certain countries than others (Muller *et al.*, 2009). Furthermore, a genotype similarity at a regional level can be used as an indicator of the amount interactions between regions; in Cameroon, a high similarity in genotypes in the north-western region suggests that cattle

movements and interactions occur more frequently between herds of that region than with others (Muller *et al.*, 2009; Egbe *et al.*, 2017).

In addition to its use in population surveys, the increased resolution of MIRU-VNTR has been extremely useful when assessing possible epidemiological links between herds (Allix *et al.*, 2006) and has highlighted the existence of a higher genetic diversity in farms containing isolates of the same spoligotype (Rodriguez-Campos *et al.*, 2013), suggesting the circulation of more than one strain of *M. bovis*. In addition, the identification of similar genotypes in isolates from cattle and wildlife species, such as wild boar or red deer, have also been important milestones in reinforcing the importance of wildlife reservoirs in the epidemiology of bTB in certain regions (Duarte *et al.*, 2010).

Region of difference (RD) typing

Regions of difference pertain to larger fragments of the mycobacterial genome when compared to other genetic markers. Although initial studies were founded on conventional PCR, other protocols based on real-time PCR technologies, either through melting-curve analysis or molecular probing, have been developed (Pinsky & Banaei, 2008; Halse *et al.*, 2011; Reddington *et al.*, 2011a; Reddington *et al.*, 2011b; Reddington *et al.*, 2012).

The discriminatory power of the major RDs is relatively low since they are shared among many different variants and, therefore, they are mostly used for phylogenetic purposes or for lineage/sublineage classification. Nevertheless, several specific RD markers have been found in different MTBC members, making these regions a suitable method for variant identification (Parsons *et al.*, 2002; Huard *et al.*, 2003; Huard *et al.*, 2006). In addition, RDs have also been used to define specific *M. bovis* clonal populations or complexes around the world, such as the RDEu1 of the European 1 clonal complex (Muller *et al.*, 2009).

In the context of bTB, RD typing can be very useful, especially in geographic regions where different MTBC members coexist, such as *M. bovis* or *M. caprae* in the Iberian Peninsula. In these cases, both MTBC members can be differentiated according to the absence or presence of the RD4 locus, respectively (Huard *et al.*, 2006). Furthermore, variations within RDs have been identified for *M. caprae* and allow for further differentiation of *M. caprae* isolates into three subtypes: Allgäu, Karwendel and Lechtal (Domogalla *et al.*, 2013). Additional variation in the RD4 was observed in one isolate of *M. caprae* from Spain (Rodriguez *et al.*, 2011), suggesting that other strains may exist in other geographical regions.

Regions of difference not only allow for variant identification in culture isolates, but can also be used directly in samples. For example, RD4-specific qPCRs have been designed for the detection of *M. bovis* in environmental samples in the UK and France (Sweeney *et al.*, 2007; Barbier *et al.*, 2016).

One intrinsic limitation of RD typing is the potential for independent strain-specific deletion events that can lead to certain RDs to overlap between different strains or variants (Huard *et al.*, 2006). For example, *M. microti*, the dassie bacillus and *M. bovis* BCG have differing, but overlapping, deletions in the RD1 region. As a result, possible variations in RD profiles should be taken into account when designing RD typing primers and algorithms.

Single nucleotide polymorphism (SNP) typing

Before the introduction of WGS, most of the SNPs were identified as a result of investigating differential phenotypes between MTBC isolates. The majority of the initial SNPs used in genotyping studies were, therefore, found in genes related to antimicrobial resistance, such as the *pncA* or *gyrB* genes (Scorpio & Zhang, 1996).

The comparison of available reference genomes allowed the identification of different SNPs that are shared or unique for specific MTBC members. The identification of variant-specific SNPs led to the development of the Hain MTBC genotype assay, which is frequently used for variant identification in human TB (Richter *et al.*, 2003; Loiseau *et al.*, 2019).

Sequencing of an increasing number of MTBC isolate collections has revealed different sets of phylogenetically important (informative) SNPs, ranging between 13-93 different SNPs (Filliol *et al.*, 2006; Comas *et al.*, 2009; Homolka *et al.*, 2012). Furthermore, a large-scale whole genome sequencing study identified 62 robust SNP markers for the classification of MTBC isolates into their respective lineages through a barcode system, which could be used for the development of genotyping tests (Coll *et al.*, 2014).

Although to a limited extent, SNP typing has also been used in the context of bTB to define specific sublineages of *M.* bovis or to differentiate between certain MTBC members. For example, a SNP in the *guaA* gene has been proposed as a specific marker of the Eu2 complex, prevalent in western continental Europe (Rodriguez-Campos *et al.*, 2012b), whereas a SNP in the *lepA* gene has been used to differentiate between *M. caprae* and the rest of the MTBC (Reddington *et al.*, 2011b). Despite the utility of SNP profiling, genotyping of MTBC members and lineages through a reduced number of sites

entails certain risks. For example, the Hain test misclassified *M. orygis* isolates as *M. africanum* due to the presence of an homoplasic SNP in the *gyrB* gene (van Ingen *et al.*, 2012). Nevertheless, SNPs have been shown to have the lowest homoplasy index when compared to other molecular markers and, therefore, are usually preferred in evolutionary and epidemiology studies (Comas *et al.*, 2009).

Traditionally, molecular analysis of SNPs in MTBC has been carried through a variety of methods, which usually involved DNA amplification followed by sequencing or restriction enzyme analysis (Stucki *et al.*, 2012; Gormley *et al.*, 2014). This could be especially laborious and time-intensive when a large number of SNPs are evaluated. Simpler platforms, such as LPAs, have been used for MTBC isolate genotyping (Richter *et al.*, 2003). However, LPAs are based on a limited number of genes, which are usually involved in antimicrobial resistance (Richter *et al.*, 2003; Chen *et al.*, 2019). On the other hand, DNA microarray technologies have also been used for the characterisation of MTBC members based on differential SNP profiles (Srilohasin *et al.*, 2014), but their routine use has been limited. Due to the limitations in the capacity of available molecular techniques, the resolution of SNP genotyping methods is only useful in large scale studies.

1.5.3.3. Whole genome sequencing (WGS)

Whole genome sequencing encompasses a wide range of techniques, mainly next generation sequencing (NGS) platforms, and bioinformatics analyses by which the complete genetic sequence of an organism is determined at the nucleotide level. From this foundation, additional analyses can be carried out to assess the phylogenetic relatedness between organisms or to discover the functional properties of the information encoded within the genomes.

As monomorphic microorganisms with little to no genetic recombination nor LGT, the MTBC is remarkably suited for WGS analyses with different purposes, such as diagnostics, evolutionary studies, surveillance and epidemiological investigations or drug treatment outcome predictions based on antimicrobial resistance signatures (Meehan *et al.*, 2019; Perea *et al.*, 2021). The availability of whole genome data from different MTBC members not only has revolutionised our understanding of the origins and evolution of the MTBC, but has also achieved the highest resolution with respect to other molecular characterisation methods.

Sequencing platforms and workflows for WGS analyses of MTBC members

According to the length of the DNA sequence (read) that is obtained from the sequencing process, current WGS technologies can be divided into short- and long-read based platforms (Metzker, 2010). Short-read technologies are very accurate methods that generate millions of reads of up to a couple hundred nucleotides. Although different platforms exist, such as MGI DNBSEQ and Illumina, the latter is nowadays the most frequently used platform and is able to generate reads between 150 and 300 bp. This technology requires an initial library preparation step in which the DNA is processed to be compatible with the sequencing technology used. These libraries are usually generated by DNA fragmentation, which can be obtained through different methods such as transposon-based fragmentation. Then, specific adapters are added to the extremes of the obtained fragments so that these may bind to the sequencing flow cell in order to be amplified and sequenced. Long-read sequencers, also called third-generation sequencers, are able to routinely produce reads longer than 10 kb and have achieved maximum lengths of 2.3 Mb (Amarasinghe *et al.*, 2020). There are currently two dominant long-read sequencing platforms: Pacific Biosciences (PacBio) single-molecule real-time sequencers and Oxford Nanopore Technologies sequencers. These platforms do not usually require a preliminary amplification procedure and can be used directly on a DNA sample. However, a common trade-off of long-read sequencing platforms is their reduced accuracy when compared with short read platforms (Amarasinghe *et al.*, 2020), which entails an additional error correction phase in which long-read alignments are usually combined with short-reads (hybrid methods) or corrected using different informatics algorithms (non-hybrid methods).

Despite the technical differences between the sequencing platforms, the process by which the DNA sequence is reconstructed is similar; the nucleotide that is read at each moment during the sequencing process is established either by differences in fluorescent signals (Illumina and PacBio) or electrical potentials (Oxford Nanopore). The quality of this signal is also recorded by the sequencer and stored along with the nucleotide sequence as a raw read file, usually in FASTQ (Illumina and Oxford Nanopore) or FAST5 (Oxford Nanopore) format. Raw reads are usually post-processed using specific softwares according to the quality and length of the reads, and the adapters used during library preparation are removed.

In general, two types of analyses can be considered the working ground of posterior analyses: reference mapping (genome resequencing) and *de novo* assembly. Genome resequencing consists of the alignment of the reads output by sequencing platforms, usually short-read platforms, to a reference genome using one or several mapping softwares. Contrarily, *de novo* assembly consists of the alignment of reads with each other in order to reconstruct the chromosome based on the similarity observed between reads. This reconstruction process is progressive and leads to the generation of large sequences of DNA called contigs, which are later reorganised into chromosomal scaffolds of larger size that will, if possible, be combined to generate a closed or complete chromosome.

The choice of method and platform to be used depends on the target organism, the type of downstream analyses that will be carried out and the accuracy required. Genome resequencing using short read sequencing is usually the assembly method of choice for most surveillance and epidemiological investigations due to its high throughput capacity and relatively low computing requirements in comparison to *de novo* assemblies. In contrast, *de novo* assemblies are better suited for structural and comparative genomic studies but are not usually employed in molecular epidemiology studies of MTBC members. This is mainly due to the fact that, until the advent of long read technologies, most *de novo* assemblies were based on short read technologies, which faced limitations in obtaining complete genomes due to the reduced capacity of current bioinformatics algorithms in resolving high repetitive regions or the appearance of misassemblies due to high GC content regions (Sohn & Nam, 2018), such as those coding for PE/PPE family proteins. Although long reads can circumvent these issues, their use in TB studies has only recently been introduced (Tafess *et al.*, 2020; Peker *et al.*, 2021).

Variant calling pipelines

From the perspective of molecular characterisation of tuberculous bacilli isolates, the most common WGS analysis is the detection of genetic variants through variant calling. This technique consists of different bioinformatics algorithms that are able to detect genetic variations on the reads or alignments when compared to a reference sequence. The most common variants are usually SNPs, but other variants, such as short insertions and deletions (indels) can be detected.

Variant calling based on short read genome resequencing is nowadays the method of choice for WGS studies in human and animal TB, and procedures in both settings follow similar workflows (Meehan *et al.*, 2019; Crispell *et al.*, 2020; Guimaraes & Zimpel, 2020).

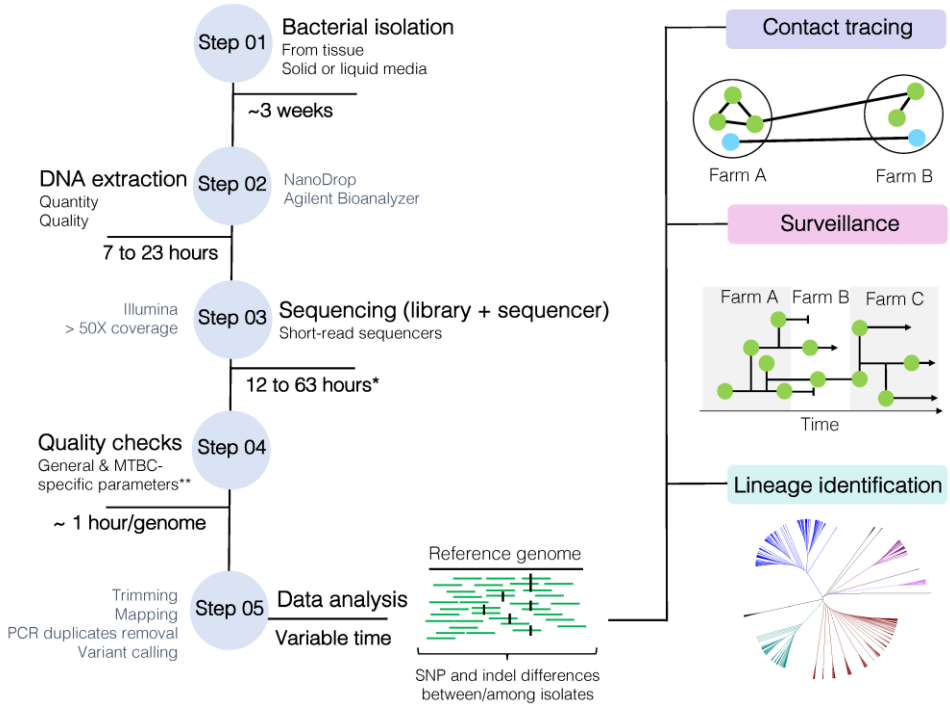


Figure 8. Whole-genome sequencing (WGS) workflow for MTBC variant calling. Image obtained from Guimaraes *et al.*, 2020, with permission from Microorganisms.

The first part of the process consists of obtaining pure MTBC cultures from samples, such as those obtained through bacteriological culture from animal tissues (figure 8, step 01). DNA is then extracted from these pure cultures (step 02), fragmented and processed into DNA libraries, for downstream sequencing on short-read platforms (step 03). Output reads are then pre-processed to assess both read quality and length using specific software (step 04), such as read-trimming tools, and these are aligned to a reference genome using a read mapping aligner, such as the Burrow-Wheelers Aligner (BWA) (Li & Durbin, 2009) (step 05). Mapping of reads to the reference genome (e.g. *M. bovis* AF2122/97) can occur with varying efficiency (mapping quality) based on the amount of identity between the read and the target sequence or the presence of secondary alignments. Variant calling softwares are then used to detect variants in the aligned reads and low quality variants are filtered out using

specific criteria, such as read depth, base quality or strand bias (Meehan *et al.*, 2019). In addition to quality filters, variant calling softwares usually include an option by which specific regions of the genome can be masked from the analysis. Masking usually involves regions that give rise to errors during variant calling, such as low complexity regions or repetitive sequences, which can present a lower accuracy as well as mapping issues. In the case of MTBC genomes, these include mobile genetic elements or *pe/ppe* genes, among others (Walter *et al.*, 2020).

Identified variants are frequently stored in a standardised format called variant call format (VCF) (Danecek *et al.*, 2011), which can then be processed to obtain different layers of information. The most common analyses involve the concatenation of SNPs from each sample into a multi-FASTA file, which can then be used to obtain SNP distance matrixes for cluster analyses or can be processed using phylogenetic reconstruction softwares to infer phylogenies for multiple purposes, such as contact tracing, surveillance or lineage identification (figure 8). Although, phylogenetic relationships between strains are usually represented in phylogenetic trees, they can also be represented using minimum spanning trees, which make data visualisation more straightforward, especially in epidemiological investigations (figure 9). Additional analyses include the annotation of VCF files in order to identify the regions of the genome that are affected by the variants and to predict their possible functional consequences, such as antibiotic resistance or virulence (Hauer *et al.*, 2019; Perea *et al.*, 2021).

The above mentioned steps can be carried out using a variety of bioinformatics tools. A selection of such tools is usually packed together and interconnected into what is termed a pipeline, so that output results from one tool are input into the next. Pipelines have the advantage of standardising the procedure for all of the samples under analysis and reducing the amount of manual processing required. Variant calling pipelines have been developed for many types of organisms, including the MTBC, and are frequently used nowadays in human TB around the world (Kohl *et al.*, 2018c; Bogaerts *et al.*, 2021). There has been an increased interest in the development of variant calling pipelines in the context of bTB, and several alternatives are available, such as BovTB, SNiPgenie or vSNP (Price-Carter *et al.*, 2018).

Core genome multi-locus sequence typing (cgMLST)

As previously mentioned, the majority of variant calling methods are based on the detection of SNPs. However, one of the main difficulties of integrating WGS genotyping based on SNP analysis is the difficulty in standardising and integrating the generated data into an easily accessible and expandable classification scheme (Kohl *et al.*, 2014). Core genome multi-locus sequence typing (cgMLST) has been proposed as an interesting alternative to SNP typing for molecular epidemiology of TB, which partly resolves the problems observed in SNP typing, such as difficulty in result standardisation.

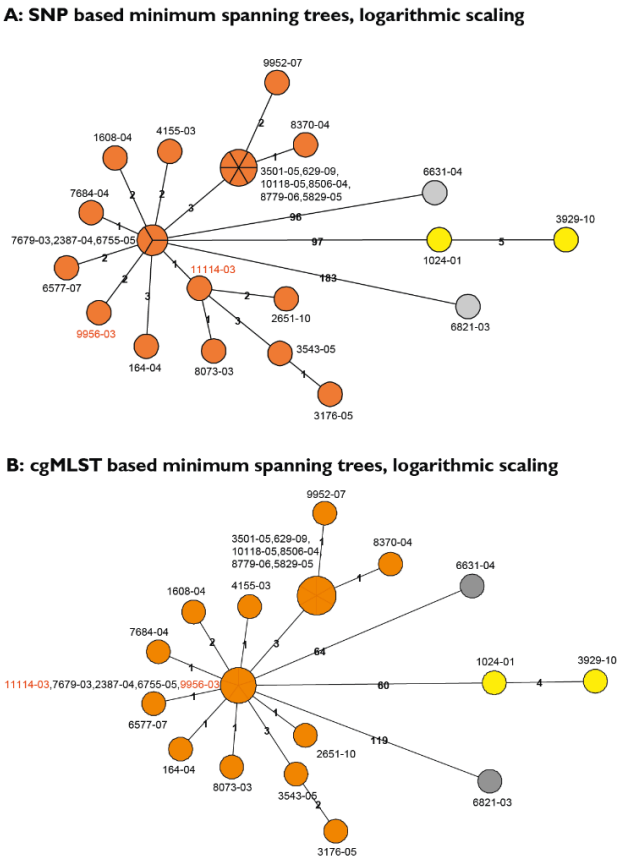


Figure 9. Maximum spanning trees obtained from a cluster (cluster 80) of *M. tuberculosis* isolates in a surveillance study in Hamburg, Germany, using A) SNP and B) cgMLST analyses. Adapted from Kohl et al., 2018a, with permission from EBioMedicine.

Similar to MLST, cgMLST uses gene alleles as the unit of variance, independent of the number of variants present. However, instead of only a few loci, cgMLST extends the analysis to thousands of genes through WGS. Gene lists used in cgMLST schemes have been selected from genome assemblies of different MTBC lineages according to specific properties; for example, the cgMLST proposed by Kohl *et al*. does not consider *pe/ppe* genes, genes with internal stop codons or partial overlaps with other genes (Kohl *et al.*, 2014). The different alleles found for each gene are given a specific nomenclature and the combination of these are summarised into sequence types, which can be used both to identify an isolate or to compare it against others. The resulting MLST is therefore not only simple to analyse and to standardise, but can also be easily exchanged between laboratories and stored in online databases for different laboratories to use.

Core genome MLST has been used in several studies for phylogenetic lineage classification of MTBC members as well as for delineating clinical outbreaks of *M. tuberculosis* (Kohl *et al.*, 2014; 2018a; Jones *et al.*, 2019). Overall, cgMLST achieves a high concordance to SNP genotyping, both through SNP barcoding or whole genome SNP analysis. Nevertheless, compared with SNP genotyping, cgMLST can present a reduction in resolution in certain cases (figure 9), which is related to the presence of phylogenetically informative SNPs outside of coding regions (i.e. in intergenic regions), in genes not classified as part of the core genome or for which quality criteria are not sufficient for inclusion in the cgMLST scheme (Kohl *et al.*, 2014). However, core genome lists can be easily expanded to include accessory genes that can be useful to further separate isolates (Kohl *et al.*, 2018c).

In silico characterisation tools

In addition to variant calling pipelines, other tools have been developed for specific analysis of MTBC members. Some of these tools mimic common molecular characterisation methods discussed in the previous section. For example, SpoTyping is a software that predicts the spoligotyping pattern based on raw read data, whereas RD analyser uses raw reads to detect the presence or absence of specific RDs (Faksri *et al.*, 2016; Xia *et al.*, 2016). In addition to specific tools, certain variant calling pipelines, such as vSNP, include specific modules that are able to detect the spoligotype of the analysed samples *in silico*.

In contrast, the inference of MIRU-VNTR profiles from WGS data has proven to be a challenge due to the difficulty to reconstruct tandem repeats using short read data. Nevertheless, several tools have been developed to infer the MIRU-VNTR profile of MTBC members using assembly data, such as the Comprehensive Analysis Server for the *Mycobacterium tuberculosis* complex or the MIRU-profiler (Iwai *et al.*, 2015; Rajwani *et al.*, 2018).

Despite their utility, discrepancies between *in silico* and *in vitro*-based techniques have been identified. This could be due to limitations in the WGS process, such as a low quality read data in the case of Spotyping or low quality assemblies in the case of MIRU-profiler (Xia *et al.*, 2016; Rajwani *et al.*, 2018). Alternatively, misclassifications can also take place in the *in vitro* techniques, such as microevolution events that affect the binding of PCR primers and probes used in *in vitro* techniques, or errors in the original gel or capillary-based fragment analysis, leading to false negative results which can be detected from sequence data (Rajwani *et al.*, 2018).

Whole genome sequencing and M. bovis epidemiology

The first comprehensive analyses of *M. bovis* genomes confirmed the clonal population structure of *M. bovis* into clonal complexes (Loiseau *et al.*, 2020; Zimpel *et al.*, 2020). However, the discovery of novel monophyletic groups of strains without clonal complex-specific markers indicates that the diversity of extant *M. bovis* strains is higher than previously expected (Loiseau *et al.*, 2020; Zimpel *et al.*, 2020). In a context in which clonal complex-specific markers may not be stable enough to grasp the overall diversity of *M. bovis* strains around the world, newer classification schemes have been proposed.

Zimpel *et al*. 2020 proposed a cluster-based classification in which a combination of signature SNPs and/or clonal complex specific markers are used to define *M. bovis* clades, in a similar manner to what was achieved with *M. tuberculosis* lineages (figure 10). Under this classification, *M. bovis* strains can be divided into at least four major clusters with a varying global distribution, defined as Lb1, Lb2, Lb3 and Lb4 (Zimpel *et al.*, 2020).

Some of these clusters agree with specific clonal complexes; for example, Lb2 and Lb4 correspond to Af1 and Eu1, respectively. However, while the cluster classification in Lb4 is specified by the presence of RDEu1, in the case of Lb2 it also includes a set of 133 signature SNPs. In contrast, other clusters may be composed from a mixture of SNPs and clonal complex markers; Lb1 contains Af2 strains with the RDAf2 region, as well as closely related strains with which

they share 19 signature SNPs. In the case of Lb3, which includes the Eu2 complex, a total of five additional SNPs were defined as cluster-specific.

Additional groups of isolates that did not fit into this classification (unknown clusters 1, 2 and 3)were identified in different parts of the world (Zimpel *et al.*, 2020). It is important to mention that the diversity of *M. bovis* has only been sampled systematically both from cattle and wildlife in only some countries, which implies that our knowledge of the global diversity of this pathogen may be biased towards culture collections built on limited sampling strategies. This is probably due to economic and/or ecological constraints, and indicates that more regions and hosts need to be sampled in order to detect specific markers with which to include the overall diversity of extant *M. bovis* strains (Zimpel *et al.*, 2020).

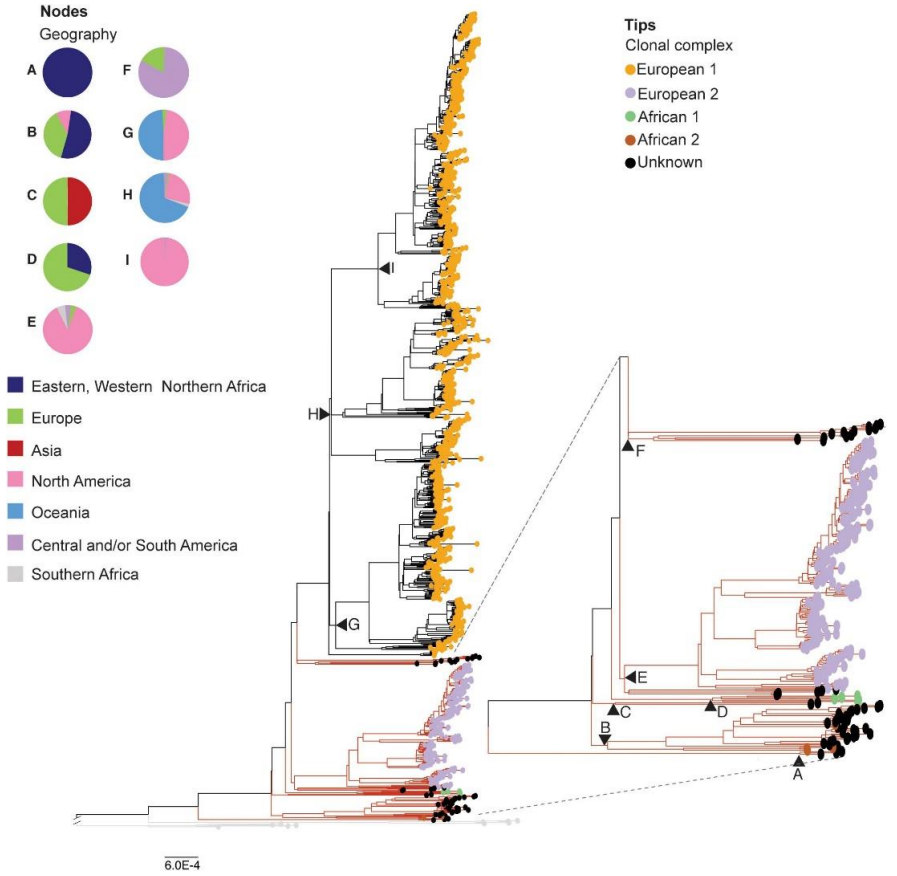


Figure 10. Phylogenetic reconstruction based on SNPs of *M. bovis* genomes (n = 1,201). Tree nodes are colored based on the genotype (clonal complexes), whereas pie charts represent the geographic origin of the genomic sequences used for analysis. Source: Zimpel et al., 2020. With permission from Frontiers in Microbiology.

The tree topology inferred by Loiseau *et al*. 2020 largely agrees with the abovementioned model, with classical clonal complexes also representing significant monophyletic groups within the phylogeny (Loiseau *et al.*, 2020). However, in this case, the detection of additional RDs as well as differences in the thresholds used to define the phylogenetic clusters have led to an alternative classification scheme which divides *M. bovis* strains into the classical clonal complexes as well as into eight unknown clades (unknown 1 – 8). Interestingly, under this classification, Lb clusters can be subdivided into the ones proposed by Loiseau *et al.*; for example, Lb1 can be subdivided into Af2 and unknown 2, whereas Lb3 can be subdivided into Eu2 and unknown 5. Recently, Zwyer *et al*., 2022 described a similar nomenclature system to the one proposed by Loiseau *et al.,* 2020, in which *M. bovis* was subdivided into eight sublineages (La1.1 to La1.8), some of which corresponded to previously known CCs (e.g. La1.8.1 = Eu1), whereas others were unknown (Zwyer *et al.*, 2021).

The implementation of novel modelling methods, such as Bayesian and molecular clock analyses, in the abovementioned studies has allowed for the first evolutionary predictions along with dating estimates for the spatial and temporal expansion of *M. bovis* around the world. However, due to the very limited WGS data available for *M. bovis* isolates from Africa, Continental Europe and, especially, Asia, as well paleozoological data of bovine and zoonotic TB, our ability to decipher the true geographic origin of this pathogen is limited (Loiseau *et al.*, 2020). Nevertheless, according to these phylogenies, isolates from northern, western and eastern Africa emerge from the oldest nodes, indicating that *M. bovis* may have originated from the African continent (Loiseau *et al.*, 2020; Zimpel *et al.*, 2020), in agreement with previous studies that also suggest this continent as the origin of the MTBC (Comas *et al.*, 2013; Supply *et al.*, 2013). Loiseau *et al*. 2020 further suggested that *M. bovis* may have originated in East Africa, since the most divergent *M. bovis* isolates are found in Malawi. Interestingly, these isolates are classified by both models as a monophyletic unknown cluster (unknown 1), and lack the p*ncA* H57D SNP that is responsible for the intrinsic resistance of *M. bovis* strains to pyrazinamide (Loiseau *et al.*, 2020). The fact that *M. bovis* genomes in South Africa are mostly clustered within Lb3 and Lb4, whereas the most ancient lineages Lb1 and Lb2 are restricted to other regions, could also suggest that *M. bovis* may have originated and diversified outside of the continent with subsequent reintroductions during time, such as during European colonisation of South Africa (Brites *et al.*, 2018; Loiseau *et al.*, 2020; Zimpel *et al.*, 2020).

Interestingly, one of the temporal models developed by Zimpel *et al*. 2020 established the possible origin of *M. bovis* at a maximum of 3,500 years before present, which agrees with the detection of an RD4-deleted genome in the remains of approximately 2,000-year-old humans from southern Siberia (Taylor *et al.*, 2007a). The absence of the RD17 marker in these remains indicates that the Lb4 cluster originated later on, in agreement with posterior dating estimates (338-1638 years BP). In contrast, Loiseau *et al*. 2020 modelled the origins of *M. bovis* between the third and twelfth century AD, which suggests a potential underestimation based on available archaeological findings and highlights the potential pitfalls of molecular dating analyses from ancient DNA (Loiseau *et al.*, 2020). Despite these discordances, both models proposed similar dating estimates for more recent introductions of *M. bovis* in different geographic regions, such as Eu2 (Lb3) and Eu1 (Lb4) in the Americas during the 16th century AD from the Iberian Peninsula and the UK, respectively, as well as Eu1 subclades in New Zealand in the 19th century.

Spoligotype patterns generally correlate well with the genetic lineages in the abovementioned studies, supporting the utility of this molecular technique as a preliminary characterisation method (Zimpel *et al.*, 2020). In addition, the identification of BCG-like spoligotypes among different clades strongly suggests that this classification of strains is not valid from the phylogenetic point of view and that caution should be taken when inferring relatedness between isolates from these spoligotypes (Loiseau *et al.*, 2020). Furthermore, due to the possibility of microevolution events and homoplasies, apparently distant strains by traditional characterisation methods can mask genetically related isolates. This has been demonstrated, for example, in a *M. caprae* isolate from the SB1908 spoligotype that presented only 9 SNP differences with isolates from SB0415 (Ciaravino *et al.*, 2021). Similarly, WGS revealed that phylogenies inferred from *M. bovis* isolates with the VNTR-10 genotype in Northern Ireland were interspersed by several VNTR-1 isolates, considered a more ancestral genotype (Trewby *et al.*, 2016).

Whole genome sequencing has also been used at a more local level to establish the population structure of *M. bovis* in different countries and regions, including New Zealand (Crispell *et al.*, 2017), France (Hauer *et al.*, 2019), Germany (Kohl *et al.*, 2020) and Ethiopia (Almaw *et al.*, 2021). These studies not only achieved a higher resolution than traditional molecular techniques, but allowed for the detection of possible introductions from countries with commercial ties (Perea *et al.*, 2021). In the context of Spain, WGS data has been limited to a certain genotype (e.g. SB0339) across several regions or

different genotypes within the same region (e.g. Catalonia), which partly limits the overall picture of *M. bovis* diversity in this country (Pozo, 2020; Perea *et al.*, 2021). Therefore, more research is required to unravel the population structure of *M. bovis* in Spain.

Similar to traditional molecular characterisation studies, whole genome analyses have shown that *M. bovis* strains tend to cluster geographically (Crispell *et al.*, 2017; Dippenaar *et al.*, 2017; Price-Carter *et al.*, 2018; Salvador *et al.*, 2019; Pozo, 2020). In this context, WGS has shown that the genetic relatedness of *M. bovis* isolates is limited among epidemiologically related herds, with pairwise genetic distances generally increasing with the temporal and geographic distance between the analysed isolates (Biek *et al.*, 2012; Orloski *et al.*, 2018; Crispell *et al.*, 2019b; Pozo, 2020). However, reduced genetic divergence between isolates in distant locations has also been described in scientific literature, which could indicate unknown epidemiological links (Kohl *et al.*, 2018b; Perea *et al.*, 2021).

The limited genetic diversity observed in many outbreak scenarios restricts to a certain extent the capacity of WGS to resolve the transmission dynamics of *M. bovis* at the animal-to-animal scale in short periods of time, being better suited for herd-level studies (Biek *et al.*, 2012). In this context, different thresholds of SNP differences are used to approximate the time of transmission; three to five SNPs are indicative of recent transmission whereas 12 SNPs are considered the upper limit for an epidemiological linkage (Walker *et al.*, 2013; Crispell *et al.*, 2020). However, these thresholds are usually extrapolated from human TB studies involving *M. tuberculosis* and may therefore not be representative of *M. bovis* biology and epidemiology.

The increased resolution of WGS, coupled with epidemiological data, has also allowed to trace back possible origins of infection and greatly reduced the associated costs of outbreak investigations (Biek *et al.*, 2012; Glaser *et al.*, 2016; Bruning-Fann *et al.*, 2017; Orloski *et al.*, 2018). Bruning-Fann *et al.* 2017 used WGS to disclose an outbreak of bTB involving an index dairy herd that led to the infection of other herds through the movement of infected animals and/or the consumption of infected unpasteurised milk by calves (Bruning-Fann *et al.*, 2017). Importantly, WGS revealed a possible epidemiological link between several herds that were investigated as a result of the outbreak but that did not have registered animal movements with the index herd. A similar finding was described by Price-Carter *et al*. 2018 when four isolates from animals housed in herds separated by 4 km from each other, and without recorded animal

movements, were only differentiated by one unique SNP, suggesting an unknown epidemiological link (Price-Carter *et al.*, 2018). Furthermore, by comparing feature SNPs between isolates and their isolation dates, the authors could reconstruct with varying levels of confidence the possible geographical and temporal dynamics of the strains in the regions under study and highlight possible origins of the outbreaks.

In addition to outbreak investigations, the application of statistical models on WGS and epidemiological data has allowed the study of transmission dynamics in multi-host systems (both domestic and sylvatic) in what is termed as phylodynamics. Genomic analyses have revealed that genetic diversity overlaps between *M. bovis* isolates from cattle and wildlife species, reflecting no species-specific clustering (Salvador *et al.*, 2019; Crispell *et al.*, 2020). In this sense, the isolation of *M. bovis* with a homogenous genetic background from sympatric cattle and wild life species can be used as evidence for recent contact events or introduction in a region (Glaser *et al.*, 2016; Salvador *et al.*, 2019). In addition, a high genetic diversity could also indicate that the strain has been circulating for a prolonged period of time or that several *M. bovis* strains were introduced in the past (Glaser *et al.*, 2016; Salvador *et al.*, 2019).

Whole genome sequencing could also be a useful tool to assess the potential risk for different wildlife species to become maintenance hosts in a region. In addition to highlighting the importance of deer in the transmission and maintenance of bTB in Michigan, the geographic overlap of *M. bovis* isolates with different genotypes suggested that several lineages of the pathogen had been circulating in the region for some time (Salvador *et al.*, 2019). Furthermore, the less frequent isolation and the reduced genetic diversity of *M. bovis* in elk (wapiti) indicated that this animal species is probably a spillover host and may not have a significant role in the infectious cycle of *M. bovis* in this region, which contrasts to its epidemiological relevance in Canada (Shury & Bergeson, 2011; Salvador *et al.*, 2019). Nevertheless, identification of closely related isolates between elks could also indicate recent intra-species transmission, which could in turn suggest an incipient transmission cycle. This could help animal health and environmental agencies in implementing different management strategies to avoid the dissemination of *M. bovis* in a new species, such as avoiding human-caused aggregations of animals that increase the contact between deer and elk (e.g. recreational feeding or bait sites). In a similar manner, although badgers are considered an important reservoir of bTB in the UK and the RoI (Crispell *et al.*, 2019b; Crispell *et al.*, 2020), the limited role of badgers in the persistence of bTB in Cumbria (UK) highlighted the need for tailored control

strategies depending on the endemicity of bTB in the affected region (Rossi *et al.*, 2021).

In spite of the utility of WGS in outbreak investigations and epidemiological studies involving different animal species, the direction of transmission is difficult to establish in many scenarios due to several reasons. Firstly, the fixation of mutations in *M. bovis* genomes is low and highly variable over short periods of time, which limits the detection of epidemiologically close cases if transmission has taken place recently (Orloski *et al.*, 2018; Price-Carter *et al.*, 2018). Secondly, in order to obtain a representative number of specimens from the animal populations involved in the region under study, appropriate sampling strategies need to be enforced (Crispell *et al.*, 2017), which may not be feasible in many situations due to economical and ethical implications of sampling apparently healthy wildlife and the limitations in the performance of available diagnostic methods in identifying all of the infected animals during an outbreak. Nevertheless, several statistical models based on genomic data have been proposed that can be used to establish the rate and probability of transmission events between different animal species. Salvador *et al.* 2019 showed that there was a stronger support for the transmission between deer and cattle and deer and elk in Michigan, than there was for cattle and elk (Salvador *et al.*, 2019). Similarly, several studies in the UK have reported that the transmission of bTB may be more likely between badgers and cattle than the reverse, and that intra-species transmission may be more frequent than inter-species transmission (Crispell *et al.*, 2019b; Rossi *et al.*, 2021; van Tonder *et al.*, 2021). However, the differences in the rates of transmission between studies suggest that the directionality of transmission could vary under different epidemiological situations. While bTB interspecies transmission is higher in endemic regions (Southwest England) (Crispell *et al.*, 2019b; van Tonder *et al.*, 2021), it was shown to be lower in an outbreak scenario in a low-risk area in Northwest England (Rossi *et al.*, 2021). A plausible mechanism for this difference could be related to the amount of time in which the transmission clusters have been occurring after the introduction and expansion in the cattle population within a region, the infection could spill over to the local badger population, leading to its amplification through intra-species transmission events and, finally, to spillback events to the cattle population (Rossi *et al.*, 2021; van Tonder *et al.*, 2021).

In addition to its application in bTB outbreaks, WGS has also proven to be an important technique in the study of zTB. Its high resolution is able to put in evidence possible human-to-cattle transmission events in the USA (Lombard *et al.*, 2021). In addition, WGS revealed the high genetic relatedness between *M. bovis* isolates obtained from unpasteurised cheese, dairy cattle and human cases in the region of Baja California, Mexico (Ortiz *et al.*, 2021), expanding its utility to food safety investigations.

As previously mentioned, the accumulation of mutations in the genome of *M. bovis* follows a sequential manner which can be used to establish a molecular clock to define the rate of mutations per genome in a given amount of time, usually a year. Using statistical modelling, the substitution rate of *M. bovis* has been suggested to range between 0.15 – 0.53 events per genome per year, which contrasts with the one established for *M. tuberculosis* (0.4 – 0.5 events per genome per year) (Biek *et al.*, 2012; Crispell *et al.*, 2017; Salvador *et al.*, 2019). Despite this low substitution rates in *M. bovis* genomes, it remains to be demonstrated whether substitution rates differ by *M. bovis* lineages (van Tonder *et al.*, 2021).

Using these substitution rates and Bayesian molecular clock analyses, the earliest possible time of introduction of the MRCA can be calculated. Using this approach, the MRCA for the *M. bovis* strains involved in a multispecies outbreak in Minnesota was established between 1999 and 2004 and additional analyses including unrelated herds from Texas infected with identical strains suggested that both outbreaks originated from a MRCA in the 1990s (Glaser *et al.*, 2016). However, time scale analyses can be affected by the sampling strategy; a narrow sampling window can have a negative effect on the precision of the estimated times, especially when substitution rates are so low, although accuracy is not affected significantly (Crispell *et al.*, 2017).

1.6. Surveillance, control and eradication of bTB

1.6.1. Surveillance and eradication of bTB around the world

Bovine TB eradication programmes were introduced in many countries around the world during the start of the 20th century, and their reinforcement in the following decades led to a significant reduction in bTB prevalence and to the declaration of officially bTB-free countries. One example is the eradication programme established in the USA, initiated in 1917, which decreased the prevalence of the infection in animals from 5% to less than 0.5% by 1941, when the country was declared bTB-free according to USA regulations (Palmer & Waters, 2011). An additional milestone in the fight against bTB was achieved with the establishment of Australia, one of the largest exporters of beef and meat products, as an officially bTB-free country in 1997 (Cousins & Roberts, 2001).

Eradication programmes for bTB are usually based on a test-and-slaughter strategy in which animals are routinely tested with an official immunological test, such as the SIT test. Reactor animals are then culled and restrictions to animal movements or derived-milk or meat products, are usually enforced in the affected herd until the presence of the infection has been ruled out. Although the goal of eradication programmes is the same around the globe, their implementation differs between countries and regions. The OIE Terrestrial Animal Health Code establishes a series of criteria before a country or region can be officially declared bTB-free (OTF), mainly a prevalence inferior to 0.1% for three consecutive years in domestic bovids or a maximum of 0.2% herd prevalence during the same period (OIE, 2021).

The direct costs deriving from the use of diagnostic tests and epidemiological studies, as well as the economic compensations paid to farmers for the culling of positive animals make eradication programmes an extremely expensive endeavour that not all countries are able to afford. In addition, eradication programmes have an indirect effect on a country's economy due to the restrictions imposed on the movement of animals and their products, which ultimately leads to a loss of competitiveness. Interestingly, only 23% of the countries affected by bTB in 2019 applied all of the relevant control measures recommended by the OIE during 2017 – 2018 (Murai *et al.*, 2019).

In the last decade, there has been an increased interest in the search of alternative or complementary measures that may be used together with test-and-slaughter policies in order to aid in the eradication of bTB. In addition to

vaccination of cattle, which has recently started to be evaluated as an option in different countries, such as the UK or Chile (Jones *et al.*, 2012; Srinivasan *et al.*, 2019; Abalos *et al.*, 2022), other alternatives include the genetic selection of cattle towards bTB-resistant phenotypes. The interest in precision breeding of cattle towards bTB resistance is based on previous quantitative studies that have shown that there is certain variation in the susceptibility of different cattle breeds to bTB, such as Holstein and Zebu (Allen *et al.*, 2010; Raphaka *et al.*, 2018) and that such resistance can be associated with specific genetic traits (Raphaka *et al.*, 2017). Such an interest has led countries like the UK to establish genetic evaluations of dairy cattle for bTB resistance since 2016 (Raphaka *et al.*, 2018), leading to the development of the TB Advantage index.

1.6.2. **Legal frame of bTB surveillance and eradication in the EU**

Eradication programmes for bTB have been in place in the majority of European countries since the first half of the 20th century (Palmer & Waters, 2011). With the establishment of the European Economic Community (EEC), EEC coordinated bTB eradication campaigns from a European perspective with the enforcement of several legal initiatives (Reviriego Gordejo & Vermeersch, 2006). The first act to be implemented at a European level was Council Directive 64/432/CEE, which established the rules, diagnostic methods and requisites required for Intra-Community trade of bovine and porcine animals, as well as the definition, among others, of "OTF bovine herds".

Despite the initial efforts, bTB was still a significant trade barrier in the EU by the mid-1970s, since some Member States still applied their old national eradication programmes in cattle destined to internal markets (Caffrey, 1994). This led to the enforcement of Directive 78/52/CEE, which set the applicable criteria for the creation of bTB eradication programmes in the EU. This act not only established common requirements for eradication programmes, but also gave to Member States access to financial aid for the accelerated eradication of bTB (Caffrey, 1994). The implementation of unified strategies allowed the declaration of several EU Member States as OTF during the late 20th century, such as Denmark in 1980, the Netherlands in 1995, Germany and Luxembourg in 1997, Austria in 1999, France in 2001 and Belgium in 2003 (Reviriego Gordejo & Vermeersch, 2006). However, the success of bTB eradication was not limited to the EU; for example, when Finland and Sweden, or the Czech Republic, joined the EU in 1995 and 2004, respectively, they received the OTF status as they had already achieved the eradication of bTB in previous years (Reviriego Gordejo & Vermeersch, 2006). These advances have led to a significant

decrease of bTB prevalence in the EU, with 17 OTF countries and three non-OTF countries with OTF regions in 2020 (see section 1.3.1).

On the 21st of April of 2021, Regulation 2016/429, "on transmissible animal diseases and amending and repealing certain acts in the area of animal health", also known as the 'Animal Health Law' (AHL), entered into force in the EU. The regulation empowers the Commission to adopt implementing and delegating acts in the field of animal health, such as define measures for the prevention and control of diseases, adopt biosecurity measures, establish lists of diseases and modify them, etc.

1.6.2.1. Listed diseases

The first major difference with previous acts was set by the Commission's Implementing Regulation (IR) 2018/1882, which established that prevention and control measures are applied to animals according to the disease in question, the presence/absence of a given disease in the EU, its importance and the epidemiological role of the different animal species considered. Based on these factors, diseases are classified into five different categories that define which measures are required for their control and eradication (table 3).

Category of listed disease	Required measures
Category A disease	Disease not normally present in the EU. Requires immediate eradication measures.
Category B disease	Disease must be controlled in all Member States with the objective of eradicating it in the EU.
Category C disease	Disease is prevalent in some Member States and measures need to be implemented to prevent its spread to other Member States, either because they are officially disease-free or have established disease-specific eradication programmes.
Category D disease	Disease for which measures are needed to prevent its spread as a result of its entry in the EU or as a result of movements between Member States.
Category E disease	Disease for which there is a need for surveillance within the EU.

Table 3. Definitions of categories of listed diseases in Commission's Implementing Regulation 2018/1882.

Under this regulation, what is defined as bTB (i.e. TB in bovids of the genus *Bison* ssp., *Bos* ssp. and *Bubalus* ssp.) is considered a disease of categories B, D and E, while infection in *Artiodactyla* other than the previously mentioned, and in other terrestrial *Mammalia*, TB is listed as a disease of categories D and E, respectively.

This act introduces two major novelties: a) explicitly states that infection in animals can be caused by three members of the MTBC (i.e. *M. bovis, M. caprae* and *M. tuberculosis*; therefore, includes *M. caprae* in the causative agents of TB in animals that require the adoption of measures); and b) extends the need to apply measures for the control and/ or surveillance of the disease in other animal species respect to bovines.

1.6.2.2. **Requirements for animal movements and eradication of bTB**

According to Delegated Regulation 2020/688, "as regards animal health requirements for movements within the Union of terrestrial animals and hatching eggs", Union movement of bovines is only allowed for OTF herds. If the animal being moved is younger than 6 weeks of age, or the OTF herd of origin is located within an OTF MS, then no testing is required. Otherwise, a negative result to an official diagnostic test is required during the 30 days prior to departure.

Delegated Regulation 2020/689, "as regards rules for surveillance, eradication programmes, and disease-free status for certain listed and emerging diseases" establishes the requirements for a country or herd to achieve or maintain the OTF status, as well as the reasons for suspension and/or withdrawal of this status. In this context, a MS may be declared OTF if at least 99.8% of its bovine herds, representing 99.9% of the total bovine population, are free from infection during the past three years.

In the EU, the official diagnostic methods used for pre-movement testing, and the granting and maintaining of disease-free status for bTB, are the skin test and the IGRA, as indicated in DeR 2020/688

Although several situations are considered in the legislation, DR 2020/689 establishes the use of official diagnostic methods as the main method for the detection of bTB infections in bovines. Importantly, it differentiates cases into two groups (suspected and confirmed cases) based on the type of diagnostic method used and the epidemiological information of the analysed herd. Suspected cases are those in which 1) clinical or *post-mortem* signs of infection

are detected; 2) an official diagnostic method indicates the likely presence of the infection; and 3) the studied animal has been in contact with a confirmed bTB case. In turn, confirmed cases are those in which the presence of the pathogen has been confirmed using an official ancillary method (table 4), or the infection was detected in an animal with clinical signs or with an epidemiological link to a confirmed case.

In the EU, the official diagnostic methods used for pre-movement testing and the granting and maintaining of disease-free status for bTB are the skin test and the IGRA, as indicated in Annex I part 2 of DeR 2020/688 and Annex III section 2 of DeR 2020/689 (table 4).

<table>
<tr><th>Type</th><th>Application</th><th>Method</th><th>Use</th></tr>
<tr><td rowspan="2">Indirect</td><td rowspan="2">Ante-mortem</td><td>Tuberculin skin test</td><td rowspan="2">Suspected case:
- Surveillance
- Granting and maintenance of OTF status</td></tr>
<tr><td>Gamma-interferon</td></tr>
<tr><td>Post-mortem</td><td>Gross & histopathology (host response)</td><td>Suspected case:
- Surveillance
- Suspension of OTF status</td></tr>
<tr><td rowspan="3">Direct</td><td rowspan="3">Post-mortem</td><td>Histopathology (pathogen detection)</td><td rowspan="3">Confirmed case:
- Surveillance
- Withdrawal of OTF status</td></tr>
<tr><td>Culture isolation</td></tr>
<tr><td>PCR</td></tr>
</table>

Table 4. Official diagnostic methods used for the diagnosis of bTB in the EU (DeR 2020/689).

The competent authority shall classify an animal or a group of animals as a confirmed case of MTBC infection when: 1) the aetiological agent has been isolated (e.g. microbiological culture); 2) an antigen or nucleic acid specific to the aetiological agent has been detected (e.g. qPCR); or 3) a positive result to an indirect diagnostic method was obtained from an animal with obvious clinical signs or that had an epidemiological link with a confirmed case.

The appropriate establishment of these cases is crucial, since it will determine if the OTF status of a herd may be suspended (suspected case) or withdrawn. The suspended OTF status may be restored if the affected herd tests negative to an official immunological test not later than 12 months, whereas a withdrawn status may be restored after two consecutive negative results, with an interval not longer than 12 months. This difference can have a significant impact in the economy of the affected herds, since during this stage animal movement is restricted.

According to DeR 2020/689, the collection of samples as well as the methodology used in the establishment of suspected and confirmed bTB cases must comply with different levels of criteria in what is known as 'the diagnostic cascade' (Article 6. Section 1). In this article, it is established that the collection of samples, techniques, validation and interpretation of the diagnostic methods implemented by the competent authorities must comply with the guidance made available by the current legislation, the Commission and the EU-RL for bTB, for the appropriate use and validation of bTB diagnostic methods. For example, the protocols and interpretation criteria of the skin tests used in the EU are defined in SOP/001/EURL. If no EU legislation or guidance is available for a specific diagnostic method, then the guidelines and recommendations in the Manual of Diagnostic Tests and Vaccines for Terrestrial Animals of the OIE must be followed. Finally, if no guidance is available in any of the previously mentioned cases, sampling and diagnostic methods should comply with Article 34 of EU Regulation 2017/625.

1.6.2.3. **Slaughterhouse surveillance and gross pathology**

Animals that react to official *ante-mortem* tests are sent to the abattoir, and their carcases are inspected for lesions compatible with TB. Inspection of animal carcases is governed by the Commission's IR 2019/627. Some lesions may not be readily visible or palpable, and therefore incisions are needed in order to inspect the interior of tissues. When tuberculous-like lesions are observed, carcases are condemned according to the grade of affection (localised *vs.* generalised).

Abattoir inspections are mainly focused on LNs, but other organs, such as the lungs or the liver, and surface cavities, such as the diaphragm or the mesentery, are also evaluated (table 5). However, many reactor animals may present lesions in only one site, or not present VLs at all, especially in early stages of infection (Corner, 1994; Whipple *et al.*, 1996). For this reason, inspection of

reactor animals should not be limited to VLs, but also apparently healthy tissues in order to detect incipient lesions or macroscopically NVLs.

Head	Mandibular LNs
	Parotid LNs
	Medial and lateral retropharyngeal
	Tonsils
Thorax	Mediastinal LNs
	Tracheobronchial LNs
	Lungs
Abdomen	Liver
	Hepatic LNs
	Spleen
	Mesenteric LNs
	Kidneys
Carcase	Caudal cervical LNs
	Subiliac LNs
	Internal, medial and lateral iliac LNs
	Gluteal or ischiatic LNs
	Sacral LNs
	Suprammary or scrotal LNs
	Udder or scrotum

Table 5. Tissues and organs that should be inspected in reactor animals by incision and palpation (Corner, 1994).

When bTB prevalence is high and animals with advanced stages of TB are frequent (e.g. early stages of eradication programmes), abattoir inspections are a reliable method detect infected animals (Corner, 1994). However, as eradication programmes progress, the frequency in which animals are tested and culled not only decreases bTB prevalence, but also the probability of an infected animal of progressing to more advanced stages of disease in which lesions can be visually detected (Cassidy, 2006). Despite this decrease in Se, passive surveillance through abattoir inspection is still critical in many bTB-free regions or countries, where TB-free herds may no longer be tested with

ante-mortem tests and breakdowns are usually identified after the detection of a lesion compatible with TB as part of passive surveillance (Schiller *et al.*, 2011; Sergeant *et al.*, 2017).

Despite the extensive list tissues and organs that can be affected, scientific literature has revealed that the majority of VLs are found in the lungs and LNs from the thorax and head (Corner, 1994; Whipple *et al.*, 1996), especially in medial retropharyngeal mediastinal and tracheobronchial LNs (table 5). These findings are in line with the respiratory route being the main route of transmission for bTB, and have also been observed in experimentally infected animals (Serrano *et al.*, 2018). However, in certain scenarios, the digestive route could lead to a different distribution of lesions (Corner, 1994; Serrano *et al.*, 2018), and other LNs, such as mesenteric LNs should also be inspected.

The complete inspection of a carcase requires a large amount of time and effort that most probably is not feasible for official abattoirs, in which large numbers of animals are processed on a daily basis. This limitation has led to routine inspections concentrating in the most probable affected tissues, as previously mentioned. In the EU, inspection of animal carcases is governed by the Commission's IR 2019/627.

1.6.3. **Surveillance and eradication of bTB in Spain**

The first measures against bTB in Spain were introduced during the early 1950s, but it was not until the 1960s that surveillance programmes were firstly introduced in dairy herds, and 1978 that an eradication campaign was implemented. With the entry of Spain in the EU in 1985, the European regulations on eradication programmes and animal movements were adopted through the Royal Decrees 2611/1996 and 1716/2000, respectively. At present, eradication of bTB is governed by the national eradication programme, the Royal Decree 2611/1996, as well as the AHL.

The eradication programme in Spain is devised for periods of five years, in which the goals are set and strategies are implemented according to the evolving epidemiological situation. Once enforced, the eradication programme is implemented in the different regions based on their bTB prevalence. Under the current eradication programme, regions are divided into OTF regions (prevalence zero), low prevalence regions (< 1% prevalence) and high prevalence regions (> 1% prevalence) (MAPA, 2022).

The targeted animals of the Spanish eradication campaign are all bovine herds with exception of certain fattening units in regions of non-zero prevalence. As previously mentioned, caprines may be included in the eradication programme if they have epidemiological links with cattle herds. Additionally, in order to control the spread of TB in wildlife and its recurrence in cattle, a complementary surveillance and control programme for wildlife has been in force since 2017 (Royal Decree 138/2020).

Detection of bTB cases is achieved through the use of official diagnostic methods. These methods may be used both during pre-movement testing or during status granting, maintenance or suspension/removal. The SIT test is the method of choice used in the eradication programme in Spain, although under certain epidemiological situations, such as OTF herds, the CIT may be used (MAPA, 2022). In this case, animals negative to the CIT test but with a reaction to bPPD are considered follow-up animals, and samples are collected for microbiological culture when the animal is culled at the end of its productive cycle (MAPA, 2022).

In order to guarantee a standardised and appropriate use of diagnostic methods, all veterinarians are required to carry out specific training when being tasked with bTB control, and every three to five years thereafter. Routine testing of animals is carried out at a maximum interval of six months in high prevalence regions, or up 24 months in low prevalence or OTF regions, respectively. In addition to diagnostic methods, the Spanish eradication programme enforces passive slaughterhouse surveillance independently of the bTB status of the herd.

Samples collected for confirmation must be obtained from at least one LN from one of the following sites: head (retropharyngeal and mandibular), thoracic cavity (mediastinic and bronchial), thorax (superficial cervical and prescapular), abdominal cavity (mesenteric and hepatic) and mammary glands (supramammary LNs) (MAPA & VISAVET, 2017).

1.6.4. Significance of NTM in bTB eradication

Infection or sensitisation to NTMs may negatively affect skin test performance either by a decreased sensitivity or specificity (de la Rua-Domenech *et al.*, 2006; Allen *et al.*, 2018). Indeed, several NTM species, such as *M. avium* subsp. *avium*, *M. intracellulare, M. flavescens* or *M. scrofulaceum*, have been shown to produce non-specific reactions in animals inoculated with bPPD at different time-points after inoculation (Pearson *et al.*, 1977; Corner & Pearson, 1978; Hope *et al.*, 2005). Infection and vaccination against MAP has also been shown to

significantly affect the results of the skin test in certain scenarios (Dunn *et al.*, 2005; Aranaz *et al.*, 2006; Garrido *et al.*, 2013; Seva *et al.*, 2014), not only in cattle but also in goats affected by caprine TB (Alvarez *et al.*, 2008; Roy *et al.*, 2018). This can lead to false-negative results due to the masking of bPPD responses by high aPPD skin reactions in the CIT, or a decrease in specificity by unspecific cross-reactivity in the SIT. Due to the use of PPDs for lymphocyte stimulation, cross-reactivity with NTMs can also occur in the IGRA (de la Rua-Domenech *et al.*, 2006). Similarly, vaccination against and infection with MAP have been shown to interfere with the interpretation of the test results both in cattle and goats, especially when using a more severe threshold (Alvarez *et al.*, 2008; Alvarez *et al.*, 2009b; Bezos *et al.*, 2012a; Roupie *et al.*, 2018). Although vaccine-induced cross-reactivity to bPPD in cattle has been detected during long periods of time (Muskens *et al.*, 2002), the effect of vaccination in goats seems to wane over time (Roy *et al.*, 2018). Finally, NTM can also produce granulomatous lesions, which can negatively affect the performance of slaughterhouse surveillance (Hughes *et al.*, 2005).

The interference of NTM with the diagnosis of bTB not only can limit the success of eradication programmes, but can negatively affect the trust of stakeholders in the success of the surveillance and eradication measures. Due to the significant effect that these limitations in specificity can have on the diagnosis of bTB, it is important to assess the extent of this cross-reactions and identify the different NTM species that can interfere with bTB diagnosis in order to improve current diagnostic methods. Recently, a cocktail of antigens has been shown to elicit no or very little cross-reactivity when used on both skin tests and IGRA in MAP-vaccinated animals (Middleton *et al.*, 2021), which could allow to increase the diagnostic performance of *ante mortem* tests in the future. In addition to improved testing, it is important for animal health authorities to be able to identify NTM species that are detected *post mortem* in order discard false positive results and assess the extent of their cross-reaction with current and future methods.

2. Thesis hypotheses and objectives

2.1. Thesis hypotheses

Bovine TB is a chronic infectious disease of bovids that is produced by members of the MTBC, predominantly *M. bovis* and *M. caprae*. Due to its significant effect in animal health and production, as well as its zoonotic potential, many countries have established control and eradication programmes for bTB, especially in high-income settings. In the EU, bTB surveillance, control and eradication programmes are regulated by the AHL (Regulation 2016/429) and related legislation, and through national legislation on the matter. Bovine TB control and eradication in Spain is established through the annual bTB eradication programme, which is founded on the Royal Decree 2611/1996.

Eradication of bTB is based on a 'test and slaughter' strategy by which the reduction of the risk of transmission is achieved through the detection and culling of infected animals. Eradication programmes have been in place for decades resulting in a significant reduction of bTB prevalence. However, this infection still remains a problem in many regions, as a result of the complex epidemiology of bTB, in which several wildlife and domestic species can act as reservoirs, as well as due to the limitations of current diagnostic and characterisation tools.

Eradication efforts are being hampered due to a series of limitations in current tools used in eradication programmes. Among these: 1) the limitations of diagnostic tools used for the confirmation of bTB infections; 2) the limitations of molecular characterisation methods used in bTB epidemiological investigations; and 3) the interference of NTM with the diagnosis of bTB.

2.1.1. Limitations in bTB diagnostic methods

The DeR 2020/689 regulates the official diagnostic methods and procedures for the control, surveillance and eradication of bTB in the EU. In this context, the intradermal tuberculin test and IGRA are used in animal movements and the granting and maintenance of the OTF status. Animals that test positive to these tests are considered suspected cases which, according to EU legislation, need to be confirmed in order for the OTF status of their herd to be withdrawn. The methods that can be used by authorities to classify a case as confirmed include the isolation of the aetiological agent or the detection of pathogen-specific antigens or nucleic acids, among others.

Culture is still the most frequently used method for the confirmation of suspected cases. However, this method presents important limitations. Firstly, due to the fastidious growth of the MTBC, time to detection and confirmation can be extremely slow (up to 12 weeks), which impacts the time by which the suspension status can be lifted from a herd. In addition, diagnostic Se of microbiological culture is significantly affected by the bacterial load and viability in the analysed samples. Therefore, it is less suited to detect the infection in animals with low bacterial loads, such as those present in early stages of infection.

Hypothesis 1: the development of sensitive and rapid methods for the detection of MTBC members would allow the detection of more positive animals, especially in early stages of infection, in a shorter period of time, resulting in a significant increase in the efficacy of eradication campaigns.

2.1.2. Limitations in molecular characterisation methods

Assessing the genetic relatedness between isolates of MTBC members is crucial in order to establish epidemiological links within and between herd breakdowns or to define transmission dynamics of infection, such as contact with infected wildlife. Epidemiological studies in bTB have largely benefited from the availability of a series of molecular genetic methods, such as DVR-spoligotyping or MIRU-VNTR, which have resulted in valuable insights into the global distribution and evolution of bTB. However, the resolution of these tools can be limited when

used to study events that are circumscribed in time and/or space, or in areas where the genetic diversity of circulating strains is limited, making them less suitable for the study of herd breakdowns.

Hypothesis 2: the development of molecular characterisation methods with a higher resolution and throughput capacity would greatly enhance epidemiological studies of bTB. This would contribute considerably to the identification of risk factors and the definition of targeted protection measures associated with the main operations within an eradication campaign.

2.1.3. **Diagnostic interference with NTM**

An additional challenge for bTB eradication campaigns are infections with NTM that interfere with *in vivo* methods, as well as limit the performance of culture. The isolation of NTM from reactor animals not only hampers eradication efforts, but can also reduce the trust of veterinary workers and farmers in eradication programmes. The thorough study of the role of NTMs and their appropriate identification in the eradication of bTB is of great importance in order to detect possible diagnostic interferences. Currently, identification of NTM isolates is based on Sanger sequencing, which is considered the reference method for the characterisation of mycobacterial species. However, identification of NTM species through this method is time consuming and costly. Moreover, variations in the taxonomy of the genus *Mycobacterium*, the high genetic similarity of certain mycobacterial groups and the existence of a large amount of redundant and low quality data in public repositories further complicate this process.

Hypothesis 3: the development of cost-effective and easily standardisable methods for the rapid and accurate identification of mycobacterial species would greatly contribute to the success of eradication campaigns by providing laboratories with tools capable of detecting the diagnostic interferences produced by NTM.

2.2. Thesis objectives

The overall aim of this doctoral thesis was to develop novel and powerful tools that could assist current and future bTB control, surveillance and eradication programmes. The main aim was achieved through three specific objectives, related to key aspects of bTB eradication and control: 1) development and validation of sensitive and rapid laboratory diagnostic methods for bTB confirmation; 2) evaluation of high-resolution molecular genetic methods for epidemiological studies of bTB; and 3) development of new tools for the rapid identification of NTM species in clinical samples.

2.2.1. Objective 1: development of novel diagnostic tools for the detection of MTBC members

With the entry into force of the AHL, other methods may be implemented in bTB eradication programmes as confirmatory methods. The possibility of using methods that can detect the genetic material of MTBC members paves the way for the use of molecular methods, such as qPCRs.

In this aspect, qPCRs are highly sensitive and specific genetic amplification methods that have shown very promising results for the direct detection of MTBC DNA in animal tissue samples (direct qPCR). In addition, they allow for a rapid detection, requiring only a few hours until results are obtained. Due to these advantages over microbiological culture, there is an increased interest in the development, evaluation and potential deployment of qPCRs as a confirmatory method.

The first objective of this thesis was the development and validation of direct qPCR protocols for the detection of MTBC genetic material in fresh bovine tissue samples.

2.2.2. Objective 2: development of novel tools for the typing of MTBC members

Whole genome sequencing is a method that allows the study of microorganisms at the nucleotide-level, offering a much higher resolution than traditional molecular genetic methods. It could therefore help unveil additional genetic diversity within circulating strains in a given geographical region or outbreak event.

The use of this technology in the study of bTB is increasingly used around the world. However, there is little information regarding the performance of available tools used for the analysis of bTB WGS data.

In addition, the implementation of WGS in herd breakdown investigations in Spain remains limited to date. In a country with a high genetic diversity of circulating *M. bovis* strains and complex bTB epidemiology, WGS could be extremely useful to identify the causes of herd breakdowns and to better assess possible risk factors of infection, thus allowing the development of enhanced and targeted contingency measures in the future.

The second objective of this thesis was the evaluation of WGS as a molecular characterisation method for bTB epidemiological studies. Within this objective, two lines were pursued:

1. Comparison of the performance of available bTB variant calling methods for the study of MTBC members.
2. Implementation of WGS as a tool for the assessment of within-herd genetic variation of *M. bovis* strains in chronically infected herds.

2.2.3. Objective 3: development of novel tools for the identification of NTM species in clinical samples

Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) has advanced significantly the identification of bacterial species, including NTM, in the human clinical setting. This methodology presents a series of advantages that makes it a promising alternative for the identification of NTM isolates with respect to the currently available techniques. Firstly, MALDI-TOF MS is a high-throughput method that allows for cost-effective and rapid identification. Secondly, the identification process, based on the comparison of spectral similarities with validated reference spectra allows for a standardised analysis, without the need of data post-processing and curation that is required in Sanger sequencing.

The implementation of MALDI-TOF MS in veterinary mycobacteriology has been limited to date and its usefulness in bTB eradication programmes has not been assessed yet. The third objective of this thesis was to evaluate the performance of MALDI-TOF MS as a screening method for the identification of NTM species.

3. Research plan and methodology

The research plan for this doctoral thesis was divided into three sections focusing on 1) the development and validation of qPCR protocols for the detection of the MTBC that could be used as bTB confirmation methods; 2) use of WGS for molecular epidemiology investigation in bTB outbreaks; and 3) identification of NTMs with MALDI-TOF MS.

3.1. Development and validation of direct qPCRs for the confirmation of bTB infections in bovine tissues

The development of any diagnostic method follows a series of steps in which the goal of the protocol is defined, the optimal parameters of the reaction are set (optimisation) and the fitness of the assay for its intended purpose is determined (validation). With this information at hand, laboratories can understand the suitability of a test to carry out a specific task and address its functional capacities and limitations.

This section focuses on the development of direct qPCR protocols for the detection and quantification of specific fragments of DNA from the MTBC in bovine tissue samples, with the objective of evaluating their performance as a confirmation method in comparison with microbiological culture. In order to be able to quantify the genomic equivalents in a sample, a qPCR needs to detect a genetic target with a stable number of copies, usually a single-copy gene. However, multiple-copy elements are considered a more suitable target for qPCR detection due to the higher Se they offer. As a result, two different

approaches were pursued; one describing the development of a qPCR for the detection of MTBC DNA based on a multiple-copy target, which could offer a high Se and therefore be of use for screening purposes, whereas the other focused on the development of a qPCR targeting a single-copy target, which could offer a high specificity and therefore could be of use for both the detection and quantification of MTBC DNA.

The general organisation and methodology used for the design, optimisation and validation of the abovementioned qPCRs will be briefly presented in this section. More detailed descriptions of the procedures can be found in the published studies:

Lorente-Leal, V, Liandris E, Castellanos E, Bezos J, Domínguez L, de Juan, L & Romero, B, 2019. Validation of a Real-Time PCR for the Detection of *Mycobacterium tuberculosis* Complex Members in Bovine Tissue Samples. Frontiers in Veterinary Science 6. https://doi.org/10.3389/fvets.2019.00061.

Lorente-Leal, V, Liandris E, Pacciarini, M, Botelho, A, Kenny, K, Loyo, B, Fernández, R, Bezos J, Domínguez L, de Juan, L & Romero, B, 2021. Direct PCR on Tissue Samples To Detect *Mycobacterium tuberculosis* Complex: an Alternative to the Bacteriological Culture. Journal of Clinical Microbiology 59:e01404-20. https://10.1128/JCM.01404-20.

3.1.1. **Literature research and target selection**

The choice of an adequate genetic target is a crucial element of any PCR protocol. Upon review of the available scientific literature on molecular detection of MTBC members in bovine tissues, the IS*6110* element and the *mpb70* gene were selected. The IS*6110* element was selected due to its conservation in the MTBC and its multiple-copy nature. In contrast, the *mpb70* gene was selected due to its single-copy nature, making it suitable both for the detection and quantification of MTBC genomes in a sample, which could allow the establishment of the bacillary load in the analysed tissues.

Since the IS*6110* element has been previously used in qPCR studies, a set of primers were selected from available literature (Michelet *et al.*, 2017a; Michelet *et al.*, 2018b). On the other hand, the *mpb70* gene has frequently been used in end-point PCRs, and therefore a specific set of primers were designed for their use in qPCR. Sequence data of the *mpb70* gene from representative MTBC members was retrieved from the NCBI database and used to assess their sequence identity and compare it with that available for NTM. Oligonucleotides targeting a 133 bp-long fragment of the *mpb70* gene were then designed using Oligo primer analysis software.

An initial *in silico* specificity evaluation was carried out for both primer sets using the Basic Local Alignment Service Tool (BLAST) from the NCBI, in order to rule out any possible cross-reactions with available sequence data, especially from NTM.

3.1.2. Optimisation and validation of qPCR protocols

The qPCR conditions used in the two studies were optimised in order to obtain 100 ± 5% reaction efficiency. An exogenous heterologous IC, included in the qPCR kit, was added to the reaction mix in order to detect inhibition phenomena. Reaction efficiency was calculated by running the qPCR in triplicate in a 10-fold serially diluted standard curve of MTBC genomic DNA (*M. tuberculosis* H37Rv and *M. bovis* BCG Danish), according to equation (1) (Gingeras *et al.*, 2005):

$$Efficiency\ per\ cycle = \left(10^{\frac{-1}{slope}} - 1\right) x\ 100\% \qquad (1)$$

In addition, the linear range of the reaction was established, which corresponded to the dilution range in which the reaction efficiency was 100% and the coefficient of linearity (R^2) was higher than 0.997.

The analytical and diagnostic validation of the qPCR protocols was carried out according to the recommendations available in the 2018 OIE's Manual of Diagnostic Tests and Vaccines for Terrestrial Animals (OIE, 2018). The analytical validation included the assessment of the analytical sensitivity or limit of detection (LOD), repeatability, inclusivity and specificity.

The LOD and repeatability were assessed by analysing 20 replicates of the last dilution point in the standard curve for which all three replicates were positive, as well as a 1:5 and 1:10 dilution thereof. The LOD was established at the dilution in which 95% of the replicates were positive (n = 19). The *in vitro* analytical inclusivity and specificity was assessed on a panel of DNA samples (n = 76 and 108 for the *mpb70* and IS*6110* studies, respectively) obtained from pure cultures of different MTBC members, NTM and other bacteria.

The diagnostic performance of the qPCRs was validated using microbiological culture as a comparison. For the *mpb70* qPCR study, 200 tissue samples were randomly selected from samples processed between the years 2013-2017, whereas 985 tissue samples were used for the IS*6110* qPCR. Diagnostic performance estimates (diagnostic Se, Sp, agreement, likelihood ratios and predictive values) were calculated using the MedCalc statistical software 19.2.0

and WinEpi 2.0 (Blas *et al.*, 2006). The Ct cut-off for a positive reaction was calculated for the IS*6110* using the ROCR 1.0 package in R (Sing *et al.*, 2005).

The IS*6110* qPCR included two additional validation studies. In the first study, diagnostic Sp was further evaluated in 100 bovine samples from 78 bTB-free herds obtained from a low-prevalence region in northwestern Spain. The second validation corresponded to an interlaboratory comparison study in which the performance of the qPCR protocol was compared between three different national reference laboratories (RL) and one regional RL in the EU. Each laboratory processed and cultured a selection of bovine tissue samples (n = 866) according to their standard operating procedures (SOPs), whereas the DNA extraction and qPCR protocols were the same as the ones presented in this thesis, although minor modifications were carried out in certain cases due to biosafety and technical requirements of the laboratory.

3.1.3. **Tissue processing, microbiological culture and DNA extraction**

All samples were processed under BSL3 conditions for DNA extraction and microbiological culture according to the SOPs of VISAVET Health Surveillance Centre. The selected tissues were visually inspected for lesions compatible with TB and 2-2.5 g were homogenised in 12 mL of sterile PCR-grade water. One mL of the tissue homogenate was collected for DNA extraction and 10 mL were decontaminated using HPC (0.75% w/v) and cultured in Coletsos and LJ media supplemented with sodium pyruvate. The remaining homogenate was frozen as backup material.

The intracellular nature of tuberculous bacilli, as well as their rigid cell wall and encapsulation within granulomas, makes the extraction of genetic material a complex procedure that requires very aggressive DNA extraction protocols. In this aspect, a modified commercial DNA extraction protocol (QIAGEN DNeasy Blood & Tissue Kit) was used for DNA extraction; in order to achieve a high DNA extraction, this protocol included two mechanical lysis steps (bead-beating) and one biochemical lysis step in between (proteinase K).

3.1.4. **Analysis of discrepant results**

When DNA amplification was detected in culture-negative samples, the samples were further tested using another method. In the *mpb70* qPCR study, DVR-spoligotyping and PCR amplification and Sanger sequencing of the 16S rRNA gene with universal primers were used as the resolver test, whereas the *mpb70* qPCR was used in the IS*6110* qPCR study. Furthermore, the DNA sample was diluted five-fold (*mpb70* qPCR study) or ten-fold (IS*6110* qPCR study) when PCR inhibition was detected [either high Ct values (i.e. ≥ 33) or absence of amplification of the IC]. Samples were considered true-positives if amplification was detected after dilution or if the presence of MTBC DNA was confirmed using the resolver technique.

3.2. **Molecular epidemiology studies of bTB using WGS**

Several variant calling pipelines are available for the analysis of TB in animals, but their overall performance has not been evaluated in the scientific literature. As a result, there was a need to assess the currently available tools for this purpose. Therefore, this section was divided into two studies; one focusing on the performance comparison of variant calling tools for bTB, and another one focusing on the application of one of the tools in a molecular epidemiology study of bTB in Spain.

3.2.1. **Performance evaluation of variant calling tools for bTB**

Similar to the development of diagnostic methods, WGS workflows require a validation process in order to ensure that they function in accordance with their intended purposes. In the context of variant calling tools, this means that the available pipelines should be able to detect all of the variants (SNPs) in a genome and provide the same results.

The first requirement for the validation of variant calling pipelines is a well characterised reference material. The use of a group of epidemiologically related isolates in which all of the variants have been identified is considered the best possible starting scenario. However, in the absence of such a reference material, laboratories frequently rely on simulated or artificial datasets. Due to the lack of a well-characterised *M. bovis* dataset, the latter approach was selected for this study.

Once the dataset was generated, the validation process included a comparison of the performance (recall and precision), agreement between the SNPs detected by the different pipelines and the congruency of the phylogenetic trees inferred from the output files. Due to the impact that different filters can have on the final results, the analyses were repeated after the application of several filtering steps.

This section will briefly describe the procedures carried out for the evaluation of bTB variant calling pipelines. More detailed descriptions can be found in the published study:

Lorente-Leal V, Farrell D, Romero B, Alvarez J, de Juan L, Gordon SV. Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. Front Vet Sci. 2021;8:780018. https://doi.org/10.3389/fvets.2021.780018.

3.2.1.1. **Pipeline selection**

In order to identify currently available variant calling tools for bTB, a literature search was carried out in the Pubmed database and Github, an open-source repository for informatics tools. Three different variant calling pipelines were selected; vSNP (USDA-APHIS, USA), BovTB (APHA, UK) and SNiPgenie (University College Dublin, RoI). An additional pipeline, MTBseq (Research Center Borstel - Leibniz Lung Cancer, Germany), which is frequently used in the study of human TB isolates, was selected as a comparator tool.

3.2.1.2. **Artificial dataset generation**

In order to obtain a dataset that reflected a real-life situation as much as possible, an artificial set of reads was generated from an already published phylogeny. This phylogeny had been inferred from a collection of *M. bovis* isolates obtained from a bTB high prevalence setting in the RoI (Crispell *et al.*, 2020). The raw VCF files containing all of the identified variants in the original dataset were extrapolated using SimuG to their corresponding positions in the reference genome of *M. bovis* AF2122/97 (NC_02945.4) to generate a set of artificial genomes representing the isolates in the dataset (Yue & Liti, 2019). These artificial genomes were then fragmented into an artificial set of FASTQ reads using ArtificialFASTQGenerator (Frampton & Houlston, 2012).

3.2.1.3. Pipeline performance and agreement evaluation

The artificial set of reads was analysed with the four pipelines using default settings except for SNiPgenie. Performance was assessed by comparing the number of SNPs in the VCF files output by the pipelines and the SNPs present in the VCF files generated in the simulation, which was considered the reference. Relative recall rate [true positives/(true positives + false negatives)] and precision [true positives/(true positives + false positives)] with respect to the reference VCFs were evaluated using the Haplotype Comparison Tool (Som.py). False positive (FP) SNPs were defined as those SNPs identified by the pipelines that were not present in the simulated genome, whereas false negative (FN) SNPs corresponded to those SNPs in the artificial genome that were not detected by the pipelines.

In addition to the above, the agreement between the pipelines and the reference, as well as the number of homoplasies, were evaluated using VennDiagram in R (Chen, 2022) and HomoplasyFinder (Crispell *et al.*, 2019a; Chen, 2022).

3.2.1.4. Evaluation of phylogenetic congruency between pipelines

The variants detected by variant calling pipelines are usually translated into phylogenetic trees to obtain a visual representation of the genetic relatedness between isolates. In this study, RAxML was used to infer maximum likelihood trees from the multi-FASTA alignments output by the different pipelines and from the simulation (Stamatakis, 2014).

Discrepancies in variant calling are of importance since they could alter the phylogenetic relationships inferred in downstream analyses. Furthermore, due to the large amount of positions analysed through WGS, even small FN and FP rates can be relevant, considering the limited genetic diversity of *M. bovis* and other MTBC members. As a result, the congruency of the phylogenetic trees inferred from the files output by the different pipelines and from the simulation was evaluated using Treespace and Phytools (Revell, 2012; Jombart *et al.*, 2017). Treespace compares the Robinson-Foulds metric between different phylogenetic trees, which is a simplified metric used to evaluate topological differences between phylogenetic trees. The software then represents these pairwise distances using a Principal Component Analysis. In contrast, Phytools was used to compare the different trees in a facing manner, allowing for the visualisation of the topological differences suggested with Treespace. Due to

the large amount of bootstrap replicates (n = 100) per analysis, the comparison with Phytools was only carried out on the ML best trees output by RAxML.

3.2.1.5. Effects of filtering in performance and phylogenetic inference

In addition to the sequencing process, erroneous calls (FN and FP) can be introduced as a result of the inability of sequence aligners to map certain regions of the genome, such as repetitive sequences or low complexity regions. In addition, the presence of recombination or highly polymorphic regions can lead to homoplasies, which can alter phylogenetic conclusions drawn from sequence alignments (Nakanishi *et al.*, 2013).

In order to correct these limitations, variant calling pipelines usually include two filtering steps: 1) proximal SNP removal and 2) annotation-based filtering. The former consists of the removal of SNPs that are located within a certain window length from each other in order to identify SNP-dense regions, which could be indicative of recombination events or highly polymorphic regions (Crispell *et al.*, 2019a; Saltykova *et al.*, 2019). Since vSNP and BovTB do not include proximal filtering in their code by default, this process was carried out using an *in house* script. The proximal window was set at 10 bp for all pipelines (Crispell *et al.*, 2020).

Annotation-based filtering consists in the masking of regions of the genome in which erroneous base calling and misalignments are considered to be more frequent (Modlin *et al.*, 2021). In order to identify regions prone to erroneous variant calling, the output VCFs were annotated using SnpEff. These regions were then divided into three categories: 1) PE/PPE encoding genes, 2) mobile genetic elements and 3) other repetitive elements (including Direct Repeats or the *pks12* gene).

Performance, agreement and phylogenetic congruence was re-evaluated after applying the above mentioned filters, as well as the pipeline-specific default filters.

3.2.2. Evaluation of the genetic diversity of bTB chronically-infected herds in Spain using whole genome sequencing

The aim of this research project was the application of WGS in molecular epidemiology studies of bovine herds chronically infected with bTB as a means of establishing possible causes of infection. A set of bovine herds with a long history of bTB infection and *M. bovis* isolation was selected. Since transmission of bTB can take place from wildlife reservoirs, samples from wildlife animals were

also obtained in order to assess the genetic similarity between the domestic and wildlife strains. The selected samples were processed and cultured to obtain pure mycobacterial isolates for WGS analyses. Short read sequencing was carried out on the extracted genomic DNA and the obtained reads were analysed using a variant calling pipeline (vSNP). The observed genetic differences were then evaluated in order to assess *M. bovis* diversity in the affected herds and compare it with different risk factors, such as animal movements or wildlife contacts.

The following section will briefly describe the methodology used in this study. A more detailed description of the methods can be found in the results section, where the findings of this study have been made available in article format (under preparation).

3.2.2.1. Sample selection

A set of bovine herds (n = 22) with a history of isolation of *M. bovis* were selected. The genotype of the *M. bovis* isolates in this study was spoligotype SB0121, which is the most prevalent genotype in Spain (Rodriguez *et al.*, 2010). In order to assess the variation of *M. bovis* during the breakdowns, samples were selected at a minimum interval of three months to allow for consecutive tests during the same year.

Samples from wildlife species from the same areas in which *M. bovis* SB0121 had been previously isolated were selected in order to assess genetic similarity of these strains with those obtained from cattle herds.

3.2.2.2. Sample processing and microbiological culture

Ninety-two and 73 samples from bovines and wildlife species selected for this study were collected in the context of the Spanish bTB eradication campaigns between the years 2003 to 2018. The isolation of *M. bovis* was carried out from archived frozen tissue homogenates using the SOP for the isolation of MTBC members from bovine tissues of VISAVET. Once MTBC growth was detected, a loop-full of bacteria was inoculated in 10 mL of Middlebrook 7H9 liquid medium supplemented with OADC and sodium pyruvate and incubated for 4-6 weeks at 37 ºC. Seventy and 29 *M. bovis* isolates were recovered from the processed cattle and wildlife samples, respectively.

3.2.2.3. DNA extraction and whole genome sequencing

Five hundred µL of culture were washed and DNA was extracted using bead-beating and PCI, and purified using ethanol. The obtained genomic DNA was then sent to the National Veterinary Services Laboratory (NVSL) in Ames, Iowa (USA) for Illumina short read sequencing using a Miseq sequencer (2 x 250 bp reads).

3.2.2.4. Read data processing, variant calling and phylogenetic inference

The quality of the raw read files was assessed and low quality and duplicated reads were removed using an *in house* read processing pipeline (see section 4.2.2). Spoligotyping data was verified using SpoTyping (Xia *et al.*, 2016), and the processed reads were then input in vSNP for variant calling, using the same filtering approach as indicated in the pipeline performance evaluation. During the second script, vSNP will recover a position from all samples if one good quality SNP is detected in at least one of the samples. Therefore, there is a chance that a low quality position (e.g. with low mapping quality or MQ) will be introduced from the remaining samples. Therefore, the obtained SNP files were manually curated and low quality positions or problematic regions were added to the filtering files.

The multi-FASTA alignment files were analysed using RAxML to obtain ML phylogenetic trees, and tree visualisation and editing was carried out using the Interactive Tree of Life 6.5.2 (Letunic & Bork, 2021). The pairwise SNP differences between isolates were extracted using MEGA X (Kumar *et al.*, 2018)and used to assess the genetic diversity in the different herds and municipalities, as well as to evaluate the genetic relationship with wildlife species.

3.3. Identification of NTM through MALDI-TOF MS

A comparative study was designed to evaluate the capacity of MALDI-TOF MS to identify NTM species compared to Sanger sequencing. Samples from different animal species with suspected mycobacteriosis were processed and cultured. The obtained NTM isolates were processed in parallel using a protein extraction protocol devised for MALDI-TOF MS analysis and a DNA extraction protocol for PCR amplification and Sanger sequencing. The identification obtained by the two methods for each isolate were compared to establish the percentage of high quality identifications and the agreement between MALDI-TOF MS and the reference method (Sanger sequencing).

The following section will briefly describe the methodology used for the fulfilment of this objective. More detailed descriptions of the procedures can be found in the published study:

> Lorente-Leal V, Liandris E, Bezos J, Pérez-Sancho M, Romero B, de Juan L. MALDI-TOF Mass Spectrometry as a Rapid Screening Alternative for Non-tuberculous Mycobacterial Species Identification in the Veterinary Laboratory. Front Vet Sci. 2021;8:780018. https://doi.org/10.3389/fvets.2022.827702

3.3.1. Sample selection

The samples included in this study were selected from a collection of samples obtained from animals with suspected mycobacteriosis (i.e. tuberculosis and other mycobacterial infections) that were submitted to the mycobacterial unit in VISAVET for routine diagnosis between the years 2011 and 2015. A total of 6 reference strains and 69 isolates from 20 different animal species were included, with the majority of samples corresponding to bovines, wild boar and red deer.

3.3.2. Sample processing and culture

Samples were processed and cultured in the BSL3 facilities of VISAVET in accordance with the protocols mentioned in previous sections. Once growth was identified, colonies were tested using an end-point PCR that amplifies a fragment of the *mpb70* gene and through DVR-spoligotyping in order to discard the presence of MTBC members.

3.3.3. DNA extraction, amplification and Sanger sequencing

Samples collected for DNA extraction and Sanger sequencing were washed twice in HPLC-grade water and heat-inactivated at 100 ºC for 30 minutes. DNA extracts were used for PCR amplification and Sanger sequencing of selected genes; the 16S rRNA gene was used for both RGM and SGM, the *hsp65* short fragment was used for SGM, the *hsp65* long fragment was used for *M. avium* and *M. intracellulare*, and the *rpoB* gene was used for RGM.

All sequence data were manually curated and compared against the NCBI nucleotide archive using BLAST for species identification. An identification was considered valid when a single taxon presented a sequence identity higher than 99% for the 16S rRNA and *hsp65* genes, and higher than 98% for the *rpoB* gene. When sequence identity of an isolate was low or more than one taxon showed a similar identity, the *hsp65* and *rpoB* genes were also sequenced. If species identity could not be established with confidence, the identification of these isolates was approximated to the genus or complex level.

3.3.4. **Protein extraction and MALDI-TOF MS**

The samples collected for proteomic analyses were centrifuged, resuspended in HPLC-grade water and heat-inactivated at 100 ºC for 30 minutes, following the recommendations for the manipulation of NTM isolates indicated by global advisory bodies such as the Centers for Disease Control and Prevention (Alcaide *et al.*, 2018). The inactivated samples were incubated overnight at -20 ºC after the addition of 96% ethanol. After centrifugation, samples were let to air-dry at room temperature until residual ethanol was evaporated. A combination of sonication and vortexing was then used to further enhance protein extraction. Finally, an equal volume of 70% HPLC-grade formic acid was added to the samples and, after centrifugation, one µL of the supernatant was spotted on a polished steel target plate and let to air-dry.

The inoculated samples were then covered with HPLC-grade HCCA matrix and plates were analysed on a Bruker Daltonik UltrafleXtreme MALDI TOF/TOF system. Mycobacterial spectra were compared against the commercial Mycobacteria Library version 3.0. The quality of the identifications (IDs) was based on both the MALDI-TOF score as well as the consistency of the ID (i.e. the taxon ID and score of the first five results). The thresholds used to consider IDs as of high-confidence, low-confidence or as non-reliable were 1.8 for high-confidence IDs, between 1.6 and 1.8 for low-confidence IDs and below 1.6 for non-reliable IDs (Alcaide *et al.*, 2018).

4. Results

4.1. Objective 1: development of novel diagnostic tools for the detection of MTBC members

4.1.1. Development and validation of a real-time PCR for the detection and quantification of MTBC members in bovine tissue samples

In this study, a qPCR targeting the *mpb70* gene was designed and validated for its use as a direct detection and quantification method of MTBC members in bovine tissues. After optimisation of the qPCR conditions, the diagnostic performance was evaluated in 200 fresh bovine tissue samples which were tested in parallel using microbiological culture and qPCR.

The LOD with 95% confidence of the *mpb70* direct qPCR was established at 50 fg/reaction (10 genomic equivalents) and the linear range of the reaction spanned between 5 ng/reaction to 10 fg/reaction, with an R^2 of 0.999. No cross-reaction was detected during the analytical Sp validation. Diagnostic Se and Sp values were, respectively, 88.41% and 92.37%, with respect to culture and agreement between methods was very good ($\kappa = 0.802$). Samples positive to both methods were more frequent among samples with VLs (n = 58 out of 61), whereas samples negative to both methods were mainly found in samples with NVLs (n = 111 out of 121).

The true positive status was determined on half (n = 9) of the samples with discordant results between direct qPCR and culture, either by re-extraction of the remaining tissue homogenate or by the detection of MTBC DNA using different molecular methods, such as DVR-spoligotyping. After confirmation, diagnostic Se and Sp increased to 94.59% and 96.03%, respectively and agreement between PCR and true positive status increased to 0.90.

The *mpb70* real-time PCR offers a comparable diagnostic performance to that obtained by culture in a significantly reduced amount of time, and therefore has the potential to be used as a valuable confirmation method in bTB eradication campaigns.

This article was accepted for publication on the 12th of February 2019 in Frontiers in Veterinary Science.

Journal impact factor in 2019: 2.245

Position in JCR index 2019: Veterinary Sciences 19/141 (Q1).

Lorente-Leal V, Liandris E, Castellanos E, Bezos J, Domínguez L, de Juan L & Romero B, 2019. Validation of a Real-Time PCR for the Detection of *Mycobacterium tuberculosis* Complex Members in Bovine Tissue Samples. Frontiers in Veterinary Science 6. https://doi.org/10.3389/fvets.2019.00061.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 04 March 2019
doi: 10.3389/fvets.2019.00061

Validation of a Real-Time PCR for the Detection of *Mycobacterium tuberculosis* Complex Members in Bovine Tissue Samples

Victor Lorente-Leal[1,2], *Emmanouil Liandris*[1], *Elena Castellanos*[3], *Javier Bezos*[1,2], *Lucas Domínguez*[1,2], *Lucía de Juan*[1,2] *and Beatriz Romero*[1]*

[1] *VISAVET Health Surveillance Center, Complutense University of Madrid, Madrid, Spain,* [2] *Animal Health Department, Veterinary Faculty, Complutense University of Madrid, Madrid, Spain,* [3] *Exosome Diagnostics Inc., Waltham, MA, United States*

OPEN ACCESS

Edited by:
Adrian Allen,
Agri-Food and Biosciences Institute (AFBI), United Kingdom

Reviewed by:
Katarina Oravcova,
University of Glasgow, United Kingdom
Isobella Honeyborne,
University College London, United Kingdom

***Correspondence:**
Beatriz Romero
bromerom@visavet.ucm.es

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *29 June 2018*
Accepted: *12 February 2019*
Published: *04 March 2019*

Citation:
Lorente-Leal V, Liandris E, Castellanos E, Bezos J, Domínguez L, de Juan L and Romero B (2019) Validation of a Real-Time PCR for the Detection of Mycobacterium tuberculosis Complex Members in Bovine Tissue Samples. Front. Vet. Sci. 6:61. doi: 10.3389/fvets.2019.00061

Although the post-mortem diagnosis of bovine tuberculosis is mainly achieved through microbiological culture, the development of other techniques to detect *Mycobacterium tuberculosis* complex (MTBC) members directly from tissue samples has been pursued. The present study describes the development, optimization and validation of a Real-Time PCR based on the *mpb70* gene to detect MTBC members in clinical tissue samples from cattle. Specific primers and a hybridization probe were used to amplify MTBC-specific sequences in order to avoid cross-reaction with non-MTBC species. An Internal Amplification Control (IAC) was included in order to assess the presence of PCR inhibitors in the samples. The PCR was optimized to achieve maximum efficiency, and the limit of detection, limit of quantification and dynamic range of the reaction were determined. The specificity of the reaction was tested against 34 mycobacterial and non-mycobacterial species. The diagnostic sensitivity, specificity and positive and negative predictive values (PPV and NPV) of the method were assessed on 200 bovine tissue samples in relation to bacteriological culture. The dynamic range of the reaction spanned from 5 ng/reaction (10^6 genome equivalents) to 50 fg/reaction (10 genome equivalents). The efficiency of the reaction was 102.6% and the achieved R^2 was 0.999. The limit of detection with 95% confidence was 10 genome equivalents/reaction. No cross-reactions with non-MTBC species were observed. The diagnostic sensitivity and specificity values of the *mpb70* specific Real-Time PCR respect to culture were 94.59% (95% CI: 86.73–98.51%) and 96.03% (95% CI: 90.98–98.70%), respectively, with a PPV of 93.33% (95% CI: 85.55–97.07%) and a NPV of 96.80% (95% CI: 92.10–98.74%). The concordance of the Real-Time PCR based on *mpb70* is comparable to that of culture (K = 0.904) showing a great potential for the detection of members of the MTBC in animal tissues.

Keywords: real-time PCR, *Mycobacterium tuberculosis* complex, tuberculosis, detection, bovine tissue

INTRODUCTION

Bovine tuberculosis (bTB) is a chronic infectious disease caused by members of the *Mycobacterium tuberculosis* complex (MTBC), which affects certain species of mammals including cattle (1). Within this group of bacteria, *M. bovis* followed by *M. caprae* are the most frequent species in bovines. Due to its zoonotic potential and to the economic importance of cattle in the EU, this disease is subject to well-established national eradication campaigns in Member States. According to the legislation in force, i.e., 64/432/ECC, the intradermal tuberculin test is the official test in order to classify TB free herds, areas or countries, and microbiological culture is the method of confirmation of MTBC infections in bovine tissues.

The reported recovery rates for culture in general oscillate between 30 and 95% (2–5) while in a recent study using a Bayesian approach, the diagnostic sensitivity and specificity of culture was 78.1 and 99.1%, respectively (6). This variation between studies can be explained by different factors associated with the technique and the samples, which can affect the performance of the method. Firstly, the choice of tissue samples at the abattoir is a key for culture. Abnormal lymph nodes and parenchymatous organs with bTB-compatible lesions must always be included when present. If pathological lesions are not detected then, specific lymph nodes (retropharyngeal, bronchial, mediastinal, supramammary, mandibular, and mesenteric) should be taken for examination and culture. Secondly, the preservation of samples until culture by refrigeration or freezing, together with the step of chemical decontamination, is mandatory in order to decrease the risk of contamination with other microorganisms. Inadequate storage and sample treatment influence the viability of MTBC and can promote the growth of contaminating microorganisms (2). Thirdly, the type of culture media chosen to grow mycobacteria may influence the recovery rate of microbiological culture. MTBC growth can be detected either by colony formation in agar and egg-based solid media (such as Middlebrook 7H10/7H11, Stonebrink or Löwenstein-Jensen with sodium pyruvate), or fluorescence or pressure differences in liquid media (BACTEC 460 TB and MGIT 960 and VersaTREK system). In studies comparing both culture systems, the recovery rates for liquid media are higher than those reported for solid media with values of 80 to 95% and 65 to 82%, respectively (3, 5). The highest recovery rates within liquid systems are recorded for the BACTEC 460 TB system, which is no longer commercially available. In addition, there is a suspected decrease in selectivity of the MGIT 960, which in turn makes liquid media more prone to overgrowth by rapidly growing microorganisms (4, 5). Members of the MTBC are grouped within the slow growing mycobacteria due to their slow replication cycle. As a result, culture detection of MTBC is extremely slow; around 28 days for liquid media and 43 days for solid media for a positive result (2, 3).

In order to overcome the problems associated with the recovery of MTBC by culture, detection of mycobacterial DNA from animal tissue samples using PCR is being considered as an alternative or complementary test to microbiological culture. Since the early 90's, many conventional PCRs have been developed and used for the direct detection of members of the MTBC in bovine samples (7). In those studies including fresh bovine tissue samples from animals with visible and non-visible lesions (VL and NVL), the reported sensitivity and specificity values of PCR with respect to culture showed great variability, ranging from 63 to 97%, and 50 to 97%, respectively (8–10). After the introduction of Real-Time PCR for the detection of MTBC species, sensitivity values increased with respect to conventional PCR. In those studies implementing Real-Time PCR in which bovine tissue samples with VL and NVL were analyzed, diagnostic sensitivity and specificity by Real-Time PCR ranged between 74 to 100% and 97 to 100%, respectively (6, 11–14). The variability in the values between studies depends not only on the type of PCR (conventional, nested or Real-Time PCR), but also on the PCR target (single- or multiple copy) and reagents, the type and number of samples included in the studies, and the DNA isolation methods. The largest study to date assessing the diagnostic performance of Real-Time PCR for the detection of MTBC using a Bayesian approach reported a diagnostic sensitivity of 87.7% and a specificity of 97% (6).

In this study, we describe the development and validation of a Real-Time PCR based on the *mpb70* gene, which encodes for a major antigenic protein conserved in all MTBC species. In addition, we assess its diagnostic performance in fresh bovine tissue samples obtained within the Spanish national eradication campaign.

MATERIALS AND METHODS

Real-Time PCR

PCR Design and Optimization

The *mpb70* gene was the target of this PCR since it encodes for a majorly expressed antigenic protein in *M. bovis*, which is conserved in all members of the MTBC. An *in silico* specificity analysis was carried out, in order to rule out any sequence homologies between other bacterial species, with the Basic Local Alignment Tool (BLAST) from the NCBI, using the *mpb70* CDS from *M. bovis* AF2122/97 (NC_002945.4). The *mpb70* sequence was then used to obtain *mpb70* homologs from the available MTBC genomic sequences deposited in the genbank (NCBI): *M. tuberculosis* H37Rv (NC_000962.3), *M. africanum* strain 25 (CP010334.1), *M. caprae* Allgeau (CP016401.1), *M. microti* strain 12 (CP010333.1), *M. mungi* strain BM22813 (LXTB01000090.1), *M. orygis* strain 112400015 (APKD01000057.1) and *M. canetti* CIPT 140010059 (NC_015848.1).

Oligonucleotides targeting the *mpb70* gene, specific for members of the MTBC, were designed to target a 133bp conserved amplicon with Oligo primer analysis software 6.0 (Molecular Insights, West Cascade, CO, USA): *mpb70*-forward: 5′-CTCAATCCGCAAGTAAACC-3′, *mpb70*-reverse: 5′-TCAGCAGTGACGAATTGG-3′ (15), and *mpb70*-probe: 5′- FAM-CTCAACAGCGGTCAGTACACGGT-BHQ1-3′. The amplicon sequences were obtained and aligned against the available MTBC sequences, as well as with the closest similarities obtained in the *in silico* specificity analysis (e.g., *M. kansasii, M. indicus pranii,* or *M. marinum*).

The Real-Time PCRs were carried out using the QuantiFast® Pathogen PCR + IC Kit (QIAGEN, Hilden, Germany). This kit includes an Internal Amplification Control (IAC), as well as specific reagents and primers/probes required for its amplification. It employs MAX™NHS Ester as a reporter dye. Different primer/probe concentrations and extension temperatures were tested in order to achieve maximum replication efficiency.

M. tuberculosis H37Rv DNA was used for the generation of the standard curve and positive controls. Ultra-pure distilled water was used as negative controls. This strain was grown in Löwenstein-Jensen slants in the BSL3 facilities at VISAVET Health Surveillance Center. A loop full of colonies was collected and heat inactivated (100°C) in 200 µl of ultra-pure distilled water during 15 min.

The efficiency and dynamic range of the reaction were assessed in triplicates using a standard curve prepared from a stock of 10 ng/µl of *M. tuberculosis* H37Rv genomic DNA, 10-fold serially diluted to a range of 1 ng/µl to 0.1 fg/µl. DNA concentration and quality of the DNA solution were measured in ten replicates using a nano-drop spectrophotometer (ThermoFisher, Waltham, MA, USA). The dynamic range of the reaction was established as the range of standard curve concentrations at which the coefficient of linearity was >0.997 and the cycle separation between the 10-fold dilutions was close or equal to 3.32 cycles. The limit of quantification was established as the lowest concentration point of the dynamic range of the reaction.

The optimized setup with a final 25 µl volume per reaction, including the Internal Amplification Control (IAC) was: 5 µl of 5x Quantifast Pathogen Master Mix, 2.5 µl of 10x IAC assay, 2.5 µl of 10x Internal Control DNA, 2 µl of 10 pmol/µl *mpb70*-Forward primer, 2 µl of 10 pmol/µl *mpb70*-Reverse primer, 0.75 µl of 10 pmol/µl *mpb70*-probe, 5.25 µl of ultrapure sterile distilled water (Sigma-Aldrich, St. Louis, MO, USA), and 5 µl of DNA sample. Primers and probe were obtained from Eurofins Genomics (Ebersberg, Germany).

All PCR reactions were carried out in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) according to the following optimized cycling conditions; 95°C for 5 min followed by 45 2-step cycles of 95°C for 15 s and 60°C for 30 s, with data acquisition at this step.

Analytical Specificity and Sensitivity

The inclusivity of the PCR was tested against seven species of the *M. tuberculosis* complex: *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. bovis* BCG, *M. caprae*, *M. microti*, and *M. pinnipedii*. Selectivity was assessed using a panel of 69 strains from 24 Non-Tuberculous Mycobacteria (NTM) species and 10 non-mycobacterial species (OM: Other Microorganisms); *M. avium* subsp. *hominissuis* ($n = 7$), *M. avium* subsp. *avium* ($n = 3$), *M. avium* group X ($n = 10$), *M. chitae*, *M. colombiense*, *M. europeum*, *M. flavescens*, *M. fortuitum* ($n = 3$), *M. gordonae*, *M. hibernae*, *M. intracellulare*, *M. kansasii* ($n = 9$), *M. marinum* ($n = 2$), *M. neoaurum*, *M. nonchromogenicum* ($n = 4$), *M. parascrofulaceum*, *M. peregrinum* ($n = 2$), *M. phlei* ($n = 2$), *M. scrofulaceum*, *M. seoulense*, *M. shimodei*, *M. smegmatis* ($n = 2$), *M. terrae*, *M. thermoresistible*, *M. vaccae*, *Brucella mellitensis*, *Brucella abortus*, *Salmonella enterica* Sv. Typhimurium, *Serratia maucencens*, *Rhodococcus equi*, *Enterococcus hirae*, *Lysteria monocytogenes*, *Nocardia* sp., *Streptomyces* sp., and *Corynebacterium pseudotuberculosis*. DNA for these bacteria was obtained from reference strains and clinical isolates from the VISAVET Health Surveillance Center (Complutense University of Madrid).

The analytical sensitivity or limit of detection (LOD) and intra-assay repeatability were estimated using a new standard curve that was prepared from a 10-fold serially diluted stock of *M. tuberculosis* H37Rv genomic DNA ranging from 1 ng/µl to 10 fg/µl, and one 1:5 dilution thereof to a concentration of 2 fg/µl (10 fg/reaction or 2 genomic equivalents). The reaction was carried out using 20 replicates per concentration and the LOD was established as the concentration in which at least 95% of the replicates were positive. Inter-assay repeatability was assessed in 20 replicates of 10 fg/µl *M. tuberculosis* DNA in a period of 6 months.

M. tuberculosis H37Rv genomic equivalents were obtained from the amount of DNA used for each point of the standard curve using the equation previously described (16): [ng of DNA $\times$ 6.023 $\times$ 10^{23} molecules/mol]/[bp length of genome $\times$ 10^9 ng/g $\times$ 660 g/mol]. The genome size recorded at the NCBI Genome entry of *M. tuberculosis* H37Rv (NC_000962.3) was used as a reference (i.e., 4.41 Mb). Genomic equivalents for each sample were obtained by extrapolating the Ct values with the quantities from the standard curve.

Selection, Preparation and Culture of Clinical Samples

Two-hundred fresh tissue samples from cattle were randomly selected from samples processed as part of the Spanish national eradication campaign against bTB during the period 2013–2017, based on the Royal Decree 727/2011. Processing took place within the BSL3 facilities of VISAVET Health Surveillance Center. Simple randomization was carried out by assigning a random value to each sample and by sorting them by increasing order. The first 200 samples of this list were included in the study. The selected tissue samples originated from 11 out of the 17 autonomic regions of Spain. Almost half ($n = 99$) of the samples were obtained from Madrid, followed by Castile-La Mancha ($n = 34$), Aragon ($n = 18$), Extremadura ($n = 13$), Valencia ($n = 12$), Murcia ($n = 8$), La Rioja ($n = 6$), Andalusia ($n = 3$), Canary Islands ($n = 3$), Balearic Islands ($n = 3$), and Castile and Leon ($n = 1$).

From the total amount of samples, 118 came from cattle that were positive to the single intradermal tuberculin (SIT) test (bovine PPD $\geq$ 4 mm), whereas 63 were from SIT-negative animals (bovine PPD $\leq$ 2 mm) and 4 had inconclusive results (2 mm $>$ bovine PPD $<$ 4 mm) according to the Royal Decree 727/2011. Following the regulation in force, in regions with high prevalence of bTB, SIT inconclusive results were considered as positive. Fifteen animals showed bTB-compatible lesions during routine abattoir inspection of carcasses and were also sent for sample collection and processing. Lymph nodes (retropharyngeal, mandibular, mediastinal, bronchial,

prescapular, mesenteric, hepatic, and/or supramammary) and/or organs were then collected for processing, culture and direct PCR.

Once in the laboratory, all tissue samples were visually inspected for lesions and sliced. A total of 78 samples had VL, whereas 122 had NVL. Approximately 2–2.5 g of tissue sample from the same animal were pooled and homogenized in 12 ml sterile distilled water in a Masticator (IUL, Barcelona, Spain) at max speed for up to 5 min. One ml of the homogenized sample was collected for DNA isolation, whereas the remainder of the homogenate was decontaminated with an equal volume of 0.75% (w/v) hexadecyl pyridinium chloride solution in agitation during 30 min (17). Samples were centrifuged during 30 min at 1,300–1,500 g. Pellets were collected with swabs and cultured in Löwenstein-Jensen with sodium pyruvate and Coletsos media (Difco, Spain) at 37°C for a maximum of 3 months. Culture was considered positive when isolates were identified as MTBC by conventional PCR (18) and /or DVR-spoligotyping (19).

Tissue DNA Extraction

DNA from tissues was obtained using the DNeasy Blood & Tissue Kit (QIAGEN) with a few modifications. Briefly, one ml of the homogenized tissue sample was added in a tube containing 100 mg of 0.5 mm and 50 mg of 0.1 mm glass beads and centrifuged for 5 min at 9,000 g. The supernatant was removed from the samples and 200 µl of sterile distilled water and 180 µl of ATL Buffer were added. Samples were then lysed in a Fastprep® FP120 homogenizer (MP Biomedicals, Santa Ana, CA, USA) using 3 cycles of 40 s at a speed of 6.5 m/sec. After an overnight chemical treatment with 20 µl of proteinase K at 56°C, the mechanical lysis step was repeated. Samples were then centrifuged briefly at maximum speed and 300 µl of supernatant were transferred to a new 1.5 ml Eppendorf tube and mixed with 400 µl of a mixture of AL buffer and 96% ethanol (equal volumes). The lysate was transferred to a spin column and was processed according to the manufacturer's instructions. DNA elution was carried out using 200 µl of AE buffer.

Diagnostic Performance

The diagnostic performance of the Real-Time PCR targeting the *mpb70* gene was assessed on 200 randomly selected tissue-extracted DNA samples. The exogenous heterologous IAC supplied with the kit was used to assess the presence or absence of inhibition phenomena. According to the manufacturer, the IAC should show Ct values of 30 ± 3. As a result, complete inhibition was defined when no IAC was amplified and partial inhibition was defined as a $Ct > 33$ for the IAC. If inhibition was detected, samples were diluted 5-fold and PCR was repeated. Results were compared against microbiological culture, and diagnostic sensitivity and specificity, Positive and Negative Predictive Values (PPV/NPVs) as well as Positive and Negative Likelihood Ratios (PLR and NLRs) were calculated using MedCalc 18.2.1 (MedCalc, Ostend, Belgium). Agreement between culture and Real-Time PCR results was assessed using Cohen's Unweighted Kappa in WinEpi 2.0 (20).

Samples with culture-negative and PCR-positive results were further analyzed by DVR-spoligotyping (detection of spacers) and sequencing of the 16S rRNA gene. Sequencing of a 1,030 bp fragment of the 16S rRNA gene (18) was carried out externally by STABvida (Lisbon, Portugal). The obtained sequences were analyzed using the Bioedit software version 7.1.3.0 (21). Samples that gave a positive result to either of the above two techniques were considered as true positives and were used to re-calculate the diagnostic performance of the Real-Time PCR. On the other hand, for samples with a culture-positive and PCR-negative results DNA extraction and PCR were repeated.

RESULTS

In silico Analysis

Sequence similarity between *mpb70* homologs in members of the MTBC is 99.7–99.8% (data not shown). Even though some non-MTBC species -such as *M. kansasii, M. marinum* or *M. gilvum*- have homologous *mpt70/mpb70* sequences (22), sequence similarity with these species is limited (data not shown). Alignments of the *mpb70* amplicons with MTBC species showed 100% identity, with exception of *M. canetti* that had a T/C substitution at position 360 (*M. bovis* AF2122/97 numbering from *mpb70* CDS start). Although this substitution falls within the length of the reverse primer, it did not affect the ability of the primer to anneal to its target in *M. canetti. M. indicus pranii, M. kansasii* and *M. marinum,* had a considerably lower identity, indicating that specificity issues would be unlikely (data not shown).

Optimization and Analytical Sensitivity and Specificity

For optimization of the PCR reaction and repeatability studies, two 10-fold diluted standard curves were prepared from a 10ng/µl stock of *M. tuberculosis* H37Rv (DNA stock concentrations with Standard Deviations or SDs of 0.35 and 0.97, respectively).

The lowest concentration of DNA detected in the standard curve by the Real-Time PCR was 10 fg/µl (50 fg/reaction or ~ 10 genomic equivalents) with all three replicates showing an amplification curve. The dynamic range of the reaction spanned from 1 ng/µl (5 ng/reaction or approx. 10^6 genome equivalents) to 10 fg/µl (50 fg/reaction or ~ 10 genome equivalents), with an R^2 of 0.999. The upper and lower Ct values of the dynamic range were 20.06 and 36.33, respectively. The quantification limit was set to 10 genome equivalents/ reaction. Replication efficiency was 102.60% with a slope of −3.27.

All 20 replicates with a concentration of 50 fg/reaction were positive for this PCR, whereas only 14/20 of the 10 fg/reaction aliquots were positive. The Ct values of both dilutions were, respectively, 37.07 (SD 0.98) and 38.92 (SD 1.28). Therefore, the limit of detection for this Real-Time PCR with a 95% confidence was 10 fg/µl (50 fg/reaction or 10 genomic equivalents) and the cut-off was set to a $Ct < 40$.

The Real-Time PCR reacted positively only against members of the MTBC and no cross-reactions were detected against any of the NTMs or non-mycobacterial species tested.

Diagnostic Performance Compared to Microbiological Culture

Two hundred DNA samples obtained from bovine tissues were analyzed using this PCR and microbiological culture. A total of 69 samples were MTBC positive for culture, whereas 131 were negative (**Table 1**). Ten out of the 131 culture-negative samples showed growth of non-tuberculous mycobacteria ($n = 4$) or other microorganisms ($n = 6$) (NTM/OM). The Real-Time PCR detected 71 positive samples, with a minimum and maximum Ct values of 24.39 and 39.35, respectively and a median Ct value of 33.48. Sixty-one out of 69 positive culture samples were also positive for the Real-Time PCR targeting *mpb70*, resulting in a sensitivity relative to culture of 88.41% (95% CI: 74.3 to 94.86%). Ten of the 131 culture-negative samples were positive for the Real-Time PCR, and the specificity value was 92.37% (95% CI: 86.41 to 96.28%). Of the 10 cultures showing growth of NTM/OM, one reacted positively to the direct Real-Time PCR.

The exogenous heterologous IAC used in this PCR detected complete inhibition in only 4 out of 200 samples (2%) and partial inhibition (IAC Ct > 33) in 15 out of 200 samples (7.5%). After dilution, all these samples were PCR negative. One of the completely inhibited samples and 3 of the partially inhibited samples were positive to culture.

PPVs and NPVs were 85.92% (95% CI: 76.97 to 91.76%) and 93.80% (95% CI: 88.72 to 96.67%), respectively. The positive and negative likelihood ratios were, respectively, 11.58 (95% CI: 6.34–21.14) and 0.13 (95% CI: 0.07–0.24). There was a very good correlation between culture and PCR results (Cohen's Unweighted Kappa = 0.802).

Samples with discording results between the two methods used were further analyzed. DNA isolation was repeated for the 8 culture-positive PCR-negative samples. Of these, half ($n = 4$) gave a positive result. For samples with culture-negative and PCR-positive results ($n = 10$), spoligotyping and 16S RNA sequencing were applied and the presence of MTBC DNA was confirmed in 5 of them. Of these, one presented growth by an actinomycete and 4 were negative to culture. These samples, in addition to all culture-positives, were considered to be true positives. As a result, the corrected relative sensitivity and specificity of PCR was calculated to be 94.59% (95% CI: 86.73% to 98.51%) and 96.03% (95% CI: 90.98–98.70%), respectively (**Table 1**). PPVs and NPVs were, then, 93.33% (95% CI: 85.55–97.07%) and 96.80 % (95% CI: 92.10–98.74%). PLRs and NLRs increased to 23.84 (95% CI: 10.08–56.37) and 0.06 (95% CI: 0.02–0.15), respectively. Correlation between culture and PCR increased to 0.904.

Among samples with VL ($n = 78$), 65 and 61 were positive to PCR and culture, respectively (**Table 2**). Three out of 7 culture-negative and PCR-positive samples were shown to contain MTBC DNA by sequencing or spoligotyping. Although 3 culture-positive samples were negative for this PCR, they became positive after the extraction protocol was repeated. Regarding NVL samples (n=122), a total of 6 samples were positive to PCR whereas 116 were found to be negative. In contrast, 8 NVL samples were culture-positive and 114 samples were culture-negative. Of these culture-negative samples, 3 were positive for the Real-Time PCR, of which 2 were confirmed as true positives by sequencing or spoligotyping. On the other hand, 5 culture-positive samples were negative for the *mpb70*-specific PCR. However, one of them was positive after the repetition of the extraction protocol. After confirmation of the true positives, Cohen's Unweighted Kappa between culture and PCR for VL and NVL samples was, respectively, 0.804 and 0.685.

Intra and Inter-Assay Variation

The intra-assay repeatability at a concentration of 10 fg/μl showed an average Ct value of 37.07 with a standard deviation of 0.98 and a coefficient of variation of 2.63%. Inter-assay repeatability using 20 replicates from a stock of 10 fg/μl in a 6 month period showed an average Ct value of 36.70 with a SD of 1.40 and a CV of 3.82%.

DISCUSSION

The purpose of this study was the design, optimization, and validation of the *mpb70* Real-Time PCR for the detection of members of the *M. tuberculosis* complex directly from animal tissue samples. In addition, this study compared the diagnostic performance of this PCR and bacteriological culture using a large number of bovine tissue samples ($n = 200$) collected in the framework of the Spanish bTB eradication program.

The Real-Time PCR targeting the *mpb70* gene showed 100% of inclusivity and selectivity. Moreover, it shows good replication efficiency (102.6%), and an analytical sensitivity of at least 10 genome equivalents with 95% confidence. Furthermore, very little variation was seen at the LOD both within and between assays (CV=2.63 and 3.82%, respectively). In addition, the linear range of the reaction spans from 5 ng/reaction (approximately 10^6 genomic equivalents) to 50 fg/reaction (~ 10 genomic equivalents). Although this PCR was developed for the detection of MTBC, the single-copy nature of the target and the wide linear range of the reaction make this PCR a suitable candidate for absolute quantification studies of MTBC in tissues. In fact, 59 out of the 71 *mpb70* PCR-positive samples showed a Ct value within the dynamic range of the reaction (data not shown). Although the quantification was not possible due to the absence of the standard curve in all runs, the range of concentrations was estimated to be between 2.29×10^5 and 63 genomic equivalents, with an average Ct value of 33.33 (~ 415 genome equivalents).

Overall, there was a good correlation between microbiological culture and PCR results in this study. Furthermore, diagnostic sensitivity and specificity values were very good when compared to microbiological culture (88.41 and 92.37%, respectively). Eight samples were negative to the direct Real-Time PCR but positive to microbiological culture. After repetition of the DNA extraction protocol, half of them became positive to the PCR. This implies that the DNA extraction protocol is very important and directly affects the sensitivity of the PCR. Several factors influence the DNA yield and quality obtained through DNA extraction protocols.

TABLE 1 | Comparison of results obtained by analyzing 200 randomly selected cattle samples by microbiological culture and Real-Time PCR.

		Culture/True positives*				Diagnostic performance	
		Result	+	−	**Total**	**Sensitivity**	**Specificity**
Raw results	PCR	+	61	10	71	88.41% [95% CI: 78.43–94.86%]	92.37% [95% CI: 86.41–96.28%]
		−	8	121	129		
		Total	69	131	200		
Corrected results	PCR	+	70	5	75	94.59% [95% CI: 86.73% to 98.51%]	96.03% [95% CI: 90.98–98.70%]
		−	4	121	125		
		Total	74	126	200		

Corrected results consider as true positives: (1) those samples that were culture positive, (2) samples that were culture-negative but PCR-positive, and for which MTBC presence was demonstrated by 16S sequencing and/or spoligotyping, and (3) culture-positive and PCR-negative samples that became positive after the DNA extraction was repeated.

TABLE 2 | Comparison of results obtained by analyzing 200 randomly selected veterinary samples by microbiological culture and Real-Time PCR, according to the presence or absence of anatomic lesions.

		Culture (True positives)		
		+	−	**Total**
VL	PCR +	58 (64)	7 (4)	65 (68)
	PCR −	3 (0)	10	10
	Total	61 (64)	17 (14)	78
NVL	PCR +	3 (6)	3 (1)	6 (7)
	PCR −	5 (4)	111	116 (115)
	Total	8 (10)	114 (112)	122

Culture negative and PCR-positive samples were considered true positives (in brackets) after the confirmation of the presence of MTBC DNA by 16S rRNA gene sequencing and/or spoligotyping.

Firstly, the amount and type of processed tissue could determine the bacterial load in the sample. The extraction protocol used in this study uses a volume of sample that is 1/10 the amount of sample used for microbiological culture, which could produce a loss of sensitivity due to the decreasing amount of bacteria available for extraction. In addition, the presence or absence of lesions can affect the amount of bacteria in the sample which in turn could determine the quantity of available DNA. In this study, the four remaining culture-positive and PCR-negative samples had NVLs, of which 3 were positive to the SIT test. This suggests that the animal may have been at early stages of infection and, therefore, have low bacterial loads. On the other hand, the recovered samples after the second extraction ($n = 4$) had mostly VLs ($n = 3$).

Secondly, the type of disruption technique used can have an important effect in the DNA extraction process. Even though the protocol in this study has been optimized to obtain a high amount of DNA through two mechanical and one overnight chemical lysis steps, improvements in the extraction protocol may reduce the number of discording results. Park et al. showed that increasing the incubation time before mechanical lysis with ATL buffer up to 3 h increased the DNA yield in *M. avium* subsp. *paratuberculosis* when compared to no pre-treatment (23). On the other hand, an 8-h pre-treatment was detrimental to the amount of extracted DNA, achieving the same amount of DNA as the no pretreatment controls. The effect of reduction in the pre-treatment incubation time should be assessed in the future. Another improvement could include the use of a homogenizer instead of a masticator in the tissue homogenization step, which could release a larger amount of bacteria from tissue samples for extraction.

Furthermore, several factors associated with the extraction protocol may introduce inhibitors in the sample, such as organic compounds or excess host DNA. In order to detect the inhibition of the PCR, the reaction mix includes an exogenous heterologous IAC, with a randomly generated DNA supplied by the manufacturer. By using the IAC, 4 and 15 samples were found to be completely or partially inhibited, respectively. Of these, 1 inhibited and 3 partially inhibited samples were culture-positive, and they remained PCR-negative after a 1:5 dilution. After repeating the extraction protocol on these samples, the inhibited and one partially-inhibited sample became positive, indicating that their dilution may have caused the further dilution of the target DNA and, therefore, may have resulted in the loss of sensitivity in the PCR. Furthermore, this could imply that the inhibitor was not present in the sample and was introduced as a result of the DNA extraction procedure, or that the extraction protocol failed to remove it in the first place. Other reported PCRs also include IACs, but only a few include information regarding the presence or absence of inhibition (11–13). Although no cases of inhibition were detected in these publications, they used endogenous or exogenous homologous IACs, which may present some disadvantages with respect to exogenous heterologous IACs. For instance, the amount of endogenous IAC template (i.e., bovine β-actin gene) varies depending on the type of sample or extraction method used, which means that readouts vary between samples and there is no indication of the level of inhibition present in the sample. In addition, they can overcome inhibitory effects in the sample as they are usually in higher concentrations than the target. Exogenous homologous IAC (i.e., *M. bovis* DNA), on the other

hand, are recognized by the target's primers and can, therefore, give rise to competition events that can hinder the amplification of the target DNA in low-concentrated samples, such as those close to the LOD. Exogenous heterologous IACs, such as the ones used in this study, use a consistent amount of control template, different to the target of interest, with a set amplification cycle. As a result, it allows the detection of complete or partial inhibition phenomena and minimizes competition, since the primers and the control target sequence are completely different to those of the target of interest.

Ten samples were negative to culture but positive to the Real-Time PCR. The use of spoligotyping and/or 16S rRNA sequencing on these discording samples showed the presence of MTBC DNA in 5 of them. The inability of culture to detect MTBC in these samples may be due to sample processing issues in which bacterial integrity is hampered and growth is impeded. In addition, very advanced granulomatous lesions may contain lower numbers of viable bacteria than early granulomas (24). Nevertheless, MTBC DNA can still be present in non-viable bacteria in enough quantity to be detected by PCR after purification. Although no histopathological evaluation was done on these samples, 7 of the culture-negative and PCR-positive samples were obtained from animals with VLs whereas 3 were obtained from animals with NVLs. Finally, growth of NTM/OM could be another reason for these discrepancies. In fact, 1 of the 10 tissue samples that showed growth of NTM/OM during culture was positive to this PCR, indicating that the growth of MTBC in culture could have been hampered by the growth of other NTM/OM. The detection of MTBC DNA in this sample by 16S rRNA sequencing and spoligotyping supported the analytical specificity of the *mpb70* oligonucleotides, indicating that the presence of other microorganisms in the sample will not interfere with this PCR.

When the presence of MTBC DNA was confirmed in the discording samples, these were considered as true positives. Therefore, 70 positive samples were correctly identified by PCR, increasing the diagnostic sensitivity and specificity values with respect to culture (from 88.41 to 94.59% and from 92.37 to 96.03%, respectively). PPVs and NPVs were 93.33 and 96.80%, respectively. Furthermore, the PLR was 23.84 indicating a high probability of correctly identifying a bTB-positive tissue sample. In addition, the NLR was very low (0.06), indicating a low probability of a negative result being positive.

The most commonly used genetic target for PCR detection of MTBC species is the IS*6110* transposon (25). Other targets used in the detection of MTBC members in veterinary samples through PCR include the 16S-23S rRNA Internally Transcribed Spacer or ITS (14), *hupB* (26), TbD1 (11), *rv2807* (12), and *devR* (16). The high sequence similarity between the different MTBC species and the single-copy nature of the *mpb70* gene make it also a suitable target for both detection and quantification through Real-Time PCR. Since the early 1990's, the *mpb70* gene has been used extensively for the detection of MTBC species through conventional PCR (18, 27–29). Additionally, it has been used as a target for Real-Time PCR quantification of MTBC members in infected cell culture extracts (15). However, in this study hybridization probes were added to increase specificity.

Although the diagnostic specificity of this PCR was similar (96.03 vs. 97%) to that seen for the Real-Time PCR used by Courcoul et al. targeting the IS*6110* element (6), diagnostic sensitivity was higher in this study (94.59% vs. 87.7%). When compared against a Real-Time PCR detecting the IS*6110* element based on melting curve analysis and hybridization probes (30), the *mpb70*-targeting PCR showed a better correlation with culture results and increased diagnostic sensitivity. A semi-nested Real-Time PCR targeting the IS*6110* showed very similar diagnostic sensitivity, specificity and predictive values to those obtained in this study; 100% diagnostic sensitivity, 97.7% diagnostic specificity, 96.3% PPV and 100% NPV (13). Even though the LOD is lower for this semi-nested Real-Time PCR (1.5fg vs. 50fg), the requirement of two PCR steps increases the risk of cross-contamination. In addition, a Real-Time PCR targeting the 16S-23S ITS showed a moderate diagnostic sensitivity of 73.87% (14).

It is important to consider that the diagnostic performance of this PCR in this study does not give information about the infection status of all animals included in this study, as it only compares culture and PCR on tissue samples. Based on the results of this study and previous publications, direct PCR has some advantages compared to culture for the detection of MTBC species in animal tissue samples. In the first place, PCR takes a few hours to complete in comparison to the weeks required for microbiological culture. Secondly, analytical specificity can be extremely high if the appropriate oligonucleotides are designed, limiting cross-reaction with contaminating microorganisms. This removes the requirement for a decontamination step, decreasing the hazardous conditions applied to the sample. Furthermore, it would reduce the risk of exposure to mycobacteria as it decreases the processing time of tissues with suspected MTBC infections before inactivation, the amount of time spent at BSL3 facilities and the bacterial load to which the user is exposed to. In addition, the *mpb70* PCR showed a comparable limit of detection and diagnostic sensitivity to that seen in IS*6110* PCRs. One disadvantage of the IS*6110* target over the *mpb70* is the risk of horizontal transfer of mobile elements between mycobacterial species, as has been recorded for IS*1245* and *M. kansasii* (31). Moreover, the IS*6110* is present in a variable number of copies within the genome of certain MTBC species, which limits its use in quantitative studies, unlike the *mpb70* gene, which is a single-copy gene.

The results obtained in this study open the possibility of using the direct Real-Time PCR as an alternative to microbiological culture in the short term. Although microbiological culture is still needed for bacterial isolation and molecular characterization with epidemiological purposes, PCR could decrease considerably the time needed until results are obtained, improving the decision making capacity during the eradication campaigns.

In conclusion, the Real-Time targeting the *mpb70* gene is a time-effective and efficient method for the detection of MTBC members in veterinary tissue samples, which shows improved diagnostic performance with respect to culture. In addition, it has a low detection limit of 10 genomic equivalents/reaction and of MTBC species. Furthermore, being a single copy gene and

having a dynamic range of 10^6-10 genomic equivalents/reaction, it could be used for quantification studies of as little as 10 genomic equivalents.

AUTHOR CONTRIBUTIONS

VL-L and EL performed all experiments in this study and the *in silico* specificity analysis. EC participated in the design of the *mpb70* specific oligonucleotides used in this study. BR, EL, and LdJ designed the study. LD and LdJ are responsible for the obtaining of samples. VL-L wrote the manuscript with the invaluable insights of EC, JB, EL, LD, BR, and LdJ. BR directed and supervised the complete study.

FUNDING

VL-L was funded with a predoctoral grant from the Complutense University of Madrid and Banco Santander 2017–2018.

ACKNOWLEDGMENTS

We would like to thank the excellent work performed by laboratory technicians F. Lozano, T. Alende, A. Gutiérrez, N. Moya, C. Viñolo, D. de la Cruz, and L. Jimenez, in culturing and spoligotyping. We would also like to thank the work carried out by Susana Gómez during the design of the primers and probe, and Pilar Pozo with sequencing. In addition, we would like to appreciate valuable statistical insights given by María Luisa de la Cruz and Julio Álvarez. This work was supported by the Área de Ganadería de la Comunidad de Madrid, the Ministerio de Agricultura, Pesca, Alimentación y Medio Ambiente, and the Programa de Tecnologías Avanzadas en Vigilancia Sanitaria (TAVS) de la Comunidad de Madrid (S2013/ABI2747). This work was supported by the Programa de Tecnologías Avanzadas en Vigilancia Sanitaria (TAVS) from the Comunidad de Madrid (ref. S2013/ABI-2747).

REFERENCES

1. Bezos J, Álvarez J, Romero B Juan L, Dominguez L. Bovine tuberculosis: historical perspective. *Res Veterin Sci.* (2014) 97:S3–4 doi: 10.1016/j.rvsc.2014.09.003
2. Corner LAL, Gormley E, Pfeiffer DU. Primary isolation of *Mycobacterium bovis* from bovine tissues: Conditions for maximising the number of positive cultures. *Vet Microbiol.* (2012) 156:162–71. doi: 10.1016/j.vetmic.2011.10.016
3. Hines N, Payeur JB, Hoffman LJ. Comparison of the recovery of *Mycobacterium bovis* isolates using the BACTEC MGIT 960 system, BACTEC 460 system, and Middlebrook 7H10 and 7H11 solid media. *J Vet Diag Invest.* (2006) 18:243–50. doi: 10.1177/104063870601800302
4. Price-Carter GFYM, Bland K, Joyce MA, Khan F, Surrey M, Lisle GW, et al. Comparison of the BBL mycobacteria growth indicator tube, the BACTEC 12B, and solid media for the isolation of *Mycobacterium bovis*. *J Vet Diag Invest.* (2017) 29:508–12. doi: 10.1177/1040638717697763
5. Robbe-Austerman S, Bravo DM, Harris B. Comparison of the MGIT 960, BACTEC 460 TB and solid media for isolation of *Mycobacterium bovis* in United States veterinary specimens. *BMC Vet Res.* (2013) 9:74. doi: 10.1186/1746-6148-9-74
6. Courcoul A, Moyen J-L, Brugère L, Faye S, Hénault S, Gares H, et al. Estimation of sensitivity and specificity of bacteriology, histopathology and PCR for the confirmatory diagnosis of bovine tuberculosis using latent class analysis. *PLoS ONE.* (2014) 9:e90334. doi: 10.1371/journal.pone.0090334
7. Costa P, Botelho A, Couto I, Viveiros M, Inácio J. Standing of nucleic acid testing strategies in veterinary diagnosis laboratories to uncover *Mycobacterium tuberculosis* complex members. *Front Mol Biosci.* (2014) 1:16. doi: 10.3389/fmolb.2014.00016
8. Liébana E, Aranaz A, Mateos A, Vilafranca M, Gomez-Mampaso E, Tercero JC, et al. Simple and rapid detection of *mycobacterium tuberculosis* complex organisms in bovine tissue samples by PCR. *J. Clin. Microbiol.* (1995) 33:33–6.
9. Stewart LD, McNair J, McCallan L, Gordon A, Grant IR. Improved detection of *mycobacterium bovis* infection in bovine lymph node tissue using immunomagnetic separation (IMS)-Based Methods. *PLoS ONE.* (2013)8:e58374. doi: 10.1371/journal.pone.0058374
10. Wards BJ, Collins DM, Lisle GW. Detection of *Mycobacterium bovis* in tissues by polymerase chain reaction. *Vet Microbiol.* (1995) 43:227–40. doi: 10.1016/0378-1135(94)00096-F
11. Araújo CP, Osório ALAR, Jorge KSG, Ramos CAN, Filho AFS, Vidal CES, et al. Detection of *Mycobacterium bovis* in bovine and bubaline tissues using NEsted-PCR for TbD1. *PLoS ONE.* (2014) 9:e91023. doi: 10.1371/journal.pone.0091023
12. Araújo CP, Osório ALAR, Jorge KSG, Ramos CAN, Filho AFS, Vidal CES, et al. Direct detection of *Mycobacterium tuberculosis* complex in bovine and bubaline tissues through nested-PCR. *Brazil J Microbiol.* (2014) 45:633–40.
13. Costa P, Ferreira AS, Amaro A, Albuquerque T, Botelho A, Couto I, et al. Enhanced Detection of tuberculous mycobacteria in animal tissues using a semi-nested probe-based real-time PCR. *PLoS ONE.* (2013) 8:e81337. doi: 10.1371/journal.pone.0081337
14. Parra A, García N, García A, Lacombe A, Moreno F, Freire F, et al. Development of a molecular diagnostic test applied to experimental abattoir surveillance on bovine tuberculosis. *Vet Microbiol.* (2008) 127:315–24. doi: 10.1016/j.vetmic.2007.09.001
15. Beltran-Beck B, Fuente J, Garrido JM, Aranaz A, Sevilla I, et al. Oral vaccination with heat inactivated *mycobacterium bovis* activates the complement system to protect against tuberculosis. *PLoS ONE.* (2014) 9:e98048. doi: 10.1371/journal.pone.0098048
16. Sevilla IA, Molina E, Elguezabal N, Pérez V, Garrido JM, Juste RA. Detection of Mycobacteria, *Mycobacterium avium* Subspecies, and *Mycobacterium tuberculosis* Complex by a Novel Tetraplex Real-Time PCR Assay. *J Clin Microbiol.* (2015) 53:930–40. doi: 10.1128/JCM.03168-14
17. Corner LA, Trajstman AC. An Evaluation of 1-Hexadecylpyridinium Chloride as a decontaminant in the primary isolation of *mycobacterium bovis* from bovine lesions. *Vet Microbiol.* (1988) 18:127–34. doi: 10.1016/0378-1135(88)90058-2
18. Wilton S, Cousins D. Detection and Identification of Multiple Mycobacterial Pathogens by DNA Amplification in a Single Tube. *Genome Res.* (1992) 1:269–73. doi: 10.1101/gr.1.4.269
19. Kamerbeek J, Schouls L, Kolk A, Agterveld MV, Soolingen D, et al. Simultaneous Detection and strain differentiation of *mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol.* (1997) 35:907–14.
20. Blas ID, Ruiz-Zarzuela I, Vallejo A. WinEpi: working in epidemiology. an online epidemiological tool. In: *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics.* (Cairns, QLD) (2006).
21. Hall TA. BioeEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symposium Series.* (1999) 41:95–8.
22. Veyrier F, Saïd-Salim B, Behr MA. Evolution of the mycobacterial sigk regulon. *J Bacteriol.* (2008) 190:1891–9. doi: 10.1128/JB.01452-07
23. Park KT, Allen AJ, Davis WC. Development of a novel DNA extraction method for identification and quantification of *Mycobacterium avium* subsp. *paratuberculosis* from tissue samples by real-time PCR. *J Microbiol Methods.* (2014) 99:58–65. doi: 10.1016/j.mimet.2014.02.003
24. Menin Á, Fleith R, Reck C, Marlow M, Fernandes P, Pilati C, et al. Asymptomatic cattle naturally infected with *Mycobacterium bovis* present

exacerbated tissue pathology and bacterial dissemination. *PLoS ONE.* (2013) 8:e53884. doi: 10.1371/journal.pone.0053884

25. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, et al. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucl Acids Res.* (1990) 18:188. doi: 10.1093/nar/18.1.188
26. Mishra A, Singhal A, Chauhan DS, Katoch VM, Srivastava K, Thakral SS, et al. Direct Detection and Identification of *Mycobacterium tuberculosis* and *Mycobacterium bovis* in Bovine Samples by a Novel Nested PCR Assay: Correlation with Conventional Techniques. *J Clin Microbiol.* (2005) 43:5670–8. doi: 10.1128/JCM.43.11.5670-5678.2005
27. Cousins DV, Wilton SD, Francis BR. Use of DNA amplification for the rapid identification of *Mycobacterium bovis*. *Vet Microbiol.* (1991) 27:187–95. doi: 10.1016/0378-1135(91)90010-D
28. Santos N, Geraldes M, Afonso A, Almeida V, Correia-Neves M. Diagnosis of Tuberculosis in the Wild Boar (*Sus scrofa*): a comparison of methods applicable to hunter-harvested animals. *PLoS ONE.* (2010) 5:e12663. doi: 10.1371/journal.pone.0012663
29. Pereira-Suárez AL, Estrada-Chávez Y, Zúñiga-Estrada A, López-Rincón G, Hernández DUM, Padilla-Ramírez FJ, et al. Detection of *Mycobacterium tuberculosis* Complex by PCR in Fresh Cheese from Local Markets in Hidalgo, Mexico. *J Food Protect.* (2014) 77:849–52. doi: 10.4315/0362-028X.JFP-13-389
30. Taylor MJ, Hughes MS, Skuce RA, Neill SD. Detection of *Mycobacterium bovis* in bovine clinical specimens using real-time fluorescence and fluorescence resonance energy transfer probe rapid-Cycle PCR. *J Clin Microbiol.* (2001) 39:1272–8. doi: 10.1128/JCM.39.4.1272-1278.2001
31. RabelloMC, Matsumoto CK, Paula de Almeida LG, Menendez MC, Oliveira RS, et al. First Description of natural and experimental conjugation between mycobacteria mediated by a linear plasmid. *PLoS ONE.* (2010) 7:e29884. doi: 10.1371/journal.pone.0029884

Conflict of Interest Statement: EC is a full-time employee of Exome Diagnostics Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.1.2. Direct PCR on tissue samples to detect *Mycobacterium tuberculosis* complex: an alternative to the bacteriological culture.

This study described the validation of direct qPCR targeting the IS*6110* element, a multiple-copy target highly specific to the MTBC complex. Diagnostic performance of the qPCR was compared against culture in 985 fresh bovine tissue samples obtained from the Spanish bTB eradication campaign, 100 tissue samples from 78 Spanish bTB-free herds and 866 tissue samples among four bTB RLs.

The LOD with 95% confidence for the direct qPCR was established at 65 fg/reaction (13 genomic equivalents). None of the tissue samples from bTB-free herds were positive to this PCR, and diagnostic Se and Sp values for the 985 tissue samples were 96.45% and 93.66%, respectively, with respect to microbiological culture. Agreement between tests was high ($\kappa = 0.88$). The use of the *mpb70* direct qPCR confirmed the presence of MTBC DNA in 85.4% (n = 35) of the culture-negative/PCR-positive samples. When these samples were considered true positive, diagnostic Se and Sp of the qPCR increased to 96.78% and 99.02%, respectively. Diagnostic performance varied considerably between RLs; diagnostic Se and Sp ranged between 66.67-100% and 64.29-99.05%, respectively. This variation could be related to differences between laboratory procedures due to biosafety requirements (e.g. heat-inactivation of tissue homogenates).

No cross-reactions were identified during the analytical Sp validation, except for one isolate of '*M. avium* subsp. *hominissuis*' (MAH). No MAH isolates were identified on culture-negative and PCR-positive samples from the Spanish bTB eradication campaign. Whole genome sequence analysis of the isolate revealed an IS*6110*-like element with a high sequence identity to the IS*6110* element from MTBC members.

The high diagnostic performance of this direct qPCR indicates that this method can be effectively incorporated in the diagnostic algorithms to confirm MTBC infections in the Spanish eradication campaign. Technical requirements need to be verified first in order for the method to be implemented by other laboratories. The use of complementary targets could be of use when MAH cross-reactivity is suspected, such as in herds with a previous history of MAH isolation.

This article was accepted for publication on the 21st of January 2021 in the Journal of Clinical Microbiology.

Journal impact factor in 2020: 5.948

Position in JCR index 2020: Microbiology 22/137 (Q1).

Lorente-Leal V, Liandris E, Pacciarini M, Botelho A, Kenny K, Loyo B, Fernández R., Bezos J, Domínguez L., de Juan L. & Romero B., 2021. Direct PCR on Tissue Samples to Detect *Mycobacterium tuberculosis* Complex: An Alternative to the Bacteriological Culture. Journal of Clinical Microbiology 59:e01404-20. https://10.1128/JCM.01404-20.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

MYCOBACTERIOLOGY AND AEROBIC ACTINOMYCETES

Direct PCR on Tissue Samples To Detect Mycobacterium tuberculosis Complex: an Alternative to the Bacteriological Culture

V. Lorente-Leal ,[a,b] E. Liandris ,[a] M. Pacciarini ,[c] A. Botelho ,[d] K. Kenny ,[e] B. Loyo ,[f] R. Fernández ,[g] J. Bezos ,[a,b] L. Domínguez ,[a,b] L. de Juan ,[a,b] B. Romero [a]

[a]VISAVET Health Surveillance Center, Complutense University of Madrid, Madrid, Spain
[b]Veterinary Faculty, Animal Health Department, Complutense University of Madrid, Madrid, Spain
[c]Istituto Zooprofilattico Sperimentale of Lombardia and Emilia-Romagna (IZSLER), National Reference Centre for Bovine tuberculosis, Brescia, Italy
[d]IP National Institute for Agrarian and Veterinary Research (INIAV), Oeiras, Portugal
[e]Central Veterinary Research Laboratory (CVRL), Department of Agriculture, Food and the Marine, Celbridge, Ireland
[f]Laboratorio Regional de Sanidad Animal, Servicio de Sanidad Animal, Dirección General de Producción Agropecuaria, Consejería de Agricultura, Ganadería y Desarrollo Rural de la Junta de Castilla y León, Villaquilambre, Spain
[g]Laboratorio Central de Sanidad Animal. Ministerio de Agricultura, Pesca y Alimentación, Granada, Spain

ABSTRACT Bovine tuberculosis (bTB) is an ongoing issue in several countries within the European Union. Microbiological culture is the official confirmation technique for the presence of Mycobacterium tuberculosis complex (MTBC) members in bovine tissues, but several methodological issues, such as moderate sensitivity and long incubation times, require the development of more sensitive and rapid techniques. This study evaluates the analytical and diagnostic performance, comparative to culture, of a real-time PCR targeting the MTBC-specific IS6110 transposon using a panel of bovine tissue samples sourced from the Spanish bTB eradication campaign. Robustness and repeatability were evaluated in an interlaboratory trial between European Union National Reference Laboratories. The limit of detection with 95% confidence was established at 65 fg/reaction of purified genomic equivalents. Diagnostic sensitivity (Se) and specificity (Sp) were, respectively, 96.45% and 93.66%, and the overall agreement (κ) was 0.88. Cross-reactivity was detected against two mycobacterial isolates identified as Mycobacterium marinum and "Mycobacterium avium subsp. hominissuis" and whole-genome sequencing (WGS) analysis of the latter isolate revealed an IS6110-like sequence with 83% identity. An identical IS-like element was found in other Mycobacterium avium complex species in the NCBI nucleotide and WGS databases. Despite this finding, this methodology is considered a valuable alternative to culture, and the strategy of use should be defined depending on the control or eradication programs.

KEYWORDS PCR, bovine tuberculosis, direct PCR, tissue samples, validation

Bovine tuberculosis (bTB) is a chronic granulomatous infectious disease that mainly affects lymph nodes and lung tissue in cattle but can extend to other organs. This infection is mainly caused by Mycobacterium bovis and, to a lesser extent, by Mycobacterium caprae, but can also be produced by other members of the Mycobacterium tuberculosis complex (MTBC).

As bTB is a communicable disease, bTB surveillance or eradication campaigns are ongoing within the European Union (1, 2). Although 21 member states had officially tuberculosis-free (OTF) regions in 2018, bTB is still an ongoing problem in many parts of the European Union, with 11 non-OTF countries (3).

The difficulty in eradicating the disease could partially be explained by the complex

Citation Lorente-Leal V, Liandris E, Pacciarini M, Botelho A, Kenny K, Loyo B, Fernández R, Bezos J, Domínguez L, de Juan L, Romero B. 2021. Direct PCR on tissue samples to detect Mycobacterium tuberculosis complex: an alternative to the bacteriological culture. J Clin Microbiol 59:e01404-20 https://doi.org/10.1128/JCM.01404-20

Editor Brad Fenwick, University of Tennessee at Knoxville

Address correspondence to B. Romero, bromerom@visavet.ucm.es.

Received 5 June 2020
Returned for modification 28 June 2020
Accepted 17 November 2020
Accepted manuscript posted online 25 November 2020
Published 21 January 2021

interactions between the pathogen and different domestic and wildlife reservoirs (4). However, the success of the control measures used for the eradication ofinfectious diseases is also largely dependent on the performance of official diagnostic methods.

According to the European Union Council Directive 64/432/EEC (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:01964L0432-20150527) on animal health problems affecting intracommunity trade in bovine animals, detection of bTB is based on immunological techniques that detect the cellular response of the host (single or comparative intradermal tuberculin test [SIT or SCIT] and supplementary testing as the gamma interferon assay) and on the identification of the agent. The presence of the members of the MTBC is confirmed by culture of the microorganism and is carried out in either liquid or solid media platforms. The fastidious growth of this microorganism can result in very long incubation times of up to 12 weeks (5–7), making this diagnostic method extremely slow. Furthermore, microbiological culture requires the use of a decontamination step, which negatively affects bacterial viability and can reduce recovery rates (5, 8). Although the diagnostic specificity (Sp) of microbiological culture is high, its diagnostic sensitivity (Se) is considered moderate (9). Consequently, the development of more sensitive and rapid tests for the confirmation of bTB could greatly improve the control of the disease.

Real-time PCR has been evaluated as a potential first-line technique for the detection of MTBC species in animal tissues across the globe (10). It is a simple, rapid, and robust technique with a performance comparable to that shown by microbiological culture (9, 11, 12). In addition to the reduced analysis times required (a few days versus several weeks in microbiological culture), it decreases the exposure to high bacterial loads and avoids contact with the hazardous chemicals used during decontamination. In comparison to microbiological culture, PCR can be affected by variations in the pathological status of the analyzed tissues, the reagents used, the genetic target, and, most importantly, the DNA extraction method (13).

The diagnostic performance of PCR has mostly been evaluated using microbiological culture as a reference. Despite the differences between studies in those, including animal tissues with visible lesions (VLs) and nonvisible lesions (NVLs), diagnostic Se of real-time PCR appears to be slightly better than that described for conventional PCR (74% to 100% versus 63% to 97%, respectively). On the other hand, diagnostic Sp is considerably higher in real-time PCR than conventional PCR (96% to 100% versus 55% to 100%, respectively) (9, 12, 14–19). The use oflatent class analysis (LCA) for the evaluation of real-time PCR showed that diagnostic Se was higher than that of microbiological culture (87.7% versus 78.1%, respectively), and Sp was similar (97% versus 99.1%, respectively) (9).

This study describes the optimization and validation of a real-time PCR based on the IS6110 element, considered specific to the MTBC species. The performance of the IS6110-specific PCR is compared to that of microbiological culture in tissue samples obtained from the Spanish bTB eradication campaign. Additionally, this study describes the identification of an IS6110-like sequence in a "Mycobacterium aviumsubsp. hominissuis" strain and its conservation in different M. avium complex (MAC) species.

MATERIALS AND METHODS

PCR optimization. The following oligonucleotides targeting 68 bp of the transposon IS6110, specific for MTBC species, were selected from the literature (20, 21): 59-GGTAGCAGACCTCACCTATGTGT-39 (F6110), 59-AGGCGTCGGTGACAAAGG-39 (R6110), and 59-FAM-CACGTAGGCGAACCC-MGB-NFQ-39 (S6110).

Analytical specificity of the oligonucleotides, as well as the IS6110 element (GenBank accession no. X17348.1), was evaluated in silico using the Basic Local Alignment Tool (BLAST) from the NCBI. Oligonucleotides were purchased from Eurofins Genomics (Ebersberg, Germany), and real-time PCRs were carried out using the Quantifast pathogen PCR 1 IC kit (Qiagen, Hilden, Germany). A heterologous exogenous internal amplification control (IAC), labeled with the MAX NHS Ester reporter dye, is included in the kit and was used to detect partial or complete inhibition in this study. According to the manufacturer, IAC cycle threshold (C_T) values should range between 30 6 3. Therefore, when an amplification curve was detected and IAC C_t values were $ 33, samples were considered partially inhibited. Absence of amplification for the IAC and sample was considered complete inhibition. When inhibition was

detected, samples were diluted 1:10 in ultrapure distilled water (Sigma-Aldrich, St. Louis, MO, USA), and the PCR was repeated.

M. bovis BCG strain Danish (CCUG 27863) genomic DNA was used to generate a standard curve. Genomic DNA was extracted from a pure culture grown in Middlebrook 7H9 medium supplemented with oleic albumin dextrose catalase (OADC) using phenol-chloroform-isoamyl alcohol (PCI). Briefly, 500 *m*l of culture was centrifuged, and the pellet was homogenized in ultrapure distilled water. The cell mixture was then transferred to a bead beater tube containing 500 *m*l of PCI 25:24:1 (Sigma-Aldrich) and incubated for 30 min. The tube was beaten in a Precellys Evolution homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) two times at 9,000 rpm for 40 s with 10-s intervals, and the genomic DNA was purified using ethanol as previously described (22). The obtained DNA pellet was dried at room temperature and resuspended in 50 to 100 *m*l of AE buffer (Qiagen). The original DNA stock concentration was measured, and quality was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). For standard curve generation, DNA concentration was set to 1.32 ng/*m*l using a Qubit 4 fluorometer (Thermo Fisher Scientific) and was serially diluted 1:10 from 1.32 ng/*m*l to 0.13 fg/*m*l. The standard curve was then used to evaluate the PCR efficiency as well as the linear range of the reaction in triplicates, which was defined as the range of dilutions where the reaction efficiency was 100% (6 10%) and the coefficient oflinearity was . 0.997.

Following optimization, the setup for 25 *m*l of reaction volume was 7.5 *m*l of ultrapure distilled water, 5 *m*l of sample, 5 *m*l of Quantifast pathogen master mix, 2.5 *m*l ofinternal control assay (ICA), 2.5 *m*l ofinternal control DNA, 1 *m*l of F6110 (10 pmol/*m*l), 1 *m*l of R6110 (10 pmol/*m*l), and 0.5 *m*l of S6110 (10 pmol/*m*l). The run protocol was set as follows: an initial activation step of 95°C for 5 min followed by a denaturation step of 95°C for 15 s and an annealing/extension step of 60°C for 1 min. The denaturation and annealing/extension steps were repeated 45 times. Data acquisition was done at the annealing/extension step.

Analytical speci ficity, inclusivity, and limit of detection. Inclusivity was tested in vitro using DNA from 6 MTBC species, including M. tuberculosis M. africanum, M. bovis (including M. bovis strain BCG), M. caprae, M. microti, and M. pinnipedii Analytical Sp was tested in vitro using 102 DNA preparations obtained from 24 nontuberculous mycobacteria (NTM) species, including M. avium subsp. avium (n = 12), M. avium subsp. hominissuis (n = 6), M. avium subsp. paratuberculosis (n = 3), M. avium complex (nonspecific strains) (n = 12), M. chitae (n = 2), M. elephantis (n = 1), M. europaeum (n = 2), M. flavescens (n = 1), M. fortuitum (n = 8), M. gordonae (n = 4), M. holsaticum (n = 1), M. intermedium (n = 1), M. intracellulare (n = 5), M. kansasii (n = 3), M. marinum (n = 2), M. neoaurum (n = 1), M. nonchromogenicum (n = 10), M. palustre (n = 1), M. parascrofulaceum (n = 1), M. peregrinum (n = 1), M. phlei (n = 5), M. seoulense (n = 1), M. shimodei (n = 1), M. smegmatis (n = 12), M. terrae (n = 2), M. thermoresistibile (n = 3), and M. vaccae (n = 1). Additionally, specificity was tested against DNA isolates of Corynebacterium spp. (n = 5), Streptomyces spp. (n = 5), and Lactobacillus brevis (n = 1). All DNA isolates were obtained from the strain repository at VISAVET Health Surveillance Center.

The analytical Se, or limit of detection (LOD), was defined as the lowest DNA concentration in which at least 95% of the replicates were positive. Twenty replicates of the lowest concentration in the standard curve in which all three replicates showed amplification, as well as 1:5 and 1:10 dilutions thereof, were evaluated.

Diagnostic validation of the IS 6110-speci fic real-time PCR. Diagnostic Sp of the direct real-time PCR was assessed in a panel of 100 tissue samples obtained from 78 bTB-free herds (SIT negative and originating from a region with a bTB prevalence of , 1%). Diagnostic performance was further evaluated on 985 tissue samples obtained within the scope of the Spanish national bTB eradication campaign during the period from 2013 to 2017. The sampling population included positive (n = 608) and negative (n = 272) animals to the skin test, according to Spain's Royal Decree 2611/1996, as well as carcasses with bTB-compatible lesions at the abattoir (n = 105). In all cases, different lymph nodes (retropharyngeal, mandibular, mediastinal, bronchial, prescapular, mesenteric, hepatic, and/or supramammary), and, if affected by VL, parenchymatous organs were sent for processing. Comparative diagnostic Se and Sp of the direct real-time PCR were evaluated against microbiological culture.

All fresh bovine tissue samples were sent to VISAVET Health Surveillance Center and processed within its biosafety level 3 (BSL3) facilities. Samples were collected from 12 autonomic regions in Spain, including Andalusia (n = 15), Aragon (n = 82), Canary Islands (n = 18), Castile-La Mancha (n = 160), Castile and Leon (n = 8), Basque Country (n = 1), Extremadura (n = 53), Balearic Islands (n = 57), Madrid (n = 465), Murcia (n = 44), La Rioja (n = 23), and Valencia (n = 59).

All samples were inspected for macroscopic lesions before processing. Four hundred and eight samples showed VL, while 577 showed NVL. Sample processing and culture were carried out as previously described (23) with minor modifications. After decontamination with hexadecylpyridinium chloride (HPC; 0.37% final concentration), centrifugation was carried out for 30 min at 2,500 × g. Löwenstein-Jensen medium with sodium pyruvate and Coletsos media (Difco, Madrid, Spain) were incubated at 37°C for a maximum of 3 months and were considered positive when growth compatible with MTBC was detected and confirmed using conventional PCR (24) and/or direct variable repeat (DVR) spoligotyping (25). For tissue DNA extraction, 1/10 of the volume (1 ml) of tissue homogenate used for culture was processed as previously described (11). Briefly, samples were lysed mechanically and chemically, and DNA was purified using a modified protocol from the DNeasy blood and tissue kit (Qiagen).

Culture-negative and direct real-time PCR-positive samples were further analyzed by the mpb70 real-time PCR as previously described (11), and an IS1245-specific PCR was used to confirm the absence of M. avium subspecies DNA in these samples (26). Positive samples by the mpb70 PCR were used as an indicator of presence of MTBC-specific DNA and, together with the culture results, were incorporated as

adjusted results and compared with the IS6110 real-time PCR as a nonreference standard. In order to validate the agreement between the IS6110 and the mpb70 PCRs, a random selection of 200 samples were analyzed using both methods. In all reactions, double-distilled water and heat-inactivated M. bovis BCG isolate were used as negative and positive controls, respectively.

Statistical analyses. The C_T cutoff used to determine if a tissue sample was positive or negative for MTBC DNA was established with a receiver operating characteristic (ROC) curve analysis using the ROCR package in R. Diagnostic Se, Sp, and predictive values, as well as their corresponding 95% confidence intervals (CI), were calculated using MedCalc statistical software 19.2.0 (MedCalc Software bv, Ostend, Belgium). Positive and negative percent agreements (PPA and NPA, respectively), as well as predictive values, were calculated using the same software for the IS6110 and mpb70 comparison and the adjusted results. Overall agreement between tests was evaluated using Cohen's unweighted kappa in WinEpi 2.0 (27). Mann-Whitney two-tailed tests were used to assess the differences in the distributions of C_T values or DNA concentrations in different groups (negative versus positive culture, samples with VLs versus with NVLs, and inhibited versus noninhibited samples).

Interlaboratory performance evaluation. An interlaboratory comparison trial was set up with four collaborating national and regional reference laboratories (RLs) for bTB. Laboratories A, B, C, and D tested 192, 93, 80, and 501 samples, respectively (n = 866). Each laboratory processed samples obtained from animals that had reacted positively to the official diagnostic techniques (skin test and/or interferon gamma [IFN-*g*]) or had VLs compatible with bTB at the slaughterhouse inspection. The tissue samples were lymph nodes and parenchymatous organs with VLs or NVLs, except for laboratory C, which tested only MTBC culture-positive samples obtained from tissues with VL. The ratios of samples with VL and NVLs in the other laboratories were 84:108 (laboratory A), 53:40 (laboratory B), and 80:421 (laboratory D).

Microbiological culture was carried out in the different RLs according to their own validated procedures based on the OIE Manual for Terrestrial Animals with minor modifications. All laboratories used the same procedure for tissue DNA extraction, as previously described (11), and IS6110 real-time PCR protocol. Due to technical and biosafety reasons, two laboratories heat inactivated the samples prior to the DNA extraction protocol. Laboratory B incubated the samples at 95°C for 20 min, and laboratory C heat inactivated the samples at 80°C for 60 min. Additionally, laboratory B included the IAC DNA during the extraction protocol, whereas the rest added the IAC DNA directly to the master mix. Laboratories A and B carried out the real-time PCRs in CFX96 thermal cyclers (Bio-Rad, Hercules, CA, USA), whereas laboratory C and D used an Aria MX (Agilent, Santa Clara, CA, USA) and a StepOne Plus thermal cycler (Applied Biosystems, Foster City, CA, USA), respectively.

Whole-genome and Sanger sequencing. Genomic DNA from an M. avium (CECT 7407) isolate that cross-reacted with the IS6110-PCR during specificity testing was sent to FISABIO (Valencia, Spain) for whole-genome sequencing. Libraries were generated using Nextera XT library prep kit (Illumina, San Diego, CA, USA) and sequenced using a Miseq reagent kit v3 (600 cycles) in a Miseq sequencer. Raw reads were adapter and quality trimmed using Trimmomatic v0.39 (slidingwindow,15:30; minlen, 36), clumped using clumpify v38.49 from BBTools, and deduplicated using PRINSEQ-lite v0.20 (28). Processed reads were aligned against the phiX-174 reference genome (GenBank accession no. NC_001422.1) using BWA (29). Unmapped reads were assembled de novo with SPAdes v3.13 (30), and assembly quality was assessed using QUAST (31) with default parameters and Mycobacterium aviumsubsp. hominissuis strain MAC109 (GenBank accession no. CP029332.1) as a reference.

Contigs were deposited in a local BLAST database using makeblastdb from the BLAST1 package from the NCBI. The IS6110 sequence from M. tuberculosis H37Rv (GenBank accession no. NC_000962.3; positions 1541952 to 1543306) was used as a query for a local BLASTn against the contig database. The aligned sequence was extracted from the contig using bedtools (32) for further analysis. Plasmid sequences were identified using BLASTn.

Further confirmation of the IS6110-like element was achieved through Sanger sequencing (Stab Vida, Caparica, Portugal) and the use of primers 59-TCAGGTGGTGCGTCGAAGA-39 (ISlike-F) and 59-AATCCAGCACCCCTTCTTGT-39 (ISlike-R).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the Whole Genome Shotgun accession no. JAAILH000000000. The version described in this paper is version JAAILH000000000.1. The nucleotide sequence of the IS6110-like element obtained through Sanger sequencing was deposited at the NCBI GenBank under accession no. MT818214.

RESULTS

Optimization and LOD. A stock of 13.21 ng/*m*l (standard deviation [SD], 0.24) of genomic DNA, obtained from a pure culture of M. bovis BCG Danish, was serially diluted 1:10 from an initial concentration of 1.32 ng/*m*l to 0.13 fg/*m*l. After optimization, the range of C_T values spanned from an average of 18.83 cycles (SD, 0.09) at 6.5 ng/reaction to 39.18 cycles (SD, 0.27) at 6.5 fg/reaction (Fig. 1a). An overall efficiency of 100% was obtained from 6.5 ng/reaction to 650 fg/reaction, with an R^2 of 0.999 (Fig. 1b).

The LOD was established at 65 fg/reaction (average C_T value, 35.44, and coefficient of variation, 1.68%) with the 95% confidence limit, and all replicates (n = 20) were positive.

Analytical speci ficity and inclusivity. The in silico specificity analysis using the 68-bp amplicon obtained from the IS6110 sequence (33) yielded no significant alignments.

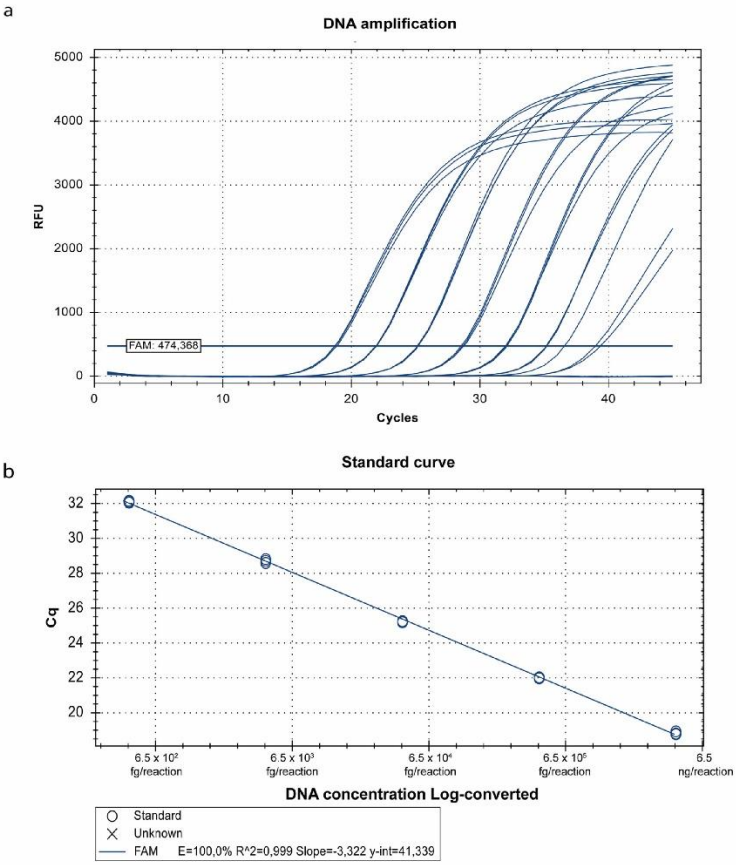


FIG 1 (a) IS6110 real-time PCR amplification range and efficiency. The standard curve ranged from 6.5 ng/reaction to 6.5 fg/reaction. (b) Dynamic range of the reaction that shows best efficiency and coefficient oflinearity (R^2).

All tested MTBC species were positive for this real-time PCR. Five NTM isolates (2 M. fortuitum, 1 M. kansasii, 1 M. intermedium, and 1 M. nonchromogenicum) produced a positive amplification curve with a C_T value . 38. Analysis of these isolates indicated a much higher concentration of mycobacterial DNA than that expected in tissue extracts (. 400 ng/*m*l), and dilutions to expected working concentrations (, 1 ng/*m*l) of mycobacterial DNA in tissue extracts produced negative results. Two additional NTM isolates (M. marinum and M. avium CECT 7407) produced amplification curves at C_T values of # 38 and remained positive at lower concentrations of DNA.

Diagnostic validation. None of the tissue samples obtained from 78 bTB-free herds (n = 100) produced an amplification curve.

In order to determine the diagnostic performance of the real-time PCR relative to microbiological culture, 985 bovine tissue samples were analyzed using both methods.

TABLE 1 Comparison of diagnostic performance between IS6110 real-time PCR and microbiological culture in 985 bovine samples

	Microbiological culture result			Diagnostic performance				
						Predictive value (% [95% CI])		
PCR result	No. positive	No. negative	Total no.	Sensitivity (% [95% CI])	Specificity (% [95% CI])	Positive	Negative	Agreement (95% CI)
No. positive	326	41	367	96.45 (93.90–98.15)	93.66 (91.51–95.29)	88.83 (85.19–91.66)	98.06 (96.64–98.89)	0.88 (0.85–0.91)
No. negative	12	606	618					
Total no.	338	647	985					

The ROC analysis established that a cutoff of 38.70 produced the highest diagnostic Se and Sp relative to culture, with an area under the curve (AUC) of 96% (95% CI, 94.6% to 97.3%) (Fig. S1 in the supplemental material). As a result, 326 samples were MTBC positive by both microbiological culture and real-time PCR, and 12 were culture-MTBC positive and real-time PCR negative (Table 1). When other microorganisms were cultured from 63 samples (NTM, 23; Actinomycetales 10; other microorganisms, 30), they were considered MTBC negative. There were 606 samples negative by both microbiological culture and real-time PCR, and 41 were positive only by the real-time PCR. Diagnostic Se and diagnostic Sp relative to culture were 96.45% (95% CI, 93.90% to 98.15%) and 93.66% (CI, 91.51% to 95.29%), respectively. Positive and negative predictive values (PPV, 88.83%; NPV, 98.06%) as well as agreement ($\kappa = 0.88$) between tests were good (Table 1).

Cycle threshold distribution was analyzed in MTBC culture-positive and PCR-positive samples and MTBC-culture negative and PCR-positive samples. C_T values were statistically higher (Mann-Whitney two-tailed test; P , 0.01) in the culture-negative group than the culture-positive group (Fig. 2a).

When considering the presence or absence of VLs in the analyzed tissues, 303 of the 326 culture-positive and PCR-positive samples were obtained from tissues with VL (Table 2). On the other hand, 28 out of 41 culture-negative and PCR-positive samples had VL. Diagnostic Se and Sp were, respectively, 97.74% (95% CI, 95.40% to 99.09%) and 71.43% (95% CI, 61.42% to 80.10%) in tissues with VLs and 82.14% (95% CI, 63.11% to 93.94%) and 97.63% (95% CI, 95.98% to 98.73%) in tissues with NVLs. Overall agreement was moderate (0.70 to 0.75) in both groups, whereas predictive values were generally high (90.91 to 99.8%), with the exception of the positive predictive value in NVLs (63.89%) (Table 2). Cycle thresholds were significantly lower (P , 0.01) in samples with VLs than samples with NVLs (Fig. 2b).

From the total of analyzed samples, two showed complete inhibition, whereas 44 were partially inhibited. One of the two inhibited samples gave a positive reaction after a 1:10 dilution. Fifteen out of 44 partially inhibited samples resulted in target amplification despite a partial inhibition of the IAC. One of the 29 remaining partially inhibited samples turned IS6110 positive after dilution. These positive samples were included among the real-time PCR-positive samples.

The analysis of 200 randomly selected samples through the IS6110 and mpb70 PCRs showed a very high overall agreement ($\kappa = 0.91$ [95% CI, 0.85 to 0.97]) (Table S1). More specifically, positive and negative agreement (PPA and NPA) values were also very high (93.10% [95% CI, 85.59% to 97.43%] and 97.35% [95% CI, 92.44% to 99.45%], respectively). Therefore, discordant results between microbiological culture and IS6110 real-time PCR were further investigated using the mpb70 PCR. This mpb70 PCR confirmed the presence of MTBC DNA in 35 out of 41 culture-negative and IS6110 PCR-positive samples (Table S2). Of these, 27 proceeded from samples with VLs (Table S3). After detection of MTBC-specific DNA in culture-negative and real-time PCR-positive samples, PPA and NPAs with respect to the adjusted results were 96.78% (95% CI, 94.46% to 98.15%) and 99.02% (95% CI, 97.88% to 99.55%), respectively (Table S2). The overall agreement and PPV of the IS6110 PCR increased to 0.96 and 98.37%, respectively. Additionally, the presence of MTBC DNA was detected in 27/28 (VL) and 8/13

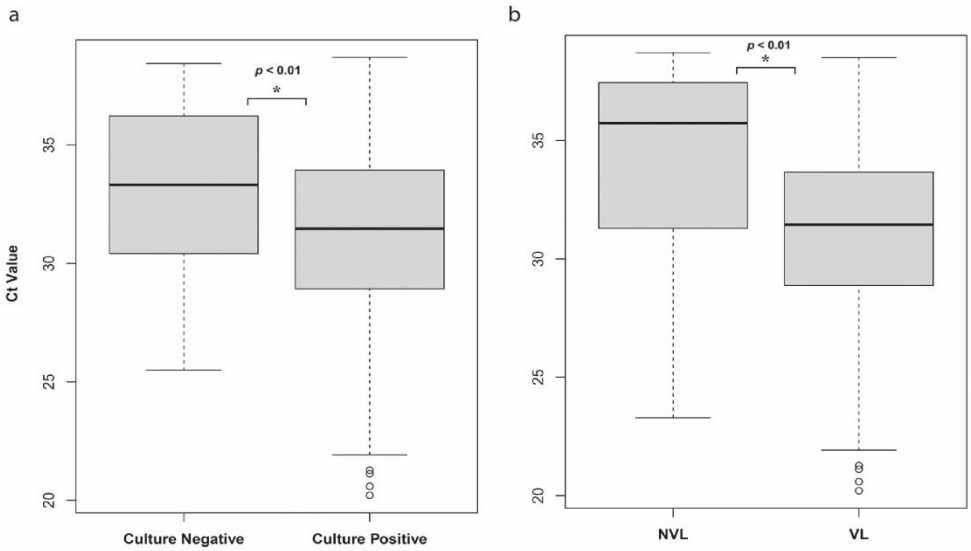


FIG 2 Cycle threshold comparison between culture-positive and culture-negative samples (a) and samples with VLs and NVLs (b). Statistical significance was carried out using a Mann-Whitney two-tailed test.

(NVL) positive samples (Table S3), and agreement between real-time PCR and the adjusted results was high when analyzing samples with VLs ($\kappa = 0.93$ [95% CI, 0.89 to 0.98]) and NVLs ($\kappa = 0.85$ [95% CI, 0.76 to 0.94]).

After the adjustment of discordant results, the C_T cutoff recommended through the ROC analysis was very similar (C_T, 39), and, when used, diagnostic performance did not vary significantly (data not shown).

Interlaboratory performance evaluation. Diagnostic performance of the real-time PCR relative to microbiological culture was assessed in four RLs as part ofinterlaboratory testing. The same C_T cutoff as specified previously was used for the analysis. In laboratory A, 78 samples were positive by both the PCR and microbiological culture, whereas 2 were only positive to culture (Table 3). Forty out of 112 culture-negative

TABLE 2 Comparison between the IS6110 real-time PCR and microbiological culture in relation to the presence or absence of anatomic lesions compatible with bTB

	Culture result			Diagnostic performance				
						Predictive value (% [95% CI])		
PCR result [a]	No. positive	No. negative	Total no.	Sensitivity (% [95% CI])	Specificity (% [95% CI])	Positive	Negative	Agreement (95% CI)
VL				97.74 (95.40–99.09)	71.43 (61.42–80.10)	91.54 (88.78–93.67)	90.91 (82.63–95.46)	0.75 (0.67–0.82)
No. positive	303	28	331					
No. negative	7	70	77					
Total no.	310	98	408					
NVL				82.14 (63.11–93.94)	97.63 (95.98–98.73)	63.89 (50.16–75.67)	99.08 (97.98–99.58)	0.70 (0.58–0.83)
No. positive	23	13	36					
No. negative	5	536	541					
Total no.	28	549	577					

[a]VL, visible lesions; NVL, nonvisible lesions.

TABLE 3 Comparison between microbiological culture results and real-time PCR results in the interlaboratory diagnostic validation study[a]

	Culture result		Diagnostic performance				
					Predictive value (% [95% CI])		
PCR result	No. positive	No. negative	Sensitivity (% [95% CI])	Specificity (% [95% CI])	Positive	Negative	Agreement (95% CI)
Laboratory A			97.50 (91.26–99.70)	64.29 (54.68–73.12)	66.10 (60.27–71.48)	97.30 (90.10–99.30)	0.58 (0.48–0.68)
No. positive	78	40					
No. negative	2	72					
Laboratory B			66.67 (52.53–78.91)	89.74 (75.78–97.13)	90.00 (76.95–96.04)	66.04 (52.59–77.31)	0.54 (0.39–0.69)
No. positive	36	4					
No. negative	18	35					
Laboratory C			98.75 (93.25–99.94)	NA	NA	NA	NA
No. positive	79	ND					
No. negative	1	ND					
Laboratory D			100.00 (95.55–100.00)	99.05 (97.58–99.74)	95.29 (88.42–98.17)	100	0.97 (0.94–1.00)
No. positive	81	4					
No. negative	0	416					

[a]ND, not determined; NA, not applicable.

samples were positive by the PCR. Diagnostic Se, Sp, and agreement between tests was, respectively, 97.50% (95% CI, 91.26% to 99.70%), 64.29% (95% CI, 54.68% to 73.12%), and 0.58% (95% CI, 0.85% to 0.91%) (Table 3). In laboratory B, 36 out of 54 culture-positive samples were also positive by the real-time PCR, whereas 4 out of 39 culture-negative samples were positive by the real-time PCR. Diagnostic Se, Sp, and agreement between tests were 66.67% (95% CI, 53.36% to 77.76%), 89.74% (95% CI, 76.42% to 95.94%), and 0.54% (95% CI, 0.39% to 0.69%), respectively. In laboratory C, 79 out of 80 culture-positive samples were also positive by PCR, and diagnostic Se was 98.75% (95% CI, 93.25% to 99.94%). Finally, 81 samples were culture and real-time PCR positive in laboratory D, whereas 416 were negative by both techniques. Four samples were positive only by culture, and no samples were positive by the real-time PCR alone. Diagnostic Se, Sp, and agreement between tests were, respectively, 100.00% (95% CI, 95.55% to 100.00%), 99.05% (95% CI, 97.58% to 99.74%), and 0.97% (95% CI, 0.94% to 1.00%). These results were translated into low to very good PPVs (66.10% to 95.29%) and NPVs (66.04% to 100.00%).

Identification of an IS*6110*-like element in *Mycobacterium avium* subsp. *hominissuis*. Due to the possibility of detecting M. avium from cattle, a more careful analysis of the M. avium isolate that cross-reacted during the analytical Sp validation was carried out using WGS.

After processing of the raw reads, 3,462,172 reads remained. The de novo assembly of M. avium CECT 7407 isolate (Whole Genome Shotgun project accession number: JAAILH000000000) yielded 133 contigs (≥ 500 bp) and an N_{50} of 108,694 bp. Compared to the reference using QUAST, the assembly represented 94.96% of the genome and had a duplication ratio of 1.00 and an NGA of 88,010 bp.

Alignment of the IS6110 sequence against these contigs using local BLASTn produced a hit with 83% similarity (Fig. S2), which corresponded to contig 83 (5,095 bp). Although a misalignment with the reference was detected in this contig (positions 1 to 2734 and positions 2735 to 5095), this did not include any part of the IS6110-like sequence (positions 3919 to 4934). Oligonucleotide sequence identity with the IS6110-like element was 86.40% (19/22) for F6110, 82.30% (14/17) for R6110, and 85.70% (12/14) for S6110.

An online search using BLASTn from the NCBI against the nucleotide and WGS

databases was carried out to assess the existence of similar sequences. Apart from IS6110 elements from members of the MTBC, alignment against the nucleotide archive produced a hit with 100% query coverage and identity with Mycobacterium aviumsubsp. hominissuis MAC109 (GenBank accession no. CP029332.1; positions 89000 to 90959). Annotation of this region indicated the presence of two truncated sequences from the IS256 family flanking the IS6110-like element.

In the case of the WGS contig database, alignments with 99 to 100% coverage and identity were found in 31 Mycobacterium avium subsp. hominissuis strains, 6 M. avium strains, as well as in 1 Mycobacterium bouchedurhonensand 1 Mycobacterium timonense strain (GenBank accession nos. MVIL01000161.1 and MVHL01000039.1, respectively).

Further analysis of the IS6110-like element with ISFinder produced partial alignments with IS6110 in the IS3 family ofinsertion sequences. The alignment had an approximate identity of 83% and was characterized by an incomplete left inverted repeat (IRL), which contained a 5-bp-long insertion at the 59 end (Fig. S2), as well as a truncated 39extreme missing the right inverted repeat (IRR).

DISCUSSION

Real-time PCRs have become an important technique in the rapid detection and identification of pathogens around the world. Although bTB detection by direct PCR has been evaluated previously with remarkable results (9, 12), microbiological culture is still considered the reference technique for the detection of tuberculosis in bovine tissues. This study described the optimization and validation of a real-time PCR targeting the IS6110 element, considered specific to members of the MTBC, and evaluated its diagnostic performance with tissue samples obtained from the Spanish bTB eradication campaign. An interlaboratory performance comparison was pursued through a collaboration with four reference laboratories from the European Union.

The IS6110-specific real-time PCR detected, during analytical validation, at least 13.6 genome equivalents/reaction (65 fg/reaction) with 95% confidence (34), taking into account that IS6110 is only present in one copy in the M. bovis BCG strain. In M. bovis strains and MTBC species with multiple copies of the IS6110, the LOD could be lower.

The IS6110-specific real-time PCR, in combination with the DNA extraction method, resulted in a rapid technique with good diagnostic performance (Se, 96.45%, and Sp, 93.66%) compared to culture. These results, including the high agreement between tests ($\kappa = 0.88$), highlight the potential of this real-time PCR as a first-line technique. Compared to other real-time PCR studies, the diagnostic Se for this PCR was similar (87.70 to 100%) in some cases (9, 11, 12) and higher in others (66.70 to 76.70%) (14–16, 35, 36). Specificity was, in general, slightly higher in these studies (96.03 to 100%). Nevertheless, the individual comparison between studies can be affected by the differences in molecular targets used (single copy versus variable copy), DNA extraction protocols, experimental design (type of culture isolation, number and type of samples, nested versus nonnested, etc.), and validation method (experimental versus Bayesian analysis) (9, 11, 12, 14–16, 34–36).

Although agreement between microbiological culture and the real-time PCR was very good ($\kappa = 0.88$), discordant results between the two methodologies were identified and further investigated.

The detection of MTBC DNA by mpb70 real-time PCR in 35 out of 41 culture-negative samples highlights possible limitations of culture, namely, lack of sensitivity, reduction of cell viability, and growth of other microorganisms, including NTM. The type of chemical decontamination and the time of exposure have an important effect on bacterial viability (8), which can have a significant effect in samples with already low bacterial loads. Indeed, real-time PCR C_T values revealed smaller amounts of MTBC DNA in culture-negative samples than in culture-positive ones. In addition, cellular viability could also have been affected by the pathological status of the granuloma (37). Although the majority (27/35) of these samples were obtained from tissues with VLs, no histopathological analysis was carried out on these tissues.

Diagnostic performance of the IS6110 real-time PCR compared to the adjusted results was very high irrespective of the presence or absence oflesions in the analyzed tissue (Table S3 in the supplemental material). Indeed, 27/28 (VL) and 8/13 (NVL) positive samples would have been missed if only culture had been carried out, in comparison to 7 (VL) and 5 (NVL) culture-positive and real-time PCR-negative samples. This suggests that real-time PCR could be an effective technique in the detection of bTB in tissues with NVLs and possible paucibacillary loads in comparison to microbiological culture.

Even though the DNA extraction protocol was carried out using 1/10 of the amount used for microbiological culture, only a small number (12/338) of culture-positive samples were negative by this PCR, indicating a good extraction efficiency. Repetition of the DNA extraction protocol in 8 out of 12 culture-positive and PCR-negative samples yielded two PCR positive results by the IS6110 and mpb70 PCRs (data not shown), indicating that the extraction procedure is critical for the performance of PCR and can be responsible for some of the false PCR-negative results.

The presence ofinhibitors can also have an important role in the performance of PCR for the detection of pathogen DNA. Inhibitors generally include excess host DNA and organic compounds such as urea, collagen, or bile salts, which can already be present in the sample or carried over during DNA extraction. Only 0.2% (n = 2) and 2.94% (n = 29) of the samples were completely or partially inhibited with no IS6110 amplification, respectively. Eight of the partially inhibited samples were culture positive, highlighting the importance ofincluding an inhibition control. DNA concentration in inhibited samples was higher (median [M], 353.65 ng/*m*l; interquartile range [IQR], 269.05 to 453.73 ng/*m*l) to that observed in noninhibited samples (M, 292.20 ng/*m*l; IQR, 201.7 to 391.8 ng/*m*l), indicating that excess host DNA could be responsible for the inhibition effect (P = 0.02). In order to overcome the inhibition, samples were diluted 1:10, which resulted in the amplification of MTBC DNA in 1 completely and 1 partially inhibited samples. However, in those samples with low bacterial loads, dilution can result in undetectable concentrations of target DNA. Certainly, the presence of MTBC DNA was confirmed in one of these samples using the mpb70 real-time PCR. Therefore, the DNA extraction should be repeated in all negative diluted samples, for which appropriate backup frozen material should be available.

Another reason behind discordance between culture-positive and PCR-negative samples could be the absence of the IS6110 transposon. The existence ofisolates lacking the IS6110 element, although infrequent, has been described previously (38). When a single copy of the IS6110 element is present, it is usually located in spacer 24 in the direct repeat locus (25). An initial analysis was carried out in some isolates available in our laboratory characterized by spoligotyping as sp24-deleted spoligotypes (21 out of 198 sp24-deleted patterns on Mbovis.org). No IS6110 amplification was detected in 6 of them, which include SB1263, SB1561, SB1630, SB1678, SB1888, and SB1901 (data not shown). All these patterns lost at least 3 spacers around spacer 24. From the total of SB patterns detected in the validation of this real-time PCR (n = 30), only one sp24-deleted spoligotype (SB1263) was detected, which represented two of the culture-positive and PCR-negative samples. The presence of MTBC DNA was confirmed using the mpb70 real-time PCR in these two samples.

In order to evaluate the diagnostic performance of the real-time PCR in different scenarios, an interlaboratory trial was carried out in four RLs using samples obtained from the eradication campaigns of the participating countries. Samples were processed using the standard procedures and biosafety measures of each laboratory. The interlaboratory performance evaluation revealed an overall high diagnostic Se relative to microbiological culture. However, diagnostic Se was considerably lower in laboratory B than in the rest of the laboratories. An important difference with the established protocol in this laboratory was the addition of a heat inactivation step of 95°C for 20 min before the extraction protocol. The effect of high-temperature treatment on the tissue homogenate before extraction has not been evaluated during the validation of

the protocol, and therefore, a negative effect cannot be ruled out. A pilot test in our laboratory evaluated the effect of high-temperature treatment (100°C for 15 min) on 43 randomly selected tissue homogenates before DNA extraction, and a reduction in PCR Se was reported (19%; data not shown). Interestingly, laboratory C also heat inactivated samples at 80°C and obtained higher diagnostic Se, but only samples with visible lesions were analyzed. Therefore, other factors can also be related to this reduction, such as a paucibacillary bacterial load in the animal tissues and/or the subsequent mechanical disruption steps. Diagnostic Sp with respect to culture was moderate to high in 3 laboratories (A, B, and D). Discrepancies in Sp could be explained by differences in isolation efficiency due to different decontamination procedures and sample processing before culture, which could have affected bacterial viability and microbiological yield. In conclusion, these results indicate that the performance of the real-time PCR can vary depending on the technical requirements of each facility and that these must be carefully evaluated before its implementation.

Few diagnostic specificity discrepancies could be related to cross-reactivity of the IS6110 probes with other bacteria. Although none of the 100 tissue samples obtained from 78 bTB-free herds were positive to this real-time PCR, giving an absolute diagnostic Sp of 100%, two cross-reactivity issues were detected during the validation of this PCR. Cross-reactivity with M. marinum has limited significance in the diagnosis of bTB, as this mycobacterial species is not a common pathogen of cattle. On the other hand, cross-reactivity with M. avium CECT 7407, identified as Mycobacterium avium subsp. hominissuis by PCR and hsp65 sequencing (data not shown), required further analysis. Mycobacterium avium subsp. hominissuis is mostly isolated in pigs and represents between 6 and 14% of NTM isolates in cattle (21, 39–41), being present as an opportunistic microorganism in bTB diagnosis (40). However, the real prevalence of Mycobacterium avium subsp. hominissuis in cattle is probably underestimated, as this pathogen is not searched for actively. Most isolations have been recorded in mesenteric lymph nodes, in accordance with the suspected oral-fecal transmission of MAC species, but isolates have also been detected in cattle respiratory samples (21, 42).

Other non-MTBC microorganisms were isolated in nine IS6110 real-time PCR-positive samples. Seven of these samples were confirmed as MTBC infections with the mpb70 direct real-time PCR. The remaining isolates were confirmed as Mannheimia granulomatis and one NTM species using 16S sequencing and conventional PCR, respectively (24). Interestingly, the latter isolate was negative to the IS1245-specific PCR (data not shown), indicating that it probably did not contain MAC DNA but other NTM DNA.

The IS6110 has been generally considered specific for the MTBC complex, although cross-reactivity with certain primer pairs or probes has been identified previously (36, 43, 44). The WGS analysis of the cross-reacting M. avium strain resulted in the discovery of an IS6110-like element with a sequence similarity of 83% with IS6110. The presence of this element was further confirmed through PCR and Sanger sequencing (Fig. S2). The truncation of the DDE motif and its IRs deemed this element inactive (Fig. S2). The IS6110-like sequence identified in this study, along with the rest of the contig's sequence, was identical to that of Mycobacterium avium subsp. hominissuis MAC109 (45), which had not been deposited in the NCBI at the time of the initial in silico specificity analysis. Another inactive IS6110-like element has been previously described in M. smegmatis strain MKD8 (46), although it only shares 50% and 70.79% coverage and identity with the IS6110-like element described in this study.

Interestingly, 100% identity and coverage were detected between the IS6110-like sequence and other contigs from Mycobacterium avium complex strains stored within the NCBI WGS database, including several Mycobacterium avium subsp. hominissuis strains, M. bouchedurhonense and M. timonense (a selection is aligned in Fig. S2). However, the last two species have not been detected in cattle in any NTM surveillance study so far (21, 39–41, 47, 48). Genomic comparative studies of mycobacterial species

showed that these two species cluster together with M. avium subspecies separately from the rest of the MAC (49, 50).

Despite the identification of an IS6110-like element, the 68-bp amplicon from CECT 7407 could not be sequenced, and therefore, a causal relationship between this element and the unspecific fluorescence signal could not be established (data not shown). Additionally, the absence of amplification in the rest of MAC isolates tested during the specificity study (n = 33), including 5 Mycobacterium aviumsubsp. hominissuisisolates, indicates that the cause of cross-reactivity may not be conserved in all MAC species or that the accumulation of mutations may hamper oligonucleotide binding. Furthermore, no cross-reactivity issues were detected against MAC isolates in a recent study published in France when the same primers and probe were used (21). A possible reason for this could be related to lineage-specific recombination and inversion events in M. avium genomes (51, 52). Further WGS comparisons between MAC species are required in order to confirm the presence of the IS6110-like sequence and understand the possible origin of this cross-reactivity.

Nevertheless, the appearance of an IS6110-like element in the genomes of several MAC species is an interesting finding and could be a result of horizontal gene transfer events (HGT) between a species of the MTBC and a MAC ancestor. Several HGT phenomena have been described in mycobacteria, such as the transfer and recombination of chromosomal DNA through distributive conjugal transfer (DCT) in M. smegmatisand M. canetti (53, 54). Although DCT has not been experimentally demonstrated in the MTBC (54), transitory conjugation has been observed in M. tuberculosisand M. bovis BCG (55), and recombination events have been identified in M. tuberculosis, indicating that DNA exchange and recombination is still possible even after the clonal expansion and genomic deletion events that led to MTBC speciation (56). Certainly, further research is required to understand the importance of HGT between MAC and MTBC species.

Based on the results, the application of the IS6110 real-time PCR on tissue samples is a very promising first-line molecular technique for the detection of MTBC species in bovine tissues. The rapidity of this technique, together with the high diagnostic sensitivity and specificity, makes this protocol an alternative to microbiological culture in some scenarios. Quick results would allow animal health authorities to make future decisions regarding bTB-infected herds in a shorter time than microbiological culture. However, the isolation of MTBC will still be required on most of the PCR-positive samples in order to carry out molecular epidemiological studies. Therefore, the strategic use of the direct PCR should be defined specifically considering the epidemiological situation, prevalence of bTB in the area of study, or the presence/absence of compatible TB lesions and could lead to cost-effective management of bTB diagnostic resources. Although the effect of the abovementioned specificity issues is limited, the possibility of cross-reactivity with other NTMs cannot be ruled out. Therefore, the additional use of an MTBC-specific PCR, such as the mpb70 PCR, and/or MAC-specific real-time PCR, as a confirmation of positive samples could further improve the effectivity of direct molecular detection of bTB in animal tissues. This would be especially relevant in those situations where speed, high specificity, and sensitivity are required (such as a low bTB prevalence area or a bovine OTF herd with a suspended status). In conclusion, the implementation of this protocol could aid in the eradication of this persevering infectious disease.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1 , PDF file, 1 MB.

ACKNOWLEDGMENTS

We thank the work performed by laboratory technicians A. Gutiérrez, F. Lozano, T. Alende, N. Moya, C. Viñolo, D. de la Cruz, and L. Jimenez in culturing, DNA extractions,

and spoligotyping. We also thank the French National Reference Laboratory for bTB for facilitating the primers and probes for this study.

V.L.-L., E.L., and laboratory technicians of the European Union RL for Bovine Tuberculosis and the Mycobacteria Unit of VISAVET carried out all the experimental procedures. V.L.-L. carried out the WGS analysis. B.R., L.D.J., L.D., J.B., A.B., M.P., K.K., B.L., and R.F. are responsible for obtaining samples and corresponding experimental data. V. L.-L. and E.L. wrote the manuscript with the insights of the rest of the authors. B.R., L.D.J., and E.L. designed and supervised the whole study.

V.L.-L. was funded with a predoctoral grant (reference CT17/18) from the Universidad Complutense de Madrid and Banco Santander 2017 to 2018. This work was funded by the European Union Reference Laboratory for Bovine Tuberculosis, the Ministerio de Agricultura, Pesca y Alimentación, and the Área de Ganadería de la Comunidad de Madrid.

REFERENCES

1. European Commission. 1977. Council Directive 77/391/EEC of 17 May 1977 introducing community measures for the eradication of brucellosis, tuberculosis and leucosis in cattle. 77/391/ECC. Off J Eur Communities 20:44–47.
2. European Commission. 1977. Council Directive 78/52/EEC of 13 December 1977 establishing the Community criteria for national plans for the accelerated eradication of brucellosis, tuberculosis and enzootic leukosis in cattle. 78/52/EEC. Off J Eur Communities 21:34–41.
3. European Food Safety Authority and European Centre for Disease Prevention and Control. 2019. The European Union One Health 2018 zoonoses report. EFSA J 17:e05926. https://doi.org/10.2903/j.efsa.2019.5926.
4. Palmer MV. 2013. Mycobacterium bovis: characteristics of wildlife reservoir hosts. Transbound Emerg Dis 60(Suppl 1):1–13. https://doi.org/10.1111/tbed.12115.
5. Corner LA, Gormley E, Pfeiffer DU. 2012. Primary isolation of Mycobacterium bovis from bovine tissues: conditions for maximising the number of positive cultures. Vet Microbiol 156:162–171. https://doi.org/10.1016/j.vetmic.2011.10.016.
6. Gormley E, Corner LA, Costello E, Rodriguez-Campos S. 2014. Bacteriological diagnosis and molecular strain typing of Mycobacterium bovis and Mycobacterium caprae Res Vet Sci 97 Suppl:S30–43. https://doi.org/10.1016/j.rvsc.2014.04.010.
7. Hines N, Payeur JB, Hoffman LJ. 2006. Comparison of the recovery of Mycobacterium bovis isolates using the BACTEC MGIT 960 system, BACTEC 460 system, and Middlebrook 7H10 and 7H11 solid media. J Vet Diagn Invest 18:243–250. https://doi.org/10.1177/104063870601800302.
8. Corner LA, Trajstman AC, Lund K. 1995. Determination of the optimum concentration of decontaminants for the primary isolation of Mycobacterium bovis N Z Vet J 43:129–133. https://doi.org/10.1080/00480169.1995.35871.
9. Courcoul A, Moyen J-L, Brugère L, Faye S, Hénault S, Gares H, Boschiroli M-L. 2014. Estimation of sensitivity and specificity of bacteriology, histopathology and PCR for the confirmatory diagnosis of bovine tuberculosis using latent class analysis. PLoS One 9:e90334. https://doi.org/10.1371/journal.pone.0090334.
10. Costa P, Botelho A, Couto I, Viveiros M, Inacio J. 2014. Standing of nucleic acid testing strategies in veterinary diagnosis laboratories to uncover Mycobacterium tuberculosis complex members. Front Mol Biosci 1:16. https://doi.org/10.3389/fmolb.2014.00016.
11. Lorente-Leal V, Liandris E, Castellanos E, Bezos J, Dominguez L, de Juan L, Romero B. 2019. Validation of a real-time PCR for the detection of Mycobacterium tuberculosis complex members in bovine tissue samples. Front Vet Sci 6:61. https://doi.org/10.3389/fvets.2019.00061.
12. Costa P, Ferreira AS, Amaro A, Albuquerque T, Botelho A, Couto I, Cunha MV, Viveiros M, Inácio J. 2013. Enhanced detection of tuberculous mycobacteria in animal tissues using a semi-nested probe-based real-time PCR. PLoS One 8:e81337. https://doi.org/10.1371/journal.pone.0081337.
13. Radomski N, Kreitmann L, McIntosh F, Behr MA. 2013. The critical role of DNA extraction for detection of mycobacteria in tissues. PLoS One 8:e78749. https://doi.org/10.1371/journal.pone.0078749.
14. Araújo CP, Osório ALAR, Jorge KSG, Ramos CAN, Filho AFS, Vidal CES, Roxo E, Nishibe C, Almeida NF, Júnior AAF, Silva MR, Neto JDB, Cerqueira VD, Zumárraga MJ, Araújo FR. 2014. Detection of Mycobacterium bovis in bovine and bubaline tissues using nested-PCR for TbD1. PLoS One 9:e91023. https://doi.org/10.1371/journal.pone.0091023.
15. Araújo CP, Osório ALAR, Jorge KSG, Ramos CAN, Souza Filho AF, Vidal CES, Vargas APC, Roxo E, Rocha AS, Suffys PN, Fonseca AA, Silva MR, Barbosa Neto JD, Cerqueira VD, Araújo FR. 2014. Direct detection of Mycobacterium tuberculosis complex in bovine and bubaline tissues through nested-PCR. Braz J Microbiol 45:633–640. https://doi.org/10.1590/s1517-83822014000200035.
16. Parra A, Garcia N, Garcia A, Lacombe A, Moreno F, Freire F, Moran J, Hermoso de Mendoza J. 2008. Development of a molecular diagnostic test applied to experimental abattoir surveillance on bovine tuberculosis. Vet Microbiol 127:315–324. https://doi.org/10.1016/j.vetmic.2007.09.001.
17. Liebana E, Aranaz A, Mateos A, Vilafranca M, Gomez-Mampaso E, Tercero JC, Alemany J, Suarez G, Domingo M, Dominguez L. 1995. Simple and rapid detection of Mycobacterium tuberculosis complex organisms in bovine tissue samples by PCR. J Clin Microbiol 33:33–36. https://doi.org/10.1128/JCM.33.1.33-36.1995.
18. Wards BJ, Collins DM, de Lisle GW. 1995. Detection of Mycobacterium bovis in tissues by polymerase chain reaction. Vet Microbiol 43:227–240. https://doi.org/10.1016/0378-1135(94)00096-f.
19. Stewart LD, McNair J, McCallan L, Gordon A, Grant IR. 2013. Improved detection of Mycobacterium bovis infection in bovine lymph node tissue using immunomagnetic separation (IMS)-based methods. PLoS One 8:e58374. https://doi.org/10.1371/journal.pone.0058374.
20. Michelet L, Cruz K, Karoui C, Hénault S, Boschiroli ML. 2017. Mycobacterium microti un agent tuberculoux méconnu. Épidémiologie et Santé Animale 71:129–138.
21. Michelet L, de Cruz K, Karoui C, Tambosco J, Moyen JL, Henault S, Boschiroli ML. 2018. Second line molecular diagnosis for bovine tuberculosis to improve diagnostic schemes. PLoS One 13:e0207614. https://doi.org/10.1371/journal.pone.0207614.
22. Ghebremariam MK, Hlokwe T, Rutten VPMG, Allepuz A, Cadmus S, Muwonge A, Robbe-Austerman S, Michel AL. 2018. Genetic profiling of Mycobacterium bovis strains from slaughtered cattle in Eritrea. PLoS Negl Trop Dis 12:e0006406. https://doi.org/10.1371/journal.pntd.0006406.
23. Corner LA, Trajstman AC. 1988. An evaluation of 1-hexadecylpyridinium chloride as a decontaminant in the primary isolation of Mycobacterium bovis from bovine lesions. Vet Microbiol 18:127–134. https://doi.org/10.1016/0378-1135(88)90058-2.
24. Wilton S, Cousins D. 1992. Detection and identification of multiple mycobacterial pathogens by DNA amplification in a single tube. PCR Methods Appl 1:269–273. https://doi.org/10.1101/gr.1.4.269.
25. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. 1997. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35:907–914. https://doi.org/10.1128/JCM.35.4.907-914.1997.
26. Guerrero C, Bernasconi C, Burki T, Bodmer T, Telenti A. 1995. A novel insertion element from Mycobacterium avium, IS1245, is a specific target for analysis of strain relatedness. J Clin Microbiol 33:304–307. https://doi.org/10.1128/JCM.33.2.304-307.1995.

27. Blas I, Ruiz-Zarzuela I, Vallejo A. 2006. WinEpi: working in epidemiology. An online epidemiological tool. In Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006, Cairns, Australia.
28. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
29. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
31. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
32. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033.
33. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL. 1990. IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res 18:188. https://doi.org/10.1093/nar/18.1.188.
34. Sevilla IA, Molina E, Elguezabal N, Perez V, Garrido JM, Juste RA. 2015. Detection of mycobacteria, Mycobacterium aviumsubspecies, and Mycobacterium tuberculosiscomplex by a novel tetraplex real-time PCR assay. J Clin Microbiol 53:930–940. https://doi.org/10.1128/JCM.03168-14.
35. Taylor MJ, Hughes MS, Skuce RA, Neill SD. 2001. Detection of Mycobacterium bovisin bovine clinical specimens using real-time fluorescence and fluorescence resonance energy transfer probe rapid-cycle PCR. J Clin Microbiol 39:1272–1278. https://doi.org/10.1128/JCM.39.4.1272-1278.2001.
36. Thacker TC, Harris B, Palmer MV, Waters WR. 2011. Improved specificity for detection of Mycobacterium bovisin fresh tissues using IS6110 real-time PCR. BMC Vet Res 7:50. https://doi.org/10.1186/1746-6148-7-50.
37. Menin A, Fleith R, Reck C, Marlow M, Fernandes P, Pilati C, Bafica A. 2013. Asymptomatic cattle naturally infected with Mycobacterium bovispresent exacerbated tissue pathology and bacterial dissemination. PLoS One 8:e53884. https://doi.org/10.1371/journal.pone.0053884.
38. Steensels D, Fauville-Dufaux M, Boie J, De Beenhouwer H. 2013. Failure of PCR-based IS6110 analysis to detect vertebral spondylodiscitis caused by Mycobacterium bovis J Clin Microbiol 51:366–368. https://doi.org/10.1128/JCM.02524-12.
39. Kim BR, Kim JM, Kim BJ, Jang Y, Ryoo S, Kook YH, Kim BJ. 2014. Identification of nontuberculous mycobacteria isolated from Hanwoo (Bos taurus coreanae) in South Korea by sequencing analysis targeting hsp65, rpoB and 16S rRNA genes. Vet Microbiol 173:385–389. https://doi.org/10.1016/j.vetmic.2014.07.019.
40. Ghielmetti G, Friedel U, Scherrer S, Sarno E, Landolt P, Dietz O, Hilbe M, Zweifel C, Stephan R. 2018. Non-tuberculous mycobacteria isolated from lymph nodes and faecal samples of healthy slaughtered cattle and the abattoir environment. Transbound Emerg Dis 65:711–718. https://doi.org/10.1111/tbed.12793.
41. Biet F, Boschiroli ML. 2014. Non-tuberculous mycobacterial infections of veterinary relevance. Res Vet Sci 97(Suppl):S69–S77. https://doi.org/10.1016/j.rvsc.2014.08.007.
42. Dvorska L, Matlova L, Bartos M, Parmova I, Bartl J, Svastova P, Bull TJ, Pavlik I. 2004. Study of Mycobacterium aviumcomplex strains isolated from cattle in the Czech Republic between 1996 and 2000. Vet Microbiol 99:239–250. https://doi.org/10.1016/j.vetmic.2004.01.008.
43. McHugh TD, Newport LE, Gillespie SH. 1997. IS6110 homologs are present in multiple copies in mycobacteria other than tuberculosis-causing mycobacteria. J Clin Microbiol 35:1769–1771. https://doi.org/10.1128/JCM.35.7.1769-1771.1997.
44. Costa P, Amaro A, Ferreira AS, Machado D, Albuquerque T, Couto I, Botelho A, Viveiros M, Inacio J. 2014. Rapid identification of veterinary-relevant Mycobacterium tuberculosiscomplex species using 16S rDNA, IS6110 and regions of difference-targeted dual-labelled hydrolysis probes. J Microbiol Methods 107:13–22. https://doi.org/10.1016/j.mimet.2014.08.017.
45. Matern WM, Bader JS, Karakousis PC. 2018. Genome analysis of Mycobacterium avium subspecies hominissuis strain 109. Sci Data 5:180277. https://doi.org/10.1038/sdata.2018.277.
46. Coros A, DeConno E, Derbyshire KM. 2008. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis suggestive oflateral gene transfer among mycobacterial species. J Bacteriol 190:3408–3410. https://doi.org/10.1128/JB.00009-08.
47. Gcebe N, Rutten V, Gey van Pittius NC, Michel A. 2013. Prevalence and distribution of non-tuberculous mycobacteria (NTM) in cattle, African buffaloes (Syncerus caffer) and their environments in South Africa. Transbound Emerg Dis 60(Suppl 1):74–84. https://doi.org/10.1111/tbed.12133.
48. Rónai Z, Eszterbauer E, Csivincsik Á, Guti CF, Dencső L, Jánosi S, Dán Á. 2016. Detection of wide genetic diversity and several novel strains among non-avium nontuberculous mycobacteria isolated from farmed and wild animals in Hungary. J Appl Microbiol 121:41–54. https://doi.org/10.1111/jam.13152.
49. Gupta RS, Lo B, Son J. 2018. Phylogenomics and comparative genomic studies robustly support division of the genus Mycobacterium into an emended genus Mycobacteriumand four novel genera. Front Microbiol 9:67. https://doi.org/10.3389/fmicb.2018.00067.
50. Tortoli E, Fedrizzi T, Meehan CJ, Trovato A, Grottola A, Giacobazzi E, Serpini GF, Tagliazucchi S, Fabio A, Bettua C, Bertorelli R, Frascaro F, De Sanctis V, Pecorari M, Jousson O, Segata N, Cirillo DM. 2017. The new phylogeny of the genus Mycobacterium the old and the news. Infect Genet Evol 56:19–25. https://doi.org/10.1016/j.meegid.2017.10.013.
51. Wu CW, Glasner J, Collins M, Naser S, Talaat AM. 2006. Whole-genome plasticity among Mycobacterium aviumsubspecies: insights from comparative genomic hybridizations. J Bacteriol 188:711–723. https://doi.org/10.1128/JB.188.2.711-723.2006.
52. Krzywinska E, Krzywinski J, Schorey JS. 2004. Naturally occurring horizontal gene transfer and homologous recombination in Mycobacterium Microbiology (Reading) 150:1707–1712. https://doi.org/10.1099/mic.0.27088-0.
53. Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM. 2013. Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS Biol 11:e1001602. https://doi.org/10.1371/journal.pbio.1001602.
54. Boritsch EC, Khanna V, Pawlik A, Honore N, Navas VH, Ma L, Bouchier C, Seemann T, Supply P, Stinear TP, Brosch R. 2016. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc Natl Acad Sci U S A 113:9876–9881. https://doi.org/10.1073/pnas.1604921113.
55. Ummels R, Abdallah AM, Kuiper V, Aajoud A, Sparrius M, Naeem R, Spaink HP, van Soolingen D, Pain A, Bitter W. 2014. Identification of a novel conjugative plasmid in mycobacteria that requires both type IV and type VII secretion. mBio 5:e01744-14. https://doi.org/10.1128/mBio.01744-14.
56. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie A-S, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez M-C, Leclerc C, Bentley SD, Stinear TP, Brisse S, Médigue C, Parkhill J, Cruveiller S, Brosch R. 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosisNat Genet 45:172–179. https://doi.org/10.1038/ng.2517.

Supplementary Table 1. Agreement evaluation between the results obtained after the analysis of 200 randomly-selected samples with the IS*6110* real-time PCR and the *mpb70* real-time PCR.

Overall comparison		**IS*6110***			**Positive Percent Agreement**	**Negative Percent Agreement**	**Agreement (κ)**
	Result	+	-	**Total**			
mpb70	+	81	3	84	93.10%	97.35%	0.91 (0.85 – 0.97)
	-	6	110	116	(95% CI: 85.59% - 97.43%)	(95% CI: 92.44% - 99.45%)	
	Total	87	113	200			
Culture positive samples		**IS*6110***			**Positive Percent Agreement**	**Negative Percent Agreement**	**Agreement (κ)**
	Result	+	-	**Total**			
mpb70	+	65	2	67	97.01%	71.43%	0.68 (0.40 – 0.97)
	-	2	5	7	(95% CI: 89.63% - 99.64%)	(95% CI: 29.04% - 96.33%)	
	Total	67	7	74			
Culture negative samples		**IS*6110***			**Positive Percent Agreement**	**Negative Percent Agreement**	**Agreement (κ)**
	Result	+	-	**Total**			
mpb70	+	16	1	17	80.00%	99.06%	0.84 (0.71 – 0.98)
	-	4	105	109	(95% CI: 56.34% - 94.27%)	(95% CI: 94.86% - 99.98%)	
	Total	20	106	126			

Supplementary Table 2. Agreement between the IS*6110* real-time PCR and the adjusted results*.

		Adjusted Results *			Percent Agreement (95% CI)	Predictive values (95% CI)	Agreement (95% CI)
	Result	+	-	**Total**	Positive:	Positive:	
IS*6110*	+	361	6	367	96.78%	98.37%	
	-	12	606	618	(94.46% to 98.15%)	(96.48% - 99.25%)	0.96
					Negative:	Negative:	(0.94 – 0.98)
	Total	373	612	985	99.02%	98.06%	
					(97.88% - 99.55%)	(96.64% – 98.89%)	

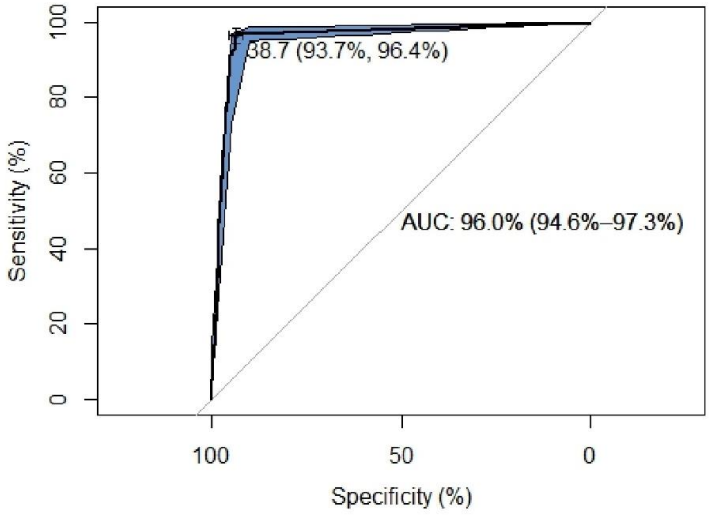
**Adjusted results include as positive: 1) culture-positive samples and 2) culture-negative and PCR-positive samples for which MTBC DNA presence was confirmed using the mpb70 real-time PCR*

Supplementary Table 3. Agreement between the IS*6110* real-time PCR and the adjusted results* in relation to the presence or absence of anatomical lesions compatible with bTB.

		Adjusted results*			
		+	-	**Total**	**Agreement (95% CI)**
VL	**PCR +**	330	1	331	0.93 (0.89 – 0.98)
	PCR -	7	70	77	
	Total	337	71	408	
NVL	**PCR +**	31	5	36	0.85 (0.76 – 0.94)
	PCR -	5	536	541	
	Total	36	541	577	

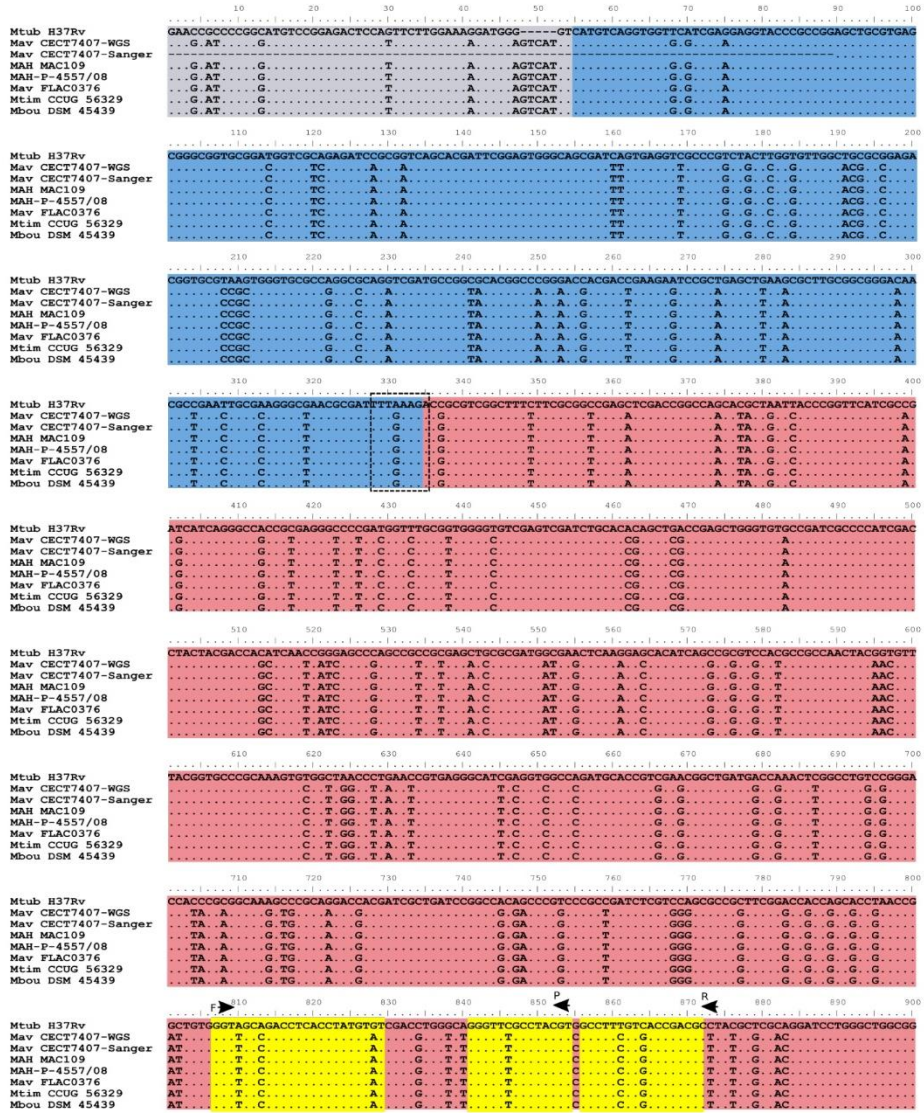
**Adjusted results include as positive: 1) culture-positive samples and 2) culture-negative and PCR-positive samples for which MTBC DNA presence was confirmed using the mpb70 real-time PCR. VL = Visible Lesions. NVL = Non-visible Lesions*

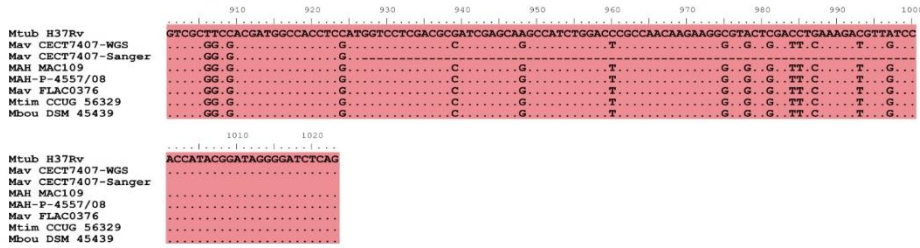
Supplementary Figure 1. Receiver Operating Curve used to determine the optimal Ct threshold value.



The ROC curve represents different diagnostic performance values according to different Ct cut-off settings. The best Ct cut-off value is represented as well as the diagnostic specificity and sensitivity (in brackets). The Area Under the Curve is represented with a 95% Confidence Interval (in brackets).

Supplementary Figure 2. Complete alignment of the IS*6110*-like element identified in this study through WGS (AN: JAAILH000000000) and Sanger sequencing (Genbank AN: MT818214) with the IS*6110* of *M. tuberculosis* H37Rv (NC_000962.3: 1541952 - 1543306) and a selection of other identical sequences in the NCBI nucleotide and WGS databases. Alignment was carried out using MUSCLE in MEGA X.





In grey: Left Inverted Repeat. In blue, red and dashed box: putative ORFa, incomplete putative ORFb and the ribosomal -1 frameshift region as predicted in (1). In yellow: alignment regions for the oligonucleotides used in this study: F (forward), P (probe) and R (Reverse). Accession numbers: "Mycobacterium avium subsp. hominissuis" strain MAC109 (CP029332.1: c90084-89062), "Mycobacterium avium subsp. hominissuis" strain MAH-P-4557/08 (PSNM01000103.1:1311-2333), *Mycobacterium avium* strain FLAC0376 (NSEU01000082.1:117870-118892), *Mycobacterium timonense* strain CCUG 56329 (MVIL01000161.1:c1134-112) and *Mycobacterium bouchedurhonense* strain DSM 45439 (MVHL01000039.1:3840-4862).

1. Gonzalo-Asensio J, Perez I, Aguilo N, Uranga S, Pico A, Lampreave C, Cebollada A, Otal I, Samper S, Martin C. 2018. New insights into the transposition mechanisms of IS*6110* and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. PLoS Genet 14:e1007282.

4.2. Objective 2: development of novel tools for the typing of MTBC members

4.2.1. Performance assessment of available variant calling pipelines for bTB.

This study evaluated the performance and the differences in the phylogenetic inferences achieved by four publicly available variant calling pipelines for *M. bovis* WGS data (vSNP, SNiPgenie, BovTB and MTBseq). The analysis was carried out on a dataset of 47 artificial *M. bovis* genomes generated from genomic data obtained from a region in the RoI with a high bTB prevalence. These genomes and their inferred phylogeny were used as a reference during the comparison of all methods.

Overall, the performance of the four pipelines was high, with recall and precision rates being higher than 99.8% when the same hard filters were applied. Homoplasies were identified in approximately 7% of the positions, and mostly removed after filtering proximal SNPs (i.e. within 10 bp) and *pe/ppe* genes. In addition, the agreement between pipelines was also very high; SNiPgenie, BovTB and MTBseq detected all of the SNPs present in the simulation, whereas vSNP was not able to detect seven SNPs.

Comparison of tree topologies through RF distances revealed that the phylogenetic inferences obtained from the alignments output by the four pipelines were filter-dependent and highly similar to the ones observed in the simulation. A more exhaustive pairwise comparison of best ML tree topology further revealed a high similarity between the trees. Similarity between trees was highest among those in which all problematic regions had been filtered out, with small topological variations being associated to highly related taxa and polytomies. Few differences were identified among the unfiltered trees, and these were related to small variations in the location of several isolates with respect to their MRCA, which altered their genetic relatedness to that observed in the simulation.

Overall, the results of this study indicate that the performance of the evaluated pipelines is highly similar as long as the same hard filters are used. Although more investigations are needed with real-life sequence data, all methods could be interesting alternatives for variant calling analyses of *M. bovis* isolates.

This article was accepted for publication on the 25th of November 2021 in Frontiers in Veterinary Science.

Journal impact factor in 2020: 3.412

Position in JCR index 2020: Veterinary Sciences 9/146 (Q1).

Lorente-Leal V, Farrell D, Romero B, Alvarez J, de Juan L, Gordon SV. Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. Front Vet Sci. 2021;8:780018. https://doi.org/10.3389/fvets.2021.780018.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fvets.2021.780018

Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization

Víctor Lorente-Leal[1,2]*, *Damien Farrell*[3], *Beatriz Romero*[1,2], *Julio Álvarez*[1,2], *Lucía de Juan*[1,2] *and Stephen V. Gordon*[3]

[1] VISAVET Health Surveillance Center, Universidad Complutense de Madrid, Madrid, Spain, [2] Animal Health Department, Faculty of Veterinary Medicine, Universidad Complutense de Madrid, Madrid, Spain, [3] UCD School of Veterinary Medicine, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:
Iryna Goraichuk,
Institute of Experimental and Clinical Veterinary Medicine, Ukraine

Reviewed by:
Lorraine Michelet,
de l'Environnement et du Travail (ANSES), France
Daniela Brites,
Swiss Tropical and Public Health Institute (Swiss TPH), Switzerland

***Correspondence:**
Víctor Lorente-Leal
vicloren@ucm.es

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 20 September 2021
Accepted: 25 November 2021
Published: 14 December 2021

Citation:
Lorente-Leal V, Farrell D, Romero B, Álvarez J, Juan Ld and Gordon SV (2021) Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. Front. Vet. Sci. 8:780018. doi: 10.3389/fvets.2021.780018

Whole genome sequencing (WGS) and allied variant calling pipelines are a valuable tool for the control and eradication of infectious diseases, since they allow the assessment of the genetic relatedness of strains of animal pathogens. In the context of the control of tuberculosis (TB) in livestock, mainly caused by *Mycobacterium bovis*, these tools offer a high-resolution alternative to traditional molecular methods in the study of herd breakdown events. However, despite the increased use and efforts in the standardization of WGS methods in human tuberculosis around the world, the application of these WGS-enabled approaches to control TB in livestock is still in early development. Our study pursued an initial evaluation of the performance and agreement of four publicly available pipelines for the analysis of *M. bovis* WGS data (vSNP, SNiPgenie, BovTB, and MTBseq) on a set of simulated Illumina reads generated from a real-world setting with high TB prevalence in cattle and wildlife in the Republic of Ireland. The overall performance of the evaluated pipelines was high, with recall and precision rates above 99% once repeat-rich and problematic regions were removed from the analyses. In addition, when the same filters were applied, distances between inferred phylogenetic trees were similar and pairwise comparison revealed that most of the differences were due to the positioning of polytomies. Hence, under the studied conditions, all pipelines offer similar performance for variant calling to underpin real-world studies of *M. bovis* transmission dynamics.

Keywords: whole genome sequencing (WGS), bioinformatics, variant calling pipeline, SNP analysis, genomic epidemiology, Bovine Tuberculosis (bTB), *Mycobacterium bovis*, *Mycobacterium tuberculosis* complex (MTBC)

INTRODUCTION

Animal tuberculosis (aTB) is a chronic infectious disease that affects a wide variety of mammalian species, which is caused by members of the *Mycobacterium tuberculosis* complex (MTBC) (1). The principal agent of TB in cattle (bovine TB, bTB) is *Mycobacterium bovis*. In this manuscript, we will use aTB to refer to TB across wild and domestic animals, and bTB to refer specifically to TB in cattle.

Bovine TB is subjected to control and eradication programmes in many countries, not only due to its economic impact, as a result of reduced yields and animal mortality, but also because of the risk of zoonotic transfer of infection from affected animals to humans (2). Eradication programmes are usually based on a test and slaughter strategy in which cattle that are positive to an official immunological test, such as the intradermal tuberculin test, are culled (3–5). In order to confirm the presence of MTBC species, tissues from the affected animals are cultured in the laboratory (6). In order to eradicate bTB, breakdown events not only need to be detected but also studied for epidemiological links, a process that is greatly facilitated by the application of molecular genetic methods. Due to the clonal structure and limited genetic variability of MTBC species, based on the observed genetic differences between the strains isolated from the breakdown herd and from other aTB episodes, authorities can establish if the outbreak originated from cattle movement, residual infection or contact with wild animal reservoirs (7).

Traditionally, molecular epidemiological studies of aTB are based on techniques that analyse small fragments of the microbial genome, such as spoligotyping or mycobacterial interspersed repeat unit-variable number of tandem repeats (MIRU-VNTR) (8, 9). Although useful in large-scale studies (10–12), some of these methods are laborious and the use of a limited number of loci entails a higher risk of homoplasies and a lack of resolution, limiting their use in the study of local transmission events (13, 14).

The advent of Whole Genome Sequencing (WGS) has revolutionized the study of microbial populations. When applied to epidemiological studies, the availability of the whole genome of the microorganism of interest allows for a much higher resolution than that obtained with previous molecular techniques (15). As a result, the use of WGS in human TB outbreak investigations has rapidly increased in the last decade (16–18).

Due to the limited genetic diversity in MTBC genomes, the standard workflow in MTBC studies is based on the alignment of genomic sequences to a reference genome followed by the detection of genomic variants, usually single nucleotide polymorphisms (SNPs) (19). The procedure starts with genomic DNA extraction, usually through phenol-chloroform or CTAB extraction, library preparation and sequencing using short read sequencing technologies, followed by short-read mapping to the reference genome and variant calling. Variants are then filtered according to certain thresholds and parameters such as proximity to other SNPs, mapping quality, base depth or strand bias. Remaining SNPs are generally concatenated into multi-FASTA files representing multiple sequence alignments and a phylogeny is reconstructed based on SNP differences.

There are several variant calling pipelines for human tuberculosis and, recently, several efforts have been made to assess their performance in human TB outbreak investigations (19–21). Regarding the veterinary field, there is a growing interest in the use of WGS for the analysis of bTB breakdowns, which has resulted in an increasing number of studies being published around the globe (22–26). Nevertheless, although several variant calling pipelines have been developed or are in the making, there are no tool-specific publications and there is a lack of information regarding their overall performance. The aim of this study was to evaluate similarities in design and performance of publicly available variant calling pipelines currently used in laboratories tasked with the application of WGS technologies for aTB eradication.

MATERIALS AND METHODS

Artificial Genome and Read Generation

In order to simulate a reference phylogeny, raw Variant Call Format (VCF) files were selected from an already published dataset from a bTB high prevalence setting in the Republic of Ireland (22). A total of 47 samples, including two outgroup isolates (isolates 161 and 182), were used to generate artificial mutant genomes by transferring the identified SNPs in the raw VCF files to the *M. bovis* AF2122/97 genome (NCBI RefSeq accession number: NC_002945.4) using simuG 1.0.0 (27) (**Figure 1**). ArtificialFASTQGenerator 1.0 was then used to generate artificial paired-end reads from the simulated genomes. Several parameters were tested to guarantee a full genome coverage and varying read depth across the whole sequence. Read length was set to 250 bp, template length mean to 650 bp (S.D. = 60), and peak coverage mean for a region was set to 250 (Standard Deviation or S.D. = 0.2) (28). Read qualities were obtained from real-life FASTQ files originating from other sequencing projects (unpublished) and sequencing errors were simulated based on these quality profiles.

Variant Calling

The artificially generated reads were analyzed with four variant calling pipelines: vSNP 2.03 (25), SNiPgenie 0.5.0, BovTB 20.4, and MTBseq 1.03 (29). The three first pipelines are used for analysis of aTB isolates in the USA (vSNP), Ireland (SNiPGenie), and UK (BovTB); MTBSeq was added as a comparator from the human TB field. Information regarding these pipelines is summarized in **Table 1** and described in Supplementary Materials and methods. All pipelines, with exception of SNiPgenie, were run using default settings in miniconda 4.9.2 in Ubuntu 18.04 and in Brigit, the HPC server of the Computer Services at Universidad Complutense of Madrid, using the default reference sequence of *M. bovis* AF2122/97 (NC_002945.4 or LT708304). In SNiPgenie, minimum mapping quality was set to 60 in variant calling and the minimum MQ was set to 30 in posterior filtering steps. Amended variant tables returned by MTBseq were converted to the VCF format using an in-house script for further comparisons. VCF files were generated from excel tables output by vSNP's step 2 script using an in-house script and including a zero coverage VCF of *M. bovis* AF2122/97 in order to include non-parsimonious SNPs.

Pipeline Performance Evaluation

As well as a FASTA file containing the artificial genome, SimuG generates a VCF file that contains all the variants included in the generated genome. These artificial VCF files were used as a reference standard to compare the VCF files output from the variant calling pipelines using the Haplotype Comparison Tools 0.3.12 (Som.py).

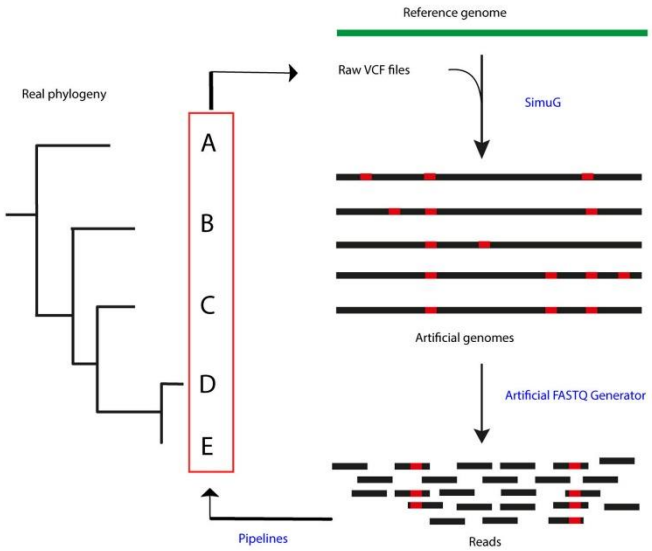


FIGURE 1 | Process summary. Raw VCF files obtained from a real-life phylogeny (22) were used to generate artificial genomes using SimuG and the *M. bovis* AF2122/97 reference genome. Artificial FASTQ Generator was then used to generate artificial reads and these were then input in the evaluated pipelines. Identified variants and output phylogenetic trees were compared between pipelines and the simulation.

Variants occurring in locations where no mutations existed in the simulated genome were considered "false positive SNPs," while mutations not detected by a given pipeline were considered "false negative SNPs".

The evaluated performance parameters were relative sensitivity or recall rate (true positives/true positives + false negatives) and relative specificity or precision (true positives/true positives + false positives). In addition, alternative (ALT) alleles were extracted from all sample VCF files obtained from each pipeline and combined to obtain the total amount of alleles identified per pipeline. The agreement between the different pipelines was then evaluated using Venn diagrams generated using VennDiagram v1.6.20 in R 3.6.3 (30).

In order to identify groups of genetic elements that usually give rise to false positive and negative calls, all VCF files were annotated using SnpEff 4.3t (31). These genetic elements were then divided into three categories: PE and PPE gene families, mobile genetic elements and other elements (including Direct Repeats and the *pks*12 gene), and their positions in the reference genome were extracted from the GFF3 annotation file available at the NCBI.

Performance was re-evaluated using different levels of hard filtering: (A) unfiltered, (B) a proximal window distance of 10 bp (22), (C) 10 bp window and pipeline default filters, (D) 10 bp window and PE/PPE family proteins, (E) 10 bp window, PE/PPE family proteins and mobile genetic elements, (F) 10 bp window, PE/PPE family proteins, mobile elements and others, and (G) PE/PPE family proteins, mobile elements and others. In order to assess the agreement between pipelines and the accuracy of these results with respect to the original simulated files, filtered positions were also removed from the simulated VCF files.

In addition, the effect of filtering on the number of identified homoplasies was assessed using HomoplasyFinder (32).

Evaluation of Phylogenetic Outputs and Epidemiological Conclusions

All pipelines, except for BovTB, generate a multi-FASTA alignment containing the concatenated variants. The SNPs in the alignment files obtained from vSNP and SNiPgenie only include polymorphic sites, whereas MTBseq alignments also include monomorphic sites. BovTB yields a consensus genome generated from the VCF files using the BCFtools consensus caller. In order to compare the different methods, core polymorphic SNPs were extracted from these consensus genomes using SNP-sites 2.5.1 (33). In addition, concatenated multi-FASTA files containing polymorphic SNPs were generated for the simulated VCFs using an in-house script.

Maximum-likelihood trees were reconstructed from the resulting multi-FASTA alignment files using RAxML 8.2.12 with

TABLE 1 | Pipeline properties of the different tools evaluated in this study.

	Pipeline			
	vSNP	**SNiPgenie**	**BovTB**	**MTBseq**
Institution	USDA-APHIS	UCD	APHA	LLI – RCB
Language	Python	Python	Nextflow	Perl
Reference	NC_002945.4	LT708304.1	LT708304.1	NC_002945.4
Parameter setup	No[a]	Yes	No	Yes
Pre-process				
Deduplication	Picard	No	FastUniq	Picard
Trimming	None	Yes[b]	Trimmomatic	None
Mapping and SNP calling				
Read aligner	BWA	BWA	BWA	BWA
SNP calling	FreeBayes	BCFtools	BCFtools	SAMtools + GATK
Phred base quality	20 (Step 1)	User defined	10	20
Normalize	No	No	Yes	Yes
SNP quality threshold	150	≥40 or User defined	None	None
Min. map quality	56	60	None	None
SNP coverage depth	None	30	5	4F and 4R
Region filter	Excel file (validated problematic positions)	BED file (PE/PPE genes)	TSV (95% similarity self-BLAST)	TSV file (repetitive sequences)
Proximality filter	None	Yes	None	Yes
Allele frequency/fraction	0.05	DP4>4	≥ 0.8	75%
Considers as diploid	Yes	No	No	No
Low coverage positions	Reference if QUAL < 50 N if 50 < QUAL < 150	Reference	Reference	Consensus base or ignore position if quality is below thresholds in >5% of samples
Alignment file	Core SNPs (polymorphic)	Core SNPs (polymorphic)	Consensus genome	Core SNPs (all)
Spoligotyping	Yes	Yes	No	No
Tree building	RAxML	RAxML	No	No
GUI	No	Yes	No	No
Other analyses	Lineage classification	INDEL analysis	Lineage classification	Lineage classification, antibiotic resistance annotation

[a] *Only allows for minor parameter settings, such as reference file or type of analysis in step 2.*
[b] *Deactivated by default.*

100 bootstraps and the GTRCATI model (34). The bipartitions and best trees obtained from each pipeline were evaluated using Robinson-Foulds (RF) distances and Ward's method for clustering through Treespace in R; briefly, RF pairwise distances between trees were decomposed into a low-dimensional space using a principal coordinate analysis (35). Trees obtained from hard filters that produced the best results in the performance evaluation were compared in a pairwise manner with the simulated phylogeny using Phytools 0.7.82 in R (36).

RESULTS

Artificial Read and Genome Simulation

An average of 2.5 x 10^6 reads (coefficient of variation or C.V. = 0.09%) were generated with an average depth of coverage of 145, with a minimum of 0 and a maximum of 310 reads per site.

Excluding the outgroup isolates, the average observed differences between isolates in the unfiltered simulation was 38.61, with an inter quartile range of 30–47, and a minimum and maximum number of 12 and 64 nucleotide differences, respectively.

Pipeline Performance

Recall rates were highest for SNiPgenie and BovTB when base parameters were employed, followed by vSNP and MTBseq (**Figure 2A**).

Sensitivity increased when increasing levels of hard filtering were applied for vSNP, SNiPgenie, and MTBseq, and remained similar for BovTB (**Figure 2A**). The positive effect was higher when pipeline-specific hard filters were used, in comparison to a proximal window alone (filter B) and a proximal window with additional PE/PPE filtering (filter D). However, recall rates of default filters (filter C) were slightly lower in comparison to the removal of combined proximal SNPs, loci encoding PE/PPE family proteins, mobile elements and other repetitive sequences (filter F) or PE/PPE family proteins, mobile elements and other repetitive sequences (filter G). This was specially the case for SNiPgenie, for which sensitivity increased to

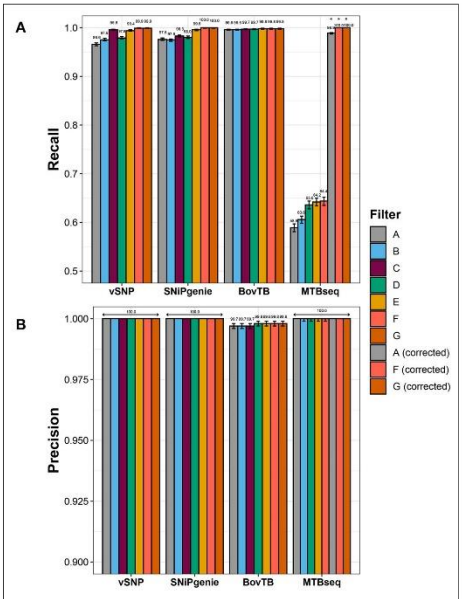


FIGURE 2 | Effect of hard filtering in the performance of the evaluated pipelines when compared to the simulated dataset as indicated in the **(A)** recall and **(B)** precision rates. Asterisks indicate filters in MTBseq for which the minimum coverage threshold was adjusted.

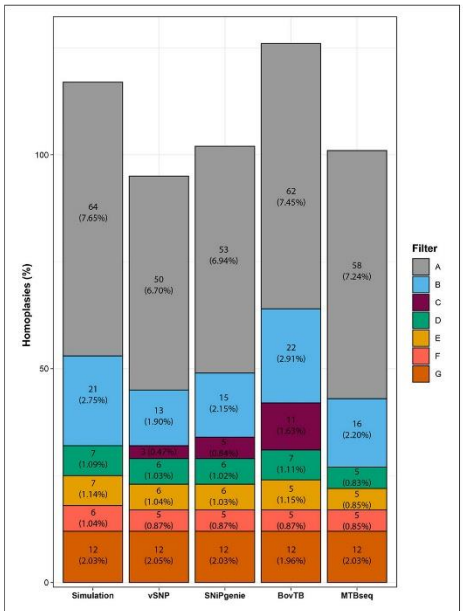


FIGURE 3 | Effect of filtering in the total number of homoplasic positions per pipeline identified by HomoplasyFinder. The percentage represents the proportion of homoplasic positions from the total number of identified positions using each specific filter.

levels similar to vSNP and BovTB when filters E, F, or G were applied.

Sensitivity remained below 65% for MTBseq despite the removal of problematic regions. Evaluation of the alignment files for this pipeline revealed that the increased amount of False Negative (FN) calls was produced by strand bias introduced by the artificial read generation, leading to forward or reverse read coverage being below the default minimum threshold ($n = 4$). Adjusting this threshold increased recall rates above 99% (**Figure 2A**).

After correction of MTBseq parameters, erroneous calls were further evaluated among the unfiltered pipeline results. FN calls were distributed unevenly among the simulated sequences (**Supplementary Figure 1**) and were mostly located within or near repetitive sequences (data not shown). More than half of the FN positions were shared by at least two of the pipelines, whereas 23 and 20% of the FN positions were identified only by BovTB and vSNP, respectively (**Supplementary Figure 2**). In addition, the majority of FN positions identified by BovTB in one sample were correctly detected as true SNPs in a varying number of samples (**Supplementary Figure 1C**).

A small proportion of false positive (FP) SNPs were identified by BovTB (43 SNPs across 37 positions) but, nevertheless, precision was high (>99%) for all of the evaluated pipelines (**Figure 2B**). Approximately 40% of FPs were located in repetitive regions and, although filtering improved precision in these cases, false positive SNPs were still detected (data not shown). Further analysis of the VCF files in BovTB revealed that the affected positions presented mixed calls caused by artificial sequencing errors. These positions were identified as both FNs and FPs by the Haplotype Caller and were appropriately removed by BovTB in later stages of the analysis. As a result, these mixed positions were ignored in the rest of the comparisons.

HomoplasyFinder identified 64 (7.65%) homoplasic positions among the generated sequences (**Figure 3**), mostly located within PE/PPE family proteins, intragenic regions or the *pks12* gene (data not shown). A similar proportion (6.70–7.45%) of homoplasies was identified in the alignments obtained from all of the evaluated pipelines. The removal of proximal SNPs reduced homoplasies to an average of 2%, similarly to what was

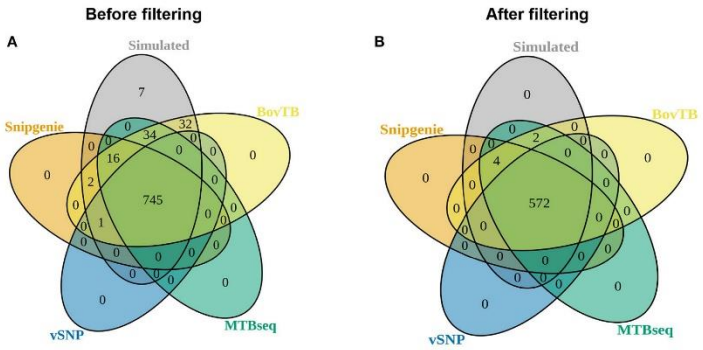


FIGURE 4 | Venn diagram showing the agreement between the positions identified by the different pipelines and the simulated dataset **(A)** before and **(B)** after hard filters (filter F) were applied.

observed for the removal of all of the problematic regions (filter G). Filtering of problematic regions with the proximity filter produced an additional reduction to 1%; most of the reduction was obtained with the removal of PE/PPE proteins alone and additional filters did not decrease the proportion significantly. Once all filters had been applied, all pipelines presented a reduced proportion of homoplasies compared to the ones present in the simulation. Finally, the use of default filters had a varying effect in the proportion of homoplasies, with vSNP and SNiPgenie obtaining the highest reduction in homoplasic positions.

Pipeline Agreement

There was a high agreement between the SNPs identified by the different pipelines and those in the simulated genomes, with the majority of simulated SNP positions being appropriately detected (**Figure 4A**). When proximal SNPs and repetitive sequences were filtered (filter F), there was an increase in the agreement between pipelines (**Figure 4B**). An identical agreement was observed when repetitive sequences were filtered without the proximity filter (filter G) (data not shown). SNiPgenie, BovTB and MTBseq were able to identify all of the SNPs from the simulation, while vSNP was not able to detect 7 SNPs (**Figure 4B**).

Tree Distance Comparison

The analysis of RF distances from best trees and bootstrap replicates revealed that trees output by the different pipelines clustered together with their simulated counterpart (**Figure 5**). In addition, cluster positioning was dependent on the type of hard filter used during the analysis. Trees obtained from the removal of problematic regions through filters D, E and F clustered together in one single group, whereas proximal filters (filter B) produced an intermediate clustering between unfiltered and filtered trees. The application of default filters (filter C) had an uneven effect in the different pipelines; BovTB trees did not separate considerably from proximally filtered trees (**Figure 5C**), whereas the trees produced by vSNP and SNiPgenie were closely related to other filtered trees.

Pairwise Phylogenetic Comparisons

In order to further evaluate topological differences among trees, a pairwise comparison of best trees obtained from each of the pipelines was carried out against their simulated counterpart (**Figure 6**; **Supplementary Figures 3, 4**). In general, there was a high level of agreement among trees and pipelines, with agreement being highest among filtered trees and, especially, among those obtained from BovTB (**Figure 6**). Among default filtered trees, those obtained from SNiPgenie and BovTB presented a higher agreement with the simulation than vSNP (**Supplementary Figure 4**).

Three major groups of taxa could be identified in all trees and no inter-cluster exchange was observed between pipelines. Among unfiltered trees, several isolates presented a small change in their relative location within the tree in the different pipelines (e.g., isolates 9, 10 and 11 in vSNP, 17 and 18 in SNiPgenie, or 28 in MTBseq and BovTB), sharing their Most Recent Common Ancestor (MRCA) with a different group of isolates to the one observed in the simulation. The main differences among filtered trees were produced by small topological variations among highly related taxa (e.g., isolates 12, 13, and 14) and the appearance of polytomies further contributed to the topological differences observed with the simulated tree. When compared against filter F, filter G resolved a small number of polytomies (**Supplementary Figure 5**, blue squares). In all cases, the filtered trees were highly congruent with the topology represented in the original publication (**Supplementary Figure 6**).

DISCUSSION

The application of WGS technologies in the study of aTB has increased in the last decade around the world. Despite its

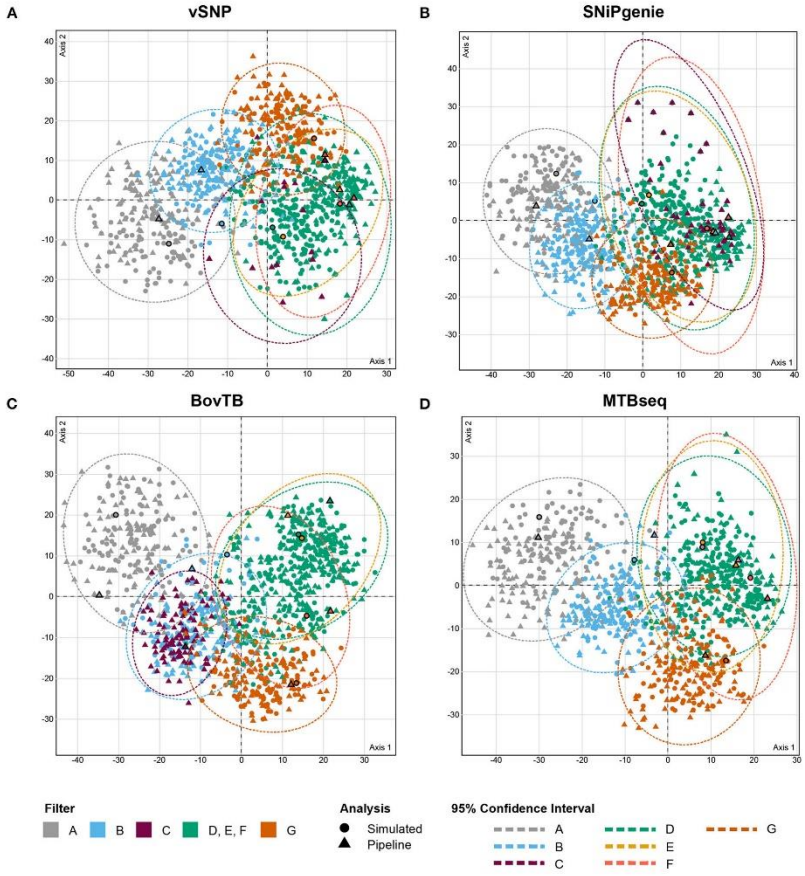


FIGURE 5 | MCA analysis of the RF distances (first two dimensions) between maximum likelihood (ML) trees obtained from core SNP multi-FASTA alignments produced by **(A)** vSNP, **(B)** SNiPgenie, **(C)** BovTB, and **(D)** MTBseq. Shapes without outlines correspond to bootstrap replicates whereas bold shapes correspond to the best ML trees output by RAxML. Color shading corresponds to the hard filtering approach used.

great promise as a higher resolution alternative to traditional molecular techniques in phylogenetic and epidemiological studies, its implementation in the eradication of bTB is still in development.

The digital nature of the data produced by WGS platforms and tools facilitates the exchange of information between laboratories, fostering collaboration between countries and organizations tasked with aTB control. However, the plethora of tools, parameters, protocols and types of analyses available may introduce variations that hamper this process of communication. Standardized procedures and parameters are needed in order to reduce the effect of these variations.

Prior to any standardization taking place, there is a need to evaluate the currently available techniques. None of the variant calling pipelines designed up to date for aTB have been benchmarked in the scientific literature, leading to uncertainty regarding the best method to implement in laboratories that are considering incorporating WGS analyses into their workflows. The aim of this study was to carry out an evaluation of the performance of the currently available aTB variant calling

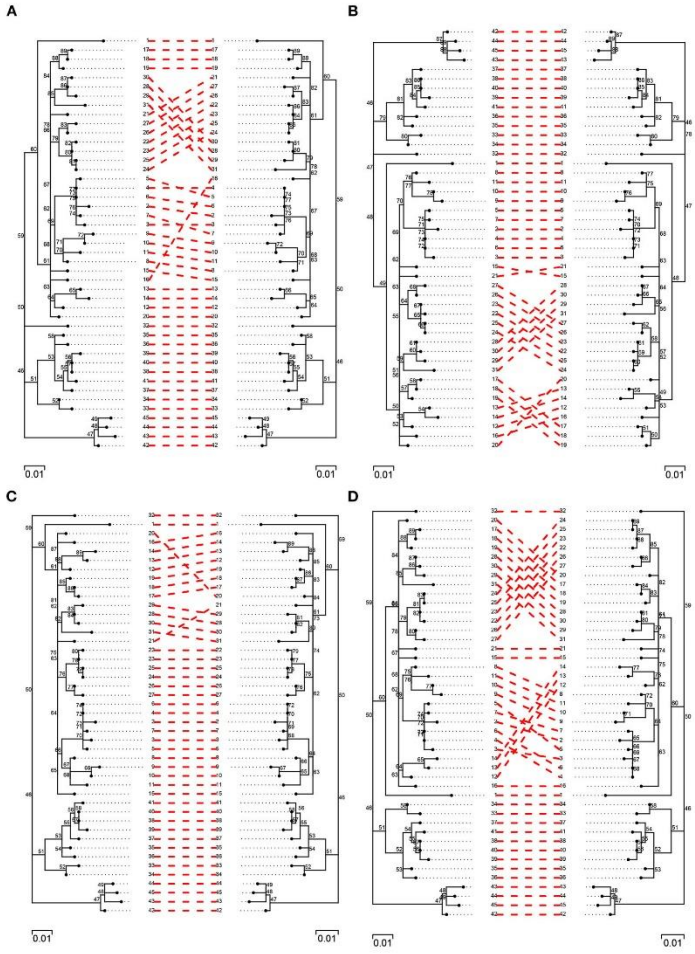


FIGURE 6 | Pairwise comparison of filtered (filter F) simulated trees (left) and trees obtained from the evaluated pipelines (right): **(A)** vSNP, **(B)** SNiPgenie, **(C)** BovTB, and **(D)** MTBseq.

pipelines and assess the degree of between-pipeline agreement in order to inform animal health authorities and laboratories.

The four pipelines evaluated in this study follow a similar procedure to other variant calling pipelines and employ bioinformatics tools widely used in the study of microorganisms. All pipelines use BWA as their sequence aligner (37), and the main procedural differences between them are related to the type of variant calling tools employed, and more importantly, the filtering process applied posteriorly.

The reduced sensitivity of vSNP, SNiPgenie, and MTBseq suggest that quality filters alone can negatively affect the performance of variant calling tools (21). The default minimum coverage settings of MTBseq, coupled with the minimum Allele Frequency of 75%, are important thresholds for the removal of

possible strand bias but were not well suited for the simulated dataset at hand in which SNPs were present in the sample but strand bias was present. This highlights the importance of evaluating and adapting these parameters to the data being evaluated, as suggested by the developers (29). Discrepancies in performance between pipelines were related to a varying proportion of erroneous calls (FP and FN SNPs). In general, a low number of FP and FN SNPs were found in this study, in contrast with previous results in which a high number of erroneous calls were identified by different caller combinations used along with BWA for the analysis of *M. tuberculosis* sequences (21). This could be due to the different approach used in this study for simulated genome and read generation and pipeline-specific filters.

All of the FNs produced by vSNP, SNiPgenie, and MTBseq, and a small proportion of those produced by BovTB, were located within or near repeat-rich regions. A small number of FN positions were identified by all four pipelines and only one of these positions was due to a reduced coverage (<10). This suggests that differences in FN calls could be due to how sample-specific filters deal with low quality regions.

BovTB was the only pipeline to identify FP SNPs and an evaluation of these positions revealed that they were a result of being identified as mixed positions. These were, in addition, partially responsible for a small proportion of the FN SNPs and were effectively removed from the analysis during the consensus calling. As a result, these positions did not have any effect in posterior analyses.

Identified variants are usually translated into phylogenetic trees as a visual aid to assess the genetic relatedness between strains, which can help to identify epidemiologically related isolates suggestive of transmission. In this study, the phylogenetic trees obtained from the different pipelines clustered together with their respective simulated trees and a pairwise inspection revealed a high level of agreement between simulated and pipeline-specific trees, especially on those obtained from BovTB. The phylogenetic trees inferred from the unfiltered SNP alignments obtained from each pipeline were compared against the reference phylogeny in order to assess the effect of pipeline performance in phylogenetic inference. Small divergences were identified in vSNP, SNiPgenie, and MTBseq with respect to the simulation, which are probably a consequence of a reduced sensitivity due to the application of stringent quality filters alone.

Despite the abovementioned effects of quality filters in performance, these are rarely the only parameter taken into consideration when carrying out variant calling in MTBC species. Repeat-rich regions, such as PE/PPE family proteins, mobile genetic elements or direct repeats, are generally considered low confidence regions either due to a higher error rate or mapping issues (19, 38), which complicates the variant calling process and could give rise to FN and FP SNPs. Indeed, the majority of erroneous calls in our simulation were identified in repeat-rich sequences, especially in *pe/ppe* genes and the *pks12* gene.

In addition to the technical constraints that repetitive regions pose to sequencing procedures and mapping algorithms, these can also have a negative impact in phylogenetic inference due to the occurrence of homoplasies. These are genetic traits that can arise independently in separate lineages due to different causes, mainly as a result of convergent evolution but also as a consequence of sequencing and mapping errors. Homoplasic events can add varying grades of background noise in phylogenetic signals and, therefore, must be taken into consideration (39). Due to the limited genetic variation in *M. bovis* strains, this could be especially relevant in closely related isolates and could potentially alter the epidemiological conclusions drawn from outbreak investigations (40). Although homoplasies can be identified anywhere in the genome, they are more frequent in repeat-rich regions such as *pe/ppe* genes (41).

In our study, a small proportion of homoplasies were identified in the simulation, probably due to the reduced number of variant positions in our dataset in comparison with published literature (32). Unfiltered alignments obtained from the different pipelines contained a similar number of homoplasies, although vSNP, and SNiPgenie presented a slightly lower proportion, probably due to their more stringent quality filters. A large reduction in homoplasies was observed when proximity filters were applied and, although the subsequent filtering of repetitive sequences decreased homoplasies further, filtering of these sequences alone (filter G) led to an increase in homoplasies. This indicates that the proximity filter could be an important feature to decrease homoplasies outside the standard repetitive sequences.

Quality filters are, therefore, usually coupled with the removal of problematic regions, an approach nowadays considered a standard procedure in WGS analyses of MTBC species (19). There is, however, no current consensus as to which of these regions should be included in the hard filtering process. In our study, filtering out a progressive amount of regions increased the sensitivity of vSNP and SNiPgenie to levels similar to those observed for BovTB or MTBseq. This positive effect in performance was especially evident for *pe/ppe* genes and mobile genetic elements, and is probably a result of the increased weight of these sequences in the overall composition of the *M. bovis* genome (7–10%) (42). Interestingly, the use of proximal filters had a strong effect in the clustering of phylogenetic trees with a clear separation of these from unfiltered trees, which in turn could be due to the large reduction in homoplasies. Agreement between pipelines also improved with filtering, indicating a difference in the stringency in which the evaluated pipelines deal with problematic regions, and their dependency on posterior masking for removal of low confidence regions.

Hard filtering also had a positive effect in the agreement between phylogenetic trees, as was reflected by the reduced differences among the best ML trees. Although topological differences were identified, these were limited to a reduced number of isolates and polytomies and did not alter the relationship between isolates as seen in the original publication (22). These topological variations are probably related to the overall low bootstrap support values of the identified clusters (43), which in turn could be due to the limited genetic diversity observed in the original dataset in which our simulation is based on. Indeed, *M. bovis* isolates in the original publication presented a maximum of 35 SNPs with respect to each other and a median distance of 14 SNPs once all filters were applied (22). Such a reduced diversity reflects a common drawback

encountered during *M. bovis* outbreak investigations, in which isolates from the same outbreak can accumulate a very small number of variants, hampering the definition of transmission events (26). Three different clusters were identified in our dataset in which genetic distances of *M. bovis* isolates were within 12 SNPs from each other; the maximum cut-off recommended for possible recent transmission of *M. tuberculosis* (44). In addition, polytomies can be resolved with increased isolate sampling, for example by including samples from wildlife or nearby breakdown events. However, this may not be a feasible option in many aTB outbreak investigations and, therefore, a removal of certain hard filters could be an interesting alternative to increase the amount of available informative SNPs. However, this alternative should be balanced to the risk of introducing possible biases or erroneous calls, such as FP SNPs. For example, although the removal of repetitive sequences without the proximity filter (filter G) increased the resolution of several polytomies, the increase in homoplasies could affect phylogenetic inference and needs to be considered.

There is little information as to how reliable low confidence regions are in phylogenetic inference, as their analysis has led to conflicting conclusions (45, 46). Nevertheless, there has been an increasing interest in the usefulness of filtering repeat-rich regions and recent data indicate that more than a half of the masked repetitive regions could be accurately identified using Illumina platforms (38). Even with the limitations of short-read sequencing platforms, the use of *de novo* assemblies or more refined masking filters may allow informative SNPs to be identified and retained (21, 38, 47). Furthermore, the introduction of long read sequencing could greatly improve the detection of variants within these regions of the genome (19). Improvements in the WGS analysis of problematic regions in MTBC species will surely benefit the field of aTB in the near future.

Pipeline choice may be based on other factors in addition to performance, and these have not been evaluated in this study. These include speed, use of disk space and memory, or ease of use, be it through the implementation of a GUI (SNiPgenie), limited command requirements (vSNP or BovTB) or by a straightforward data representation (vSNP and MTBseq) which could allow for more inexperienced users to access the bioinformatics analyses. In addition, the inclusion of additional analyses, such as antibiotic resistance profiling and cluster analysis (MTBseq), detection of INDELs and Regions of Difference (SNiPgenie), or lineage definition (vSNP, BovTB and MTBseq) could also be of interest for certain studies. However, in a similar manner to pipeline parameters, there is currently no standardized *M. bovis* lineage classification nor nomenclature based on WGS data. Although recent studies have suggested different lineages for *M. bovis* (48, 49), efforts toward this goal are still required. It is important to highlight that the results of our study are limited to simulated data and may not be representative of a real-life outbreak. The dataset used to generate our simulation does not correspond to an outbreak investigation but to a prevalence study. As a result, the capacity of each pipeline was approximated through their level of agreement with the simulation, rather than on their capacity to investigate true herd breakdown events. In addition, although this simulation partly mimics the negative impact of GC-rich sequences in genome coverage, it may be an underestimate in comparison to the actual sequencing of *M. bovis* isolates. Recent data highlight the existence of coverage blind spots in the *M. tuberculosis* reference genome which result from library preparation, sequencing as well as specific sequence attributes, such as homopolymers (50). Therefore, further work on a real-world dataset with a validated SNP profile and appropriate metadata is needed to evaluate these sources of bias.

Furthermore, the use of *M. bovis* AF2122/97 as a scaffold for the generation of simulated genomes meant that there were no sample-specific deletions, and therefore the capacity of these pipelines in calling SNPs near deletion events could not be evaluated. Furthermore, as is the case in human TB with *M. tuberculosis*, the choice of reference genome could also have an important effect in the WGS analysis of aTB due to differences in gene content between lineages, which could be masked by an inappropriate reference selection (19). This could be especially relevant when considering that traditional *M. bovis* lineages or clonal complexes are usually defined based on lineage-specific RDs, such as RDEu1 for the European 1 (Eu1) complex, or that certain genomic deletions may occur independently, such as the RD900 deletion. This study focused on the use of *M. bovis* AF2122/97, an Eu1 complex strain which is the default genome used by the evaluated aTB pipelines and the most extensively used *M. bovis* reference genome. However, the use of this reference genome in regions in which other clonal complexes are prevalent, such as the African 1 in western Africa, may lead to a loss of phylogenetic information. Therefore, other reference genomes may be better suited for different countries or regions and should be evaluated in the future.

Finally, it is important to note that, unlike other pipelines, manual and visual curation of SNPs is an important component of vSNP's design and functioning. As a result, a more detailed evaluation of this pipeline's results may have led to a reduced number of inconsistencies but would have added subjectivity to this comparison and was therefore avoided.

In conclusion, despite the above-mentioned limitations, the results of our comparison show that all evaluated pipelines perform well as long as similar hard filters are used, with minor differences amongst them with regard to performance and phylogenetic inference. This highlights the importance of standardizing and appropriately annotating filtering files when analyses are carried out between different laboratories or countries, and in particular in the context of aTB disease control.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.5281/zenodo.5179838, 5179838; https://github.com/Viloleal/bTB-pipeline-comparison-data-and-tools, None.

AUTHOR CONTRIBUTIONS

VL-L was involved in conceptualization, data curation, formal analysis, investigation, software, validation, and writing of the original draft. DF was involved in conceptualization, data curation, methodology, software, validation, and review and editing of the original draft. BR and LJ were involved in conceptualization, funding acquisition, and review and editing of the original draft. JÁ was involved in conceptualization and resources and review and editing of the original draft. SG was involved in conceptualization, funding acquisition, resources, validation, supervision, and review and editing of the original draft. All authors contributed to the article and approved the submitted version.

FUNDING

The author VL-L was funded with a predoctoral grant from the Complutense University of Madrid and Banco Santander 2017–2018 (CT17/17—CT18/17). This research project was made possible thanks to an international fellowship grant from the Complutense University of Madrid and Banco Santander 2020 (EB25/20). DF and SG acknowledge funding from the Department of Agriculture Food and the Marine (DAFM) award 2019R404 (BTBGenIE). This project was partially funded by the EU-RL for Bovine Tuberculosis.

ACKNOWLEDGMENTS

We would like to thank and acknowledge the valuable cooperation and feedback provided by Richard Ellis (APHA, BovTB), Jessica Hicks (USDA-APHIS, vSNP), and Tod Stuber (USDA-APHIS, vSNP) in the completion of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.780018/full#supplementary-material

REFERENCES

1. OIE. "Infection with *Mycobacterium tuberculosis* complex," In: *OIE—Terrestrial Animal Health Code*. (2021), 6.
2. OIE. "Bovine Tuberculosis," In: *OIE Terrestrial Manual*. (2018), 1058–1074.
3. European-Commission. "Regulation (EU) 2016/429 of the European Parliament and of the Council of 9 March 2016 on transmissible animal diseases and amending and repealing certain acts in the area of animal health ('Animal Health Law')," In: *EU 2016/429*. European Union Commision, (2016). EUR-Lex.
4. Livingstone PG, Hancox N, Nugent G, de Lisle GW. Toward eradication: the effect of *Mycobacterium bovis* infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. *N Z Vet J*. (2015) 63:4–18. doi: 10.1080/00480169.2014.971082
5. VerCauteren KC, Lavelle MJ, Campa H. Persistent spillback of bovine tuberculosis from white-tailed deer to cattle in Michigan, USA: status, strategies, and needs. *Front Vet Sci*. (2018) 5:301. doi: 10.3389/fvets.2018.00301
6. Gormley E, Corner LAL, Costello E, Rodriguez-Campos S. Bacteriological diagnosis and molecular strain typing of *Mycobacterium bovis* and *Mycobacterium caprae*. *Res Vet Sci*. (2014) 97:S30–43. doi: 10.1016/j.rvsc.2014.04.010
7. Durr PA, Clifton-Hadley RS, Hewinson GR. Molecular epidemiology of bovine tuberculosis - II. *Applicat Genotyp Revue Scientifique et Technique*. (2000) 19:689–701. doi: 10.20506/rst.19.3.1240
8. Frothingham, R., and Meeker-O'connell, W. A. (1998). Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology*. 144, 1189–1196. doi: 10.1099/00221287-144-5-1189
9. Kamerbeek, J., Schouls, L., Kolk, A., Agterveld, M., v., Soolingen, D., et al. Simultaneous Detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clinic Microbiol*. (1997) 35, 907–914. doi: 10.1128/jcm.35.4.907-914.1997
10. Rodriguez-Campos S, Navarro Y, Romero B, de Juan L, Bezos J, Mateos A, et al. Splitting of a prevalent *Mycobacterium bovis* spoligotype by variable-number tandem-repeat typing reveals high heterogeneity in an evolving clonal group. *J Clin Microbiol*. (2013) 51:3658–65. doi: 10.1128/JCM.01271-13
11. Rodriguez S, Romero B, Bezos J, de Juan L, Alvarez J, Castellanos E, et al. High spoligotype diversity within a *Mycobacterium bovis* population: clues to understanding the demography of the pathogen in Europe. *Vet Microbiol*. (2010) 141:89–95. doi: 10.1016/j.vetmic.2009.08.007
12. Smith NH. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infect Genet Evol*. (2012) 12:857–65. doi: 10.1016/j.meegid.2011.09.007
13. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE*. (2009) 4:e7815. doi: 10.1371/journal.pone.0007815
14. Rodriguez-Campos S, Aranaz A, de Juan L, Saez-Llorente JL, Romero B, Bezos J, et al. Limitations of spoligotyping and variable-number tandem-repeat typing for molecular tracing of *Mycobacterium bovis* in a high-diversity setting. *J Clin Microbiol*. (2011) 49:3361–4. doi: 10.1128/JCM.00301-11
15. Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev*. (2015) 28:541–563. doi: 10.1128/CMR.00075-13
16. Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, et al. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol*. (2014) 52:2479–86. doi: 10.1128/JCM.00567-14
17. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential? *Clin Microbiol Infect*. (2018) 24:604–9. doi: 10.1016/j.cmi.2017.10.030
18. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*. (2015) 15:1193–202. doi: 10.1016/S1473-3099(15)00062-6
19. 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. doi: 10.1038/s41579-019-0214-5
20. Jajou R, Kohl TA, Walker T, Norman A, Cirillo DM, Tagliani E, et al. Toward standardization: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Euro Surveill*. (2019) 24:1900130. doi: 10.2807/1560-7917.ES.2019.24.50.1900130
21. Walter KS, Colijn C, Cohen T, Mathema B, Liu Q, Bowers J, et al. Genomic variant-identification methods may alter *Mycobacterium*

tuberculosis transmission inferences. *Microb Genom.* (2020) 6:8. doi: 10.1099/mgen.0.000418
22. Crispell J, Cassidy S, Kenny K, McGrath G, Warde S, Cameron H, et al. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. *Microb Genom.* (2020) 6:20. doi: 10.1099/mgen.0.000388
23. Hauer A, Michelet L, Cochard T, Branger M, Nunez J, Boschiroli ML, et al. Accurate phylogenetic relationships among *Mycobacterium bovis* strains circulating in france based on whole genome sequencing and single nucleotide polymorphism analysis. *Front Microbiol.* (2019) 10:955. doi: 10.3389/fmicb.2019.00955
24. Kohl TA, Kranzer K, Andres S, Wirth T, Niemann S, Moser I. Population Structure of *Mycobacterium bovis* in Germany: a long-term study using whole-genome sequencing combined with conventional molecular typing methods. *J Clin Microbiol.* (2020) 58:e01573–20. doi: 10.1128/JCM.01573-20
25. Orloski K, Robbe-Austerman S, Stuber T, Hench B, Schoenbaum M. Whole genome sequencing of *Mycobacterium bovis* isolated from livestock in the United States, 1989–2018. *Front Vet Sci.* (2018) 5:253. doi: 10.3389/fvets.2018.00253
26. Price-Carter M, Brauning R, de Lisle GW, Livingstone P, Neill M, Sinclair J, et al. Whole genome sequencing for determining the source of *Mycobacterium bovis* infections in livestock herds and wildlife in New Zealand. *Front Vet Sci.* (2018) 5:272. doi: 10.3389/fvets.2018.00272
27. Yue JX, Liti G. simuG: a general-purpose genome simulator. *Bioinformatics.* (2019) 35, 4442–4. doi: 10.1093/bioinformatics/btz424
28. Frampton M, Houlston R. Generation of artificial FASTQ files to evaluate the performance of next-generation sequencing pipelines. *PLoS ONE.* (2012) 7:e49110. doi: 10.1371/journal.pone.0049110
29. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ.* (2018) 6:e5895. doi: 10.7717/peerj.5895
30. Chen, H. (2018). *Generate High-Resolution Venn and Euler Plots.* Available online at: https://cran.r-project.org/web/packages/VennDiagram/VennDiagram.pdf
31. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly.* 6:80–92. doi: 10.4161/fly.19695
32. Crispell J, Balaz D, Gordon SV. HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. *Microb Genom.* (2019) 5:1. doi: 10.1099/mgen.0.000245
33. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom.* (2016) 2:e000056. doi: 10.1099/mgen.0.000056
34. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
35. Jombart T, Kendall M, Almagro-Garcia J, Colijn C. treespace: Statistical exploration of landscapes of phylogenetic trees. *Mol Ecol Resour.* (2017) 17:1385–92. doi: 10.1111/1755-0998.12676
36. Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* (2012) 3:217–23. doi: 10.1111/j.2041-210X.2011.00169.x
37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
38. Marin M, Vargas R, Harris M, Jeffrey B, Epperson LE, Durbin D, et al. Genomic sequence characteristics and the empiric accuracy of short-read sequencing. *bioRxiv.* (2021) 2004:438862. doi: 10.1101/2021.04.08.438862
39. Brandley MC, Warren DL, Leache AD, McGuire JA. Homoplasy and clade support. *Syst Biol.* (2009) 58:184–98. doi: 10.1093/sysbio/syp019
40. Nakanishi N, Wada T, Arikawa K, Millet J, Rastogi N, Iwamoto T. Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages. *Infect Genet Evol.* (2013) 16:174–7. doi: 10.1016/j.meegid.2013.02.007
41. Tantivitayakul P, Ruangchai W, Juthayothin T, Smittipat N, Disratthakit A, Mahasirimongkol S, et al. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in *Mycobacterium tuberculosis*. *Sci Rep.* (2020) 10:1–10. doi: 10.1038/s41598-020-64895-4
42. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* (1998) 393:537–44. doi: 10.1038/31159
43. Pettengill JB, Luo Y, Davis S, Chen Y, Gonzalez-Escalona N, Ottesen A, et al. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. *PeerJ.* (2014) 2:e620. doi: 10.7717/peerj.620
44. Walker TM, Ip CLC, 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. doi: 10.1016/S1473-3099(12)70277-3
45. Godfroid M, Dagan T, Kupczok A. Recombination signal in *Mycobacterium tuberculosis* stems from reference-guided assemblies and alignment artifacts. *Genome Biol Evol.* (2018) 10:1920–6. doi: 10.1093/gbe/evy143
46. Phelan JE, Coll F, Bergval I, Anthony RM, Warren R, Sampson SL, et al. Recombination in *pe/ppe* genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. *BMC Genom.* (2016) 17:151. doi: 10.1186/s12864-016-2467-y
47. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respirat Med.* (2013) 1:786–92. doi: 10.1016/S2213-2600(13)70231-5
48. Loiseau C, Menardo F, Aseffa A, Hailu E, Gumi B, Ameni G, et al. An African origin for *Mycobacterium bovis*. *Evol Med Public Health.* (2020) 2020:49–59. doi: 10.1093/emph/eoaa005
49. Zimpel CK, Patane JSL, Guedes ACP, de Souza RF, Silva-Pereira TT, Camargo NCS, et al. Global Distribution and Evolution of *Mycobacterium bovis* Lineages. *Front Microbiol.* (2020) 11:843. doi: 10.3389/fmicb.2020.00843
50. Modlin SJ, Robinhold C, Morrissey C, Mitchell SN, Ramirez-Busby SM, Shmaya T, et al. Exact mapping of Illumina blind spots in the *Mycobacterium tuberculosis* genome reveals platform-wide and workflow-specific biases. *Microb Genom.* (2021) 7:3. doi: 10.1099/mgen.0.000465

Conflict of Interest: SNiPgenie was developed by DF and SG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

4.2.2. Application of Whole Genome Sequencing to Assess Within-herd Variation of *M. bovis* SB0121 Isolates in Spain

This study evaluated the use of WGS to assess the within-herd genetic variation of *M. bovis* SB0121, in chronically bTB-infected herds in Spain. This was achieved through the sequencing of 70 *M. bovis* SB0121 isolates from 22 herds located in six provinces of Spain (Madrid, Ciudad Real, Toledo, Zaragoza, Teruel and Valencia), in which infection with this genotype was confirmed during at least three different sampling dates, with a minimum of three months between samplings. Epidemiological surveys for the selected herds were collected through a collaboration with the veterinary authorities tasked with bTB control in the studied regions. Twenty-nine *M. bovis* SB0121 isolates from wildlife were included in order to assess the role of wildlife in the persistence of bTB in these herds. The sequencing analysis was carried out using Illumina short-read sequencing and the phylogenetic analysis was performed through the use of a variant calling pipeline (vSNP).

Overall, the phylogenetic analysis revealed the presence of a high genetic diversity of *M. bovis* SB0121 among the studied isolates, with three major genetic clades (11, 12 and 13) distributed across the studied regions. The within-herd genetic diversity of the isolates revealed a complex epidemiological situation in the majority of the herds, in which different factors could have contributed to the maintenance and/or recurrence of the infection. The detection of highly similar strains during short (< 1 year) and long periods (> 1 year) of time indicated the presence of an ongoing infection in approximately half of the herds (n = 12) and the persistence of genetically related strains within the herd or its surroundings. In contrast, the detection of distantly related or unrelated isolates at short and/or long time intervals in 19 herds indicated the presence of a high genetic diversity within the herd or its surroundings, as well as multiple re-introductions of the infection. Isolates with less than 12 SNP differences were detected between different herds (n = 9), as well as between different species (cattle-wildlife and wildlife-wildlife), indicating a possible epidemiological link between herds and/or wildlife reservoirs. The epidemiological data revealed possible causes of persistence in several herds, such as the presence of shared pastures, the presence of wildlife in the surroundings or the introduction of animals from infected herds.

The results of this study reflect the complex epidemiology of bTB in Spain and highlight the potential of WGS technologies in establishing the possible causes of persistence within a herd, which could enhance outbreak investigations and aid in the eradication of the infection in Spain. Further studies are required that should include more detailed epidemiological data, as well as a more representative sample of cattle and wildlife, in order to assess the diversity of *M. bovis* in Spain and the possible directionality of transmission between hosts.

This study has not been published in any scientific journal at the time this thesis was written. Nevertheless, the results and conclusions have been included in the following sections.

Application of Whole Genome Sequencing to Assess Within-herd Variation of *M. bovis* SB0121 Isolates in Spain

Víctor Lorente-Leal[1,2], Pilar Pozo[1], Julio Álvarez[1,2], Javier Bezos[1,2], José Luis Sáez[3], Joaquín Vicente[4], Thyler Thacker[5], Lucía de Juan[1,2], Beatriz Romero[1,2]

1. VISAVET Animal Health Surveillance Centre, Universidad Complutense de Madrid, Spain.

2. Animal Health Department, Faculty of Veterinary Medicine, Universidad Complutense de Madrid, Spain.

3. Subdirección General de Sanidad e Higiene Animal y Trazabilidad, Dirección General de Sanidad de la Producción Agraria, Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain

4. Health and Biotechnology group (SaBio), Instituto de Investigaciones en Recursos Cinegéticos, Universidad de Castilla la Mancha, Spain.

5. National Veterinary Services Laboratories (NVSL) Diagnostic Bacteriology and Pathology Laboratory, Animal and Plant Health Inspection Service, Veterinary Services, Ames, IA, USA.

4.2.2.1. Abstract

The presence of herds persistently infected with bTB is an important limitation of bTB eradication programmes, since they require the use of increased resources and are a risk factor for the infection of other herds or wildlife. Whole genome sequencing is a promising technique for the study of bTB due to its high throughput and resolution. This study evaluated the use of WGS for the analysis of the within-herd genetic diversity of *M. bovis* SB0121 in chronically infected herds as a means of assessing possible causes of persistence. The study included 70 *M. bovis* isolates from 22 cattle herds from six Spanish provinces, as well as 29 *M. bovis* isolates obtained from wildlife species sampled in the surroundings of the herds. Isolates were classified in three major clades (11, 12 and 13), revealing a high genetic diversity for the *M. bovis* isolates included in this study. The analysis of the sequences using vSNP revealed a within-herd diversity ranging from 5-310 median SNP differences. The presence of *M. bovis* isolates with less than six SNP differences in 12 herds indicated the circulation of highly similar isolates within the herd and/or its surroundings, either through a common source of infection or the presence of undetected infected animals. The presence of more distantly related (6-12 SNPs) and/or unrelated (> 12 SNPs) isolates in 19 herds suggested the presence of multiple reintroductions of the infection, as well as a potentially high genetic diversity in the surroundings of the herds. Furthermore, isolates with less than 12 SNP differences were identified between nine herds, as well as between different animal species (cattle and wildlife), indicating the presence of possible epidemiological links, such as common sources of infection between

herds and wildlife hosts. The findings of this study highlight the fitness of WGS for the study of bTB outbreaks and pinpoints its potential in unravelling sources of bTB persistence in cattle herds.

4.2.2.2. Introduction

Bovine TB is a chronic infectious disease of bovids caused by the members of the MTBC, mainly *M. bovis* and *M. caprae* (OIE, 2018). Due to the relevance of this disease in cattle production and its zoonotic potential, bTB is subjected to control and eradication programmes in many countries around the world. In the EU, the basis of these programmes resides in the detection of infected animals through the use of official diagnostic tests, such as the intradermal tuberculin skin test and/or the interferon-gamma test, and the culling of reactor animals to prevent the spread of the infection.

Persistently infected herds are an important obstacle in bTB eradication, since they require the use of increased resources and are a potential source of infection for other herds and the environment (Karolemeas *et al.*, 2011; Doyle *et al.*, 2016). Persistence of bTB may be caused by the re-introduction of the infection (reinfection) from different exogenous sources, including shared pastures with infected herds, contact with infected wildlife or the introduction of infected animals from other herds. In addition, persistence may be caused by infection recurrence or relapse, which relates to endogenous sources that maintain the infection within the herd, such as the presence of undetected infected animals due to a lack of performance of current *ante-mortem* diagnostic tools (Alvarez *et al.*, 2014), or the presence of other domestic reservoirs in the herd, such as goats (Zanardi *et al.*, 2013).

In Spain, herd prevalence has slowly, but steadily decreased in the last decade as a result of the implementation of the National bTB eradication programme, reaching 1.61% in 2020 (MAPA, 2022). Nevertheless, the distribution of the infection in Spain is heterogeneous, with herd prevalence ranging between 0-10.97% in 2020, and is generally restricted to the south-central regions of the country (MAPA, 2022). This difference is the result of a combination of factors associated with the difficulties in detecting the infection and the complex epidemiology of bTB in these regions, in which multiple risk factors associated with bTB persistence coexist, such as high local bTB prevalence in cattle and wildlife (Guta *et al.*, 2014; Martinez-Lopez *et al.*, 2014; Doyle *et al.*, 2016; Pozo *et al.*, 2020), the presence of wild game hunting estates in the vicinity of the herds (Martinez-Lopez *et al.*, 2014), different breeds (e.g. bullfighting) or management

types (intensive, extensive and shared pastures) (Karolemeas *et al.*, 2011; Alvarez *et al.*, 2014; Guta *et al.*, 2014; Pozo *et al.*, 2020), among others.

Molecular characterisation of MTBC isolates is critical to evaluate the level of relatedness between the involved strains and establish possible epidemiological links. Currently, the most common molecular characterisation techniques used to study bTB are DVR-spoligotyping and MIRU-VNTR analysis, which evaluate a small portion of the mycobacterial genome and establish the relationship between isolates according to the observed similarities in the molecular patterns, either through the presence or absence of spacers (DVR-spoligotyping) or the number of tandem copies (MIRU-VNTR) (Gormley *et al.*, 2014). Much of what is known about *M. bovis* population structure and global distribution today has been based on these methods (Smith *et al.*, 2006a; Smith, 2012). The majority of *M. bovis* isolates in the Iberian Peninsula are characterised by the absence of spacer 21 and a SNP in the *guaA* gene, which are molecular hallmarks of the European-2 clonal complex (Rodriguez-Campos *et al.*, 2012b). The use of these techniques in Spain revealed a high genetic diversity among *M. bovis* strains, with more than 200 different molecular patterns (Rodriguez *et al.*, 2010; Rodriguez-Campos *et al.*, 2013). The most frequent *M. bovis* genotype is SB0121, which represented 27.94% of the cultured *M. bovis* strains in Spain between 1997 to 2007, followed by SB0134 (11.21%) and SB0339 (8.06%) (Rodriguez *et al.*, 2010). High prevalence *M. bovis* genotypes are well distributed across the country, although in certain cases they can be frequently isolated in certain regions (e.g. SB0339 in the Community of Madrid) (Pozo, 2020). Less prevalent genotypes have been found to be more geographically restricted, such as SB1142 in Madrid or SB1230 in the South of Spain (Rodriguez *et al.*, 2010). In addition, specific *M. bovis* genotypes have been identified both in domestic animals, such as cattle and goats, as well as in different wildlife species, mainly wild boar and red deer (Rodriguez *et al.*, 2010).

Due to the highly clonal genomes of MTBC members and the reduced number of evaluated loci, currently used molecular characterisation techniques suffer from a limited resolution and high homoplasy rate that makes them less suitable for outbreak investigations or evolutionary studies (Comas *et al.*, 2009; Rodriguez-Campos *et al.*, 2012b; Trewby *et al.*, 2016). In the last decade, whole genome sequencing (WGS) has been proposed as a very promising tool for the molecular characterisation of bTB thanks to its high throughput and nucleotide-level resolution in comparison to conventional molecular methods (Hauer *et al.*, 2019; Meehan *et al.*, 2019; Reis *et al.*, 2021). This technique has been used to assess *M. bovis* diversity and evolution around the world (Hauer *et*

al., 2019; Loiseau *et al.*, 2020; Zimpel *et al.*, 2020), as well as to perform outbreak investigations in different countries, such as the USA (Bruning-Fann *et al.*, 2017; Orloski *et al.*, 2018), New Zealand (Price-Carter *et al.*, 2018) or the UK (Rossi *et al.*, 2021). Genomic data of *M. bovis* isolates from Spain have revealed a high genetic diversity in cattle and wildlife species, both at a global and regional level (Pozo, 2020; Perea *et al.*, 2021). Genetic similarity between isolates obtained from cattle and wildlife in these studies suggest the transmission of TB between different animal hosts, reflecting the complex epidemiology of bTB in Spain.

Due to the presence of multiple risk factors associated with bTB persistence in high-prevalence regions in Spain, there is a need to evaluate new techniques with which to assess the genetic diversity of chronically infected herds. Despite the increasing use of WGS to examine bTB transmission, its application in the study of bTB persistence has been limited to date (Biek *et al.*, 2012; Pozo, 2020). A study in Northern Ireland revealed that subsequent outbreaks in the same herd were caused by the same genetic lineage isolated in the same region, suggesting the persistence of *M. bovis* within the farm or the surrounding environment (Biek *et al.*, 2012). In contrast, the evaluation of several persistently infected herds in Spain revealed a significant variation in within-herd diversity of *M. bovis* SB0339 isolates in Madrid, which suggests the presence of multiple sources of infection (Pozo, 2020). The aim of the study presented herein was to assess the within-herd diversity of *M. bovis* SB0121 isolates originated from chronically-infected herds and to assess the suitability of WGS in establishing possible causes of persisting bTB infections at the herd level.

4.2.2.3. Materials and methods

Sample selection

The target population for this study consisted of cattle herds included in the Spanish National Mycobacteria Database (mycoDB.es) in which bTB was confirmed by culture for at least three different sampling dates, usually three years, and in which the obtained genotype was spoligotype SB0121 (Rodriguez-Campos *et al.*, 2012a). The minimum interval between samples was three months in order to include animals from consecutive testing performed during the same year.

The study included 90 frozen bovine tissue samples of animals infected with *M. bovis* SB0121 obtained from 22 farms from 20 municipalities in six provinces of Spain; Ciudad Real (n = 9), Madrid (n = 3), Teruel (n = 1), Toledo (n = 2), Valencia (n = 1) and Zaragoza (n = 6) (figure 11). After processing (see below), a total of 70 *M. bovis* SB0121 isolates were recovered. In half of the herds (n = 11), other spoligotypes were also retrieved, especially at the end of the study period (figure 12). The median number of *M. bovis* SB0121 isolates analysed per herd was 3.0 (interquartile range or IQR: 2.5-3.5), with a maximum of six (herd 15) and a minimum of two (herds 3, 7, 8, 16 and 18). The sequenced isolates were recovered throughout a median period of three years (IQR: 2.25-4 years) (figure 12; red dots), with a median interval between years in which the sequenced isolates were retrieved from a given herd of 0.89 years (IQR: 0.41 – 2.12 years) (figure 12, red dots).

Figure 11. Province of origin of the 22 herds analysed in this study.

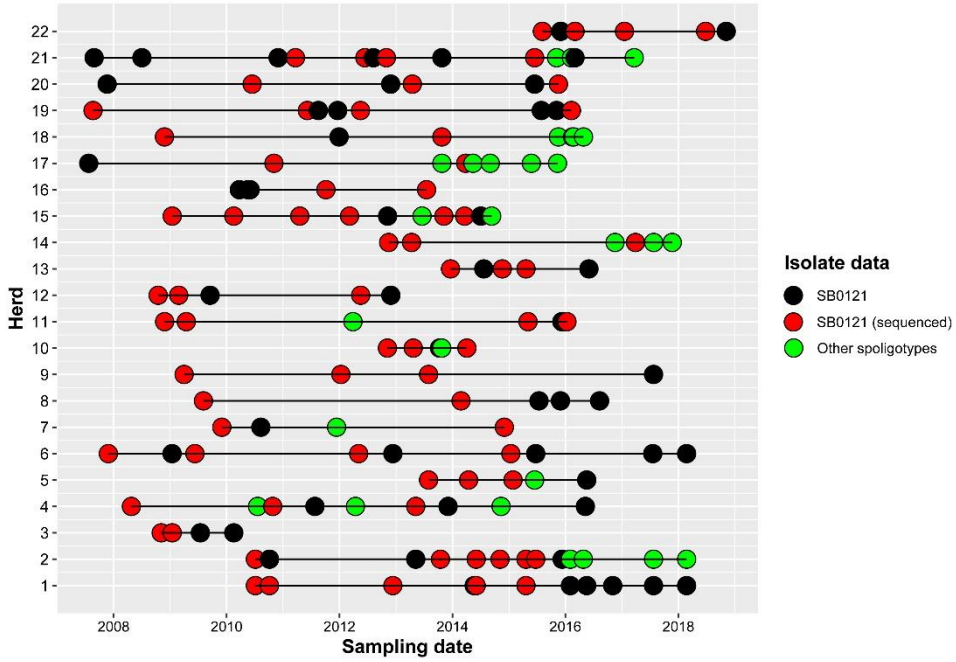


Figure 12. Sampling dates for the tissue samples obtained from the different herds included in this study. Dots represent the isolation date of *M. bovis* in the analysed herds.

Epidemiological surveys that were available for the analysed herds were remitted by the veterinary authorities tasked with bTB control in the Autonomous Regions included in this study. These surveys included questions about possible animal movements, as well as common risk factors associated with bTB persistence, such as, the management type (extensive, intensive or mixed), the type of breed (bullfighting or beef), the use of shared pastures with other herds, the presence of a fence surrounding the herd, or the presence of wildlife in the surroundings (either through shared pastures or nearby hunting estate) (Guta *et al.*, 2014).

In order to establish possible epidemiological links between cattle and wildlife reservoirs, tissue samples obtained from wild animal species and registered in mycoDB were selected based on their geographical location (same or surrounding municipality) and genotype (SB0121). When the samples were not available at the tissue sample collection in VISAVET Health Surveillance Centre, these were remitted by the National Reference Laboratory for bTB in Santa Fe, Granada. A total of 29 *M. bovis* isolates were recovered from 73 tissue samples from wildlife species (15 wild boars, 12 red deer, 1 fallow deer and 1 fox) obtained from Ciudad Real (n = 23), Madrid (n = 4) and Toledo (n = 2).

Sample processing and microbiological culture

The selected samples were re-cultured in the BSL3 facilities at VISAVET. Two mL aliquots of frozen tissue homogenate were thawed and decontaminated in two mL of HPC (final concentration = 0.375%) for 30 minutes. After incubation, samples were centrifuged for 30 minutes at 2,500 x g. The supernatant was then discarded and a sterile swab was used to seed one Coletsos and LJ slant supplemented with sodium pyruvate, respectively. The culture tubes were incubated at 37 ºC for a maximum of 12 weeks or until growth compatible with MTBC was detected.

When MTBC growth was detected in the tubes, one loop-full of culture was transferred to 15 mL of Middlebrook 7H9 broth supplemented with casein, sodium pyruvate and OADC and incubated for 4 to 6 weeks at 37 ºC.

DNA extraction and whole genome sequencing

After incubation, 500 µl of liquid culture was centrifuged at 2,500 x g and the pellet was re-suspended in five mL of phosphate buffered saline (PBS). The resuspended culture was centrifuged again at 2,500 x g and the pellet was re-suspended in 4 mL of PBS. Then, 500 µL of the cell suspension were transferred to two mL screw cap bead beater tubes containing approximately 50 µL of 0.1 mm silica beads and 500 µL of PCI (Sigma-Aldrich). Cell suspensions were then incubated for 30 minutes and, posteriorly, they were mechanically lysed in a Precellys Evolution homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) twice at 9,000 rpm for 40 seconds, with a 10 second interval. After lysis, samples were centrifuged for five minutes at 16,000 x g and approximately 300 µL were collected from the top aqueous solution and mixed with 700 µL cold (-80 ºC) absolute ethanol and 40 µL of 3M sodium acetate buffer. The mixture was inverted ten times, vortexed for one second and incubated overnight at -20 ºC. The mixture was then centrifuged at 16,000 x g for 15 minutes at 4 ºC. After centrifugation, the supernatant was poured off and 1 mL of cold 70% ethanol was carefully added. Samples were centrifuged at 16,000 x g for 5 minutes at 4 ºC and the ethanol was discarded. Finally, samples were centrifuged again at 8,000 x g for 15 seconds and any excess ethanol was removed through pipetting. The DNA pellet was then let dry at room temperature for 20 minutes and resuspended in 50-100 µL of double-distilled water (Sigma Aldrich). The quality and concentration of the DNA was assessed using a NanoDrop spectrophotometer and a Qubit4 fluorometer (ThermoFisher Scientific).

Mycobacterial DNA was submitted to the National Veterinary Services Laboratory (NVSL) in Ames, Iowa (USA) for Illumina short read sequencing. The genomic DNA was processed using the Nextera XT library preparation kit and the prepared libraries were sequenced on a MiSeq instrument (Illumina, San Diego, CA, USA), producing 2 x 250 bp reads.

Read data processing and variant calling

Raw FASTQ files were adapter- and quality-trimmed using Trimmomatic v0.39 (sliding window 15:30; minimum length 36 bp) (Bolger *et al.*, 2014), clumped using clumpify v38.49 from BBtools (Bushnell, 2014) and de-duplicated using PRINSEQ-lite v0.20. De-duplicated reads were then analysed using vSNP 2.03 (USDA-APHIS) according to the pre-established parameters (https://github.com/USDA-VS/vSNP).

Briefly, processed reads were aligned to *M. bovis* AF2122/97 reference genome (NC_002945.4) using BWA (Li & Durbin, 2009), and duplicate mapped reads were removed using Picard MarkDuplicates 2.22.2. Variant calling was then carried out using FreeBayes 1.3.2 (Garrison & Marth, 2012) and SNPs with a base quality below 20 were removed from the obtained VCF files. Zero coverage VCFs output by the pipeline, which include positions for which there was no read coverage in the reference genome, were then used to create a sample-based SNP table. Single nucleotide variants were reported in a position if at least one variant in the dataset had an allele count (AC) of 2 (indicating the absence of a mixed call), a quality (QUAL) value above 150 and MQ higher than 56. In contrast, variants with an AC of 1 were reported as IUPAC ambiguous if quality was above 150, as indeterminate (N) if quality was between 50 and 150; as the reference if quality was below 50; or as "-" if the alternate allele was missing. Variants that were shared among all of the samples (monomorphic SNPs) were removed from the analysis. Positions located in variant-dense regions were removed using a 10 SNP window through an in house script, and repetitive regions of the reference genome were removed with previously defined masking files (Lorente-Leal *et al.*, 2021), which were modified to include additional positions in which mapping was problematic. In addition, spoligotyping was inferred for all isolates using SpoTyping v2.1 (Xia *et al.*, 2016).

Phylogenetic inference, SNP and geographic distance extraction

Maximum likelihood (ML) trees were inferred from the concatenated multi-FASTA alignment files using Randomized Axelerated Maximum Likelihood (RAxML) 8.2.12, the GTR model of substitution and the CAT model for rate heterogeneity across sites (Stamatakis, 2014). The reference genome (root) was used as an outgroup. The phylogenetic trees were then annotated using the Interactive Tree of Life 6.5.2 (Letunic & Bork, 2021).

The MRCA encompassing all of the isolates in each herd was identified and herds were classified according to the genetic relatedness of the isolates as follows (figure 13):

1. Monophyletic herd: the MRCA was only related to isolates from one herd.
2. Paraphyletic herd: the MRCA was related to isolates from one or more herds.
3. Polyphyletic herd: the MRCA was related to isolates from more than one clade or subclade as defined by vSNP.

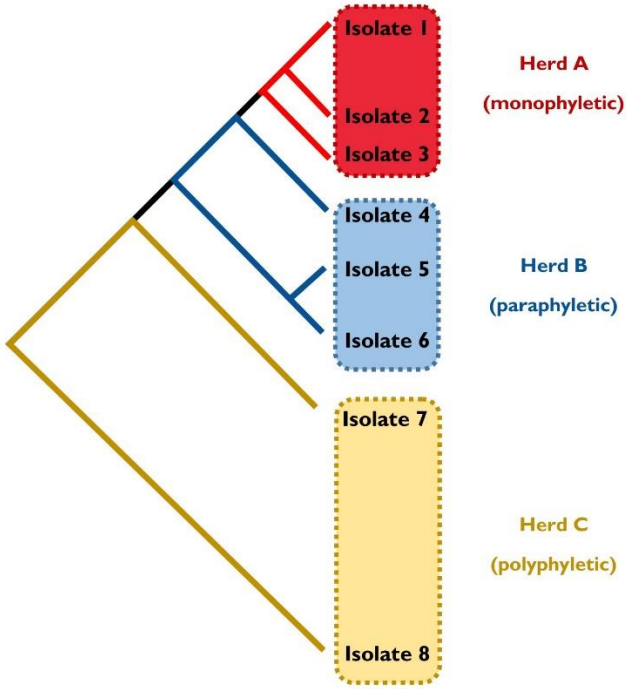


Figure 13. Schematic representation of the different systematic categories used to describe the gross phylogenetic relationships between isolates from the same and different herds.

The number of differences between *M. bovis* pairs of isolates was extracted using MEGA-X (Kumar *et al.*, 2018). The geographic distances between isolates were calculated using the distHaversine function from geosphere 1.5-10 package in R 4.1.2, based on centroid coordinates from the municipalities in which they were located (Hijmans *et al.*, 2021). Informed thresholds published in literature were used to estimate the possible epidemiological relationships between *M. bovis* isolates (Walker *et al.*, 2013; Crispell *et al.*, 2019b); a genetic distance between 0-6 SNPs was used to indicate highly similar strains with a close epidemiological relationship; a genetic distance between 7-12 SNPs was used to indicate similar strains with a distant epidemiological relationship; whereas a threshold > 12 SNPs was used to delineate epidemiologically unrelated isolates.

4.2.2.4. **Results**

An average of 1.88 x 10^6 reads was obtained per sample (standard deviation or s.d. = 6.58 x 10^5), with an average length of 225 bp (s.d. = 8.3) and 212 bp (s.d. = 8.1) for reads 1 and 2 (supplementary table 1), and an average quality per read of 36 and 34, respectively. The average coverage with respect to the reference genome was 99.6% (s.d. = 0.2%), whereas the average depth of coverage was 92 x (s.d. 30 x). Finally, a median of 700 SNPs (IQR = 683-723 SNPs) were identified with respect to the reference genome. Isolates were confirmed as SB0121 through SpoTyping, with the exception of samples 25 and 34 for which the presence of spacer 18 could not be confirmed using this method. However, spoligotyping was repeated for these two samples and they were confirmed as SB0121.

The analysed *M. bovis* sequences obtained from the 22 persistently infected herds were grouped into three genetic clades; clade 11 (40 isolates), clade 12 (26 isolates) and clade 13 (4 isolates) (figure 14). The median genetic distance in each of the clades was, 139 SNPs (IQR: 77-201), 211 SNPs (IQR: 29-244) and 126 SNPs (IQR: 11-146), respectively. These clades were, in turn, divided into four subclades (11.B.1, 11.B.2, 12.A and 12.D) and one undefined subclade (12-undefined) (figure 14).

The monophyletic group was composed of ten herds (3, 4, 8, 10, 11, 15, 18, 19, 20 and 22). In this group, *M. bovis* isolates from the same herd clustered together and shared a MRCA (table 6). In contrast, the paraphyletic group included isolates from six herds (1, 2, 5, 6, 9 and 21). In this group, at least one isolate from one herd shared a MRCA with isolates from other herds. Finally, the polyphyletic group of herds was composed of six herds (7, 12, 13, 14, 16 and 17).

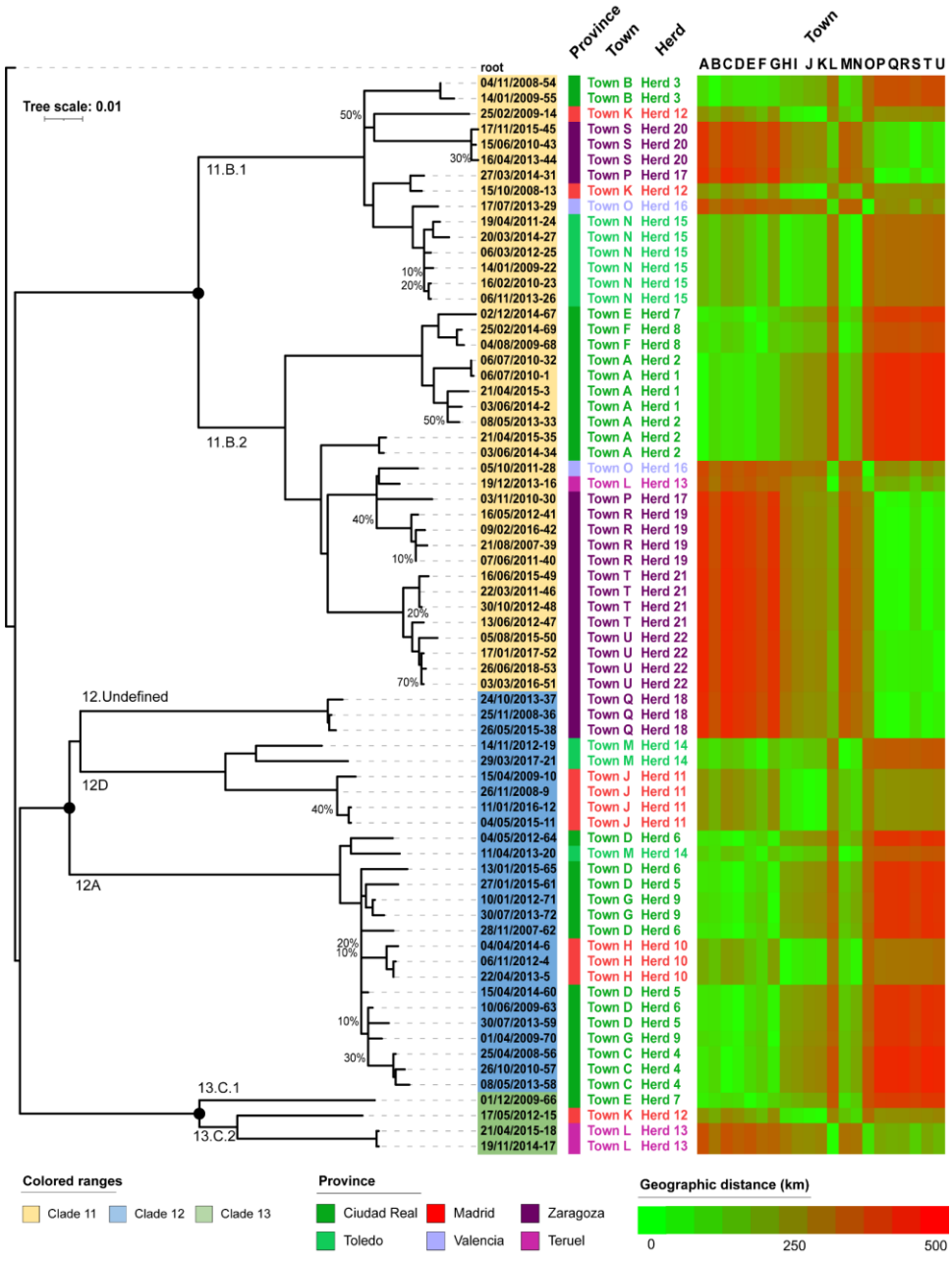


Figure 14. ML tree showing the phylogenetic relationships observed between 70 *M. bovis* isolates obtained from bTB chronically infected herds. The heatmap represents the geographic distance between the centroids of the municipalities in which the herds were located in km. Only bootstrap values below 90% are indicated. The black circles represent the MRCA of the detected clades with 100% bootstrap values.

Pro.	M-H	Diversity (median SNPs [range min - max])	Sampling time difference (median years [IQR]	Pairs of isolates within threshold (%)		
				≤ 6 SNPs	7-12 SNPs	> 12 SNPs
CRL	A-1	29 [15-32]	3.9 [2.4-4.35]	0 (0)	0 (0)	3 (100)
	A-2	116 [5-121]	2.4 [1.3-3.6]	1 (17)	0 (0)	5 (83)
	B-3	5	0.3	1 (100)	0 (0)	0 (0)
	C-4	8 [6-10]	2.5 [2.5-3.8]	1 (33)	2 (67)	0 (0)
	D-5	18 [15-27]	0.8 [0.75-1.5]	0 (0)	0 (0)	3 (100)
	D-6	36 [19-52]	3.7 [2.7-5.3]	0 (0)	0 (0)	6 (100)
	E-7	331	5.0	0 (0)	0 (0)	1 (100)
	F-8	5	4.6	1 (100)	0 (0)	0 (0)
	G-9	15 [6-19]	2.8 [2.2-3.6]	1 (33)	0 (0)	2 (67)
MAD	H-10	8 [1-9]	1 [0.7-1.2]	1 (33)	2 (67)	0 (0)
	J-11	7 [2-14]	6.2 [2.0-6.7]	3 (50)	1 (17)	2 (33)
	K-12	310 [70-330]	3.2 [1.8-3.4]	0 (0)	0 (0)	3 (100)
TER	L-13	310 [2-310]	0.9 [0.7-1.1]	1 (33)	0 (0)	2 (67%)
TOL	M-14	241 [67-252]	4.0 [2.2-4.2]	0 (0)	0 (0)	3 (100)
	N-15	10 [2-15]	2.3 [1.4-3.4]	3 (20)	8 (53)	4 (27)
VAL	O-16	194	1.8	0 (0)	0 (0)	1 (100)
ZAR	P-17	194	3.3	0 (0)	0 (0)	1 (100)
	Q-18	8 [4-10]	4.9 [3.3-5.7]	1 (33)	2 (67)	0 (0)
	R-19	7 [4-10]	4.2 [3.8-4.7]	3 (50)	3 (50)	0 (0)
	S-20	5 [5-6]	2.8 [2.7-4.1]	3 (100)	0 (0)	0 (0)
	T-21	11 [1-20]	2.1 [1.3-2.9]	3 (50)	3 (50)	0 (0)
	U-22	7 [2-11]	1.4 [1.0-2.1]	3 (50)	3 (50)	0 (0)

Table 6. Genetic diversity of isolates observed in the 22 cattle herds included in this study. Pro. = Province; M-H = Municipality-Herd; CRL= Ciudad Real; MAD = Madrid; TER = Teruel; TOL = Toledo; VAL = Valencia; ZAR = Zaragoza. Colours indicate the herd clustering category: white = monophyletic herds; light grey = paraphyletic herds; and dark grey = polyphyletic herds.

The genetic diversity was lowest in the monophyletic group, with the median pairwise genetic distance ranging between 5-10 SNPs per herd (table 6, supplementary figure 1). The median time difference between isolations in this group of herds was 2.65 years (IQR: 1.6-4.5 years), with a minimum of 0.3 years (herd 3) and a maximum of 6.2 years (herd 11). All herds had 1-3 pairs of highly similar isolates (≤ 6 SNPs), corresponding to 20-100% of their total number of pairs of isolates, whereas seven herds also had pairs of isolates with 7-12 SNP differences (17-67%) (table 6). Only two herds from this group (11 and 15) had pairs of isolates with more than 12 SNP differences (14 and 15 SNPs, respectively). In contrast, the paraphyletic group of herds presented a higher genetic diversity, with median pairwise distances ranging between 11-116 SNPs (table 6). The median time distance between isolations in these herds was 2.6 years (IQR: 2.2-3.5 years), with a minimum of 0.8 (herd 5) and a maximum of 3.9 years (herd 1). All herds, with the exception of herd 21, had pairs of isolates with a genetic distance higher than 12 SNPs, corresponding to 0-100% of the total number of pairs in each herd. Herds 2 and 9 had a single pair (17 and 33%, respectively) of highly similar isolates, whereas herd 21 had 3 pairs of highly similar and 3 pairs of distantly related isolates (50%), respectively. Finally, the genetic distance of the isolates in the polyphyletic herds was significantly higher than the distance of the isolates in the rest of the groups, with the median SNP difference ranging between 194-331 SNPs (table 6). The median time distance between the first and last isolations in this group was 3.3 years (IQR: 2.1-3.8 years), with a minimum of 0.9 years (herd 13) and a maximum of five years (herd 7). Only one herd (13) had a pair of highly similar isolates, corresponding to 33% of its total number of pairs of isolates, whereas the remaining herds had pairs of isolates with more than 12 SNP differences.

Nine herds had *M. bovis* isolates with 1-12 SNP differences with respect to isolates from another herd (table 7). The isolates were obtained 0-6 years apart and were detected within 43 km of each other, with the exception of a pair of isolates from herds 12 and 17, for which the distance was 235 km. Two herds (1 and 2) had a pair of isolates, obtained during the same year and in the same town, with one SNP difference.

M-H	Geographic distance (Km)	Pairs of isolates	Pairwise SNP distance	Maximum time distance (Years)
A-1 *vs.* A-2	0	1	1	0
		1	11	1.1
D-5 *vs.* D-6	0	1	8	4.8
		1	11	4.1
D-5 *vs.* G-9	24	1	9	2.3
		1	12	5.0
D-6 *vs.* G-9	24	1	8	4.8
		1	11	2.6
K-12 *vs.* P-17	235	1	11	5.4
T-21 *vs.* U-22	43	2	10	4.6
		1	11	6.0

Table 7. Herds with *M. bovis* isolates in which the pairwise genetic distance was lower than or equal to 12 SNP differences with respect to isolates from another herd. M-H = Municipality-Herd.

Epidemiological surveys were available for 17 out of the 22 herds, and the majority of them reported at least one risk factor for bTB persistence (supplementary table 2). The most common risk factor was extensive farming, which was reported in 16 out of the 17 herds with available surveys, followed by the presence of wildlife in the surroundings (n = 13 herds), either in pastures or in a nearby hunting estate. The presence of a fence surrounding the herd was reported in the majority of cases, with the exception of farm 8. Five of the herds had bullfighting animals and one herd reported the presence of shared pastures with other herds in the same region but that were not included in this study.

The wildlife isolates included in the analysis were also distributed in the same three genetic clades as the cattle isolates (figure 15); 11, 18 and four isolates corresponded to clades 11, 12 and 13, respectively. The median genetic distance between wildlife isolates in each of the clades was 194 SNPs (IQR: 98.5-203), 33 SNPs (IQR: 26.75-42) and 143.5 (IQR: 101.5-148.75) for clades 11, 12 and 13, respectively. In contrast, the genetic distance between wildlife and

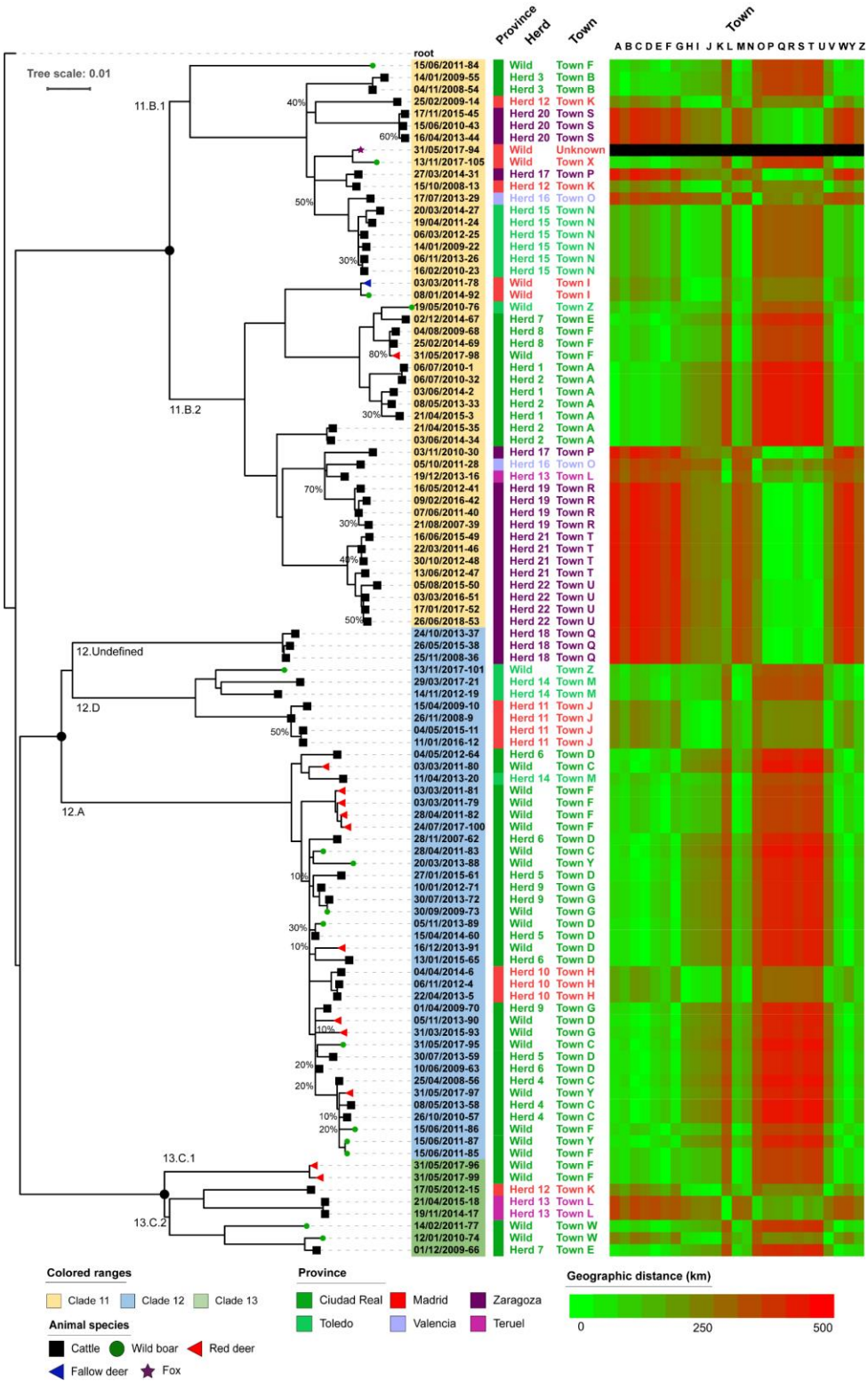

Province
Herd
Town
Town
A B C D E F G H I J K L M N O P Q R S T U V W Y Z
root
Tree scale: 0.01
11.B.1
11.B.2
12.Undefined
12.D
12.A
13.C.1
13.C.2
40%
60%
50%
30%
80%
30%
70%
30%
40%
50%
50%
10%
30%
10%
10%
20%
20%
10%
20%
15/06/2011-84 Wild Town F
14/01/2009-55 Herd 3 Town B
04/11/2008-54 Herd 3 Town B
25/02/2009-14 Herd 12 Town K
17/11/2015-45 Herd 20 Town S
15/06/2010-43 Herd 20 Town S
16/04/2013-44 Herd 20 Town S
31/05/2017-94 Wild Unknown
13/11/2017-105 Wild Town X
27/03/2014-31 Herd 17 Town P
15/10/2008-13 Herd 12 Town K
17/07/2013-29 Herd 16 Town O
20/03/2014-27 Herd 15 Town N
19/04/2011-24 Herd 15 Town N
06/03/2012-25 Herd 15 Town N
14/01/2009-22 Herd 15 Town N
06/11/2013-26 Herd 15 Town N
16/02/2010-23 Herd 15 Town N
03/03/2011-78 Wild Town I
08/01/2014-92 Wild Town I
19/05/2010-76 Wild Town Z
02/12/2014-67 Herd 7 Town E
04/08/2009-68 Herd 8 Town F
25/02/2014-69 Herd 8 Town F
31/05/2017-98 Wild Town F
06/07/2010-1 Herd 1 Town A
06/07/2010-32 Herd 2 Town A
03/06/2014-2 Herd 1 Town A
08/05/2013-33 Herd 2 Town A
21/04/2015-3 Herd 1 Town A
21/04/2015-35 Herd 2 Town A
03/06/2014-34 Herd 2 Town A
03/11/2010-30 Herd 17 Town P
05/10/2011-28 Herd 16 Town O
19/12/2013-16 Herd 13 Town L
16/05/2012-41 Herd 19 Town R
09/02/2016-42 Herd 19 Town R
07/06/2011-40 Herd 19 Town R
21/08/2007-39 Herd 19 Town R
16/06/2015-49 Herd 21 Town T
22/03/2011-46 Herd 21 Town T
30/10/2012-48 Herd 21 Town T
13/06/2012-47 Herd 21 Town T
05/08/2015-50 Herd 22 Town U
03/03/2016-51 Herd 22 Town U
17/01/2017-52 Herd 22 Town U
26/06/2018-53 Herd 22 Town U
24/10/2013-37 Herd 18 Town Q
26/05/2015-38 Herd 18 Town Q
25/11/2008-36 Herd 18 Town Q
13/11/2017-101 Wild Town Z
29/03/2017-21 Herd 14 Town M
14/11/2012-19 Herd 14 Town M
15/04/2009-10 Herd 11 Town J
26/11/2008-9 Herd 11 Town J
04/05/2015-11 Herd 11 Town J
11/01/2016-12 Herd 11 Town J
04/05/2012-64 Herd 6 Town D
03/03/2011-80 Wild Town C
11/04/2013-20 Herd 14 Town M
03/03/2011-81 Wild Town F
03/03/2011-79 Wild Town F
28/04/2011-82 Wild Town F
24/07/2017-100 Wild Town F
28/11/2007-62 Herd 6 Town D
28/04/2011-83 Wild Town C
20/03/2013-88 Wild Town Y
27/01/2015-61 Herd 5 Town D
10/01/2012-71 Herd 9 Town G
30/07/2013-72 Herd 9 Town G
30/09/2009-73 Wild Town G
05/11/2013-89 Wild Town D
15/04/2014-60 Herd 5 Town D
16/12/2013-91 Wild Town D
13/01/2015-65 Herd 6 Town D
04/04/2014-6 Herd 10 Town H
06/11/2012-4 Herd 10 Town H
22/04/2013-5 Herd 10 Town H
01/04/2009-70 Herd 9 Town G
05/11/2013-90 Wild Town D
31/03/2015-93 Wild Town G
31/05/2017-95 Wild Town C
30/07/2013-59 Herd 5 Town D
10/06/2009-63 Herd 6 Town D
25/04/2008-56 Herd 4 Town C
31/05/2017-97 Wild Town Y
08/05/2013-58 Herd 4 Town C
26/10/2010-57 Herd 4 Town C
15/06/2011-86 Wild Town F
15/06/2011-87 Wild Town Y
15/06/2011-85 Wild Town F
31/05/2017-96 Wild Town F
31/05/2017-99 Wild Town F
17/05/2012-15 Herd 12 Town K
21/04/2015-18 Herd 13 Town L
19/11/2014-17 Herd 13 Town L
14/02/2011-77 Wild Town W
12/01/2010-74 Wild Town W
01/12/2009-66 Herd 7 Town E
Colored ranges
Clade 11
Clade 12
Clade 13
Animal species
Cattle
Wild boar
Red deer
Fallow deer
Fox
Province
Ciudad Real
Madrid
Zaragoza
Toledo
Valencia
Teruel
Geographic distance (km)
0
250
500

Figure 15. ML tree representing the phylogenetic relationships observed between 99 *M. bovis* isolates obtained from chronically infected herds and wildlife species in the same or surrounding municipalities. The heatmap represents the geographic distance between the centroids of the municipalities in which the herds were located in km. Only bootstrap values below 90% are indicated. The black circles represent the MRCA of the detected clades with 100% bootstrap values.

cattle isolates in each group was 172 SNPs (IQR: 119-200), 40 SNPs (IQR: 29-243) and 146 (IQR: 143-150), respectively.

Eight wildlife isolates with a maximum of 12 SNP differences (of which five were highly similar) with respect to nine cattle isolates from five herds (4, 5, 6, 8 and 9) were identified (table 8). All of the above wildlife isolates were obtained from Ciudad Real and were geographically close to the cattle herds (< 30 km), with the exception of two wild boar isolates from town F (> 110 km). The median time distance between cattle and wildlife samples was 3.1 years (IQR: 1.9-3.9) (data not shown), whereas the minimum and maximum time distances were 0.6 and 9.1 years, respectively (table 8).

M-H	Town-Species	WL isolates	Geographic distance (Km)	SNP distance* (min – max)	Time distance (min-max)
G-9	G-WB	2	0	1 – 5	2.3 – 3.8
D-5	D-WB	1	0	4	1.1
D-6	D-WB	1	0	12	3.8
F-8	F-RD	2	0	5 – 6	3.3 – 7.8
D-5	G-WB	1	24.1	12	4.5
C-4	Y-WB	1	27.2	6 - 11	3.1 – 9.1
C-4	Y-RD	1	27.2	8 – 9	0.6 – 6.6
D-5	C-WB	1	27.6	10	3
D-6	C-WB	1	27.6	12	1.9
C-4	F-WB	2	117.7	6 – 12	0.6 – 3.1

Table 8. Pairwise genetic distances between *M. bovis* isolates obtained from cattle and wildlife species with less than 12 SNP differences in the province of Ciudad Real. M-H = Municipality-Herd; WL = Wild life.; WB = Wild boar; RD = Red deer. * SNP distance with respect to cattle isolates from the indicated herd.

Finally, 12 of the wildlife isolates differed by a maximum of 12 SNPs from other wildlife isolates, of which six were highly similar (data not shown). These isolates were located in the province of Ciudad Real, with the exception of two wildlife isolates obtained in Madrid (wild boar and fallow deer; seven SNPs). The majority of isolates were located within the same municipality, with the exception of four isolates that were located in the same region (Ciudad Real), but separated by 97 km.

4.2.2.5. Discussion

Distribution of* M. bovis *isolates into three genetic clades reveals a very high genetic diversity of SB0121 in Spain

This study evaluated the use of WGS to assess the within-herd genetic diversity of *M. bovis* SB0121 isolates in chronically infected herds in order to identify possible causes of persistence. The phylogenetic analysis of the sequenced cattle isolates revealed that they belonged in three genetic clades (figure 13), with the majority of isolates clustering within clades 11 or 12 (41 and 26 isolates, respectively). Isolates from all three clades were detected across the majority of the studied regions, reflecting the wide distribution of genetically diverse SB0121 isolates (Rodriguez *et al.*, 2010). In addition, the genetic diversity of isolates within these clades was very high, ranging between 126-211 median SNPs, which agrees with previous findings for this spoligotype using MIRU-VNTR in Spain and Portugal, in which 65 and 44 genotypes were identified, respectively (Duarte *et al.*, 2010; Rodriguez-Campos *et al.*, 2013). Similarly, a recent WGS study revealed a high genetic diversity of the SB0121 spoligotype in Catalonia (northeastern Spain), in which this genotype was distributed across four different genetic clades (Perea *et al.*, 2021), which further confirms the high genetic diversity of this spoligotype in Spain. In contrast, results from a recent study that analysed 139 *M. bovis* isolates from the third most prevalent spoligotype in Spain (SB0339) revealed a significantly lower genetic diversity for this genotype, as all isolates differed by 49-88 SNPs from their MRCA (Pozo, 2020). Considering that SB0339 misses four spacers (spacers 26-29) that are present in SB0121 and could, therefore, have evolved from SB0121 (Rodriguez *et al.*, 2010), a possible explanation for the reduced genetic diversity of SB0339 in comparison to the former spoligotype could be that this genotype may have been circulating for a shorter period of time.

Phylogenetic clustering and SNP distances reveal a complex epidemiological scenario in which genetic diversity varies within chronic infected herds

In order to visualise the possible epidemiological relationships between herds, these were divided into three different categories according to their gross phylogenetic relationships.

The monophyletic group of herds comprised 10 herds in which all *M. bovis* isolates from each herd shared a MRCA and clustered in a monophyletic fashion. The median genetic distance of the isolates in each of these herds was low, ranging between 5-10 SNPs (table 6). Three herds had isolates with less than six SNP differences with isolates from the same herd (herds 3, 8 and 20), which could indicate that these breakdowns were caused by closely related strains that were in circulation in the herds or their surroundings. In the case of herd 3, the isolations were separated by only three months, corresponding to two consecutive bTB tests, which could indicate that infected animals could have remained undetected between tests. In contrast, the isolation data in herds 8 and 20 indicated that highly similar *M. bovis* strains (five SNPs) were present at longer time periods of 2.5-5.4 years. These findings, together with the limited substitution rate of *M. bovis*, which has been estimated to be an average of 0.31 (range: 0.1-0.94) base pairs x genome x year (Biek *et al.*, 2012; Crispell *et al.*, 2017; Rossi *et al.*, 2021), could suggest the persistence and microevolution of closely-related isolates within these herds or their surroundings.

A higher within-herd genetic diversity was detected in the remaining herds of this group. All herds had at least one pair of highly similar *M. bovis* isolates and a varying proportion of distantly related (50-67%) and unrelated (0-33%) *M. bovis* pairs of isolates. In addition to these isolates, other spoligotypes were detected in all herds during the evaluated period, with the exception of herds 19 and 22. These findings revealed a complex epidemiological situation in these herds, in which bTB persistence was likely to have been caused by multiple factors, both endogenous and exogenous. The fact that highly similar isolates were detected both within the same year and at longer intervals in some of these herds (herds 11, 19 and 22) could suggest a persisting infection with closely-related *M. bovis* strains, as well as a relapse or the re-introduction of the infection from a common source. Similar findings were observed in a different study for *M. bovis* SB0339 in two herds from Madrid, in which little genetic variation was observed despite the long time elapsed between isolations (Pozo, 2020). Furthermore, the detection of three almost identical isolates, with one or two SNP differences, in

three herds (10, 11 and 22) within five to 11 months could suggest a recent within-herd transmission event.

In contrast, the presence of more distantly related or unrelated isolates in all of the herds could indicate the presence of different re-introductions of the infection from external sources, especially for those isolates that were obtained at longer time intervals. In herd 18, the detection of an isolate with 8-9 SNP differences in 2013 (isolate 37) with respect to two isolates obtained in 2008 and 2015 could indicate the presence of an external introduction or a long-time recurrence of infection (figure 16). The epidemiological survey for this herd reported the replacement of animals from positive mothers, which is considered a risk factor for the persistence of infection (Guta *et al.*, 2014), and could also have contributed to the maintenance and microevolution of *M. bovis* strains in the herd (isolates 36 and 38).

Interestingly, *M. bovis* isolates with 8-14 SNP differences were obtained in four herds (10, 11, 15 and 22) within one year, indicating that isolates with a high genetic diversity were present in the herd during consecutive tests. This contrasts with another study that evaluated within-herd genetic diversity in cattle and deer herds in the USA, in which *M. bovis* isolates obtained within a single year did not differ by more than 6 SNPs (Orloski *et al.*, 2018). This difference in the within-herd diversity could be related to the fact that bTB is eradicated in the USA and that most bTB outbreaks are caused by imported cattle, with limited between-herd transmission (Orloski *et al.*, 2018). Although wildlife reservoirs have been described in the USA, their involvement in bTB maintenance is usually restricted to certain regions, such as Michigan or Minnesota (Glaser *et al.*, 2016; Salvador *et al.*, 2019). In contrast, the long presence of bTB in Spain, coupled with its complex epidemiology and higher prevalence of infection could be possible reasons for the occurrence of simultaneous infections with multiple genotypes. For example, the epidemiological survey from herd 22 revealed that this herd was created in April 2012 with animals from different bullfighting herds. Three animals of that herd reacted positive in June of the same year. The samples included in this study corresponded to a fraction of the animals that tested positive almost three years later. The detection of two different SNP

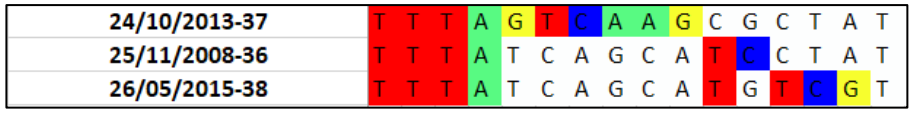


Figure 16. SNP patterns observed for *M. bovis* isolates obtained from herd 18.

patterns in herd 22 in a reduced amount of time, with 2-8 SNP differences with respect to the MRCA, strongly indicates the presence of a diverse pool of *M. bovis* isolates in the herd at the time of the breakdown. This would be in agreement with the mixed origin of the animals. The fact that the animals were located in a closed facility limited the possibility of interaction of the herd with exogenous factors, such as wildlife or neighbouring herds. Although animal movements were reported at later sampling dates, the fact that the sequences were highly similar (≤ 3 SNPs) to one of the previous strains suggested that persistence was probably caused by a relapse from the original outbreak. Another herd (11) showed two clearly distinct SNP patterns during the evaluated period (figure 17). The first pattern was detected between 2008-2009 and involved two isolates that differed by eight SNPs, whereas the remaining isolates were isolated in 2015-2016 and differed by 6-14 SNPs from the previous isolates and by a unique SNP between them (figure 17). The epidemiological information of this herd indicated that it shared pastures with three other non-OTF herds. Therefore, in this case, bTB persistence could be attributed to shared circulation of genetically related *M. bovis* isolates between different herds, as well as a relapse from previous infections in the herd. This speculation could be in agreement with the detection of two isolates with six SNP differences with respect to the first isolate, more than six years later (figure 17).

The presence of *M. bovis* isolates that descended from the same MRCA but were obtained from different herds indicated a possible epidemiological link between the involved herds (paraphyletic herds). The genetic diversity of the isolates in these herds was higher than in the monophyletic group, with median pairwise distances of 11-116 SNPs (table 6). This finding strongly suggests the presence of multiple introductions during the studied period and/or the existence of a diverse pool of *M. bovis* strains in the surroundings. Despite the high genetic diversity of the isolates, all of the herds in this group had *M. bovis* isolates that were distantly related (8-12 SNP differences) with isolates from another herd, with the exception of herds 1 and 2, which had a pair of highly similar isolates (1 SNP difference) (table 7). A direct epidemiological link was identified between the latter herds, since they were located in the same

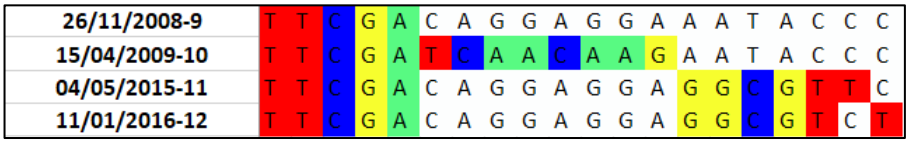


Figure 17. SNP patterns observed for *M. bovis* isolates obtained from herd 11.

municipality and owned by the same farmer. Furthermore, epidemiological data indicated the movement of animals between these herds two years before the first isolations were obtained. In the other cases, the affected herds were located within 43 km of each other, which suggests that related *M. bovis* strains were in circulation in the surroundings. This agrees with previous results in Northern Ireland in which subsequent outbreaks in the same farms tended to involve genetic lineages that were also detected in the surroundings (Biek *et al.*, 2012). Among the possible causes for this distribution is interspecies transmission between cattle and wildlife, since TB prevalence is high both in cattle and wildlife in the region in which these herds are located (Ciudad Real; see section below). Finally, distantly related isolates were identified between two herds in Zaragoza separated by 43 km (herd 21 and 22) (table 7). As a low TB prevalence in wildlife is reported in that region (MAPAMA & AECOSAN, 2017), the detection of related *M. bovis* strains in these herds could be attributed to animal movements or contact with neighbouring herds.

Finally, six herds (herds 7, 12, 13, 14, 16 and 17) had isolates that were polyphyletic, in which the MRCA of the isolates in a given herd included isolates from different genetic clades or subclades. The within-herd diversity was highest in this group of strains (table 6), suggesting multiple introductions as the cause of bTB persistence. These findings could indicate the introduction of infected animals from distant regions or the presence of a high genetic diversity of *M. bovis* strains in the surroundings of the herd. Interestingly, a pair of isolates from herds 12 and 17, located more than 200 km away from each other, had 11 SNP differences (5-6 SNP differences with respect to the MRCA), suggesting a possible epidemiological link (table 7), such as animal movements between herds.

Genetic data from cattle and wildlife isolates reveals the circulation of M. bovis between domestic and wildlife species in the province of Ciudad Real

The majority of herds in this study with available epidemiological surveys reported the presence of wildlife animals in the surroundings (70.5%; n = 12 out of 17), either through shared pastures or nearby hunting estates. In order to evaluate the role of sympatric wildlife in the persistence of bTB in cattle, available tissue samples from wildlife species infected with *M. bovis* SB0121 in the same and surrounding municipalities were analysed. Due to the wide dispersion capacity of some of the most common wildlife reservoirs of TB (e.g. wild boar and red deer), samples from more distant municipalities were also analysed when available.

The obtained wildlife sequences were dispersed across the phylogenetic tree, reflecting a high diversity of *M. bovis* SB0121 in wildlife species (figure 13). The majority of wildlife isolates were obtained from Ciudad Real, with wildlife isolates from municipalities in this region having between 4-88 median SNP differences (Supplementary figure 1). This reflects an overall high genetic diversity of *M. bovis* SB0121 in wild animals across the region, especially in town F in which all of the genetic clades were detected (figure 15, supplementary figure 2).

Wildlife isolates from the province of Ciudad Real generally clustered together with cattle isolates from different herds in the region, suggesting the existence of a shared pool of *M. bovis* strains with possible interspecies transmission. The pairwise comparison of *M. bovis* isolates between species of animals revealed the presence of both highly similar and distantly related *M. bovis* isolates in wild boar and red deer located within the same municipality, as well as within 30 km of four infected herds (4, 5, 6 and 9) (table 8, figure 15). Furthermore, highly similar isolates were detected in wildlife located up to 120 km from infected cattle (herd 4), reflecting the distribution of closely-related strains through long distances, either by domestic or wildlife animal movements. Three wild boar isolates were obtained approximately seven months before the detection of a cattle isolate with 8-12 SNP differences in herd 4, which indicates the presence of related *M. bovis* isolates in the surroundings of the herd at the time of the infection. In addition, the fact that related *M. bovis* isolates were detected in wildlife more than a year before and/or after their detection in cattle suggests the presence of a common source of infection both for domestic animals and wildlife that could contribute to multiple interspecies transmission events. Furthermore, the detection of highly similar *M. bovis* isolates between different

animals from the same (e.g. deer isolates 82 and 100) or different wildlife species (wild boar and deer isolates 86 and 97, respectively) could indicate the transmission of TB in sympatric wildlife. These findings indicate a wide circulation of *M. bovis* SB0121 in wildlife in the region, as well as within-species transmission events that could contribute to the maintenance of TB in domestic and/or sylvatic cycles. Certainly, additional studies are needed in order to assess the possible directionality of transmission between sympatric wildlife and cattle in Spain, as has been achieved in other countries using WGS and different statistical methods (Crispell *et al.*, 2017; Crispell *et al.*, 2019b; Salvador *et al.*, 2019; O'Hare *et al.*, 2021; Rossi *et al.*, 2021; van Tonder *et al.*, 2021; Duault *et al.*, 2022).

Finally, the number of wildlife samples available in this study did not permit further investigations on putative epidemiological links. This was a result of the limited amount of wildlife samples that were available for each other region included in the study and highlights the importance of achieving an appropriate sampling depth of wildlife in regions in which TB is prevalent in these animals.

Current challenges and next steps

Despite the promising use of WGS in the study of bTB persistence, several limitations may have affected the results in this study. Firstly, the restricted sampling strategy, which only included a single *M. bovis* isolate per sampling date and farm, may have resulted in an underrepresentation of the actual within-herd diversity by excluding other genotypes present in the herd (Orloski *et al.*, 2018). This, in turn, could have led to both an overestimation and underestimation of the within-herd pairwise distances in certain cases, especially when a low sampling size was obtained (e.g. herds 3 or herd 8).

Due to the high resolution of WGS, in which isolates may differ by hundreds of SNPs, the support of epidemiological links between isolates also demands an adequate representation of *M. bovis* isolates from different regions. The fact that no cattle samples were selected from neighbouring herds and municipalities limited the capacity of this study to establish the genetic diversity of local *M. bovis* SB0121 strains, which could have added further epidemiological context to the observed genetic diversity in the herds under analysis. The detection of a possible between-herd transmission event between two herds from the same town and owner (herds 1 and 2), as well as the reporting of a likely transmission event due to shared pastures (herd 11), reflect the importance of assessing the genetic diversity of *M. bovis* isolates from surrounding herds in

order to detect possible transmission events (Price-Carter *et al.*, 2018). Future studies evaluating the genetic diversity of *M. bovis* SB0121 in the studied municipalities as well as in other regions will help to generate a prevalence map which would greatly facilitate the establishment of more detailed transmission dynamics in these areas.

In addition, the low number of *M. bovis* SB0121 isolates from wildlife only allowed the establishment of putative interspecies transmission events in Ciudad Real, in which animal TB has amply been studied. The limited availability of wildlife isolates, which is often mentioned in literature (Price-Carter *et al.*, 2018), constitutes an important obstacle in epidemiological investigations of bTB, since it is often limited to certain regions or obtained through specific practices, such as hunter-harvested animals (Glaser *et al.*, 2016; Crispell *et al.*, 2020). The collection of wildlife isolates and their incorporation into bTB outbreak investigations and active surveillance studies would increase the representation of *M. bovis* genomic sequences from diverse environments and species and would contribute to the identification of epidemiological links between cattle and wildlife (MAPAMA & AECOSAN, 2017).

Similar to other characterisation techniques, the utility of WGS largely depends on the availability of adequate epidemiological data with which to establish possible epidemiological links between isolates. Despite the limited availability of epidemiological data in this study, its availability greatly enhanced the performance of WGS in certain cases (e.g. herd 22), being able to suggest a possible explanation for the observed genetic diversity within a herd or a possible epidemiological link between genetically similar isolates (e.g. herds 1 and 2). Future studies will benefit from including more comprehensive epidemiological data, such as the complete animal movement history of each of the herds, or the herd's bTB status history or the bTB prevalence at the town or regional level.

Conclusions

The results of this study demonstrate the potential of WGS in bTB breakdown investigations, since its high resolution can be used to uncover genetic differences below the resolution levels of conventional characterisation tools. With this additional level of information and adequate epidemiological data, authorities tasked with bTB control could rapidly assess endogenous or exogenous origin of the breakdown and focus their investigation towards the more likely cause.

4.2.2.6. **Funding and acknowledgements**

This study was financed by the Spanish Ministry of Agriculture, Fisheries and Feed (Ministerio de Agricultura, Pesca y Alimentación). Víctor Lorente-Leal was funded through a PhD grant from Banco Santander & Universidad Complutense de Madrid (CT17/18).

This study was carried out in collaboration with the veterinary authorities of the different Autonomous Regions involved; Subdirección General de Producción Agroalimentaria, Dirección General de Agricultura, Ganadería y Alimentación (Comunidad de Madrid); Sección de Coordinación de Programas de Sanidad Animal, Departamento de Agricultura, Ganadería y Medio Ambiente (Gobierno de Aragón); Dirección General de Agricultura y Ganadería, Consejería de Agricultura Agua y Desarrollo Rural (Junta de Castilla-La Mancha); Servicio de Producción y Sanidad Animal, Consejería de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica (Generalitat Valenciana); as well as the collaboration with the National Reference Laboratory for bTB in the Central Laboratory for Animal Health of Santa Fe, Granada.

4.2.2.8 Supplementary material

Supplementary table 1. Metadata and sequencing statistics for the 99 *M. bovis* SB0121 isolates that were sequenced in this study.

Sample metadata					R1 reads			R2 reads			Mapping statistics				Phylogenetic data	
Sample	Pro.	Mun.	Herd	Species	Av. Length	Av. Quality	Q30	Av. Length	Av. Quality	Q30	Total Reads	Mapped	Cov.	Depth	SNPs	Clade
06/07/2010-1	CRL	A	1	BV	232.6	36.4	96.70%	224.1	35.4	93.20%	2974448	2,976,456	99.50%	153.6	730	11
03/06/2014-2	CRL	A	1	BV	224.3	36.3	96.30%	217.2	35.5	93.50%	1897394	1,895,971	99.42%	95.3	715	11
21/04/2015-3	CRL	A	1	BV	199.9	36.3	96.10%	193.5	35.5	91.90%	3660564	3,621,096	99.55%	163	731	11
06/11/2012-4	MAD	H	10	BV	226.2	35.9	95.10%	210.1	34.2	87.20%	1777680	1,779,645	99.67%	88.2	694	12
22/04/2013-5	MAD	H	10	BV	232.2	36.4	97.00%	220.1	35.1	91.30%	1490734	1,499,241	99.57%	76.5	691	12
04/04/2014-6	MAD	H	10	BV	190.2	35.5	94.10%	180	34.6	88.10%	1211262	1,212,143	99.60%	51.3	682	12
26/11/2008-9	MAD	J	12	BV	214.7	36.1	95.90%	203.6	35	91.20%	2488640	2,483,072	99.64%	118	664	12
15/04/2009-10	MAD	J	12	BV	226.2	36.1	95.70%	211.1	34.4	88.20%	2044630	2,041,757	99.61%	101.6	667	12
04/05/2015-11	MAD	J	12	BV	229.3	36.3	96.90%	218.6	35.1	91.50%	1649812	1,650,266	99.54%	84.1	668	12
11/01/2016-12	MAD	J	12	BV	223	36.4	96.60%	217	35.7	94.80%	3415866	3,421,694	99.71%	168.6	678	12
15/10/2008-13	MAD	K	13	BV	216.5	35.8	95.40%	201.1	33.7	84.00%	2448710	2,431,953	99.72%	115.8	708	11
25/02/2009-14	MAD	K	13	BV	223.5	36.2	96.20%	218.5	35.6	94.10%	1846008	1,845,536	99.63%	92.9	728	11
17/05/2012-15	MAD	K	13	BV	225.1	36.2	96.40%	217.3	35.2	92.10%	1727880	1,728,058	99.14%	87	645	13
19/12/2013-16	TER	L	14	BV	214.4	35.7	94.60%	201.8	34.2	87.40%	2086756	2,076,857	99.73%	98.6	734	11
19/11/2014-17	TER	L	14	BV	229.2	35.5	94.00%	217.6	34.3	88.30%	1644468	1,635,812	99.06%	83	660	13
21/04/2015-18	TER	L	14	BV	224.3	36.1	96.00%	215	35.1	90.50%	2713198	2,700,263	99.23%	134.9	675	13
14/11/2012-19	TOL	M	15	BV	215.4	35.2	92.70%	200.4	33.5	84.70%	1545776	1,542,045	99.55%	73.3	651	12

11/04/2013-20	TOL	M	15	BV	218	35.7	94.70%	202.4	33.8	84.70%	1851024	1,844,590	99.66%	88.5	714	12
29/03/2017-21	TOL	M	15	BV	231.3	36.1	95.80%	223.2	35.2	92.80%	1629438	1,695,245	99.61%	82.5	676	12
14/01/2009-22	TOL	N	16	BV	223.7	36.3	96.70%	213.3	35	90.70%	1773222	1,770,374	99.68%	88.2	711	11
16/02/2010-23	TOL	N	16	BV	223.5	36.4	96.70%	217.1	35.7	94.50%	1905062	1,902,820	99.69%	95.2	707	11
19/04/2011-24	TOL	N	16	BV	223.5	36.3	96.60%	215.6	35.4	92.90%	1783216	1,781,726	99.68%	88.9	728	11
06/03/2012-25	TOL	N	16	BV	220.9	36.5	97.20%	213.6	35.6	93.60%	3266656	3,263,256	99.79%	159.6	710	11
06/11/2013-26	TOL	N	16	BV	221.2	36.2	96.20%	215.5	35.6	94.30%	1539310	1,540,288	99.64%	76.6	706	11
20/03/2014-27	TOL	N	16	BV	223.2	36.1	95.90%	208.4	34.5	88.30%	2411974	2,403,556	99.72%	117.8	725	11
05/10/2011-28	VAL	O	17	BV	230.2	36.1	96.00%	219.7	35	91.80%	1447582	1,456,434	99.68%	73.9	727	11
17/07/2013-29	VAL	O	17	BV	219.8	35.6	94.30%	209.8	34.4	89.10%	1937538	1,930,517	99.63%	94.5	709	11
03/11/2010-30	ZAR	P	18	BV	233.1	36.2	96.60%	220.3	34.8	91.40%	1702002	1,803,639	99.83%	85.4	767	11
27/03/2014-31	ZAR	P	18	BV	220.5	35.7	95.00%	208.9	34.5	89.50%	2516014	2,499,410	99.71%	122.1	725	11
06/07/2010-32	CRL	A	2	BV	218	34.3	88.80%	187.1	31.1	63.30%	807098	806,376	99.15%	37.4	684	11
08/05/2013-33	CRL	A	2	BV	213.4	35.5	93.60%	203.2	34.8	90.30%	1950804	1,944,093	99.47%	92.2	708	11
03/06/2014-34	CRL	A	2	BV	222.7	36.2	96.00%	212.5	34.9	89.60%	2534520	2,525,532	99.76%	125.2	695	11
21/04/2015-35	CRL	A	2	BV	219.2	35.6	94.50%	206.4	34.2	87.30%	2137586	2,129,415	99.74%	103.5	696	11
25/11/2008-36	ZAR	Q	19	BV	226.4	35.5	94.50%	209.1	33.7	84.20%	1811184	1,805,631	99.58%	90.1	677	12
24/10/2013-37	ZAR	Q	19	BV	220	36.6	97.60%	207.2	34.8	90.30%	3674872	3,654,392	99.74%	175.9	689	12
26/05/2015-38	ZAR	Q	19	BV	226.2	36.3	96.90%	211.7	34.6	88.50%	1713274	1,740,877	99.70%	84.6	675	12
21/08/2007-39	ZAR	R	20	BV	221	35.9	95.40%	210.7	35	91.90%	2122100	2,119,503	99.74%	103.9	747	11
07/06/2011-40	ZAR	R	20	BV	200.2	36.3	96.30%	193.1	35.2	90.60%	4679280	4,622,498	99.85%	207.1	771	11
16/05/2012-41	ZAR	R	20	BV	221.8	36.3	96.70%	213.6	35.2	91.90%	1699256	1,697,454	99.69%	84.4	725	11

09/02/2016-42	ZAR	R	20	BV	229.9	36.3	96.50%	217.6	35	91.50%	1482936	1,540,761	99.77%	74.1	757	11
15/06/2010-43	ZAR	S	21	BV	225.2	36.4	97.00%	215.5	35.2	91.80%	1564188	1,563,829	99.62%	78.5	731	11
16/04/2013-44	ZAR	S	21	BV	225.1	36.2	96.90%	205.7	33.8	85.20%	2483562	2,533,482	99.81%	119.1	754	11
17/11/2015-45	ZAR	S	21	BV	231	36.4	97.20%	221.7	35.3	93.30%	1752806	1,756,498	99.62%	90	734	11
22/03/2011-46	ZAR	T	22	BV	224.9	36.3	96.70%	209.8	34.6	88.90%	2456188	2,449,823	99.76%	121.2	747	11
13/06/2012-47	ZAR	T	22	BV	219.1	35.8	95.00%	203.7	34	85.80%	2015542	2,006,327	99.72%	96.8	733	11
30/10/2012-48	ZAR	T	22	BV	223.9	36	95.20%	215.3	35	90.50%	2470548	2,460,961	99.74%	123.5	761	11
16/06/2015-49	ZAR	T	22	BV	188.4	36.3	95.90%	184.5	35.9	94.10%	3437750	3,405,596	99.85%	144.4	757	11
05/08/2015-50	ZAR	U	23	BV	221.3	36.1	95.80%	214.9	35.4	92.70%	2648240	2,637,703	99.77%	131	765	11
03/03/2016-51	ZAR	U	23	BV	216.8	35.6	94.20%	205.1	34.2	87.70%	2470750	2,453,455	99.78%	118.3	750	11
17/01/2017-52	ZAR	U	23	BV	221.8	35.9	94.90%	209.2	34.5	88.80%	2355478	2,347,378	99.76%	115.5	744	11
26/06/2018-53	ZAR	U	23	BV	229.4	36.4	97.00%	214.5	34.5	88.50%	945710	984,775	99.72%	46.9	767	11
04/11/2008-54	CRL	B	3	BV	225.9	35.5	93.60%	216.6	34.5	88.80%	1632360	1,630,354	99.64%	82	715	11
14/01/2009-55	CRL	B	3	BV	221.9	36.3	96.80%	214.1	35.3	92.70%	1836500	1,836,478	99.67%	91	730	11
25/04/2008-56	CRL	C	4	BV	223.8	36	95.90%	210.1	34.6	88.30%	2056316	2,053,878	99.66%	101.2	698	12
26/10/2010-57	CRL	C	4	BV	225	36.3	96.80%	215.4	35.2	91.60%	1837746	1,837,545	99.66%	91.9	701	12
08/05/2013-58	CRL	C	4	BV	225	35.8	95.10%	216.2	35	91.80%	1741024	1,741,960	99.65%	87.1	705	12
30/07/2013-59	CRL	D	5	BV	217.6	35.5	93.90%	208.5	34.6	90.50%	1903054	1,898,741	99.66%	92.2	690	12
15/04/2014-60	CRL	D	5	BV	233	36.4	97.10%	216.3	34.5	88.20%	2612890	2,616,405	99.67%	132.7	691	12
27/01/2015-61	CRL	D	5	BV	226	35.9	95.10%	214.7	34.9	91.00%	2374636	2,369,264	99.68%	118.8	700	12
28/11/2007-62	CRL	D	6	BV	226.5	36.1	95.70%	214.1	34.9	90.10%	2182062	2,181,857	99.66%	109.1	698	12
10/06/2009-63	CRL	D	6	BV	226.4	36.3	97.00%	210.6	34.5	87.60%	2224736	2,218,541	99.67%	110.3	678	12

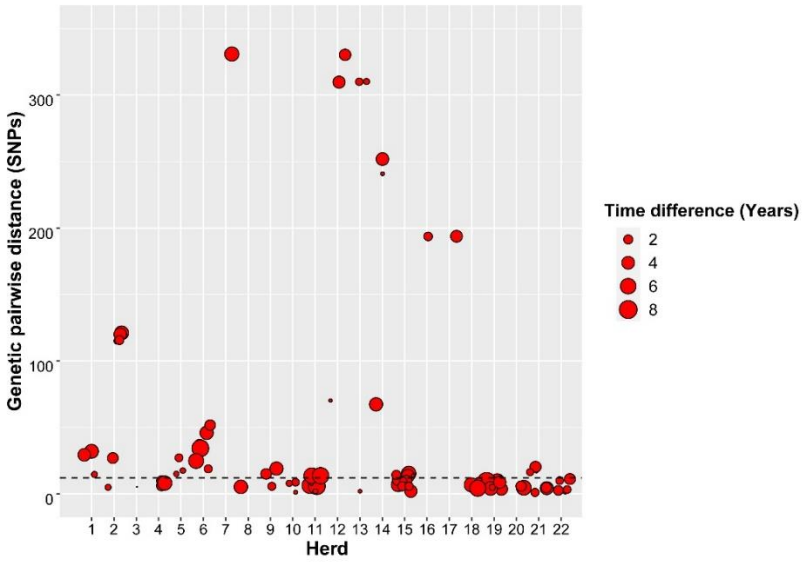
04/05/2012-64	CRL	D	6	BV	224.1	36.2	96.20%	214.4	35.2	92.10%	2231320	2,228,307	99.69%	110.7	699	12
13/01/2015-65	CRL	D	6	BV	225.1	36.2	96.70%	219.3	35.7	94.60%	2130434	2,128,660	99.65%	107.6	709	12
01/12/2009-66	CRL	E	7	BV	219.7	35.6	93.70%	207.4	34.1	86.90%	1934956	1,930,804	99.16%	93.8	654	13
02/12/2014-67	CRL	E	7	BV	217.6	35.6	94.40%	204.9	34.4	88.60%	2055124	2,046,269	99.44%	98.9	721	11
04/08/2009-68	CRL	F	8	BV	230.3	35.7	95.10%	208.2	33.1	79.80%	1211996	1,214,853	99.33%	60.6	701	11
25/02/2014-69	CRL	F	8	BV	223.3	35.6	95.10%	193.7	32	69.90%	1541566	1,536,121	99.37%	73.2	703	11
01/04/2009-70	CRL	G	9	BV	225.6	36.2	96.30%	216.6	35.2	92.00%	1292496	1,294,705	99.59%	65.2	671	12
10/01/2012-71	CRL	G	9	BV	219.2	35.7	94.80%	205.3	34.1	87.00%	1856044	1,849,321	99.63%	89.5	682	12
30/07/2013-72	CRL	G	9	BV	219.9	35.6	94.20%	205.4	34	86.10%	1722432	1,718,324	99.65%	83.5	680	12
30/09/2009-73	CRL	G	Wild	WB	231.8	36.1	96.30%	217.4	34.6	88.50%	1372788	1,379,426	99.52%	70.1	683	12
12/01/2010-74	CRL	W	Wild	WB	229.6	35.6	94.40%	217.9	34.5	89.90%	1251702	1,255,123	99.03%	63.8	642	13
19/05/2010-76	TOL	Z	Wild	WB	227.9	34.8	92.20%	206.1	32.8	80.40%	773538	777,082	99.21%	38.2	696	11
14/02/2011-77	CRL	W	Wild	WB	231	36.1	95.80%	217.4	34.6	89.10%	1395428	1,398,601	99.05%	71.2	660	13
03/03/2011-78	MAD	I	Wild	FD	229.9	35.8	95.30%	218.6	34.8	90.30%	1564040	1,568,114	99.61%	79.8	711	11
03/03/2011-79	CRL	F	Wild	RD	230.6	36.1	96.00%	217.3	34.6	88.50%	1540220	1,545,417	99.56%	78.4	678	12
03/03/2011-80	CRL	C	Wild	RD	229.8	35.9	95.40%	219.8	34.9	91.30%	1395184	1,400,385	99.54%	71.3	688	12
03/03/2011-81	CRL	F	Wild	RD	229.8	35.9	95.30%	218.3	34.6	89.70%	1687878	1,689,516	99.61%	86	699	12
28/04/2011-82	CRL	F	Wild	RD	231.7	36	95.80%	217.3	34.5	88.60%	1297198	1,303,868	99.57%	66.1	688	12
28/04/2011-83	CRL	C	Wild	WB	230.9	36.1	96.40%	220.8	35	91.10%	1605126	1,608,011	99.59%	82.5	687	12
15/06/2011-84	CRL	F	Wild	WB	231.8	35.9	95.50%	220	34.8	90.50%	1236408	1,240,527	99.57%	63.5	714	11
15/06/2011-85	CRL	F	Wild	WB	232.8	36.3	97.10%	217.3	34.6	90.00%	1352410	1,358,316	99.62%	69.1	701	12
15/06/2011-86	CRL	F	Wild	WB	231.4	36.3	97.00%	217.7	34.9	91.30%	2018720	2,023,479	99.63%	103.1	709	12

15/06/2011-87	CRL	Y	Wild	WB	232.7	36.3	96.90%	219.3	34.9	90.90%	1791850	1,795,552	99.62%	91.8	696	12
20/03/2013-88	CRL	Y	Wild	WB	230.7	36.1	96.30%	219.8	35	92.70%	1672208	1,676,035	99.60%	85.6	699	12
05/11/2013-89	CRL	D	Wild	WB	232.3	36.3	97.00%	219.6	35	91.90%	2033596	2,034,248	99.66%	104.3	692	12
05/11/2013-90	CRL	D	Wild	RD	232.7	36.3	97.20%	218.9	34.9	91.10%	1599102	1,602,165	99.57%	81.9	687	12
16/12/2013-91	CRL	D	Wild	RD	231.7	36.2	97.00%	215.5	34.5	89.20%	1255150	1,257,275	99.48%	64	677	12
08/01/2014-92	MAD	I	Wild	WB	235.1	36.2	96.60%	219.3	34.5	89.50%	1115718	1,125,065	99.59%	57.4	716	11
31/03/2015-93	CRL	G	Wild	RD	232.5	36.3	96.80%	218.1	34.6	89.70%	1819784	1,821,569	99.58%	93.2	699	12
31/05/2017-94	MAD	UNK	Wild	FX	232.9	36.2	96.70%	219	34.9	91.00%	1648880	1,651,986	99.38%	84.7	719	11
31/05/2017-95	CRL	C	Wild	WB	232.2	36.1	96.40%	212.2	34.1	86.60%	955058	957,051	99.34%	48.4	691	12
31/05/2017-96	CRL	F	Wild	RD	231.5	36.1	96.60%	214	34.2	87.70%	1152290	1,155,126	98.71%	58.6	642	13
31/05/2017-97	CRL	Y	Wild	RD	229.9	36.1	96.60%	213.4	34.4	88.50%	1225914	1,226,959	99.45%	61.8	690	12
31/05/2017-98	CRL	F	Wild	RD	230.9	36.2	96.50%	213.3	34.2	87.50%	1168344	1,170,540	99.17%	59.2	702	11
31/05/2017-99	CRL	F	Wild	RD	230	36.2	96.80%	211.3	34.1	86.80%	1139258	1,140,912	98.69%	57.4	639	13
24/07/2017-100	CRL	F	Wild	RD	233	36.3	97.10%	212.1	34.1	86.00%	995616	996,811	99.37%	50.6	676	12
13/11/2017-101	TOL	Z	Wild	WB	233.5	36.2	96.50%	215.4	34.4	88.60%	1429178	1,434,081	99.53%	72.7	658	12
13/11/2017-105	MAD	X	Wild	WB	233.1	36.2	96.50%	215.3	34.2	87.70%	1075320	1,078,575	99.20%	54.9	700	11

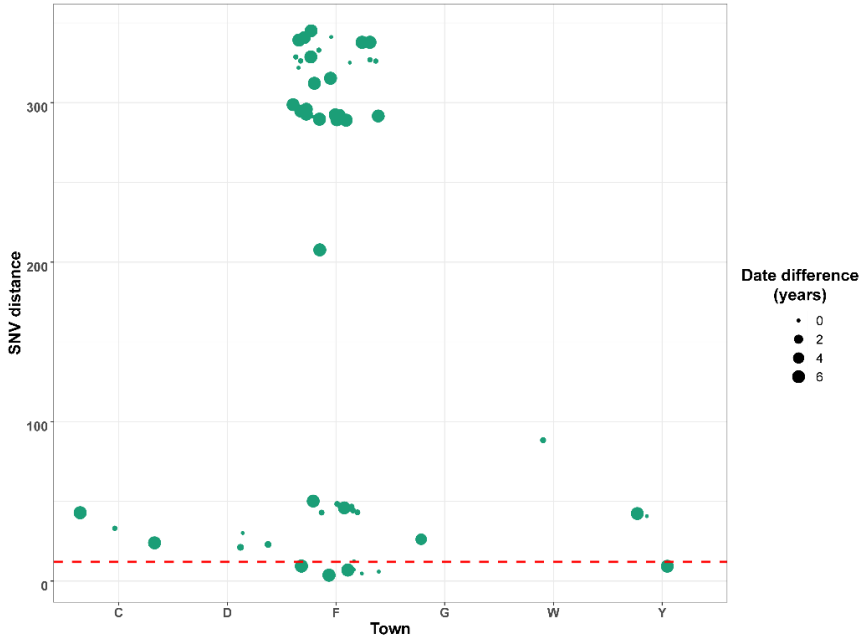
Pro = Province; MAD= Madrid; CRL= Ciudad Real; TOL = Toledo; ARA = Aragon; TER = Teruel; VAL = Valencia; Mun = Municipality; BV = Bovine; WB = Wild boar; RD = Red deer; FD = Fallow deer; FX = Fox; Av. = Average; Q30 = Percentage of reads with a Phred score higher than 30; Cov. = Coverage; SB = Spoligotype.

Town	Herd	Production; Management	Fence	Wildlife presence	Shared Pastures
A	1	Beef; semi-extensive	Partial	Yes	No
A	2	Beef; extensive	Partial	Yes	No
B	3	Beef; semi-extensive	Partial	N.A.	No
C	4	Beef; extensive	Partial	Yes	No
D	5	N.A.	N.A.	N.A.	N.A.
D	6	Beef; extensive	Partial	Yes	No
E	7	Beef; extensive	Partial	Yes	No
F	8	Beef; extensive	Absent	Yes	No
G	9	Beef; extensive	Partial	Yes	No
H	10	N.A.	N.A.	N.A.	N.A.
J	11	Beef; extensive	Unknown	Yes	Yes
K	12	N.A.	N.A.	N.A.	N.A.
L	13	Mixed; extensive	Complete	Yes	No
M	14	N.A.	N.A.	N.A.	N.A.
N	15	Beef; extensive	Partial	Yes	No
O	16	Beef; intensive	Complete	Yes	No
P	17	N.A.	N.A.	N.A.	N.A.
Q	18	BF; extensive	Complete	No	No
R	19	BF; extensive	Complete	No	No
S	20	BF; semi-extensive	Complete	No	No
T	21	BF; extensive	Complete	Yes	No
U	22	Mixed (BF); semi-extensive	Unknown	No	No

Supplementary table 2. Bovine TB persistence risk factors identified in the different herds included in this study. BF = Bullfighting; N.A. = Not Available.



Supplementary figure 1. Pairwise genetic distances observed for *M. bovis* isolates obtained from herds with chronic outbreaks of bTB. The dashed line represents the 12 SNP threshold used to indicate putative epidemiological relationships between isolates.



Supplementary figure 2. Pairwise genetic distances observed for *M. bovis* isolates obtained from wildlife species in the province of Ciudad Real. The dashed line represents the 12 SNP threshold used to indicate putative epidemiological relationships between isolates.

4.3. Objective 3: development of novel tools for the identification of NTM species in clinical samples

4.3.1. Use of MALDI-TOF MS as an alternative method to Sanger sequencing for the identification of NTM species in the veterinary practice.

This study evaluated the capacity of MALDI-TOF MS (Bruker BioTyper) to reliably identify six reference strains and 69 NTM isolates obtained from 20 animal species in comparison to Sanger sequencing, which was considered the method of reference.

Both methodologies were able to identify a similar number of NTM isolates, since 88% (n=66) of the isolates were reliably identified by both methods (n = 57) or could not be identified by either (n = 9). In the latter case, the unidentified isolates possibly represented unknown NTM species and both methodologies approximated their identifications to species from a particular complex (e.g. *M. avium* complex).

Discordant results were identified for nine additional isolates. In four cases, discordance could be attributed to the MALDI-TOF score being below the threshold set for a high-confidence identification ($1.6 \leq$ score < 1.8). In another four cases, identifications through MALDI-TOF MS also had a low confidence score and sequencing allowed only for the identification at complex level (<99% for the 16S rRNA subunit and *hsp65* gene, and <98% identity for the *rpoB* gene). Finally, one discordant result was due to a misidentification; while Sanger sequencing identified the isolate as *M. malmesburyense*, MALDI-TOF MS identified the isolate as *M. novocastrense* with high confidence, a closely-related mycobacterial species. This was due to the absence of *M. malmesburyense* main spectra profiles in the commercial database that was used for MALDI-TOF MS (Mycobacteria Library v3.0).

This study demonstrates that MALDI-TOF MS is a valuable method for the rapid and cost-effective identification of NTM isolates in veterinary practice and can also contribute to the identification of novel mycobacterial species.

This article was accepted for publication on the 28th of January 2022 in Frontiers in Veterinary Science.

Journal impact factor in 2020: 3.412

Position in JCR index 2020: Veterinary Sciences 9/146 (Q1).

Lorente-Leal V, Liandris E, Bezos J, Pérez-Sancho M, Romero B, de Juan L. MALDI-TOF Mass Spectrometry as a Rapid Screening Alternative for Non-tuberculous Mycobacterial Species Identification in the Veterinary Laboratory. Front Vet Sci. 2021;8:780018. https://doi.org/10.3389/fvets.2022.827702

frontiers in Veterinary Science

BRIEF RESEARCH REPORT
published: 28 January 2022
doi: 10.3389/fvets.2022.827702

MALDI-TOF Mass Spectrometry as a Rapid Screening Alternative for Non-tuberculous Mycobacterial Species Identification in the Veterinary Laboratory

Víctor Lorente-Leal[1,2], Emmanouil Liandris[1], Javier Bezos[1,2], Marta Pérez-Sancho[1,2], Beatriz Romero[1,2] and Lucía de Juan[1,2]*

[1] *VISAVET Health Surveillance Center, Universidad Complutense de Madrid, Madrid, Spain,* [2] *Animal Health Department, Faculty of Veterinary Medicine, Universidad Complutense de Madrid, Madrid, Spain*

OPEN ACCESS

Edited by:
Srinand Sreevatsan, Michigan State University, United States

Reviewed by:
Danijela Horvatek Tomic, University of Zagreb, Croatia
Filip Boyen, Ghent University, Belgium

***Correspondence:**
Beatriz Romero
bromerom@ucm.es

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *02 December 2021*
Accepted: *03 January 2022*
Published: *28 January 2022*

Citation:
Lorente-Leal V, Liandris E, Bezos J, Pérez-Sancho M, Romero B and Juan Ld (2022) MALDI-TOF Mass Spectrometry as a Rapid Screening Alternative for Non-tuberculous Mycobacterial Species Identification in the Veterinary Laboratory. Front. Vet. Sci. 9:827702. doi: 10.3389/fvets.2022.827702

Non-tuberculous mycobacteria (NTM) are difficult to identify by biochemical and genetic methods due to their microbiological properties and complex taxonomy. The development of more efficient and rapid methods for species identification in the veterinary microbiological laboratory is, therefore, of great importance. Although MALDI-TOF Mass Spectrometry (MS) has become a promising tool for the identification of NTM species in human clinical practise, information regarding its performance on veterinary isolates is scarce. This study assesses the capacity of MALDI-TOF MS to identify NTM isolates ($n = 75$) obtained from different animal species. MALDI-TOF MS identified 76.0% ($n = 57$) and 4% ($n = 3$) of the isolates with high and low confidence, respectively, in agreement with the identification achieved by Sanger sequencing of housekeeping genes (16S rRNA, *hsp65*, and *rpoB*). Thirteen isolates (17.3%) were identified by Sanger sequencing to the complex level, indicating that these may belong to uncharacterised species. MALDI-TOF MS approximated low confidence identifications toward closely related mycobacterial groups, such as the *M. avium* or *M. terrae* complexes. Two isolates were misidentified due to a high similarity between species or due to the lack of spectra in the database. Our results suggest that MALDI-TOF MS can be used as an effective alternative for rapid screening of mycobacterial isolates in the veterinary laboratory and potentially for the detection of new NTM species. In turn, Sanger sequencing could be implemented as an additional method to improve identifications in species for which MALDI-TOF MS identification is limited or for further characterisation of NTM species.

Keywords: MALDI-TOF MS, non-tuberculous mycobacteria (NTM), mycobacteria, Sanger sequencing, identification, veterinary samples, screening

INTRODUCTION

Mycobacteria are a diverse group of acid-fast Gram-positive bacilli that include more than 200 species differentiated into five newly emended genera (1). This group of ubiquitous bacteria can be found in a wide range of environments, with many species being important animal and human primary and secondary pathogens. Although the members of the *Mycobacterium tuberculosis* complex (MTBC), causal agents of tuberculosis (TB) in animals and humans, are probably the most widely studied mycobacterial agents in veterinary microbiology, other non-tuberculous mycobacteria (NTM), such as *Mycobacterium avium* subsp. *paratuberculosis* (MAP) in ruminants or *Mycobacterium marinum* in fish are also very relevant animal pathogens (2, 3). Importantly, NTM have been isolated from a diverse range of animals and, not only do they interfere with diagnostic methods implemented in the eradication and control of TB, but may also pose a risk for immunocompromised patients (4–6).

Molecular genetic methods greatly improved the capacity of clinical microbiology laboratories to identify NTM species with the development of several methods, such as genetic probe assays, including the INNO-LiPA Mycobacteria or the GenoType Common Mycobacteria assays (7–9). Although these methods are still used in many laboratories, they are expensive, identify a limited number of mycobacterial species and misidentify less prevalent NTM species due to probe cross-reactivity (10). By these reasons, sequence-based identification methods, such as Sanger sequencing, have become an alternative for the identification of mycobacterial species due to their nucleotide-level resolution as well as flexibility (11, 12). Advanced procedures based on Whole Genome Sequencing (WGS) have also been effectively used to identify NTM isolates and could be very useful for the description of novel NTM species or epidemiological investigations (13, 14). However, due to the increased costs and technical and training requirements of WGS, its implementation in routine diagnosis of NTM is limited. As a result, many laboratories still use a combination of probes and Sanger sequencing as a reference method to identify NTM species (8, 15, 16).

Sanger sequencing is based on the analysis of conserved genetic regions, mainly the hypervariable region of the 16S rRNA gene (11, 12, 17), the hypervariable region of the *hsp65* gene (18, 19), the *rpoB* gene (12, 20) or the 16S-23S internal transcribed spacer (21). Nevertheless, the routine use of sequencing in the microbiological laboratory is not as straightforward as it could initially seem for several reasons (11). Firstly, the ever-increasing number of mycobacterial species, taxonomical variations and high genetic similarity between species makes genetic identification a daunting task. In addition, the vast amount of sequence data in public repositories, which may not be appropriately annotated, and the rarity of certain mycobacterial species requires special caution when assigning an identification.

The introduction of MALDI-TOF MS in bacterial species identification has revolutionised clinical microbiology in the last decade, allowing a cost-effective and rapid identification of many important human pathogens (22). In the veterinary field, the application of MALDI-TOF MS has also received great interest and has been used to identify different veterinary important microbial pathogens, such as *Brucella* or *Staphylococcus,* with very promising results (23–25).

During many years, mycobacterial species have proven to be a difficult agent to identify through MALDI-TOF MS, mainly due to their complex cell wall composition and their fastidious growth. Since its first implementation in mycobacterial identification, there have been major advances in the use of MALDI-TOF MS for the identification of NTM species (26). In the first place, improvements in extraction protocols have enhanced the availability and quality of proteins for MALDI-TOF MS analysis (8, 27, 28). Secondly, an increasing number of mycobacterial spectra, which are the foundation for MALDI-TOF MS identification, have become available over the years (26, 29). Nevertheless, several mycobacterial species are still difficult to identify through proteomic methods, including veterinary important species, such as *M. bovis* or members of the *M. avium* complex (MAC). In this aspect, several efforts have been made to tackle these limitations, such as expanding Main Spectra Profiles (MSPs) of these mycobacteria for MALDI-TOF MS analyses (30, 31).

There is limited data regarding the general capacity of MALDI-TOF MS in the identification of NTM isolates in the veterinary setting, when compared to the information available in the human clinical practice (30–34). The aim of this study was to evaluate the use MALDI-TOF MS in the identification of NTM obtained from different animal species, and compare its performance against standard genetic methods.

MATERIALS AND METHODS

Sample Selection and Processing

A total of seventy-five ($n = 75$) isolates were included in this study; 69 isolates were selected from a collection of samples obtained from animals ($n = 20$ species, **Table 1**) with suspected mycobacteriosis that were originally submitted for bacteriological culture during routine diagnostics by the Mycobacterial Unit of VISAVET Health Surveillance Centre (years 2011–2015). Six isolates corresponded to established reference strains and were used as controls (**Table 1**, **Supplementary Table 1**).

All samples were originally considered as suspicious for tuberculosis and were therefore processed and cultured in the BSL3 facilities of VISAVET according to standardised mycobacteriological procedures as described elsewhere (35). Briefly, tissues were homogenised in sterile distilled water (Sigma-Aldrich, St. Louis, MO, USA), decontaminated with 0.75% (w/v) hexadecyl pyridinium chloride (Sigma-Aldrich) and centrifuged. Pellets were then swabbed on Löwenstein-Jensen and Coletsos slants supplemented with sodium pyruvate (Difco, Madrid, Spain), and incubated for a maximum of 3 months. When growth was observed, isolated colonies were analysed using PCR and DVR-spoligotyping in order to detect the presence of the MTBC (36, 37). Once the presence of MTBC members was discarded, a loop-full of mycobacterial culture was inoculated in 15 mL of Middlebrook 7H9 liquid medium, supplemented with sodium pyruvate and oleic albumin

TABLE 1 | Samples ($n = 75$) analysed in this study.

N° samples	Animal species	Sample origin
2	Alpaca	Lung, lymphnodes
1	American oystercatcher	Liver
32	Bovine	Lung, lymphnodes
1	Common shelduck	Lymphnodes
6	Deer	Lymphnodes
1	Domestic goat	Lymphnodes
2	Domestic Pig	Tissue homogenate, lymph nodes
1	Eurasian griffon	Lymphnodes
2	Ferret	Lymphnodes, liver, spleen
1	Fox	Tissue homogenate
1	Fulvous whistling duck	Liver, spleen
1	Giant Wood-Rail	Liver
1	Lesser kestrel	Lymphnodes
1	Mackerel	Necrotic granuloma
2	Malayan tapir	Trunk lavage
5	Mountain goat	Lymphnodes
1	Orangutan	Gastric lavage
1	Raccoon	Lymphnodes
6	Reference culture	Spanish type culture collection (CECT)
1	Roe Deer	Lymphnodes
7	Wild boar	Lymphnodes

The relationship between the animal species and the isolates identified within these species can be found in ***Supplementary Table 1****.*

dextrose catalase (OADC) (Becton Dickinson, Franklin Lakes, NJ, USA), and incubated for up to 30 days. Cultures were visually inspected every 5, 10, 21, and 30 days and when microbial growth was visible, 1 and 1.5 mL of liquid culture were processed for proteomic and genetic analysis, respectively.

Protein Extraction and MALDI-TOF MS Analysis

One mL of Middlebrook 7H9 medium was centrifuged at 15,500 × g for 2 min, the supernatant was discarded and the pellet was resuspended in 300 µL of HPLC-grade water (Sigma-Aldrich, St. Louis, MO, USA). Culture samples were then heat-inactivated at 100°C for 30 min and, following the addition of 900 µL of 96% ethanol (Panreac, Castellar del Vallès, Barcelona, Spain), these were stored overnight at −20°C.

Samples were centrifuged at 15,500 × g for 2 min, the supernatant was discarded and the pellet was let to air-dry at room temperature for at least 15 min. A small quantity of 0.5 mm glass beads and a volume equal to pellet size of HPLC-grade acetonitrile (Honeywell Fluka™, Charlotte, NC, USA) were added and samples were vortexed for 10 min and sonicated at 40 KHz for another 10 min on a Ultrasons sonicator (Selecta, Barcelona, Spain). An equal volume of 70% HPLC-grade formic acid (Sigma-Aldrich) was added and samples were centrifuged at 15,500 × g for 2 min. One µL of the supernatant was spotted on a polished steel target plate, let dry at room temperature and overlaid with one µL of α-cyano-4-hydroxy-cinnamic acid (HCCA) matrix HPLC grade (Bruker Daltonics, Billerica, MA, USA). Plates were then analysed on a Bruker Daltonik UltrafleXtreme MALDI TOF/TOF system (Bruker Daltonics) and spectra were compared against the Mycobacteria Library version 3.0 (38). Calibration of the Bruker BioTyper was performed using the Bruker Bacterial Test (BTS) according to the manufacturer's recommendations, and validation was achieved by running a reference strain of *E. coli* (ATCC 25922) as well as with the incorporation of mycobacterial reference strains (**Table 1**, **Supplementary Table 1**).

Final MALDI-TOF MS identifications were based on the score and consistency of the results. For the Bruker systems, MALDI-TOF scores higher or equal to 2.0 are required for a high confidence identification. Nevertheless, due to the biological features of mycobacteria, previous studies have shown that lower scores can effectively be used for NTM identification (15, 26). Therefore, MALDI-TOF scores higher or equal to 1.8 were considered high confidence identifications as described elsewhere (26) and were used to establish a positive NTM identification. Low confidence identifications were established when MALDI-TOF scores were obtained between 1.6 and 1.8, and non-reliable identifications were set when MALDI scores were below 1.6. In addition, consistency of the identification was evaluated by assessing the taxon and the ID score of the first five results returned by the MALDI-TOF MS. When more than one taxon was identified with a similar score, the results were approximated to the most related mycobacterial group. In the cases in which identifications were not reliable or had a low confidence score, extractions and MALDI-TOF MS analyses were repeated. In either case, these low confidence identifications were not considered for the final identification result.

DNA Extraction and Sanger Sequencing

One and a half (1.5) mL of Middlebrook 7H9 media were centrifuged at 11,500 × g for 10 min. The pellet was washed in 1 mL of HPLC-grade water and samples were centrifuged again at 11,500 × g for 10 min. The pellet was then resuspended in 100 µL of HPLC-grade water and heat-inactivated at 100°C for 30 min.

For species identification through Sanger sequencing, samples were analysed following an algorithm designed according to growth speed in culture and partial 16S rRNA sequence (**Figure 1**) (17). The *rpoB* gene and the *hsp65* short fragment were used to identify rapid-growing and slow-growing mycobacteria, respectively (19, 20). In the case of *M. avium* and *M. intracellulare*, the *hsp65* long fragment was also used for species and subspecies identification (18). DNA amplicons were sent to STABvida (Lisbon, Portugal) for Sanger sequencing.

All DNA sequences were manually curated and compared against the NCBI nucleotide database using the Basic Local Alignment Search Tool (BLAST). Species identification was accepted when sequence identity was above 99% for the 16S rRNA and *hsp65* genes, and above 98% for the *rpoB* gene (11). If incongruent results were obtained between sequences or against MALDI-TOF MS identification, the *hsp65* short fragment or the *rpoB* gene were sequenced in those cases where they were not

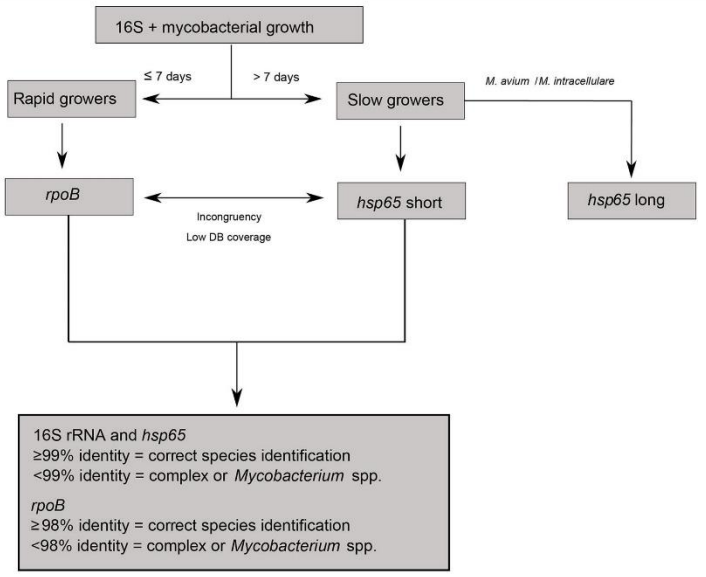


FIGURE 1 | Algorithm describing species identification through Sanger sequencing.

required initially (**Figure 1**). When sequence identity was low, identification was limited to the genus or complex level.

RESULTS

From the total number of isolates ($n = 75$), MALDI-TOF MS was able to identify 58 isolates (77.3%) with high confidence (score ≥ 1.8) to the group/species level, and 8 with low confidence (score between 1.6 and 1.8) (**Table 2**). The remaining isolates ($n = 9$) could not be identified reliably (score < 1.6). One third ($n = 25$) of the total number of isolates were identified as *M. avium*.

The majority of the isolates ($n = 62$, 82.7%) were identified with high confidence to the species or subspecies level through Sanger sequencing (**Table 2**, **Supplementary Table 1**). One third ($n = 25$) of the total number of isolates were identified as *M. avium* subspecies, and could be further differentiated into *M. avium* subsp. *avium* ($n = 17$) and "*M. avium* subsp. *hominissuis*" ($n = 8$). The remaining non-*M. avium* isolates ($n = 37$) were identified as other NTM species, with most of them corresponding to *M. fortuitum* ($n = 6$), *M. non-chromogenicum* ($n = 4$), *M. thermoresistibile* ($n = 4$), and *M. peregrinum* ($n = 3$). Thirteen isolates could not be reliably identified to the species level based on Sanger sequencing, suggesting that they could possibly represent unknown NTM species. Sequence identity was below 98% for the complementary sequencing targets (i.e., *hsp65*, *rpoB* or both) for 12 isolates, but the 16S rRNA sequence was related to the *M. avium* ($n = 8$), *M. terrae* ($n = 2$), *M. fortuitum* ($n = 1$), or *M. simiae* ($n = 1$) complexes (**Supplementary Table 1**). An additional isolate could be classified as *M. bourgelatii* or *M. intermedium* based on 16S rRNA and *hsp65*/*rpoB* sequences, respectively. However, when compared to *M. intermedium*, this isolate presented the characteristic 12 bp gap of *M. bourgelatii* (data not shown).

The ability of the two methods to identify NTM was similar, since 88% ($n = 66$) of the isolates were identified by both methods ($n = 57$) or could not be identified by either ($n = 9$). In the latter case, non-reliable identifications obtained by MALDI-TOF MS corresponded to isolates for which Sanger sequencing could not provide a clear matching identification. In a similar manner to sequencing, MALDI-TOF MS approximated the identifications to closely related groups of species, such as the *M. avium* or *M. terrae* complexes (**Supplementary Table 1**).

The remaining isolates ($n = 9$) presented discording identifications and were evaluated further. One of these isolates was identified as *M. malmesburyense* by Sanger sequencing and as *M. novocastrense* through MALDI-TOF MS. Four isolates could be identified with Sanger sequencing but the MALDI-TOF MS identification had a low confidence score. Despite their low confidence score, the identification between three of these isolates agreed with the one obtained through Sanger sequencing,

TABLE 2 | Species identification comparison between Sanger sequencing and MALDI-TOF MS (v3.0).

			Score			
Isolate	**Sanger sequencing ID**	**MALDI-TOF MS ID**	**< 1.6**	**>1.6 and < 1.8**	**≥ 1.8 and < 2.0**	**>2.0**
17	*M. avium* subsp. *avium*	*M. avium*		2	4	11
8	'*M. avium* subsp. *hominissuis*'	*M. avium*			6*	2
8	*M. avium* complex	*M. intracellulare/chimaera* group	8			
1	*Mycobacterium* spp[a]	*M. intermedium*	1			
2	*M. chitae*	*M. chitae*			1	1
1	*M. colombiense*	*M. colombiense*			1*	
1	*M. elephantis*	*M. elephantis*			1	
1	*M. engbaekii*	*M. hiberniae/engbaekii*[b]		1		
2	*M. europaeum*	*M. europaeum*			1	1
6	*M. fortuitum*	*M. fortuitum*			1	5
1	*M. fortuitum* complex	*M. fortuitum*		1		
1	*M. intracellulare*	*M. intracellulare/chimaera group*				1
1	*M. kansasii*	*M. kansasii*				1*
1	*M. malmesburyense*	*M. novocastrense*			1	
1	*M. neoaurum*	*M. neoaurum*		1		
4	*M. nonchromogenicum*	*M. nonchromogenicum*			1	3
1	*M. palustre*	*M. palustre*				1
3	*M. peregrinum*	*M. peregrinum*			1	2
2	*M. phlei*	*M. phlei*				2*
1	*M. septicum*	*M. septicum*			1	
1	*M. seoulense*	*M. seoulense*			1	
1	*M. shimoidei*	*M. shimoidei*			1	
1	*M. simiae* complex	*M. interjectum*[c]		1		
1	*M. smegmatis*	*M. smegmatis*				1
1	*M. terrae*	*M. terrae*				1*
2	*M. terrae complex*	*M. hiberniae/engbaekii*[b]		2		
4	*M. thermoresistibile*	*M. thermoresistibile*				4
1	*M. vaccae*	*M. vaccae*				1*

Sanger sequencing identifications fulfilled the required thresholds depending on the genetic target used: 99% (16S rRNA and hsp65) and 98% (rpoB).

**Includes a single reference strain (n = 6).*

[a]*Closely related to M. bourgelatii and M. intermedium.*

[b]*M. terrae complex.*

[c]*M. simiae complex.*

while the fourth isolate (*M. engbaekii*) could not be appropriately differentiated from *M. hiberniae*, a related species from the *M. terrae* complex. Finally, four isolates were identified with low confidence through MALDI-TOF MS but could only be identified to the complex level by Sanger sequencing.

DISCUSSION

The implementation of MALDI-TOF MS in the identification of mycobacteria has become an interesting alternative to genetic methods in human clinical microbiology laboratories. In veterinary mycobacteriology, MALDI-TOF has proven to be a reliable identification method of mycobacteria in a study evaluating 111 NTM isolates obtained from wild boars in Switzerland (33). In addition, other studies have focused on other veterinary important mycobacteria such as MAP in different animal hosts or *M. marinum* in fish (30, 34). Our study expands the utility of MALDI-TOF MS characterisation in a diverse array of animal species ($n = 20$), with an overall high agreement with Sanger sequencing.

Sanger sequencing is commonly used as a reference method for taxonomical identification of mycobacterial species in the clinical setting (8, 15, 16). However, its implementation is not without drawbacks due to the taxonomical complexity of the genus and the high genetic similarity of certain species or groups, which require the use of several housekeeping targets, and the lack of updated and curated sequence databases. Efforts have been made toward the automation of this process with the development of Open-Source tools for the identification of NTM species using Sanger sequencing data (12), which could greatly facilitate the use of this methodology in the future. In contrast, although WGS allow for a higher resolution than traditional sequencing of conserved genetic targets and show a

great potential for the identification of NTM species directly from samples (14, 39), its increased costs and training requirements currently limits its implementation as a routine method in many microbiology laboratories. However, this may change in the future as WGS becomes more accessible with the development of simple bioinformatics tools.

MALDI-TOF MS offers a simplified process for NTM identification based on spectral similarities with a curated MSP database and reinforced with confidence values, leading to a more standardised procedure. In our study, high confidence MALDI-TOF MS scores ($\geq$ 1.8) were used for NTM species identification, and the majority of the evaluated isolates were identified as such, indicating a good correlation with the stored MSPs in the database. A recent publication proposed the threshold for a high-confidence identification to be set at 2.0 (15). However, a threshold of 1.8 proved to be sufficient in the context of this study, since all identifications using this threshold agreed with the reference genetic method, with one exception (*M. malmesburyense*). Nevertheless, further studies including a wider range of NTM strains from different animal species should be carried out in the future to assess this threshold. Interestingly, identifications to the species or complex level could still be achieved even with low MALDI-TOF scores (1.6–1.8), indicating that these could provide valuable information that could be used by laboratory personnel to fine tune subsequent tests for isolate identification (15).

From the point of view of cost-effectiveness, MALDI-TOF MS presents several advantages to Sanger sequencing. Sanger sequencing requires at least two PCRs per isolate followed by, at least, two sequencing reactions per target. Therefore, 96 well-sequencing plates can analyse up to 48 samples per run, which in turn can take several hours. Furthermore, the obtained sequences require careful sequence curation and interpretation. In contrast MALDI-TOF MS requires just one procedure and up to 384 samples can be tested per run in a couple of minutes, allowing for a much more rapid identification and reduced cost per sample. Although our study used a liquid culture step from primary isolation until final identification, good identification results can also be achieved from primary isolations in solid media (26).

MALDI-TOF cannot effectively discern between certain closely related NTM species or subspecies of veterinary and human importance, such as *M. intracellulare* and *M. chimaera* or *M. avium* subspecies. These species represent an important group of mycobacteria in veterinary medicine that can be frequently isolated in cattle and swine (40–42). Their potential interference with routine animal TB diagnosis and their opportunistic nature make their identification an important element of eradication programmes and public health (2, 5, 43–45). Differentiation of these species can only be achieved through genetic methods, such as sequencing of the 3' end of the *hsp65* gene, PCR-based detection of specific Insertion Sequences or WGS approaches (18, 39, 46). One third of the isolates in our study corresponded to *M. avium* subspecies, and we additionally identified one *M. intracellulare-chimaera* group isolate. MALDI-TOF confidence scores in our study were, in general, high for these species, probably as a result of the large amount of MSPs stored in current libraries. This suggests that, despite its limit in resolution, MALDI-TOF MS could be used as a rapid method for initial screening of *M. avium* and therefore aid in the subsequent choice of the appropriate genetic targets for a more thorough characterisation. Recent efforts in the analysis of MSPs from this group of bacteria suggest that, with an appropriately curated database, MALDI-TOF MS identification of these, and other, challenging species could be achievable in the future (30, 31, 47). Certainly, MSP analysis for the detection of characteristic peaks between closely related mycobacteria should be carried out in the future (48). In addition, the analysis of lipid profiles through positive-ion MALDI-TOF MS or alternative MS instruments, such as High Resolution Tandem Mass Spectrometry, have been recently used to differentiate between closely related mycobacteria, such as the MTBC and *M. abscessus* subspecies (48, 49). Although these studies were carried out in a limited number of strains, their use in other closely related mycobacteria could also be an interesting alternative in the future.

Misidentifications and non-reliable identifications are two important limitations in the implementation of MALDI-TOF MS in microbiological laboratories (33). This is mainly a result of a reduced number and variety of MSPs for less prevalent mycobacterial species in current databases or the absence of MSPs for unknown NTM species.

Low confidence and non-reliable identifications in our study were mostly observed in isolates that could not be identified through Sanger sequencing either. Interestingly, although MALDI-TOF identifications in these cases are not considered reliable, they were similar to the ones obtained through Sanger sequencing. For example, the isolate that was closely related to *M. bourgelatii/intermedium* was identified as *M. intermedium* by MALDI-TOF MS, and the *M. simiae* complex isolate was identified as *M. interjectum*, a closely related species as seen by sequencing of 16S rRNA and *rpoB* genes. This indicates that MALDI-TOF may also be useful in detecting non-established NTM species and aid in the selection of complementary methods for a cost-effective characterisation (50). Further characterisation of these isolates would be needed to define these species and is outside of the scope of this publication.

Only one misidentification with a high-confidence score was observed in our study, which corresponded to *M. malmesburyense*, a rare mycobacterial species for which no MSPs are currently available. In addition to the absence of MSPs, a limited number of spectra from less prevalent mycobacterial species (e.g., *M. setense*) has also been shown to have a negative effect in MALDI-TOF MS performance (15, 29). The isolation of rare and unknown NTM species in animal samples in this study, as well as in the one published by Ghielmetti et al., strongly indicate that animals contain a large diversity of NTM species that is not being represented by current nucleotide and MSP databases, which are mainly focused on human clinical isolates (24, 33). Furthermore, as more animal and environmental sources are sampled, there is an increasing need to describe novel species of mycobacteria for a better classification and understanding of mycobacterial ecology and taxonomy. Thus, the addition of MSPs from animal sources could aid in the identification of less prevalent mycobacterial species and improve the performance of MALDI-TOF MS identification in the future.

In conclusion, despite the limited discriminatory power among certain mycobacterial groups, MALDI-TOF MS may be a suitable alternative for NTM species identification for several reasons. Unlike sequencing, in which multiple targets need to be identified, purified and reprocessed with a sequencing PCR, MALDI-TOF spectra are readily obtained after the extraction protocol. In addition, MALDI-TOF steel plates allow for the simultaneous analysis of a large number of samples, making this methodology extremely time- and cost-efficient. If adequate MSP libraries are available, data processing is minimal, and does not require thorough taxonomic or database research which could lead to erroneous species identification (11). However, there is a need to incorporate a larger number of MSPs from veterinary isolates in the available databases in order to increase MALDI-TOF performance with samples with a high diversity of mycobacterial species, such as those originating from animals or environmental sources. Nevertheless, even in the absence of curated MSPs, MALDI-TOF can be a powerful tool for the detection of potentially unknown NTM species. In these cases, or when closely-related mycobacterial species are detected, the use of more precise molecular genetic methods could be of use and refined identification algorithms could significantly improve identification turnaround times.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession numbers can be found at: https://www.ncbi.nlm.nih.gov/nuccore/, OK538881 - OK538918, OK539021 - OK539047, OK538946 - OK539020, OK538919 - OK538945.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because all animals were sampled due to suspicion of mycobacterial infection in a clinical setting or under an official eradication programme (e.g., tuberculosis). Written informed consent for participation was not obtained from the owners because all animals were sampled due to suspicion of mycobacterial infection in a clinical setting or under an official eradication programme (e.g., tuberculosis).

AUTHOR CONTRIBUTIONS

VL-L performed all the experiments in this study. MP-S, EL, and BR participated in the design of the study. LJ, JB, and BR were responsible for obtaining the animal samples. VL-L wrote the manuscript with the insights of all the co-authors. BR and MP-S supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

VL-L was funded with a predoctoral grant from the Complutense University of Madrid and Banco Santander 2017–2018. This work was partially financed and supported by the Ministerio de Agricultura, Pesca y Alimentación and the Programa de Tecnologías Avanzadas en Vigilancia Sanitaria (TAVS) de la Comunidad de Madrid (S2013/ABI2747). This work was partially funded by the European Union Reference Laboratory for bovine tuberculosis.

ACKNOWLEDGMENTS

We would like to thank the excellent work performed by N. S. Copel and laboratory technicians F. Lozano, A. Gutiérrez, and D. de la Cruz in mycobacterial sample processing and culturing. We would also like to thank the insights and help provided by A. Casamayor in MALDI-TOF MS functioning and analyses. Finally, we would like to thank Bruker Diagnostics for their insights and help during the characterisation of the isolates presented in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2022.827702/full#supplementary-material

REFERENCES

1. Gupta RS, Lo B, Son J. Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Front Microbiol.* (2018) 9:67. doi: 10.3389/fmicb.2018.00067
2. Biet F, Boschiroli ML. Non-tuberculous mycobacterial infections of veterinary relevance. *Res Vet Sci.* (2014) 97(Suppl):S69–77. doi: 10.1016/j.rvsc.2014.08.007
3. Hashish E, Merwad A, Elgaml S, Amer A, Kamal H, Elsadek A, et al. *Mycobacterium marinum* infection in fish and man: epidemiology, pathophysiology and management; a review. *Vet Q.* (2018) 38:35–46. doi: 10.1080/01652176.2018.1447171
4. Ghielmetti G, Friedel U, Scherrer S, Sarno E, Landolt P, Dietz O, et al. Non-tuberculous Mycobacteria isolated from lymph nodes and faecal samples of healthy slaughtered cattle and the abattoir environment. *Transbound Emerg Dis.* (2018) 65:711–8. doi: 10.1111/tbed.12793
5. Michelet L, de Cruz K, Karoui C, Tambosco J, Moyen JL, Henault S, et al. Second line molecular diagnosis for bovine tuberculosis to improve diagnostic schemes. *PLoS One.* (2018) 13:e0207614. doi: 10.1371/journal.pone.0207614
6. Scherrer S, Landolt P, Carroli N, Stephan R. Molecular Characterization of *Mycobacterium avium* subsp. hominissuis of two groups of lymph nodes, being intradermal tuberculin or interferon-gamma test positive and negative, isolated from swiss cattle at Slaughter. *Front Vet Sci.* (2018) 5:32. doi: 10.3389/fvets.2018.00032
7. Bouzinbi N, Marcy O, Bertolotti T, Chiron R, Bemer P, Pestel-Caron M, et al. Evaluation of the GenoType NTM-DR assay performance for the identification and molecular detection of antibiotic resistance in Mycobacterium abscessus complex. *PLoS One.* (2020) 15:e0239146. doi: 10.1371/journal.pone.0239146
8. Alcolea-Medina A, Fernandez MTC, Montiel N, Garcia MPI., Sevilla CD, North N, et al. An improved simple method for the

identification of Mycobacteria by MALDI-TOF MS (Matrix-Assisted Laser Desorption- Ionization mass spectrometry). *Sci Rep.* (2019) 9:20216. doi: 10.1038/s41598-019-56604-7

9. 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. doi: 10.1128/CMR.00038-17
10. Tortoli E, Pecorari M, Fabio G, Messino M, Fabio A. Commercial DNA probes for mycobacteria incorrectly identify a number of less frequently encountered species. *J Clin Microbiol.* (2010) 48:307–10. doi: 10.1128/JCM.01536-09
11. Tortoli E. Standard operating procedure for optimal identification of mycobacteria using 16S rRNA gene sequences. *Stand Genomic Sci.* (2010) 3:145–52. doi: 10.4056/sigs.932152
12. Kim YG, Jung K, Kim S, Kim MJ, Lee JS, Park SS, et al. SnackNTM: an open-source software for sanger sequencing-based identification of nontuberculous mycobacterial species. *Ann Lab Med.* (2021) 42:213–48. doi: 10.3343/alm.2022.42.2.213
13. Dohal M, Porvaznik I, Solovic I, Mokry J. Whole genome sequencing in the management of non-tuberculous mycobacterial infections. *Microorganisms.* (2021) 9:2237. doi: 10.3390/microorganisms9112237
14. He Y, Gong Z, Zhao X, Zhang D, Zhang Z. Comprehensive Determination of *Mycobacterium tuberculosis* and nontuberculous mycobacteria from targeted capture sequencing. *Front Cell Infect Microbiol.* (2020) 10:449. doi: 10.3389/fcimb.2020.00449
15. Rodriguez-Temporal D, Rodriguez-Sanchez B, Alcaide F. Evaluation of MALDI biotyper interpretation criteria for accurate identification of nontuberculous mycobacteria. *J Clin Microbiol.* (2020) 58:e01103-20. doi: 10.1128/JCM.01103-20
16. Kalaiarasan E, Thangavelu K, Krishnapriya K, Muthuraj M, Jose M, Joseph NM. Diagnostic performance of real time PCR and MALDI-TOF in the detection of nontuberculous mycobacteria from clinical isolates. *Tuberculosis.* (2020) 125:101988. doi: 10.1016/j.tube.2020.101988
17. Böddinghaus B, Rogall T, Flohr T, Blöcker H, Böttger EC. Detection and identification of mycobacteria by amplification of rRNA. *J Clin Microbiol.* (1990) 28:1751. doi: 10.1128/jcm.28.8.1751-1759.1990
18. Turenne CY, Semret M, Cousins DV, Collins DM, Behr MA. Sequencing of *hsp65* distinguishes among subsets of the *Mycobacterium avium* complex. *J Clin Microbiol.* (2006) 44:433–40. doi: 10.1128/JCM.44.2.433-440.2006
19. Telenti A, Marchesi F, Balz M, Bally F, Böttger EC, Bodmer T. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *J Clin Microbiol.* (1993) 31:3. doi: 10.1128/jcm.31.2.175-178.1993
20. Ben Salah I, Adekambi T, Raoult D, Drancourt M. *rpoB* sequence-based identification of *Mycobacterium avium* complex species. *Microbiology.* (2008) 154:3715–23. doi: 10.1099/mic.0.2008/020164-0
21. Frothingham R, Wilson KH. Sequence-based differentiation of strains in the *Mycobacterium avium* complex. *J Bacteriol.* (1993) 175:7. doi: 10.1128/jb.175.10.2818-2825.1993
22. Rodríguez-Sánchez B, Cercenado E, Coste AT, Greub G. Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018. *Eurosurveillance.* (2019) 24:1800193. doi: 10.2807/1560-7917.ES.2019.24.4.1800193
23. Pérez-Sancho M, Vela AI, Horcajo P, Ugarte-Ruiz M, Domínguez L, Fernández-Garayzábal JF, et al. Rapid differentiation of *Staphylococcus aureus* subspecies based on MALDI-TOF MS profiles. *J Vet Diagn Invest.* (2018) 30:7. doi: 10.1177/1040638718805537
24. Randall LP, Lemma F, Koylass M, Rogers J, Ayling RD, Worth D, et al. Evaluation of MALDI-ToF as a method for the identification of bacteria in the veterinary diagnostic laboratory. *Res Vet Sci.* (2015) 101:42–9. doi: 10.1016/j.rvsc.2015.05.018
25. Sali M, De Maio F, Tarantino M, Garofolo G, Tittarelli M, Sacchini L, et al. Rapid and safe one-step extraction method for the identification of *Brucella* strains at genus and species level by MALDI-TOF mass spectrometry. *PLoS One.* (2018) 13:e0197864. doi: 10.1371/journal.pone.0197864
26. Alcaide F, Amlerova J, Bou G, Ceyssens PJ, Coll P, Corcoran D, et al. How to: identify non-tuberculous *Mycobacterium* species using MALDI-TOF mass spectrometry. *Clin Microbiol Infect.* (2018) 24:599–603. doi: 10.1016/j.cmi.2017.11.012
27. Rodriguez-Temporal D, Perez-Risco D, Struzka EA, Mas M, Alcaide F. Evaluation of two protein extraction protocols based on freezing and mechanical disruption for identifying nontuberculous mycobacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry from liquid and solid cultures. *J Clin Microbiol.* (2018) 56:e01548-17. doi: 10.1128/JCM.01548-17
28. Mather CA, Rivera SF, Butler-Wu SM. Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *J Clin Microbiol.* (2014) 52:130–8. doi: 10.1128/JCM.01996-13
29. Rodriguez-Temporal D, Perez-Risco D, Struzka EA, Mas M, Alcaide F. Impact of updating the MALDI-TOF MS database on the identification of nontuberculous mycobacteria. *J Mass Spectrom.* (2017) 52:597–602. doi: 10.1002/jms.3944
30. Ravva SV, Harden LA, Sarreal CZ. Characterization and Differentiation of *Mycobacterium avium* subsp. paratuberculosis from other mycobacteria using matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Front Cell Infect Microbiol.* (2017) 7:297. doi: 10.3389/fcimb.2017.00297
31. Ricchi M, Mazzarelli A, Piscini A, Di Caro A, Cannas A, Leo S, et al. Exploring MALDI-TOF MS approach for a rapid identification of *Mycobacterium avium* ssp. paratuberculosis field isolates. *J Appl Microbiol.* (2017) 122:568–77. doi: 10.1111/jam.13357
32. Bacanelli G, Olarte LC, Silva MR, Rodrigues RA, Carneiro PAM, Kaneene JB, et al. Matrix Assisted Laser Desorption Ionization-Time-of-Flight mass spectrometry identification of *Mycobacterium bovis* in *Bovinae*. *J Vet Med Sci.* (2019) 81:1400–8. doi: 10.1292/jvms.19-0214
33. Ghielmetti G, Hilbe M, Friedel U, Menegatti C, Bacciarini L, Stephan R, et al. Mycobacterial infections in wild boars (*Sus scrofa*) from Southern Switzerland: diagnostic improvements, epidemiological situation and zoonotic potential. *Transbound Emerg Dis.* (2020) 68:573–86. doi: 10.1111/tbed.13717
34. Puk K, Banach T, Wawrzyniak A, Adaszek L, Zietek J, Winiarczyk S, et al. Detection of *Mycobacterium marinum, M. peregrinum, M. fortuitum and M. abscessus* in aquarium fish. *J Fish Dis.* (2018) 41:153–6. doi: 10.1111/jfd.12666
35. Lorente-Leal V, Liandris E, Castellanos E, Bezos J, Dominguez L, de Juan L, et al. Validation of a Real-Time PCR for the Detection of *Mycobacterium tuberculosis* complex members in bovine tissue samples. *Front Vet Sci.* (2019) 6:61. doi: 10.3389/fvets.2019.00061
36. Cousins DV, Wilton SD, Francis BR. Use of DNA amplification for the rapid identification of *Mycobacterium bovis*. *Vet Microbiol.* (1991) 27:187–95. doi: 10.1016/0378-1135(91)90010-D
37. Kamerbeek J, Schouls L, Kolk A, Agterveld Mv, Soolingen Dv, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol.* (1997) 35:907–14. doi: 10.1128/jcm.35.4.907-914.1997
38. Rodriguez-Sanchez B, Ruiz-Serrano MJ, Ruiz A, Timke M, Kostrzewa M, Bouza E. Evaluation of MALDI Biotyper Mycobacteria Library v3.0 for identification of nontuberculous mycobacteria. *J Clin Microbiol.* (2016) 54:1144–7. doi: 10.1128/JCM.02760-15
39. Yoon JK, Kim TS, Kim JI, Yim JJ. Whole genome sequencing of Nontuberculous Mycobacterium (NTM) isolates from sputum specimens of co-habiting patients with NTM pulmonary disease and NTM isolates from their environment. *BMC Genomics.* (2020) 21:322. doi: 10.1186/s12864-020-6738-2
40. Hulinova Stromerova N, Faldyna M. *Mycobacterium avium* complex infection in pigs: a review. *Comp Immunol Microbiol Infect Dis.* (2018) 57:62–8. doi: 10.1016/j.cimid.2018.06.005
41. Gcebe N, Michel AL, Hlokwe TM. Non-tuberculous *Mycobacterium* species causing mycobacteriosis in farmed aquatic animals of South Africa. *BMC Microbiol.* (2018) 18:32. doi: 10.1186/s12866-018-1177-9
42. Dvorska L, Matlova L, Bartos M, Parmova I, Bartl J, Svastova P, et al. Study of *Mycobacterium avium* complex strains isolated from cattle in the Czech Republic between 1996 and 2000. *Vet Microbiol.* (2004) 99:239–50. doi: 10.1016/j.vetmic.2004.01.008
43. Pollock C. Mycobacterial infection in the ferret. *Vet Clin North Am Exot Anim Pract.* (2012) 15:121–9, vii. doi: 10.1016/j.cvex.2011.09.002
44. Vaerewijck MJ, Huys G, Palomino JC, Swings J, Portaels F. Mycobacteria in drinking water distribution systems: ecology

and significance for human health. *FEMS Microbiol Rev.* (2005) 29:911–34. doi: 10.1016/j.femsre.2005.02.001

45. Biet F, Boschiroli ML, Thorel MF, Guilloteau LA. Zoonotic aspects of *Mycobacterium bovis* and *Mycobacterium avium-intracellulare* complex (MAC). *Vet Res.* (2005) 36:411–36. doi: 10.1051/vetres:2005001
46. van Ingen J, Turenne CY, Tortoli E, Wallace RJ, Jr., Brown-Elliott BA. A definition of the *Mycobacterium avium* complex for taxonomical and clinical purposes, a review. *Int J Syst Evol Microbiol.* (2018) 68:3666–77. doi: 10.1099/ijsem.0.003026
47. Pranada AB, Witt E, Bienia M, Kostrzewa M, Timke M. Accurate differentiation of *Mycobacterium chimaera* from *Mycobacterium intracellulare* by MALDI-TOF MS analysis. *J Med Microbiol.* (2017) 66:670–7. doi: 10.1099/jmm.0.000469
48. Jia Khor M, Broda A, Kostrzewa M, Drobniewski F, Larrouy-Maumus G. An improved method for rapid detection of *Mycobacterium abscessus* complex based on species-specific lipid fingerprint by routine MALDI-TOF. *Front Chem.* (2021) 9:715890. doi: 10.3389/fchem.2021.715890
49. Bajaj AO, Saraswat S, Knuuttila JEA, Freeke J, Stielow JB, Barker AP. Accurate identification of closely related mycobacterium tuberculosis complex species by high resolution tandem mass spectrometry. *Front Cell Infect Microbiol.* (2021) 11:656880. doi: 10.3389/fcimb.2021.656880
50. Angelakis E, Raoult D. Methods for the discovery of emerging pathogens. *Microbial Pathogene.* (2014) 77:114–8. doi: 10.1016/j.micpath.2014.06.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Table 1. Species identification of the NTM species isolated in this study. In cases where sequence ID was not reliable (<99% or <98%), closest related mycobacterial complex or species is specified. ID = Identification. N.D. = Not Done.

Code	Origin	16S	Hsp65 short	Hsp65 long	RpoB	Final ID
19	Mountain goat	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
20	Alpaca	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
27	Alpaca	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
45	Deer	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
58	Roe Deer	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
103	Bovine	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
110	Wild boar	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
147	Bovine	*M. avium* complex	Not reliable ID (<98%)[1]	N.D.	Not reliable ID (<98%)[2]	*M. avium* complex
9	Deer	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
14	Fulvous whistling duck	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
16	Common shelduck	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
22	Mountain goat	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
23	Mountain goat	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
24	Mountain goat	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
25	Eurasian griffon	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
29	American oystercatcher	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
51	Lesser kestrel	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
52	Mountain goat	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
53	Domestic goat	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
54	Deer	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
59	Deer	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
62	Domestic pig	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
98	Raccoon	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
129	Bovine	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
154	Bovine	*M. avium*	N.D.	*M. avium* subsp. *avium*	N.D.	*M. avium* subsp. *avium*
2	CECT 7407	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
15	Ferret	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
57	Ferret	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
81	Bovine	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
104	Domestic pig	*M. avium*	N.D.	*M. avium* subsp. *hominissuis* .	N.D.	*M. avium* subsp. *hominissuis*
114	Wild boar	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
125	Fox	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*
144	Bovine	*M. avium*	N.D.	*M. avium* subsp. *hominissuis*	N.D.	*M. avium* subsp. *hominissuis*

[1] closely related to *M. scrofulaceum*

[2] closely related to *M. bouchedurhonense*

Supplementary Table 1 (continued).

Code	Origin	16S	Hsp65 short	Hsp65 long	RpoB	Final ID
61	Bovine	*M. bourgelatii*	*M. intermedium*	N.D.	*M. intermedium*	*Mycobacterium* sp.[3]
55	Wild boar	*M. chitae*	*M. chitae*	N.D.	N.D.	*M. chitae*
75	Bovine	*M. chitae*	*M. chitae*	N.D.	N.D.	*M. chitae*
6	CECT 3035	*M. avium* complex	*M. colombiense*	N.D.	N.D.	*M. colombiense*
117	Wild boar	*M. engbaekii*	N.D.	N.D.	*M. engbaekii*	*M. engbaekii*
44	Deer	M. europaeum	M. europaeum/ scrofulaceum	N.D.	*M. europaeum*[4]	*M. europaeum*
161	Bovine	*M. europaeum*	*M. europaeum*	N.D.	N.D.	*M. europaeum*
5	Malayan tapir	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
21	Malayan tapir	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
28	Bovine	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
31	Ipecaa	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
149	Bovine	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
156	Bovine	*M. fortuitum*	N.D.	N.D.	*M. fortuitum*	*M. fortuitum*
56	Bovine	*M. fortuitum* complex	N.D.	N.D.	*M. septicum*	*M. septicum*
157	Bovine	*M. intracellulare/chimerae*	N.D.	*M. intracellulare*	N.D.	*M. intracellulare*
1	CECT 3030	*M. kansasii /gastri*	*M. kansasii*	N.D.	N.D.	*M. kansasii*
80	Bovine	*M. malmesburyense*	N.D.	N.D.	*M. malmesburyense*	*M. malmesburyense*
109	Bovine	*M. neoaurum*	N.D.	N.D.	*M. neoaurum*	*M. neoaurum*
105	Bovine	*M. nonchromogenicum*	*M. nonchromogenicum*	N.D.	N.D.	*M. nonchromogenicum*
121	Bovine	*M. nonchromogenicum*	*M. nonchromogenicum*	N.D.	N.D.	*M. nonchromogenicum*
128	Bovine	*M. nonchromogenicum*	*M. nonchromogenicum*	N.D.	N.D.	*M. nonchromogenicum*
131	Wild boar	*M. nonchromogenicum*	*M. nonchromogenicum* (98%)	N.D.	*M. nonchromogenicum*	*M. nonchromogenicum*

[3] Closely related to *M. bourgelatii* and *M. intermedium*

[4] Isolate with 99% identity and triplet nucleotide substitution

Supplementary Table 1 (continued).

N	Origin	16S	Hsp65 short	Hsp65 long	RpoB	Final ID
73	Bovine	*M. palustre*	*M. palustre*	N.D.	N.D.	*M. palustre*
159	Bovine	*M. simiae* complex[5]	Not reliable ID (<99%)[6]	N.D.	Not reliable ID (<98%)[7]	*M. simiae* complex
113	Bovine	*M. fortuitum* complex	N.D.	N.D.	*M. peregrinum*	*M. peregrinum*
68	Bovine	*M. fortuitum* complex	N.D.	N.D.	*M. peregrinum*	*M. peregrinum*
143	Bovine	*M. fortuitum* complex	N.D.	N.D.	*M. peregrinum*	*M. peregrinum*
10	CECT 3009	*M. phlei*	N.D.	N.D.	*M. phlei*	*M. phlei*
146	Bovine	*M. phlei*	N.D.	N.D.	*M. phlei*	*M. phlei*
47	Bovine	*M. fortuitum* complex	Not reliable ID (<99%)[8]	N.D.	Not reliable ID (<98%)[9]	*M. fortuitum* complex
158	Bovine	*M. shimoidei*	*M. shimoidei*	N.D.	N.D.	*M. shimoidei*
107	Bovine	*M. smegmatis*	N.D.	N.D.	*M. smegmatis*	*M. smegmatis*
12	CECT 3028	*M. terrae*	*M. terrae*	N.D.	N.D.	*M. terrae*
115	Wild boar	*M. engbaekii*	*M. arupense*	N.D.	Not reliable ID (<98%)[10]	*M. terrae* complex
118	Wild boar	*M. engbaekii*	*M. arupense*	N.D.	Not reliable ID (<98%)[10]	*M. terrae* complex
13	Orangutan	*M. thermoresistibile*	N.D.	N.D.	*M. thermoresistibile*	*M. thermoresistibile*
71	Bovine	*M. thermoresistibile*	N.D.	N.D.	*M. thermoresistibile*	*M. thermoresistibile*
79	Bovine	*M. thermoresistibile*	N.D.	N.D.	*M. thermoresistibile*	*M. thermoresistibile*
82	Bovine	*M. thermoresistibile*	N.D.	N.D.	*M. thermoresistibile*	*M. thermoresistibile*
30	CECT 3019	*M. vaccae*	N.D.	N.D.	*M. vaccae*	*M. vaccae*
169	Mackerel	*M. elephantis*	N.D.	N.D.	*M. elephantis*.	*M. elephantis*
26	Deer	*M. seoulense/ paraseoulense*	*M. seoulense/ paraseoulense*	*N.D.*	*M. seoulense*	*M. seoulense*

[5] Closely related to *M. paraense, M. interjectum* and *M. saskatchewanense*.

[6] Related to *M. paraense*.

[7] Related to *M. paraense* and *M. interjectum*.

[8] Related to *M. peregrinum, M. mortmartrense* and *M. septicum*.

[9] Related to *M. fortuitum*.

[10] Related to *M. engbaekii*

5. Discussion

Bovine TB remains an ongoing problem in many regions of the EU. This infectious disease poses a threat to both animal and human health, and has a significant effect in livestock production. In addition to the complex epidemiology of bTB, which involves both domestic and wildlife reservoirs, eradication efforts are negatively affected by the limitations of current diagnostic and characterisation techniques.

Current official methods for the detection and typing of MTBC members date back to the 1980s and 1990s. Recent developments in the fields of molecular biology, genomics and proteomics offer novel and powerful tools that could be implemented in the fight against bTB.

The main objective of this doctoral thesis was the development and evaluation of novel tools that could assist bTB eradication campaigns in three critical aspects; 1) rapid confirmation of bTB in bovine tissues; 2) high-resolution differentiation of bTB aetiological agents; and 3) ancillary identification of NTM species in clinical samples. The aims of this doctoral thesis were achieved through five different studies focusing on the development of molecular genetic tools for the confirmation of bTB in bovine tissues (studies 1 and 2), the implementation of WGS methods in molecular epidemiology studies of bTB (studies 3 and 4) and the implementation of proteomic methods for the identification of NTM (study 5).

5.1. Real-time PCRs as a confirmation method of bTB

Under the current legislative framework in the EU (i.e. Regulation 2016/429 and associated regulations), TB in bovines is considered a disease of categories B, D and E. This implies that within the EU, bTB is a disease that is subject to eradication programmes (cat. B), its spread needs to be prevented on account of its entry into the EU or animal movements between Member States (cat. D), and for which surveillance programmes need to be established (cat. E). The official diagnostic methods that may be used for case detection and bTB herd status establishment are the skin test (both single and comparative) and the IGRA. Reactor animals are considered suspected bTB cases; they are culled and their carcasses are visually inspected for macroscopic lesions compatible with TB. Suspected cases can also be detected during passive slaughterhouse surveillance or because of an epidemiological link with a confirmed case. Suspected cases can be subjected to further analyses through the sampling of specific tissues or organs, such as LNs, for their posterior processing and laboratorial analysis. An animal or group of animals may then be confirmed as a case of bTB if the aetiological agent has been isolated from the sample (microbiological culture) or a specific antigen or nucleic acid has been identified in an animal with clinical signs or epidemiological relationship with a confirmed case. Furthermore, any suspected case obtained from a positive result to an official *ante-mortem* test that has been epidemiologically linked with a confirmed bTB case may also be declared as confirmed.

In general, the official bTB-free status is granted to herds in which bTB has not been detected in the last 12 months, among other requirements. In contrast, if there is a suspected case in a herd with an official bTB-free status, the OTF status will be suspended until the results of the confirmatory method are obtained. If the presence of the aetiological agent is confirmed, the OTF status will be withdrawn, whereas if the results of the confirmatory method are negative, the OTF status will only be suspended. Herds with a suspended OTF status are not granted free movement of animals and require to pass another herd test in a maximum period of 12 months, whereas a herd with a withdrawn status will require to pass two consecutive herd tests in a maximum period of 12 months. Since the AHL is applicable throughout the EU, the same rulings are considered in the Spanish bTB eradication programme, although testing requirements may be more stringent under certain circumstances (e.g. withdrawn status).

Since confirmatory methods can have such a significant impact, it is of utmost importance that the methods used are as sensitive and specific as possible. If a method is not sensitive enough, bTB cases may not be confirmed, leading to possible infections in the herd to remain unnoticed. In contrast, if a test is not specific enough, false-positive results can lead to the withdrawal of the bTB-free status, with a considerable economic impact for the livestock owners. In both cases, the low performance of methods used for confirmation will ultimately hamper the eradication efforts against bTB and may increase the mistrust of stakeholders towards the efficiency of eradication campaigns.

The main findings and conclusions drawn from the qPCR studies presented herein will be described below, whereas general guidelines for the implementation of qPCRs in eradication programmes will be detailed at the end of this section (section 5.1.8) and in section 6.

5.1.1. Real-time PCR has a similar diagnostic performance when compared to microbiological culture

Microbiological culture is the main confirmation technique for bTB in many parts of the world, including the EU. In spite of its utility for the isolation of MTBC members, this technique presents a series of well-characterised limitations that can affect its performance and, therefore, can negatively affect the control and eradication of the disease. Real-time PCR has increasingly been used for the diagnosis of many infectious diseases and could be a powerful tool for the confirmation of bTB cases. As previously mentioned, the detection of nucleic acids as a confirmation method in RD 2020/689 paves the way for the incorporation of qPCRs in bTB diagnostic workflows. This has been the case of the Spanish bTB eradication campaign, in which histopathology and/or direct qPCR, including the IS*6110* qPCR presented in this thesis, has been implemented for the confirmation of bTB in suspected cases from OTF herds with an epidemiological link or samples with VLs obtained from slaughterhouse surveillance (MAPA, 2021).

In order for qPCR to be used as a confirmation method, its diagnostic performance should, at least, be comparable to culture. The two qPCRs described herein present an overall high diagnostic performance, with diagnostic Se and Sp values of 88.41% and 92.37% for the *mpb70* qPCR and 96.45% and 93.66% for the IS*6110* qPCR relative to culture (summarised in table 9). These values agree with previously described results from literature, in which diagnostic Se and Sp ranges between 66.7-100% and 88.7-100%, respectively (Parra *et al.*, 2008a; Thacker *et al.*, 2011; Costa *et al.*, 2013; Araújo *et al.*, 2014a;

Araújo *et al.*, 2014b; Sanchez-Carvajal *et al.*, 2021). In addition, the agreement between the two methods and microbiological culture was also very high (0.80 and 0.88, respectively). These results indicate that qPCR is a valid alternative method to culture, since it offers a similar performance but with the added value of being significantly faster.

5.1.2. Confirmation of MTBC DNA in culture-negative samples indicates a higher sensitivity of qPCR over culture

Microbiological culture is considered an imperfect reference method, which implies that any comparative study will be negatively biased towards the new diagnostic method if this is more sensitive. Certainly, the presence of MTBC DNA was confirmed in 85.36% (n = 35/41) of culture-negative/qPCR-positive samples in the IS*6110* study using a different method. When compared to the proportion of culture-positive/qPCR-negative samples (3.68%; n = 12), these results suggest a higher Se of the IS*6110* qPCR over culture. In contrast, the presence of MTBC DNA was confirmed in 50% (n = 5/10) of culture-negative/qPCR-positive samples in the *mpb70*, whereas the proportion of culture-positive/qPCR-negative samples was (11.59%; n = 8/69), which could indicate a comparable performance of both methods. These differences could probably be caused by a decreased performance of the resolver techniques used in the *mpb70* study to confirm the presence of MTBC DNA in culture-positive samples (DVR-spoligotyping and Sanger sequencing) in comparison to the IS*6110* study (*mpb70* qPCR). Sanger sequencing and DVR-spoligotyping are commonly used on pure DNA obtained from *M. bovis* isolates, and limitations in the performance of direct spoligotyping on tissue extracts has been described previously, with a Se of 50% (Taylor *et al.*, 2007b). Indeed, 90% (n = 9/10) of the culture-negative/qPCR-positive samples from the *mpb70* study were also positive when re-analysed using the IS*6110* qPCR (data not shown).

Overall, the confirmation of the presence of MTBC DNA in culture-negative/qPCR-positive samples indicates a higher sensitivity of qPCR over culture. If the results obtained from the analysis of discrepant results are combined together with the results obtained from the reference method in order to establish the true status of infection, diagnostic performance improves significantly for the two qPCRs. In the case of the *mpb70* study, diagnostic Se increased slightly to 89.19% and Sp increased to 96.03%, whereas for the IS*6110* qPCR, diagnostic Se and Sp increased to 96.78% and 99.02%, respectively.

Study	Sample	Stat	Result	Result (correction, CN only)
mpb70	All	Se	88.41% [78.43 – 94.86%]	89.19% [79.80 – 95.22%]
		Sp	92.37% [CI: 86.41 – 96.28%]	96.03% [90.98 – 98.70%]
		Ag	0.80 [0.77 – 0.89]	0.86 [0.79 – 0.93]
		LR+	11.58 [6.34 – 21.14]	22.48 [9.49 – 53.25]
		LR-	0.13 [0.07 – 0.24]	0.11 [0.06 – 0.22]
	VL	Se	95.08% [86.29 – 98.97%]	95.31% [86.91 – 99.02%]
		Sp	58.82% [32.92 – 81.56%]	71.43% [41.90 – 91.61%]
		Ag	0.56 [0.37 – 0.81]	0.69 [0.47 – 0.91]
		LR+	2.31 [1.30 – 4.09]	3.34 [1.45 – 7.65]
		LR-	0.08 [0.03 – 0.27]	0.06 [0.02 – 0.20]
	NVL	Se	37.50% [8.52 – 75.51%]	50.00% [18.71 – 81.29%]
		Sp	97.37% [92.50 – 99.45%]	99.11% [95.13 – 99.98%]
		Ag	0.40 [0.22 – 0.57]	0.60 [0.43 – 0.77]
		LR+	14.25 [3.41 – 59.59]	56.00 [7.23 – 433.83]
		LR-	0.64 [0.37 – 1.10]	0.50 [0.27 – 0.94]

IS*6110*	All	Se	96.45% [93.90 – 98.15%]	96.78% [94.46 – 98.15%]
		Sp	93.66% [91.51 – 95.29%]	99.02% [97.88 – 99.55%]
		Ag	0.88 [0.85 – 0.91]	0.96 [0.94 – 0.98]
		LR+	15.22 [11.31 – 20.48]	98.72 [44.51 – 218.92]
		LR-	0.04 [0.02 – 0.07]	0.03 [0.02 – 0.06]
	VL	Se	97.74% [95.40 – 99.09%]	97.92% [95.77 – 99.16%]
		Sp	71.43% [61.42 – 80.10%]	98.59% [92.40 – 99.96%]
		Ag	0.75 [0.67 – 0.82]	0.93 [0.89 – 0.98]
		LR+	3.42 [2.50 – 4.68]	69.53 [9.93 – 486.82]
		LR-	0.03 [0.02 – 0.07]	0.02 [0.01 – 0.04]
	NVL	Se	82.14% [63.11 – 93.94%]	86.11% [70.50 – 95.33%]
		Sp	97.63% [95.88 – 98.73%]	99.08% [97.86 – 99.70%]
		Ag	0.70 [0.58 – 0.83]	0.85 [0.76 – 0.94]
		LR+	34.69 [19.73 – 60.99]	93.17 [38.56 – 225.14]
		LR-	0.18 [0.08 – 0.40]	0.14 [0.06 – 0.32]

Table 9. Diagnostic performance estimates obtained for the *mpb70* and IS*6110* qPCRs when compared to culture, as well as a subsample of 200 samples from the IS*6110* (SUB) that were analysed with the *mpb70* qPCR. Values between brackets represent 95% Confidence Intervals. Correction of estimates using discrepant resolution were achieved through DVR-spoligotyping/Sanger sequencing (*mpb70* study) or the *mpb70* qPCR (IS*6110* study). Stat = statistic; Se = Sensitivity; Sp = Specificity; Ag = Agreement; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; CN = Culture negative samples.

Study	Sample	Stat	Result	Result (correction, CN only)
SUB	IS*6110*	Se	90.54% [81.48 – 96.11%]	N.D.
		Sp	84.13% [76.56 – 90.03%]	N.D.
		Ag	0.72 [0.58 – 0.86]	N.D.
		LR+	5.70 [3.79 – 8.58]	N.D.
		LR-	0.11 [0.06 – 0.23]	N.D.
	mpb70	Se	90.54% [81.48 – 96.11%]	N.D.
		Sp	86.51% [79.28 – 91.94%]	N.D.
		Ag	0.75 [0.61 – 0.88]	N.D.
		LR+	6.71 [4.29 – 10.51]	N.D.
		LR-	0.11 [0.05 – 0.22]	N.D.

Table 9 (continued). Diagnostic performance estimates obtained from the *mpb70* and IS*6110* studies, as well as a subsample of 200 samples from the IS*6110* (SUB) that were analysed with the *mpb70* qPCR. Values between brackets represent 95% Confidence Intervals. Correction of estimates using discrepant resolution were achieved through DVR-spoligotyping/Sanger sequencing (*mpb70* study) or the *mpb70* qPCR (IS*6110* study). CN = Culture negative samples.

Although discrepant resolution has become a popular method to calculate the accuracy of diagnostic methods in the presence of an imperfect reference standard, such as microbiological culture, this methodology is known to introduce bias towards the test being evaluated as a result of only evaluating discrepant false-positive results (Hadgu, 1996). Although the identification of two targets present in MTBC genomes in a DNA sample can still be considered a strong indicator of the presence of these pathogens, the use of a composite reference standard, which combines several methods in order to obtain a final diagnosis, could be an interesting alternative to discrepant analysis (Naaktgeboren *et al.*, 2013). This was assessed on a panel of 200 randomly-selected samples from the IS*6110* study (VL: 96; NVL: 104) in which both the

mpb70 qPCR and microbiological culture were combined in order to establish the true status of infection (data not shown). Under this context, the IS*6110* qPCR presented a very high diagnostic sensitivity and specificity of 91.21% and 96.33%, respectively. Alternatively, a diagnostic test without a perfect reference standard could be achieved using Bayesian modelling, such as Latent Class analysis (Alvarez *et al.*, 2012; Courcoul *et al.*, 2014). Using this methodology, Courcoul *et al.* 2014 described a higher diagnostic Se for the qPCR in comparison to microbiological culture (87.7% *vs.* 78.1%) and a similar diagnostic Sp (97.0% *vs.* 99.1%) (Courcoul *et al.*, 2014).

The higher Se of qPCR over culture is relevant when considering that the amount of tissue homogenate used in the qPCR protocols described herein is only 1/10 of the volume used in microbiological culture. Furthermore, the majority *of* the confirmed culture-negative/qPCR-positive samples were obtained from samples with VL: 77.14% (n = 27) for the IS*6110* qPCR when confirmed with the *mpb70* qPCR; 60% (n = 3) and 66.67% (n = 6) for the *mpb70* qPCR study when confirmed with DVR-spoligotyping/sequencing or the IS*6110* qPCR, respectively. This is interesting, considering that bacterial loads are usually higher in lesions compatible with TB. Although acid-fast bacillary counts have been found to be higher in advanced stages of granulomas in experimentally infected cattle (Wangoo *et al.*, 2005; Palmer *et al.*, 2007), this correlation remains controversial in naturally-infected animals (Liebana *et al.*, 2008; Menin *et al.*, 2013). In addition, acid-fast bacillary counts may not reflect cellular viability, since a negative correlation between cellular viability and lesion progression has also been described (Menin *et al.*, 2013). Furthermore, considering that macroscopic lesions may contain less developed granulomas with fewer AFB counts, some samples with VLs may still be paucibacillary (Palmer *et al.*, 2007).

Even if the bacterial load was higher in tissues with VLs, their decreased viability, aggravated by the decontamination process, could have led to a reduced isolation. This effect is not observed in qPCR, since it does not depend on cellular viability. This was also reflected in the capacity of qPCR to detect MTBC DNA in those samples in which the presence of other microorganisms had masked or prevented the growth of tuberculous bacilli. This was especially relevant in the IS*6110* study, in which 21.95% (n = 9) of culture-negative/qPCR-positive samples showed growth of other microorganisms.

5.1.3. Diagnostic performance varies according to the presence or absence of lesions

The diagnostic performance of both qPCRs differed according to the gross pathology of the analysed tissues (i.e. samples with VL or NVL). According to the results of both studies, qPCR presented a good performance when analysing samples with VLs, since the diagnostic Se was very high in this group of samples (95.08% and 97.74%, for the *mpb70* and IS*6110* studies, respectively) (table 9). Nevertheless, diagnostic Sp was lower in the *mpb70* (58.82%) and IS*6110* (71.43%) qPCR studies, which could partly be a result of the reduced sensitivity of microbiological culture, as previously indicated. In contrast, when samples with NVLs were analysed, diagnostic Sp relative to culture was approximately 97% in both studies, whereas diagnostic Se was lower for the *mpb70* qPCR (37%) than for the IS*6110* qPCR (82%) (table 9). These findings are in agreement with previous literature; Sánchez-Carvajal *et al.* 2021 showed a Se and Sp of 100% and 50%, respectively, for a qPCR targeting the IS*6110* on bovine tissues with VL, whereas Se and Sp values were 67.92% and 99.84%, respectively, when samples with NVL were analysed (Sanchez-Carvajal *et al.*, 2021). A decreased Sp in samples with VL could, in this case, be due to the low number of VL samples that were analysed (n = 26), coupled with differences in the DNA extraction and culture protocols. In addition, Parra *et al.* 2008 showed a high proportion (79-81%) of positive samples in animals with lesions compatible with TB, in contrast to 61% of qPCR-positives in samples with NVLs (Parra *et al.*, 2008a). Similarly, a higher proportion (41.18-61.25%) of qPCR positive samples was observed in samples with VL when compared to samples with NVL (0-36.17%) in a study evaluating the use of the TbD1 region to detect MTBC DNA in bovine tissues (Araújo *et al.*, 2014a).

One possible explanation for this decrease in performance in samples with NVLs could be the low bacterial loads in these samples, which can be accentuated by the reduced amount of tissue homogenate that is used for DNA extraction in comparison to what is used for culture (1 ml *vs.* 10 ml) and the tendency of tuberculous bacilli to form cellular aggregates, which in turn could limit their homogenous distribution across aliquots. Indeed, the higher Ct values observed in NVLs (35.73) respect to VLs (31.45) in the IS*6110* study were significant ($p < 0.01$, Mann-Whitney two-tailed test), which could reflect a lower bacterial load in these samples. Similar findings were described by Parra *et al*. 2008, in which Ct values were generally higher for samples with NVLs than in samples with VLs, especially when the latter were obtained from animals with a generalised

infection (Parra *et al.*, 2008a). Comparison of Ct values was not performed for the *mpb70* study due to the small proportion of qPCR-positive samples in NVLs (n = 6).

An increase in the amount of sample processed in the DNA extraction could improve the detection of these samples through qPCR (Leth *et al.*, 2017). Nevertheless, this would need to be carefully weighed against the risks of introducing a larger amount of host DNA or other inhibitors (see section below). Sequence capture platforms could also be an interesting alternative to concentrate the amount of MTBC DNA in low-concentrated samples (Taylor *et al.*, 2001; Taylor *et al.*, 2007b; Parra *et al.*, 2008b; Fell *et al.*, 2016). Nested qPCRs could also be an interesting alternative to detect MTBC DNA in paucibacillary samples, although the risk of cross-contamination due to the handling of highly concentrated amplicons should be considered. Nevertheless, single-tube nested qPCRs have been recently described in human TB and could be evaluated in the future for this type of samples (Leung *et al.*, 2018).

The results presented herein indicate that qPCR could be especially recommended as an alternative for microbiological culture in samples with VLs. Although diagnostic Se was moderate (82%) in samples with NVLs, its speed and versatility still make it a promising alternative to microbiological culture for the confirmation of bTB in bovine tissues.

5.1.4. Inhibitors are an important factor to consider during qPCR detection of MTBC members in bovine tissues

As previously indicated, a small proportion of culture-positive samples tested negative to both qPCRs evaluated in this thesis; 11.59% (n = 8) for the *mb70* and 3.55% (n = 12) for the IS*6110* qPCRs, respectively. This suggested that these techniques may fail to detect MTBC DNA in certain cases as a result of several factors, such as the presence of PCR inhibitors, a failure of the DNA extraction protocol, a low bacterial load in the tissue or a combination of the above.

The presence of PCR inhibitors can lead to a false negative result (complete inhibition) or to a decrease in reaction efficiency with an increase in Ct values (partial inhibition). This highlights the importance of including ICs in the qPCR reaction in order to monitor inhibition phenomena and assess the success of both the DNA extraction and qPCR protocols, as has been previously indicated (Costa *et al.*, 2013; Radomski *et al.*, 2013). The exogenous heterologous IC included in the two qPCRs was able to identify inhibition (partial or complete) in 9.5% (n = 19) and 3.15% (n = 31) of the samples analysed in the *mpb70* and

IS*6110* studies, respectively. Inhibition was present in half (n = 4) and two thirds (n = 8) of the culture-positive/qPCR-negative samples, suggesting that it is an important factor for the failure of qPCR and should be taken into consideration.

Since the DNA extraction protocol is carried out in tissue samples, the concentration of host DNA in the sample is higher than that of the pathogen. High concentrations of host DNA (> 160 ng/µl) have been shown to inhibit qPCR detection of MAP and *M. tuberculosis* (Radomski *et al.*, 2013). Therefore, the amount of overall DNA concentration was assessed in the samples using spectrophotometry. Interestingly, the median concentration of DNA was significantly higher in the inhibited samples than in the non-inhibited samples in both studies (Mann-Whitney two-tailed test; $p < 0.01$ and $p < 0.02$); in the *mpb70* and IS*6110* studies, inhibited samples had, respectively, a median concentration of 396.7 ng/µl (IQR: 259.1-497.2 ng/ul) and 353.65 ng/µl (IQR: 269.05-453.73 ng/µl), whereas the non-inhibited samples had a median concentration of 207.4 ng/µl (IQR: 134.25-314.8 ng/µl) and 292.20 ng/µl (IQR: 201.70-391.80 ng/µl). This indicates that the overall concentration of host DNA is a factor to consider in order to improve the performance of qPCR in tissue extracts.

One way of circumventing inhibitory effects is the dilution of the samples, since this may in turn dilute the effects of inhibitor molecules. Accordingly, PCR-negative samples with partial or complete inhibition of the IC were diluted five-fold and ten-fold in the *mpb70* and IS*6110* studies, respectively. Among samples with inhibition, amplification was observed after dilution in two of the samples from the IS*6110* study (one complete and one partial), indicating that this strategy may only be able to recover a small proportion of positive samples, possibly as a result of decreasing the concentration of MTBC DNA below the LOD. In order to avoid this dilution effect, the DNA extraction could be repeated from remaining tissue homogenate. Repetition of the DNA extraction was achieved in all of the discordant samples in the *mpb70* study and eight out of 12 in the IS*6110* study. Of these, half (n = 4, two previously inhibited) and one quarter (n = 2, one previously inhibited) were positive to the qPCRs.

Therefore, it is recommended that DNA concentration is evaluated after DNA extraction or if inhibition is detected (Sanchez-Carvajal *et al.*, 2021); if the qPCR is inhibited, the DNA samples should be diluted and/or the DNA extraction should be repeated.

5.1.5. The IS*6110* and *mpb70* qPCRs may present a similar performance under certain circumstances

The results of the two qPCR studies indicated that the diagnostic Se of the IS*6110* qPCR was higher than the *mpb70* qPCR, which could relate to the differences in the number of copies of both elements (multiple-copy *vs.* single-copy, respectively). However, due to differences in the number of samples analysed, as well as possible differences in the bacterial loads in the selected samples, a direct comparison between both studies is not possible. Interestingly, the diagnostic performance of both qPCRs was almost identical in the 200 randomly-selected samples of the IS*6110* study (table 9); diagnostic Se was 90.45% for both qPCRs, whereas diagnostic Sp was 84.13% for the IS*6110* qPCR and 86.51% for the *mpb70* qPCR. Despite their similar performance, the pairwise comparison of the Ct values from this subsample revealed that the median Ct values were significantly higher ($p < 0.01$; Wilcoxon Signed-Rank Test) for the *mpb70* qPCR than for the IS*6110* qPCR (34.00 *vs.* 32.86, respectively). Although this Ct difference may seem apparently small, it can still be relevant as the concentration of amplicons duplicates with each cycle. In addition, a linear regression model revealed a moderate coefficient of linearity between the Ct values of both qPCRs ($R^2 = 0.76$) (Appendix 1). Altogether, these results suggest that the Ct values in the *mpb70* qPCR are usually higher than for the IS*6110* qPCR, and that certain paired samples have a significant difference in Ct values. This could be explained by the fact that the majority of *M. bovis* strains usually harbour a single copy of the IS*6110*, whereas others may present more than one copy. It would be worthy to evaluate the performance of both methods in parallel using a larger sample population.

5.1.6. IS*6110*-like elements and absence of the IS*6110* element may reduce the performance of qPCR under certain scenarios

Although the presence of MTBC DNA was confirmed in the majority of culture-negative/qPCR-positive samples, amplification due to cross-reactivity between the primer sets and the genomes of other microorganisms remains a possibility. The *mpb70* gene is considered specific of the MTBC, but cross-reactivity for *mpb70*-specific primers against *M. kansasii* has been previously described in literature, as well as the identification of the homologous *mpk70* gene (Woolford *et al.*, 1997; Veyrier *et al.*, 2008). In turn, the IS*6110* element is also considered specific of the MTBC, although cross-reactivity of oligonucleotides targeting this element with different NTM, such as *M. wolinksy* or *M. avium-M. intracellulare,* have been described in literature (McHugh *et al.*, 1997; Thacker *et al.*, 2011). In

addition, an IS*6110*-like sequence has also been identified in *M. smegmatis* (Coros *et al.*, 2008).

The *in silico* specificity analysis carried out in the *mpb70* qPCR study revealed an overall similarity of 78% between the *mpb70* and *mpk70* coding sequences and an 84% identity between the *mpb70* qPCR amplicon and the *mpk70* gene (data not shown). Nevertheless, complementarity in the oligonucleotide regions was low, especially in the reverse primer, in which a four bp-long mismatch indicated a low probability of annealing to the *mpk70* gene. Furthermore, no cross-reactions were observed for the nine *M. kansasii* isolates that were analysed during the analytical specificity validation, suggesting that this qPCR was highly specific. Although no cross-reactivity was detected during the initial *in silico* specificity analyses of the IS*6110* primer set, one '*M. avium* subsp. *hominissuis*' and one *M. marinum* isolate, from a total of 102 DNA samples obtained from 24 different NTM species, produced a positive amplification signal. Due to the higher possibility of isolating MAH in cattle (Dvorska *et al.*, 2004; Biet & Boschiroli, 2014; Ghielmetti *et al.*, 2018; Michelet *et al.*, 2018b), the origin of this cross-reactivity was evaluated in this isolate. Whole genome sequencing and *de novo* assembly revealed an IS*6110*-like element with 83% identity to the IS*6110* element in *M. tuberculosis* H37Rv. An extended BLAST search on the WGS archive of the NCBI revealed identical IS*6110*-like sequences in several MAC species, including *M. avium, M. timonense* and *M. bouchedurhonense*. As a result, the six remaining culture-negative/IS*6110*-positive samples that were negative to the *mpb70* qPCR were reanalysed using a MAC-specific PCR based on the IS*1245* element. None of these discordant samples was positive to this PCR, suggesting that MAH was not present in the sample and was not responsible for the amplification signal. Additionally, no amplification was detected in the 100 tissue samples obtained from 78 bTB-free herds in a low prevalence region from northwestern Spain, which could indicate that cross-reactivity may not interfere significantly with the detection of MTBC. In order to evaluate this finding further, a random selection of MAH isolates (n = 124; obtained from a total of 261 MAH isolates between the years 2017 – 2019) in the same region were analysed using the IS*6110* qPCR. Ninety-two of these samples (75%) were positive to this qPCR (data not shown). These MAH isolates were obtained from a total of 12577 bovine samples, of which only 0.32% (n = 40) were positive to bTB through culture. Therefore, in the worst case scenario, the use of the IS*6110* qPCR would have detected a total of 196 false-positive samples, corresponding to 1.56% of the total number of samples. Nevertheless, this scenario does not take into account the difference in the DNA

concentration of MAH between pure cultures and tissue samples. Therefore, it is not possible to know how the IS*6110* qPCR would have performed on DNA samples obtained from the same tissues. In any case, these findings indicate that the IS*6110* qPCR should be used strategically under certain epidemiological contexts, such as in low prevalence regions or regions with a high prevalence of MAH. In these cases, additional genetic targets (see below) and/or methods should be employed.

Although rare, an additional cause for a PCR-negative result in culture-positive samples is the absence of the genetic target in MTBC strains. Certainly, the absence of the IS*6110* element has been described sporadically in literature both in single strains of *M. tuberculosis* and *M. bovis* (Howard *et al.*, 1998; Steensels *et al.*, 2013), which could indicate that some culture-negative/qPCR-positive results in the IS*6110* study could be due to this factor. The absence of this mobile element was confirmed in only two (non-inhibited) culture-positive/qPCR-negative samples in the IS*6110* study; no amplification was obtained from the cultured isolates, whereas the *mpb70* qPCR resulted positive. Due to the fact that the IS*6110* is usually present as a single copy, the deletion of this element could have led to this absence of amplification. Interestingly, a *M. bovis* isolate described by Steensels *et al*., had the SB2402 genotype, which lacks spoligotype membrane spacers 12-32. Taking into account that the IS*6110* element is usually present between genomic spacers 34 and 35 in the DR region (spoligotype membrane spacers 24-25), one possibility could have been the excision of the IS*6110* transposon during DNA replication or as a result of the recombination of the DR locus. In agreement, two of the false-negative isolates in the IS*6110* qPCR study presented the SB1263 genotype, which lacks genomic spacers 30-35 (membrane spacers 20-25). In order to assess the significance of this finding, spoligotypes with the spacer 24-25 deletion were searched for in Mbovis.org, revealing that only 8% (n = 198) of the registered spoligotypes had this deletion. Twenty-one *M. bovis* isolates with this deletion that were available in VISAVET were tested using the *mpb70* and IS*6110* qPCR protocols. Only five out of the 21 isolates tested negative for the IS*6110* qPCR, but not to the *mpb70* qPCR. In contrast, amplification was detected in the 16 other strains, indicating that this deletion does not necessarily lead to an absence of amplification. This could be explained by the presence of additional IS*6110* elements in other parts of the genome of these strains or differences in the recombination of the DR region. Nevertheless, in order to assess if these differences in amplification are a result of a true deletion of the

IS*6110*, the absence of the IS*6110* element should be proven *in vitro* through Southern blotting or WGS.

5.1.7. Interlaboratory validation confirms the high diagnostic sensitivity of the IS*6110* qPCR

The diagnostic performance values reflected in this thesis, as well as in published literature, reflect the significant effect that differences in DNA extraction protocols, reaction chemistry and parameters, or in sample selection, can have in the performance of qPCRs. This effect was demonstrated in the study by Sanchez-Carvajal *et al.* 2021, in which the same IS*6110* primers and qPCR kits as the ones used in this thesis, but a different DNA extraction protocol and sample selection, led to a significantly different performance, with diagnostic Se and Sp values of 77.1% and 99.4%, respectively (Sanchez-Carvajal *et al.*, 2021).

In order to evaluate the robustness of the IS*6110* qPCR under different conditions, its performance was further evaluated in an interlaboratory study with the participation of three different National RL in the EU and one Regional RL in Spain. The results of this study confirmed the high diagnostic Se of the IS*6110* qPCR, since three out of four participating RLs observed a diagnostic Se higher than 97%. The remaining laboratory (laboratory B) achieved a lower diagnostic Se of 66.67%, which could be related to variations in the DNA extraction protocol that were considered necessary by that laboratory for biosafety reasons, such as the inclusion of an inactivation step of 95 ºC for 15 minutes. The effect of heat-inactivation in the performance of the DNA extraction was not evaluated during the *in house* validation of the protocol and, therefore, a negative effect cannot be ruled out. A pilot study in VISAVET with 43 tissue samples showed that the inactivation of tissue homogenates at 100 ºC for 15 minutes reduced the number of positive qPCR samples by 19% (data not shown). Although laboratory C also incorporated a heat-inactivation step, the temperature used was lower and only samples with VL were analysed, which could have reduced the negative effect of heat-inactivation.

The diagnostic Sp values varied significantly between laboratories, ranging between 64.29-99.05%. This discrepancy could be due to differences in sample processing and decontamination procedures, which could have affected cellular viability. This could have been especially the case in laboratory A, in which a significant proportion of culture-negative/qPCR positive samples were identified.

Despite these differences, the results from the interlaboratory validation reinforce the potential of the IS*6110* qPCR in the detection of MTBC members in bovine tissues. Nevertheless, these results also highlight the importance of adequately verifying diagnostic protocols before their implementation into laboratory workflows in order to guarantee the best performance.

5.1.8. Towards the implementation of qPCR for the detection of MTBC members in bovine tissues

The results of the two qPCR studies presented herein indicate a higher diagnostic performance of these methods over microbiological culture, but also provide valuable information regarding the opportunities and risks of implementing qPCR for the detection of MTBC members in bovine tissues.

The high diagnostic Se of the IS*6110* qPCR, especially in samples with VLs, indicates that this technique could be a very useful screening technique for the detection of MTBC DNA in this kind of samples. This would allow for a more accurate case confirmation and an increase in the trust of the involved parties in the diagnostic methods used. The fact that qPCR can deliver results in a matter of days in comparison to the weeks needed for microbiological culture can reduce the time by which a breakdown event is confirmed or by which the bTB-free status of a herd is recovered or withdrawn.

The lower diagnostic Se of the IS*6110* qPCR in samples with NVLs reflects the challenges of selecting an appropriate sample and processing a small amount of tissue homogenate from macroscopically healthy tissues. However, since only a few culture-positive samples from NVLs were analysed in both studies, more research is required in order to fully assess the performance of qPCR in this type of samples. Nevertheless, considering that NVL samples are expected to increase as eradication programmes progress, the IS*6110* qPCR is still a promising alternative to culture, since its higher speed and throughput capacity will allow the sampling of a larger amount of reactor animals per herd, increasing the chance of detecting the pathogen, if present. Taking into consideration the financial costs and the epidemiological scenario, for example in bTB free areas or in single outbreak investigations, additional methods could be implemented to increase the chance of detecting MTBC infections in NVLs, such as culture or histopathology (Larenas-Muñoz *et al.*, 2022). For example, the Spanish government recommends the use of qPCR and/or histopathology in OTF herds under certain circumstances in order to improve case confirmation, such as in follow-up animals (see section 1.6.3), animals from OTF herds with previous

history of bTB, or samples obtained from passive slaughterhouse surveillance (MAPA, 2021).

Since the molecular characterisation of bacterial isolates is crucial for the surveillance and control of bTB, obtaining pure colonies of these isolates is, at least for now, necessary. The major advantage in the use of qPCR for the detection of bTB would be the very fast availability of results. Positive samples can be posteriorly cultured for genotyping studies. This would be a cost-effective approach to enhance epidemiological investigations of bTB breakdown events.

If qPCR protocols are implemented in bTB campaigns, it is crucial to include ICs in order to monitor the presence of inhibition phenomena. If inhibition is detected, the dilution of inhibited DNA samples alone may not be the recommended choice due to the limited amount of tissue homogenate that is used for extraction and the possibility of having a low bacterial loads in the sample. Additionally, if samples are still qPCR-negative after the dilution, the DNA extraction protocol should be repeated. In this aspect, the storage of frozen backup aliquots of tissue homogenate is very important.

The cross-reactivity of the IS*6110* primers with certain MAH strains and the existence of *M. bovis* strains lacking the IS*6110* element, although rare, can challenge the performance of the IS*6110* qPCR under certain circumstances. Nevertheless, the increased speed, Se and Sp of this qPCR over culture outweighs the risks. In addition, the choice of qPCR target may be decided according to the epidemiological scenario or the type of samples. In a high bTB-prevalence setting, the use of the IS*6110* qPCR may be a valid option for the rapid confirmation of the infection. In contrast, the use of both qPCRs could be better suited in low prevalence settings or for assessing the status of infection in OTF herds. Alternatively, if the prevalence of MAH or other NTM infections is high in the region under study, the use of other genetic targets, such as the IS*1311* for *M. avium* or the ITS for the *Mycobacterium* genus, could be used to increase the specificity of the diagnosis (Sevilla *et al.*, 2015). Although this can be achieved sequentially, the flexibility of qPCR design allows for the combination of genetic targets in a multiplex format (Sevilla *et al.*, 2015), which would decrease the cost of qPCR testing. In addition to an increased specificity, the use of a multiplex qPCR with two MTBC-specific targets could increase the detection in samples with low bacterial load. Finally, the choice of target could also depend on the type of study being conducted. For example, the *mpb70* is better suited in

quantification studies in which the bacterial load in the sample needs to be assessed, due to its single-copy nature and stability (Roy *et al.*, 2019).

Due to the increased interest in the use of vaccines in cattle in certain parts of the world, it is important to evaluate the possibility of cross-reactivity of real-time PCR with live vaccine strains, such as *M. bovis* BCG. Despite the limited replication capacity of *M. bovis* BCG, there is a possibility that this strain could be isolated from vaccinated animals if the bacteria are localised in the lymph nodes or tissues, in which case false-positives may occur with the *mpb70* and IS*6110* qPCRs. Vaccine strains have been isolated from tissue samples from vaccinated badgers, deer and goats at different time points (Palmer *et al.*, 2014; Perez de Val *et al.*, 2016; Lesellier *et al.*, 2019), but there is currently no information regarding the shedding and replication dynamics of BCG in cattle. In the case of cross-reactivity issues, a DIVA strategy should be pursued in those countries in which vaccination with live-attenuated strains is implemented, such as the deployment of PCRs targeting the RD1 deletion in *M. bovis* BCG.

The qPCRs developed in this thesis are very promising techniques that may be implemented in current diagnostic frameworks for bTB. The higher diagnostic Se of the IS*6110* qPCR suggests its use as a rapid screening and confirmation tool, whereas the higher specificity and single-copy nature of the *mpb70* qPCR, as well as its broad dynamic range and high LOD (as little as 10 genomic equivalents), indicate its suitability as a confirmation and quantification method. However, it is important to indicate that the LOD for both qPCRs was calculated using pure DNA and may not reflect matrix-specific effects in this parameter, such as the presence of host DNA or the presence of components from the matrix or the extraction protocol at low concentrations of pathogen DNA. More studies are warranted to assess these effects on spiked negative tissue and DNA samples.

Overall, the increased Se of qPCR will definitely aid eradication programmes by increasing the speed by which results are obtained and by reducing the number of samples that need to be processed through culture, therefore increasing the cost-effectiveness of current eradication campaigns.

5.2. Whole Genome Sequencing as a high-resolution method for molecular characterisation of *M. bovis*

The key element of eradication of bTB is the interruption of the spread of the infection through the detection and elimination of infected animals. However, from the perspective of surveillance and control of bTB, it is equally important to be able to establish similarities between bacterial isolates in order to unveil the sources of infection and, therefore, implement contingency measures to interrupt its spread and recurrence.

In the last years, WGS has become an increasingly demanded method for the study of MTBC infections, including bTB, due to its multiple advantages over traditional molecular characterisation techniques. Firstly, WGS achieves a higher throughput than other methods, allowing the processing and comparison of large amounts of data. Additionally, WGS offers a higher discrimination capacity, achieving a resolution down to the nucleotide level. This is extremely important in order to unveil the genetic diversity of phylogenetically related isolates in such a clonal group of pathogens as the MTBC. In addition, the informatics nature of the generated data allows for an increased portability and connectivity through the internet or cloud services, which greatly enhances the sharing of information between laboratories. Furthermore, the versatility of WGS data allows for a great flexibility in the type of tools and analyses that can be carried out, allowing scientists to generate novel applications that can be easily shared. In an increasingly connected world, and considering the free movement of animals across the EU, this connectivity between state authorities tasked with the surveillance of animal diseases is of great importance.

One of the most promising applications of WGS in the epidemiology of bTB is the use of variant calling pipelines, in which SNP differences between MTBC members are established through a series of interconnected softwares. Phylogenetic relationships are then inferred based on the concatenation of these SNPs and the use of phylogenetic tree reconstruction algorithms, usually based on maximum likelihood or Bayesian models. Based on the amount of SNP differences, the dynamic of circulation of *M. bovis* isolates can be assessed by deducing whether a group of isolates have been involved in the same or other breakdowns in a given area or timeframe.

5.2.1. **Current bTB variant calling pipelines show a high performance**

Whole genome sequencing methods, and any associated downstream analyses including variant calling pipelines, need to be carefully evaluated before they are implemented into the workflow of microbiology laboratories. This is especially relevant considering that the results obtained by WGS processing applications are largely dependent on the software and versions used, and that the use of such technologies requires a certain level of informatics knowledge and expertise that may not be available in all microbiology laboratories. It is, therefore, crucial that the performance of any bTB tools for WGS analyses is assessed. Although few studies on the use of such tools have been published for human TB (Kohl *et al.*, 2018c; Jajou *et al.*, 2019; Walter *et al.*, 2020), there was no available information regarding the capacity of variant calling pipelines for the analysis of *M. bovis* WGS data. The second objective of this thesis was the evaluation of WGS as a molecular characterisation method for bTB epidemiological studies. In this context, the performance of publicly available bTB WGS pipelines was assessed and compared against MTBseq, one of the reference methods for WGS analysis in human TB.

Four pipelines were selected for this study: vSNP, SNiPgenie, BovTB and MTBseq. Some of these pipelines, such as vSNP, have been available for years and used in published studies (Price-Carter *et al.*, 2018; Perea *et al.*, 2021), whereas others have been developed only recently (e.g. the first version of SNiPgenie was released in 2020). When the performance of the four pipelines was compared against the simulated dataset using quality filters alone, all pipelines achieved very high recall (≥ 96.6%) and precision rates (≥ 99.7%). This was also translated in a high agreement between all pipelines and the simulation, with the majority of SNPs being detected by all tools, although minor discrepancies were observed between them. Differences in recall rates between vSNP/SNiPgenie and BovTB/MTBseq indicated the presence of erroneous calls (FN and FP) that required further investigation.

5.2.2. Differences between quality filters and masking parameters can lead to discrepancies between pipelines

Erroneous calls may be introduced during the sequencing process or can be a result from different tool combinations (Walter *et al.*, 2020). Due to the artificial nature of the reads used in this study, the package composition and filtering parameters of these pipelines were compared. Mapping of reads was achieved through BWA in all four pipelines, indicating that all analyses shared the same foundation for posterior variant calling. It was in this latter step that most differences were observed, since different variant callers were used and varying levels of quality filters were applied. The main quality filters used by all pipelines to establish a probable SNP in a specific position could be simplified into four statistics: 1) the base quality score recorded by the sequencer at that position, 2) the average mapping quality of the reads aligning to that specific position, 3) the minimum read coverage required for a position to be considered, and 4) the minimum proportion of alternate alleles for a SNP to be considered, either through allele frequency (SAM/BCFtools) or minimum alternate fraction (FreeBayes). Not all of the pipelines employed the four thresholds; for example, BovTB and MTBseq did not use mapping quality thresholds, contrary to vSNP and SNiPgenie, which used thresholds of 56 and 60, respectively. Considering the overall good sequence quality and coverage of the artificial reads, and taking into account that mapping quality decreases within problematic regions, the observed recall rates could most probably be attributed to the differences in mapping quality thresholds between vSNP/SNiPgenie and BovTB/MTBseq.

Discrepancies in variant calling can negatively affect the phylogenetic relationships inferred in downstream analyses. Due to the large amount of data, even small false-negative and false-positive rates can be relevant, especially considering the limited genetic diversity of MTBC members (Meehan *et al.*, 2019). Despite these discrepancies, the overall topology of the phylogenetic trees inferred from the unfiltered pipeline results were similar, as indicated by the high level of clustering between pipeline-specific trees and the simulated trees. Similarly, the pairwise comparison between the best ML trees output by RAxML revealed a high congruence between the simulated and pipeline inferred trees. However, minor variations in isolate clustering were observed, especially in vSNP and SNiPgenie. This was probably due to the effect that the higher proportion of unrecognised variants (FN variants) had in the resolution of polytomies, which

were very frequent as a result of the high genetic similarity between the studied *M. bovis* isolates.

Further analysis of discrepancies revealed that these mainly occurred in repetitive elements and other problematic regions, such as *pe/ppe* genes. As a result, in addition to quality filters, there are two other important filtering steps that variant calling pipelines frequently incorporate in their analyses: 1) proximal SNP removal and 2) annotation-based filtering. The removal of proximal SNPs is based on the removal of highly variable regions that could be indicative of recombination events and lead to homoplasies (Crispell *et al.*, 2019a; Saltykova *et al.*, 2019). Although this approach may not seem very useful in a group of microorganisms with such a limited recombination capacity as is the MTBC, highly polymorphic regions do exist in the genomes of these microorganisms, such as *pe/ppe* genes. These could give rise to homoplasies, and could therefore alter phylogenetic inference in a similar manner to recombination events (Nakanishi *et al.*, 2013). It is important to note that this filtering step is not included in vSNP and BovTB by default, and was achieved using an *in house* script. Annotation-based filtering consists in the masking of regions of the genome in which erroneous base calling and misalignments are considered to be more frequent, such as repetitive elements or low complexity regions (Modlin *et al.*, 2021). These positions of the genome are frequently filtered in scientific literature, although their definition can be rather ambiguous, hampering the comparison between studies. In most cases these positions are defined as "repetitive elements", and may include the DR locus, mobile elements, *pe/ppe* genes, polyketide synthases or even *esx* genes (Jajou *et al.*, 2019). Different regions are included in the default masking files that are incorporated in the evaluated pipelines; for example, while SNiPgenie masks *pe/ppe* genes by default, vSNP additionally filters other positions according to the genotype classification of the isolate, which probably correspond to lineage-specific differences that can alter variant calling, such as deletions. In contrast, the annotation file used by BovTB was generated from a self-BLAST which identified those regions of the genome with a similarity higher than 95%. Therefore, in order to assess the overall effect of different genetic elements in performance, default filters, as well as increasing levels of hard filters, were applied.

In agreement with the observation that most discrepancies occurred in problematic regions, the overall performance and agreement of the evaluated pipelines increased with increasing levels of hard filters. Similarly, the discrepancies observed previously in the pairwise comparison of phylogenetic trees were limited to the positioning of clusters of highly similar isolates. This

indicates that when there is a high genetic similarity between isolates, certain topological variation is to be expected due to the inability to resolve and position polytomies (Walsh *et al.*, 1999). Performance also varied when default filters were applied, reflecting the differences in pipeline-specific masking files. It is crucial to appropriately annotate masked positions in variant calling analyses in order to assess the level of filtering that is being applied to each analysis.

Although performance increased when all problematic regions were removed, independently of the application of proximal filters, the removal of proximal SNPs had a significant effect in tree topology, as evidenced by the clear separation of the inferred trees after the applications of filters C and G with respect to the rest. Accordingly, the proportion of homoplasies decreased significantly with the application of the abovementioned hard filters, especially with the removal of proximal SNPs and the masking of *pe/ppe* genes. This reflects the effect that homoplasies can have in the establishment of phylogenetic relationships between isolates and highlights the importance of evaluating problematic regions beyond their effect in performance. Interestingly, when hard filters were used without the proximal filter, the homoplasy rate increased slightly, in accordance to the intermediate clustering of trees between proximal filtered trees and those inferred from the removal of problematic regions alone. This is interesting, since it indicates that homoplasies may be present in other regions of the genome that may not be considered if only problematic regions are masked.

5.2.3. The increasing use of WGS in the field study of bTB calls for future international standardisation efforts

As the use of WGS is expected to increase for the study of *M. bovis*, efforts towards the standardisation of available tools is necessary to guarantee a similar performance and exchangeability of results across the different laboratories. The results of this pipeline comparison are a first step towards standardisation, but more research is warranted in order to achieve this goal. Although available variant calling tools have a similar performance when the same masking files are used during filtering, the results of this study are based on an artificial scenario in which the challenges of sequencing a biological sample using different library generation and sequencing platforms were not considered. Given that biases, such as coverage blind spots, can be introduced depending on the type of sequencing platform used (Modlin *et al.*, 2021), it would be important to generate a validated dataset of reads obtained from a real-life

scenario in which these effects could be assessed. In addition, the dataset used in this study corresponds to a prevalence study in which the epidemiological relationship of the isolates was not established. As a result, the performance was estimated through the degree of agreement with the simulation, rather than on the capacity of investigating a true herd breakdown event. In order to assess this capacity, the dataset used should include epidemiologically related isolates as well as their metadata.

In view of the significant effect of hard filtering in the performance and phylogenetic inference of variant calling pipelines, the definition of a standardised masking file could be of great use for the future standardisation of WGS analyses around the world. The file generated in this evaluation is a starting point towards this direction, since it includes well defined genetic elements, including *pe/ppe* genes, mobile genetic elements or direct repeats. Other genetic elements, such as the *pks12* gene, were also incorporated due to the observation of a decreased mapping quality of artificial reads in certain regions. Based on this, the genetic elements included in masking files could also be adjusted by assessing the mapping quality of artificial reads towards certain regions of the reference genome. Furthermore, recent data shows that a proportion of masked repetitive regions could be accurately mapped using Illumina platforms (Marin *et al.*, 2021), which entails that these elements may be incorporated in variant calling, leading to a better resolution. This could be especially relevant considering that the high similarity between *M. bovis* isolates is a common drawback in the epidemiological investigation of many bTB breakdowns (Price-Carter *et al.*, 2018).

In addition to pipeline performance, other factors related to WGS analysis of bTB should be considered when working towards a standardised methodology. One important factor to consider is the threshold used to establish when a group of isolates can be considered part of the same breakdown event or indicate if the isolate was recently transmitted or not. In this aspect, only MTBseq included an option to establish transmission clusters based on a predefined threshold. Such approach would certainly be interesting to apply in bTB epidemiological investigations. Several thresholds have been proposed in different studies and have frequently been used in the study of bTB (including study 4), ranging from 3 SNPs for recent transmission scenarios to 12 SNPs in more distant ones (Walker *et al.*, 2013; Crispell *et al.*, 2020). However, it is important to highlight that these thresholds have been extrapolated from human TB studies. Therefore, their use in the study of herd breakdown events or in the transmission dynamics in

complex ecosystems with different hosts or *M. bovis* genotypes should be assessed in the future.

There are three major published classification schemes of *M. bovis* lineages based on WGS data (Loiseau *et al.*, 2020; Zimpel *et al.*, 2020; Zwyer *et al.*, 2021). However, there is currently no recommendation as to which of these should be used. Consequently, all but SNiPgenie included some kind of nomenclature scheme. The nomenclature used in BovTB is based on the one established by the APHA for *M. bovis* isolates in the UK and may not be suitable for other geographical areas. MTBseq was developed based on human TB data and therefore, the animal-adapted variants are not resolved further (Homolka *et al.*, 2012; Coll *et al.*, 2014). Finally, in the case of vSNP, the lineage classification is based on certain defining SNP patterns, with more than 30 major *M. bovis* genotypes, including the Af1 and Af2 CCs, as well as several *M. caprae* and *M. orygis* genotypes. It is important that a unified nomenclature scheme is put in place in order to make data exchange and comparison between laboratories easier. To this direction, the nomenclatures proposed by Loiseau *et al*., Zwyer *et al.,* may be a good alternative, since they incorporate the classical classification through CCs as well as novel unknown groups, which are in turn included within the Lb1-4 classification proposed by Zimpel *et al.* (Loiseau *et al.*, 2020; Zimpel *et al.*, 2020; Zwyer *et al.*, 2021).

Variant calling is traditionally carried out using *M. bovis* AF2122/97 as a reference genome. However, due to differences in gene content between genetic lineages, mapping to a reference genome from a different lineage could result in loss of information (Meehan *et al.*, 2019). This has been shown to have a negligible effect in a well-defined *M. tuberculosis* outbreak with recent transmission, in which the use of different reference strains, including *M. bovis* and *M. canettii,* did not alter the epidemiological conclusions drawn from the variant calling analysis (Lee & Behr, 2016). Nevertheless, the extent to which this finding holds true for the animal-adapted strains, such as *M. bovis*, should be evaluated, especially considering that the main *M. bovis* lineages are defined based on RDs. Recently, the complete genome of *M. bovis* Mb3601, a strain belonging to the Eu3 complex, which is the most representative clonal group in France, has been made available (Branger *et al.*, 2020). Future studies should follow to include other strains from different clonal complexes. Furthermore, considering that bTB can be caused by other members of the MTBC, such as *M. caprae, M. orygis* or *M. microti*, reference genomes for these variants may result useful to finely tune WGS studies in more complex epidemiological scenarios in which other MTBC members are also present. Furthermore, the

increased use of long reads could allow for the generation of *ad hoc* reference genomes specifically suited for a certain geographical region or the use of *de novo* assemblies for a reference free approach in which multiple genome alignments are carried out (Meehan *et al.*, 2019).

It is important to mention that the choice of a variant calling pipeline may, at times, go beyond the performance of a given tool. An important limitation of some of the evaluated pipelines was their overall automation, which implied that the quality filters and types of analyses could not be modified. In this aspect, only MTBseq and SNiPgenie presented a modular structure, allowing for the modification of parameters and choice of analyses. Certainly, the majority of MTBC pipelines made available in recent years follow a modular approach, in which different levels of analyses can be run at different time points (Kohl *et al.*, 2018c; Bogaerts *et al.*, 2021; van Heusden *et al.*, 2022). A modular program can easily be modified to include new types of analyses that may arise as WGS technologies advance. Therefore, a modular approach, in which different levels of analyses can be run at different times, should be pursued.

Given that most of the available pipelines are command line tools, users must have a deeper knowledge of bioinformatics, which may be challenging in certain scenarios. In order to overcome this, the user accessibility of WGS analyses tools should also be considered. Several commercial platforms have been developed that allow for an easy selection of tools and parameters that can accommodate different analyses, such as Geneious Prime or Bionumerics. Despite their utility, these platforms were not included in this comparison since the focus of the study was on open source pipelines. While the designers of SNiPgenie invested time to improve the accessibility of the platform, this was not the case for the other tools considered in the study. The user friendliness of analyses would certainly be a worthy feature to work for (van Heusden *et al.*, 2022); future pipelines may benefit from incorporating Graphical User Interfaces so that less proficient users may be able to use of such platforms (van Heusden *et al.*, 2022).

5.2.4. Whole genome sequence analysis reveals a high genetic diversity of the most frequent *M. bovis* spoligotype in Spain

The analysis of 99 genome sequences of *M. bovis* SB0121 isolates from cattle (n = 70) and wildlife species (n = 29) revealed the presence of three major genetic clades of the SB0121 genotype isolates included in this study. Another WGS study in northeastern Spain evaluated the genetic diversity of *M. bovis* strains from different spoligotypes, revealing the distribution of SB0121 into four different genetic clades, along with other genotypes such as SB0339 or SB0265 (Perea *et al.*, 2021). These results agree with previous studies in the Iberian Peninsula which described a high genetic diversity of the SB0121 spoligotype through MIRU-VNTR typing, with 45 and 65 genotypes in Portugal and Spain, respectively (Duarte *et al.*, 2010; Rodriguez-Campos *et al.*, 2013). Based on the observed high genetic diversity, and considering the overall limited variability of the MTBC, these results support a long lasting circulation of this genotype in Spain. More studies are needed to assess the genetic diversity of SB0121 and other related Eu-2 complex genotypes, such as the SB0339, with a more representative sample in order to shed light on the evolution of bTB in Spain.

5.2.5. Within-herd diversity reveals a complex epidemiology in herds chronically infected with *M. bovis* SB0121

The results from study 4 revealed a remarkably high genetic diversity of the isolates originated from the beef and bullfighting herds included in the study, with the majority (n = 19/22) harbouring isolates with a median pairwise genetic distance higher than six SNPs, a common threshold used to indicate a close epidemiological relation between isolates. Nevertheless, the genetic diversity of isolates varied considerably across the herds, with the minimum and maximum SNP difference ranging between 1-310 SNPs. This variability in the within-herd diversity probably reflects the complex epidemiology of bTB in Spain, in which different factors may lead to bTB persistence. For example, the detection of highly similar isolates in several herds could indicate the presence of an ongoing infection in a herd through undetected infected animals from previous rounds of testing (relapse) or the contact with similar *M. bovis* strains from a common source, most probably in close vicinity to the herd (re-introduction). In contrast, the detection of distantly related or unrelated strains in long periods of time indicates that chronic breakdowns of bTB in these cases was probably caused by separate introductions from external sources. Interestingly, *M. bovis* isolates with more than 6 and 12 SNP differences were also detected in several herds within a single year, suggesting either the presence of isolates with a high

genetic diversity in the surroundings of the herd and/or the introduction of animals from diverse sources in multiple occasions in the past. Unfortunately, movement information for these herds was not available at the time of this study and was limited to what was included in the epidemiological surveys. More studies are warranted in the future to assess the movement of animals in these herds, in order to assess the risk of transmission and/or the presence of local or non-local movements that could have contributed to the differences in the observed genetic diversity.

The overall high within-herd genetic diversity observed in this study contrasts with the findings of previous bTB outbreak investigations in other countries. In the case of the USA, Orloski *et al.,* 2018 reported the absence of *M. bovis* isolates with more than 6 SNP differences in *M. bovis* isolates obtained from the same herd and year across the country, whereas Glaser *et al.* 2016 reported a very limited within- and between-herd genetic diversity (< 6 SNPs) during a four-year bTB outbreak among 12 cattle herds in a region in which both cattle and deer were involved (Glaser *et al.*, 2016; Orloski *et al.*, 2018). In Northern Ireland, WGS data from a cluster of bTB breakdowns in five herds, located within 5 km from each other and during a 10 year interval, revealed a low within- and between-herd genetic diversity (0 to 7 SNP differences) (Biek *et al.*, 2012), whereas very limited genetic diversity (0 to 4 SNPs) was observed within the same breakdowns and herds in another study in the same country (Trewby *et al.*, 2016). In contrast, a WGS analysis of the third most prevalent genotype in Spain (SB0339) revealed that the average within-herd SNP distances varied significantly in a selection of herds with chronic breakdowns (Pozo, 2020). These disparities are probably related to the different epidemiological situation in these countries (e.g. bTB prevalence, wildlife hosts involved, overall genetic diversity, etc.) or to differences in the study designs [lower geographical extension in the studies by Biek *et al.,* 2012 and Glaser *et al.,* 2016, or more restricted genotype selection (VNTR-10) in the study by Trewby *et al.,* 2016]. For example, since bTB is eradicated in the USA and most of the outbreaks are caused by imported cattle, or by infected wildlife in certain regions (e.g. Minnesota) (Orloski *et al.*, 2018), the introduction of multiple genotypes in a region is less likely to arise due to the constant surveillance and the lack of a sustained between-herd transmission. In contrast, although Northern Ireland has a higher incidence of bTB and a more extended cattle-wildlife transmission than the USA (Skuce *et al.*, 2010), the genotype analysed in the study by Trewby *et al.,* 2016 has a low prevalence and a probably recent emergence (Trewby *et al.*, 2016), suggesting a reduced evolutionary time. In addition, the lower diversity observed in

Northern Ireland could also be explained by the severe reduction in *M. bovis* population size that took place in the British Isles due to the success of test-and-slaughter protocols until the infection expanded in the 80s (Smith *et al.*, 2006a). In contrast, the significantly higher prevalence of SB0121 and SB0339 could reflect a longer time of circulation and expansion of these genotypes due to a later reduction in population size, which in turn could have contributed to an increased genetic diversity. This, coupled with the complex epidemiology of bTB in Spain, could have resulted in the simultaneous presence of different genotypes within the same herd observed in this study.

5.2.6. The combination of sequence and epidemiological data from multiple sources is critical for the effective investigation of bTB outbreaks

The implementation of WGS in the bTB eradication campaign can provide a resolution level that is not possible with conventional molecular characterisation methods. However, similar to any other characterisation method, the conclusions drawn from outbreak investigations will be greatly influenced by the overall quality and availability of both epidemiological and sequence data to compare with.

Despite the fact that epidemiological data in study 4 was not available for all herds, several possible sources of infection could be established, such as previous animal movements or the contact between two herds owned by the same farmer. The importance of this information was further demonstrated in a bullfighting herd (herd 22) in which two *M. bovis* isolates with ten SNP differences were detected in only seven months, which pointed to two separate introduction events. From the available epidemiological data, it was possible to establish that this herd was originally created in 2012 from different OTF bullfighting herds and suffered an explosive breakdown in 2015. No previous movements were reported at the time and contacts with other herds or wildlife were discarded due to the housing of the animals in closed facilities, which in turn could explain the extent of the within-herd transmission. Although animal movements were recorded in subsequent years, the high genetic similarity between the later isolates (≤ 3 SNPs) indicated that the persistence was probably caused by a relapse from the previous breakdown.

Another interesting finding was the definition of common pasture as the source of infection for one of the herds in this study. The herd in question (herd 11), shared pastures with several non OTF herds in the region. This finding suggests clearly that the acquisition and/or analysis of sequence data from the surroundings of an outbreak is crucial for its effective investigation. Future outbreak investigation studies of bTB outbreaks would benefit from such data, when available, by allowing the assessment of the local genetic diversity and by pinpointing possible epidemiological links between herds. Due to the relevance of different wildlife reservoirs in the maintenance of aTB in Spain, it is also crucial to analyse samples obtained from different wildlife species during outbreak investigations, especially in those cases in which prevalence of infection in wildlife is high in the region. The majority of the analysed wildlife samples in study 4 were restricted to the province of Ciudad Real, in accordance with the high aTB prevalence and intensive sampling of wildlife in the region. The analysed sequences from this region revealed a generally high genetic diversity, which could indicate a long lasting circulation of the infection in wildlife in this region, or a frequent contact with a diverse pool of *M. bovis* from cattle. Certainly, a possible epidemiological link between wildlife and cattle was identified with the detection of highly similar and distantly related *M. bovis* isolates, especially within a short geographic distance. Furthermore, the detection of genetically related isolates among wild animals from the same or different species could also point towards a common source of infection in wildlife. Therefore, sampling of wildlife should be considered as a standard practice in outbreak investigations in order to fully address the between and within species transmission dynamics of aTB. In this sense, the national programme for TB surveillance in wildlife will greatly benefit from the high resolution of WGS in establishing and controlling the risk of infection between different wildlife species and cattle (MAPAMA & AECOSAN, 2017).

The use of WGS has rapidly increased around the world and a similar trend is to be expected to occur in Spain. Its nucleotide-level resolution is likely to enhance outbreak investigations by revealing previously hidden molecular patterns between highly similar *M. bovis* strains from multiple origins. However, similarly to other molecular characterisation techniques, the implementation of WGS for outbreak investigations will be demanding, in the sense that it will require the acquisition of a representative panel of *M. bovis* sequences with which to generate accurate and up to date genetic maps to support outbreak investigations, as well as to evaluate interspecies transmission dynamics and the risk of infection from different animal reservoirs. This can only be achieved with

a structured and well distributed sample collection strategy, as well as high quality epidemiological meta-data. However, due to the complexity of bTB epizootics, in which many cases remain undetected due to limitations in current diagnostic techniques, the obtaining of an accurate picture may be a challenging endeavour. In this aspect, the implementation of WGS will most likely benefit from new developments in case confirmation, such as the use of real-time PCR as a rapid and sensitive method for the selection of samples for their inclusion in WGS workflows. Ultimately, the combination of accurate genetic and epidemiological data will enhance bTB eradication by rapidly identifying the most likely causes of infection and by implementing contingency measures with which to decrease the spread of the disease.

5.3. Rapid identification of NTM species through MALDI-TOF mass spectrometry

In the context of bTB eradication, the role of NTM species is of importance. Considering that NTM species may produce tuberculous-like lesions, the rapid identification of these species is especially relevant in the context of bTB-free regions in which control of the infection is carried out through slaughterhouse surveillance (Scherrer *et al.*, 2018). Due to the capacity of certain NTM species to produce cross reactions with *ante-mortem* tests, the identification of interfering species is crucial for their characterisation and the development of more specific methods.

However, the importance of mycobacterial identification extends beyond the field of bTB, since many NTM species are important animal pathogens, such as MAP or *M. marinum*. Additionally, considering that NTM are ubiquitous in the environment and in many different animal species, and that there is an increasing prevalence of NTM infections in humans, the adequate identification of these bacteria could be relevant in detecting possible zoonotic origins of infection.

Traditionally, mycobacterial species were identified using a series of biochemical tests which were laborious and time consuming to carry out due to the fastidious growth of certain mycobacterial species. The use of genetic molecular methods for the identification of NTM, such as molecular probes or Sanger sequencing (Tortoli, 2010; Alcolea-Medina *et al.*, 2019; Bouzinbi *et al.*, 2020), greatly enhanced the identification capacity of NTM in microbiology laboratories, revealing the complex taxonomy of the genus. In this aspect, Sanger sequencing has been proposed as a reference method due to its

versatility and resolution. However, this technique has a high cost per sample and requires a significant amount of post-processing for an accurate identification.

5.3.1. MALDI-TOF MS is a highly effective and rapid screening method for the identification of NTM isolates

In the last decade, the implementation of MALDI-TOF MS significantly increased the throughput capacity and the speed of processing, as well as decreased the costs in clinical microbiology laboratories (Tsuchida *et al.*, 2020). In recent years, MALDI-TOF MS has been increasingly used to identify NTM species in the clinical practice (Rodriguez-Temporal *et al.*, 2017; Alcaide *et al.*, 2018; Rodriguez-Temporal *et al.*, 2020; Rodriguez-Temporal *et al.*, 2022). However, in the context of veterinary microbiology, and specifically in the field of bTB, the use of MALDI-TOF MS for NTM identification is limited in comparison to human medicine. Published data regarding the use of this method in veterinary samples mainly concerns isolates from a single animal species, such as wild boar, or a specific group of mycobacteria, such as *M. avium* or the MTBC (Ravva *et al.*, 2017; Ricchi *et al.*, 2017; Bacanelli *et al.*, 2019; Ghielmetti *et al.*, 2021a). The study in this thesis expands the use of MALDI-TOF MS to the identification of isolates from a wide array of animal species.

When the identification capacity of MALDI-TOF MS was compared with Sanger sequencing, the two methods had a high level of agreement, with the majority (n = 66, 88%) of the isolates identified by both methods (n = 57, 76%) or by neither (n = 9, 12%). Considering that MALDI-TOF MS offers a simplified and rapid sample processing and analysis in comparison to Sanger sequencing, these results indicate that MALDI-TOF MS can be used as an identification method NTM.

5.3.2. The identification of a large number of NTM species in cattle indicates a high mycobacterial diversity in this species of animal

A total of twenty-eight mycobacterial species were identified in this study, of which 71.4% were cultured from bovine samples, revealing the ability of a wide range of NTM species to establish infections in bovines. Nevertheless, it is important to pinpoint that this study represents a non-structured sample selection based on the routine analysis of suspicious mycobacteriosis cases, and may therefore not represent the true mycobacterial diversity in different animal species.

The most frequently encountered species in cattle (n = 32) were *M. avium* (n = 4), *M. fortuitum* (n = 4), *M. nonchromogenicum* (n = 3), *M. peregrinum* (n = 3) and *M. thermoresistibile* (n = 3). The higher species diversity observed in bovine samples could depend on the higher number of samples originating from cattle (n = 32) with respect to the isolates originating from other animal species. In another study, a large number of NTM species was identified in wild boar (Ghielmetti *et al.*, 2021a). This could indicate a high number of NTM species that are able to establish infections in cattle and wild boar. The isolation of such a wide range of NTM species in bovines warrants further research, since some of these species could be at the origin of cross-reactivity observed with current *ante mortem* tests. The use of MALDI-TOF MS could be of great use to identify NTM species isolated from reactor animals or the carcases of animals with VL detected during slaughterhouse surveillance. This information could then be used to assess the cross-reactivity of these NTM species through experimental infections and skin testing of infected animals.

An adequate representation of MSPs is crucial for the effective identification of mycobacterial isolates (Rodriguez-Sanchez *et al.*, 2016). Given that one third of the isolates in this study were identified as *M. avium,* this high identification capacity could partly be a reflection of the overall representation of *M. avium* in MSP libraries. A similar finding was obtained by Ghielmetti *et al*., in which approximately one quarter of the identified isolates corresponded to *M. avium* and were identified with high confidence scores. In contrast to the previous study, in which *M. avium* was isolated from wildboar, the isolates evaluated in this thesis were obtained from 70% (n = 14) of the animal species sampled, including bovines, wild boar, domestic pig and deer, among others. This could reflect a high distribution of *M. avium* in many different animal species, in which case MALDI-TOF MS would allow for a rapid and effective identification. Despite the predominance of *M. avium* isolates, a diverse range of NTM species was sampled and MALDI-TOF MS was able to identify most of these species as well.

5.3.3. Discrepancies between methods reflect the complexities and limitations of NTM typing in veterinary samples

Despite the overall good agreement between both techniques, discrepant results were observed in a small proportion of samples. These were either caused by an erroneous classification of the mycobacterial isolate through MALDI-TOF MS when compared to Sanger sequencing or due to non-reliable or low-confidence identifications, as previously indicated (Rodriguez-Temporal *et al.*, 2020; Ghielmetti *et al.*, 2021a).

Due to the dependence of MALDI-TOF MS identification on reference MSP databases, one possible explanation for these discrepancies could be related to the lack of spectra available for comparison for certain mycobacterial species. Certainly, all non-reliable and most of the low-confidence identifications in the present study concerned isolates for which a reliable identification was not possible with Sanger sequencing either. This was also observed in the study by Ghielmetti *et al*. in which four NTM isolates could not be identified to the species level neither by MALDI-TOF MS nor Sanger sequencing (Ghielmetti *et al.*, 2021a). This indicates the presence of a yet unknown mycobacterial diversity in certain environments, which agrees with the fact that only a small fraction of microbial diversity is being represented today due to limitations in sampling and culturing processes (Stewart, 2012). Despite the inability of both methods of producing high confidence identifications for these NTM isolates, they both provided an identification at the complex level such as the MAC or the *M. simiae* complex. Low confidence or non-reliable identifications could therefore indicate the presence of uncharacterised species/strains. Taking this into account, MALDI-TOF MS could be used as a powerful tool for a rapid screening of possibly novel NTM species. Interestingly, the two *M. terrae* complex strains identified in two wild boar specimens in this thesis, and one of the isolates from the study by Ghielmetti *et al.* were identified with low confidence as *M. arupense* through Sanger sequencing (Ghielmetti *et al.*, 2021a). This indicates that some of these uncharacterised NTM isolates could have a wide geographical distribution. A more exhaustive biochemical and genomic characterisation is needed for the appropriate characterisation of these isolates.

Only one discrepant result in this study was caused by a misidentification (see below), which supports the evidence that this is not a frequent event (Rodriguez-Temporal *et al.*, 2020; Rodriguez-Temporal *et al.*, 2022). The causes of misidentifications could relate to variations in the experimental method, the quality and representativeness of the MSP library or the mycobacterial species under study. Certainly, misidentifications can be a result of low quality protein extraction, issues during spectra acquisition, due to cross contamination or sample interchange. However, misidentifications are commonly reported on species that are closely related, such as *M. scrofulaceum* and *M. parascrofulaceum* (Rodriguez-Temporal *et al.*, 2022). In this thesis, one isolate identified through Sanger sequencing as *M. malmesburyense* was identified with high confidence by MALDI-TOF MS as *M. novocastrense*, which is a closely-related mycobacterial species (Gcebe & Hlokwe, 2017). *Mycobacterium*

malmesburyense was not included in the Mycobacteria Library v3.0 used in this study, which explains the misidentification by MALDI-TOF MS. Although Sanger sequencing was able to identify this mycobacterial species, it is important to highlight that this identification was problematic. This was because the *hsp65* and *rpoB* sequences were 100% similar to the ones deposited at Genbank for *M. malmesburyense*, but 98.7% similar for the 16S rRNA sequence from the original publication (Gcebe *et al.*, 2017). In contrast, a BLAST search using the WGS archive at the NCBI revealed that the three genetic targets were identical to the ones deposited by the same authors for the *M. malmesburyense* type strain. This fact puts in evidence the difficulty of obtaining an accurate identification for certain species with Sanger sequencing due to inconsistencies in the available databases.

The choice of threshold for high confidence MALDI-TOF MS identification is important since it could also lead to misidentifications; Bruker Daltoniks recommends the use of a threshold higher or equal to 2.0, nevertheless, other studies have used lower thresholds (Alcaide *et al.*, 2018; Rodriguez-Temporal *et al.*, 2022). In this study, three out of the four identifications of NTM isolates with a low quality score agreed with the identification obtained through Sanger sequencing. One of these isolates corresponded to *M. neoaurum,* and the decreased confidence could therefore be related to the low number of MSPs included for this species in the spectra library. Another case corresponded to *M. avium*, for which a significant variation of log scores has been shown in different studies, ranging between 1.30 and 2.30 (Rodriguez-Temporal *et al.*, 2020; Rodriguez-Temporal *et al.*, 2022). *Mycobacterium avium* subspecies cannot be discerned with current databases, requiring the use of specific MSP libraries (Ravva *et al.*, 2017; Ricchi *et al.*, 2017). As a result, *M. avium* susbspecies are not differentiated in most MALDI-TOF MS studies. Therefore, a possible explanation for this variation in log scores within the same study could be related to proteomic differences between subspecies. One possible approach for increasing the performance of MALDI-TOF MS could be to adapt the threshold according to the identified mycobacterial species (Rodriguez-Temporal *et al.*, 2020). However, this would need to be assessed further using a larger panel of NTM isolates from different animal species.

The spectra included in this study were also sent to Bruker Daltoniks for analysis using the Mycobacteria Library 6.0. Interestingly, the identifications achieved using the newer database were the same, with no significant differences observed in the MALDI-TOF log scores.

Overall, the results of this thesis indicate that MALDI-TOF MS could be a very useful technique to implement in veterinary microbiology laboratories for the identification of NTM species. Although Sanger sequencing shows a higher resolution in certain cases, such as in *M. avium* subspecies identification, MALDI-TOF MS offers a series of advantages. Unlike sequencing, in which several PCR and sequencing reactions are required in addition to the DNA extraction protocol, MALDI-TOF MS requires only a simple protein extraction protocol. In addition, sequencing of samples can take several hours, in contrast to the seconds required for spectra acquisition. Furthermore, MALDI-TOF MS uses curated and standardised MSP databases that unify the analysis of results, whereas Sanger sequencing requires the manual curation of sequences and their comparison against publicly available or *in house* databases. Considering that the majority of MSPs included in available libraries are from human origin, the incorporation of this methodology in the workflow of veterinary laboratories involved with the detection and identification of mycobacteria, including bTB-causing agents, would significantly contribute to the increase in the number and diversity of reference spectra from animals and the environment. This would undoubtedly be beneficial beyond the veterinary field, considering that several NTM infections could have a zoonotic origin (Ghielmetti *et al.*, 2021a). Although the running costs per sample of MALDI-TOF MS are significantly lower, the initial investment for the acquisition of a MALDI-TOF mass spectrometer is high. Nevertheless, available literature indicates that the routine implementation of MALDI-TOF MS reduced the direct costs of AFB identification by 74%, which coupled to its use with other microorganisms led to an overall reduction of 51% in total costs ($73,646.18), offsetting the initial cost of the instrument in approximately three years (Tran *et al.*, 2015). Therefore, the implementation of MALDI-TOF MS could be a promising option to decrease costs in centralised veterinary laboratories in the future.

5.4. Integration of Novel tools in eradication campaigns

The research in this thesis focused on the development and implementation of novel methodologies that can be used in the surveillance, control and eradication of bTB. While these tools may not be perfect, their incorporation in the workflow of laboratories tasked with the diagnosis and characterisation of bTB will greatly enhance the eradication efforts.

The following diagram (figure 18) presents a suggested scheme for the incorporation of the abovementioned methods. In this workflow, the qPCR may be used in substitution of microbiological culture. The IS*6110* qPCR may be used alone as a rapid and sensitive method, or in combination/in parallel with the *mpb70* qPCR in order to increase the specificity of the analysis. A positive result to the qPCR could be used by the competent authorities to declare a case as confirmed, as indicated in DeR 2020/289. If considered necessary, tissue homogenates that are positive to the qPCR may then be cultured in order to obtain pure isolates with can then be characterised by WGS for epidemiological investigations. In case a NTM isolate is obtained, MALDI-TOF MS may be used to identify it.

In case of negative qPCR results, inhibition phenomena need to be excluded, therefore the inclusion of ICs is necessary. If inhibition is present, the DNA extract should be diluted and/or the DNA extraction should be repeated from a frozen back-up aliquot, and the qPCR should be repeated. If amplification is detected after retesting, the case could be declared confirmed as previously indicated. In any case, the infection is not confirmed if the qPCR is negative but inhibition was not present. In this case, other complementary methods, such as microbiological culture or histopathology, may be considered according to the epidemiological situation from which the given sample originates and the available resources for the investigation.

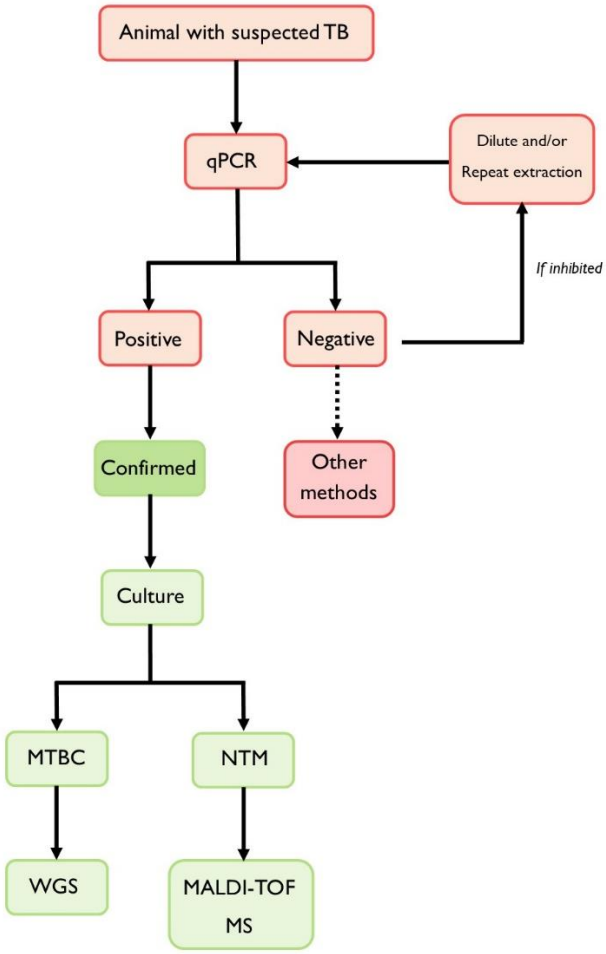


Figure 18. Proposed workflow for the incorporation of the tools developed in this thesis in the workflow of laboratories tasked with the detection and typing of bTB. The red colour indicates steps carried out prior to confirmation of the infection, whereas green indicates steps that may be carried out once infection is confirmed through qPCR.

5.5. Future perspectives in the laboratory diagnosis of bTB

As technologies advance, new methods are likely to arrive that may complement or even substitute current techniques. In addition to the methodologies described in this thesis, new and very promising tools are being developed that may facilitate the diagnosis and characterisation of MTBC members.

Digital PCR and CRISPR-based methods are two promising approaches that may see an increased use in the future. The high sensitivity of digital PCR could be especially useful in samples with low concentrations of bacterial DNA (Nyaruaba *et al.*, 2019), such as tissues with NVL. In contrast, CRISPR-based methods could be a cost-effective method for the molecular detection of MTBC members, especially in developing countries, since they do not require sophisticated equipment and have low estimated costs (Mustafa & Makhawi, 2021). Pre-amplification strategies may be required in order for the implementation of CRISPR-based methods, in which case LAMP could be an interesting approach (Kaminski *et al.*, 2021).

Sequence capturing methods may significantly advance the throughput capacity of microbiology laboratories by allowing the enrichment of pathogen-specific sequences in complex samples (Hayden *et al.*, 2022). This would increase the speed of molecular characterisation studies by allowing the direct sequencing of MTBC DNA in the sample and, therefore, avoiding the culturing step. Several studies have highlighted the potential of direct sequencing in the field of human TB (Doyle *et al.*, 2018; Goig *et al.*, 2020), which could also have applications in the molecular diagnosis of TB (Votintseva *et al.*, 2017). Such developments are yet to be seen in the field of bTB, since the higher complexity of tissue matrixes in comparison to sputum samples may require additional efforts. In addition, these approaches could also be used in next-generation metagenomics studies (Wei *et al.*, 2022; Xu *et al.*, 2022), which would allow for an improved detection and characterisation of microbial communities present in animal samples, including MTBC members and NTM.

Developments in the field of mass spectrometry have microbiology laboratories to evaluate other methods for the identification of mycobacteria, such as high-resolution tandem MS for proteomic analyses or lipid-fingerprinting through MALDI-TOF MS (Bajaj *et al.*, 2021; Jia Khor *et al.*, 2021; Bajaj *et al.*, 2022). These technologies could allow for in the future for variant or subspecies level resolution, facilitating the identification of highly similar microorganisms.

6. Conclusions

First

The high diagnostic sensitivity and specificity of the *mpb70* and IS*6110* real-time PCRs, in comparison to microbiological culture, indicate that these are powerful and rapid tools for the detection of bTB that can be used as confirmatory methods.

Second

Performance was significantly higher in samples with lesions compatible with TB with respect to samples without visible lesions. Nevertheless, the sensitivity of the IS*6110* real-time PCR in the latter remains acceptable for its use in this type of samples.

Third

The inclusion of inhibition controls, the use of multiple genetic targets and the repetition of the DNA extraction and real-time PCR in inhibited samples can minimise the risk of PCR false-negative results.

Fourth

An IS*6110*-like element has been identified in '*M. avium* subsp. *hominissuis*' which can cross react with the IS*6110* qPCR. The use of additional genetic targets in parallel or in series with the IS*6110* element must be carefully considered, according to the prevalence of bTB or NTM infection in the area under study.

Fifth

The incorporation of qPCR protocols in the workflow of laboratories tasked with bTB diagnosis across the EU requires the previous verification of the method in order to guarantee the described performance.

Sixth

Available variant calling pipelines offer comparable results for the WGS analysis of *M. bovis* strains as long as the same filtering files are used. Similar to other molecular methods, there is a need for standardising the use of variant calling pipelines.

Seventh

Whole genome sequencing is a powerful tool to evaluate within-herd diversity in cattle herds. Whole genome sequencing revealed the complex epidemiology behind bTB persistence some study herds in Spain.

Eighth

MALDI-TOF mass spectrometry is a rapid and cost-effective method for the identification of NTM, with a similar performance as Sanger sequencing.

Ninth

There is a need to enrich MALDI-TOF mass spectrometry spectral databases with NTM species of veterinary interest in order to increase the spectrum of NTM that can be reliably identified.

Tenth

The presence of a significant proportion of uncharacterised isolates by MALDI-TOF MS and Sanger sequencing reflects the presence of a high number of putative novel NTM species in animals that are worthy of further evaluation.

List of abbreviations

AHL	Animal Health Law
APHA	Animal & Plant Health Agency
APHIS	Animal & Plant Health Inspection Service
aPPD	Avian purified protein derivative
aTB	Animal tuberculosis
BC	Benzylkonium chloride
BLAST	Basic local alignment search tool
bp	Base pair
bPPD	Bovine Purified Protein Derivative
BSL-3	Biosafety level 3
bTB	Bovine tuberculosis
BWA	Burrow-Wheelers aligner
CFP	Culture filtrate proteins
cgMLST	Core genome multi-locus sequence typing
CIT	Comparative intradermal tuberculin

CMI	Cell-mediated Immunity
CPC	Cetidylpyridinium chloride
CRISPR	Clustered regularly interspersed short palindromic repeat
Ct	Cycle threshold
DCT	Distributive conjugative transfer
DeR	Delegated Regulation
DIVA	Differentiating infected from vaccinated animals
dPCR	Digital PCR
DR	Direct repeat
DVR	Direct variable repeat
EEC	European Economic Community
ELISA	Enzyme-linked immunosorbent assay
ESAT	Early secreted antigenic target
ETR	Exact tandem repeat
EU-RL	European Union Reference Laboratory
FN	False negative
FP	False positive
FPA	Fluorescence polarisation assay
FRET	Fluorescence resonance energy transfer
H_2SO_4	Sulphuric acid
HCCA	α-cyano-4-hydroxycinnamic acid
HIC	High income country
HPC	Hexadecylpyridinium chloride
HPLC	High-performance liquid chromatography
IC	Inhibition control
IFN-γ	Interferon gamma

IGRA	Interferon gamma release assay
Indel	Insertion and deletion
IR	Implementing Regulation
IS	Insertion sequence
IT	Intradermal tuberculin
ITS	Internal transcribed spacer
kb	Kilobase
KOT	Koch's old tuberculin
LAMP	Loop-mediated isothermal amplification
LCA	Latent class analysis
LGT	Lateral gene transfer
LN	Lymph node
LOD	Limit of detection
LPAs	Line probe assays
LSP	Large sequence polymorphism
MAC	*Mycobacterium avium* complex
MAH	'*Mycobacterium avium* subsp. *hominissuis*'
MALDI-TOF	Matrix-assisted laser desorption ionisation - time of flight
MAP	*Mycobacterium avium* subsp. *paratuberculosis*
MAPIA	Multi-antigen print immunoassay
Mb	Megabase
MeS	Member State
MGIT	Mycobacterial growth indicator tube
MIRU	Mycobacterial interspersed repetitive units
ML	Maximum likelihood
MOTM	Microorganisms other than mycobacteria

MPTR	Major polymorphic tandem repeat
MRCA	Most recent common ancestor
MS	Mass spectrometry
MSP	Main spectra prophile
MTBC	Mycobacterium tuberculosis complex
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
NVL	Non visible lesion
NVSL	National Veterinary Services Laboratories
OA	Oxalic acid
OADC	Oleic albumin dextrose catalase
OD	Optical density
OIE	World Organisation for Animal Health
ORF	Open reading frame
OTF	Officially bTB free
PCR	Polymerase chain reaction
PE	Proline-glutamate
PFGE	Pulsed field gel electrophoresis
PGRS	Polymorphic GC-rich repeat sequence
PPE	Proline-proline-glutamate
qPCR	Quantitative or real-time PCR
QUB	Queen's University Belfast
RADP	Random amplified polymorphic deoxyribonucleic acid analysis
RAxML	Random Axelerated Maximum Likelihood
RD	Region of difference

REA	Restriction enzyme analysis
RFLP	Restriction fragment length polymorphisms
RL	Reference laboratory
RoI	Republic of Ireland
RGM	Rapid growing mycobacteria
Se	Sensitivity
SGM	Slow growing mycobacteria
SIT	Single intradermal tuberculin
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
Sp	Specificity
TB	Tuberculosis
TLC	Thin-layer chromatography
UK	United Kingdom
USA	United States of America
VCF	Variant call format
VL	Visible lesion
VNTR	Variable number of tandem repeat
WGS	Whole genome sequencing
WHO	World Health Organisation
zTB	Zoonotic TB

Reference list

Abalos, P., Valdivieso, N., Perez de Val, B., Vordermeier, M., Benavides, M.B., Alegria-Moran, R., Saadi, K., Wistuba, M., Ortega, C., Sanchez, N., Retamal, P., 2022. Vaccination of Calves with the *Mycobacterium bovis* BCG Strain Induces Protection against Bovine Tuberculosis in Dairy Herds under a Natural Transmission Setting. Animals (Basel) 12.

Adekambi, T., Berger, P., Raoult, D., Drancourt, M., 2006. *rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. Int J Syst Evol Microbiol 56, 133-143.

Adekambi, T., Colson, P., Drancourt, M., 2003. *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J Clin Microbiol 41, 5699-5708.

Adeogun, A., Omobowale, O., Owuamanam, C., Alaka, O., Taiwo, V., van Soolingen, D., Cadmus, S., 2016. *Mycobacterium tuberculosis* and Dual *M. tuberculosis/M. bovis* Infection as the Cause of Tuberculosis in a Gorilla and a Lioness, Respectively, in Ibadan Zoo, Nigeria. Case Rep Vet Med 2016, 8568237.

Agada, C.A., Adesokan, H.K., Igwe, D., Cadmus, S.I.B., 2014. *Mycobacterium africanum* and nontuberculous mycobacteria from fresh milk of pastoral cattle and soft cheese in Oyo State--implications for public health. Afr J Med Med Sci 43, 7.

Alcaide, F., Amlerova, J., Bou, G., Ceyssens, P.J., Coll, P., Corcoran, D., Fangous, M.S., Gonzalez-Alvarez, I., Gorton, R., Greub, G., Hery-Arnaud, G., Hrabak, J., Ingebretsen, A., Lucey, B., Marekovic, I., Mediavilla-Gradolph, C., Monte, M.R.,

O'Connor, J., O'Mahony, J., Opota, O., O'Reilly, B., Orth-Holler, D., Oviano, M., Palacios, J.J., Palop, B., Pranada, A.B., Quiroga, L., Rodriguez-Temporal, D., Ruiz-Serrano, M.J., Tudo, G., Van den Bossche, A., van Ingen, J., Rodriguez-Sanchez, B., European Study Group on, G., Molecular, D., 2018. How to: identify non-tuberculous *Mycobacterium* species using MALDI-TOF mass spectrometry. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 24, 599-603.

Alcolea-Medina, A., Fernandez, M.T.C., Montiel, N., Garcia, M.P.L., Sevilla, C.D., North, N., Lirola, M.J.M., Wilks, M., 2019. An improved simple method for the identification of Mycobacteria by MALDI-TOF MS (Matrix-Assisted Laser Desorption- Ionization mass spectrometry). Sci Rep 9, 20216.

Alexander, K.A., Laver, P.N., Michel, A.L., Williams, M., van Helden, P.D., Warren, R.M., Gey van Pittius, N.C., 2010. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. Emerg Infect Dis 16, 1296-1299.

Alexander, K.A., Laver, P.N., Williams, M.C., Sanderson, C.E., Kanipe, C., Palmer, M.V., 2018. Pathology of the Emerging *Mycobacterium tuberculosis* Complex Pathogen, *Mycobacterium mungi*, in the Banded Mongoose (*Mungos mungo*). Veterinary pathology 55, 303-309.

Alexander, K.A., Pleydell, E., Williams, M.C., Lane, E.P., Nyange, J.F., Michel, A.L., 2002. *Mycobacterium tuberculosis*: an emerging disease of free-ranging wildlife. Emerg Infect Dis 8, 598-601.

Alexander, K.A., Sanderson, C.E., Larsen, M.H., Robbe-Austerman, S., Williams, M.C., Palmer, M.V., 2016. Emerging Tuberculosis Pathogen Hijacks Social Communication Behavior in the Group-Living Banded Mongoose (*Mungos mungo*). mBio 7.

Alfredsen, S., Saxegaard, F., 1992. An outbreak of tuberculosis in pigs and cattle caused by *Mycobacterium africanum*. The Veterinary record 131, 51-53.

Allen, A.R., Minozzi, G., Glass, E.J., Skuce, R.A., McDowell, S.W., Woolliams, J.A., Bishop, S.C., 2010. Bovine tuberculosis: the genetic basis of host susceptibility. Proc Biol Sci 277, 2737-2745.

Allen, A.R., Skuce, R.A., Byrne, A.W., 2018. Bovine Tuberculosis in Britain and Ireland - A Perfect Storm? the Confluence of Potential Ecological and Epidemiological Impediments to Controlling a Chronic Infectious Disease. Front Vet Sci 5, 109.

Allix, C., Walravens, K., Saegerman, C., Godfroid, J., Supply, P., Fauville-Dufaux, M., 2006. Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS*6110* restriction fragment length polymorphism analysis and spoligotyping. J Clin Microbiol 44, 1951-1962.

Almaw, G., Mekonnen, G.A., Mihret, A., Aseffa, A., Taye, H., Conlan, A.J.K., Gumi, B., Zewude, A., Aliy, A., Tamiru, M., Olani, A., Lakew, M., Sombo, M., Gebre, S., Diguimbaye, C., Hilty, M., Fane, A., Muller, B., Hewinson, R.G., Ellis, R.J., Nunez-Garcia, J., Palkopoulou, E., Abebe, T., Ameni, G., Parkhill, J., Wood, J.L.N., The Ethicobots, C., Berg, S., van Tonder, A.J., 2021. Population structure and transmission of *Mycobacterium bovis* in Ethiopia. Microb Genom 7.

Alvarez, A.H., Estrada-Chavez, C., Flores-Valdez, M.A., 2009a. Molecular findings and approaches spotlighting *Mycobacterium bovis* persistence in cattle. Vet Res 40, 22.

Alvarez, J., de Juan, L., Bezos, J., Romero, B., Saez, J.L., Marques, S., Dominguez, C., Minguez, O., Fernandez-Mardomingo, B., Mateos, A., Dominguez, L., Aranaz, A., 2009b. Effect of paratuberculosis on the diagnosis of bovine tuberculosis in a cattle herd with a mixed infection using interferon-gamma detection assay. Vet Microbiol 135, 389-393.

Alvarez, J., de Juan, L., Bezos, J., Romero, B., Saez, J.L., Reviriego Gordejo, F.J., Briones, V., Moreno, M.A., Mateos, A., Dominguez, L., Aranaz, A., 2008. Interference of paratuberculosis with the diagnosis of tuberculosis in a goat flock with a natural mixed infection. Vet Microbiol 128, 72-80.

Alvarez, J., Perez, A., Bezos, J., Marques, S., Grau, A., Saez, J.L., Minguez, O., de Juan, L., Dominguez, L., 2012. Evaluation of the sensitivity and specificity of bovine tuberculosis diagnostic tests in naturally infected cattle herds using a Bayesian approach. Vet Microbiol 155, 38-43.

Alvarez, J., Perez, A., Marques, S., Bezos, J., Grau, A., de la Cruz, M.L., Romero, B., Saez, J.L., del Rosario Esquivel, M., Martinez Mdel, C., Minguez, O., de Juan, L., Dominguez, L., 2014. Risk factors associated with negative in-vivo diagnostic results in bovine tuberculosis-infected cattle in Spain. BMC Vet Res 10, 14.

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21, 30.

Amato, B., Capucchio, T.M., Biasibetti, E., Mangano, E., Boniotti, B.M., Pacciarini, L.M., Migliore, S., Vitale, M., Fiasconaro, M., Di Marco Lo Presti, V., 2017. Pathology and genetic findings in a rare case of *Mycobacterium caprae* infection in a sow. Vet Microbiol 205, 71-74.

Amato, L., Mulatti, P., Pacciarini, M., Schiavon, E., Zanoni, M., Bonfanti, L., 2019. *Mycobacterium caprae* in a dairy farm in Northeast Italy. Veterinaria italiana 55, 375-379.

Ambrosio, S.R., de Deus Oliveira, E.M., Rodriguez, C.A., Ferreira Neto, J.S., Amaku, M., 2008. Comparison of three decontamination methods for *Mycobacterium bovis* isolation. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology] 39, 241-244.

Ameni, G., Aseffa, A., Engers, H., Young, D., Gordon, S., Hewinson, G., Vordermeier, M., 2007. High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia. Clinical and vaccine immunology : CVI 14, 1356-1361.

Ameni, G., Tadesse, K., Hailu, E., Deresse, Y., Medhin, G., Aseffa, A., Hewinson, G., Vordermeier, M., Berg, S., 2013. Transmission of *Mycobacterium tuberculosis* between farmers and cattle in central Ethiopia. PLoS One 8, e76891.

Ameni, G., Vordermeier, M., Firdessa, R., Aseffa, A., Hewinson, G., Gordon, S.V., Berg, S., 2011. *Mycobacterium tuberculosis* infection in grazing cattle in central Ethiopia. Veterinary journal (London, England : 1997) 188, 359-361.

Andersen, P., Askgaard, D., Gottschau, A., Bennedsen, J., Nagai, S., Heron, I., 1992. Identification of immunodominant antigens during infection with *Mycobacterium tuberculosis*. Scandinavian journal of immunology 36, 823-831.

Andersen, P., Askgaard, D., Ljungqvist, L., Bennedsen, J., Heron, I., 1991. Proteins released from *Mycobacterium tuberculosis* during growth. Infection and immunity 59, 1905-1910.

Anhalt, J.P., Fenselau, C., 2002. Identification of bacteria using mass spectrometry. Analytical Chemistry 47, 219-225.

Anon., 2013. Working Document on Eradication of Bovine Tuberculosis in the EU Accepted by the Bovine Tuberculosis Subgroup of the Task Force on Monitoring Animal Disease Eradication. Unit G5 -Veterinary Programmes, Brussels, p. 20.

Aranaz, A., Cousins, D., Mateos, A., Dominguez, L., 2003. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. Int J Syst Evol Microbiol 53, 1785-1789.

Aranaz, A., De Juan, L., Bezos, J., Alvarez, J., Romero, B., Lozano, F., Paramio, J.L., Lopez-Sanchez, J., Mateos, A., Dominguez, L., 2006. Assessment of diagnostic

tools for eradication of bovine tuberculosis in cattle co-infected with *Mycobacterium bovis* and *M. avium* subsp. *paratuberculosis*. Vet Res 37, 593-606.

Aranaz, A., Liebana, E., Gomez-Mampaso, E., Galan, J.C., Cousins, D., Ortega, A., Blazquez, J., Baquero, F., Mateos, A., Suarez, G., Dominguez, L., 1999. *Mycobacterium tuberculosis* subsp. *caprae* subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. Int J Syst Bacteriol 49 Pt 3, 1263-1273.

Aranaz, A., Liébana, E., Mateos, A., Domı□nguez, L., Cousins, D., 1998. Restriction fragment length polymorphism and spacer oligonucleotide typing: A comparative analysis of fingerprinting strategies for *Mycobacterium bovis*. Veterinary Microbiology 61, 311-324.

Aranaz, A., Liebana, E., Mateos, A., Dominguez, L., Vidal, D., Domingo, M., Gonzolez, O., Rodriguez-Ferri, E.F., Bunschoten, A.E., Van Embden, J.D., Cousins, D., 1996a. Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. J Clin Microbiol 34, 2734-2740.

Aranaz, A., Liebana, E., Pickering, X., Novoa, C., Mateos, A., Dominguez, L., 1996b. Use of polymerase chain reaction in the diagnosis of tuberculosis in cats and dogs. The Veterinary record 138, 276-280.

Araújo, C.P., Osório, A.L., Jorge, K.S., Ramos, C.A., Filho, A.F., Vidal, C.E., Roxo, E., Nishibe, C., Almeida, N.F., Júnior, A.A., Silva, M.R., Neto, J.D., Cerqueira, V.D., Zumárraga, M.J., Araújo, F.R., 2014a. Detection of *Mycobacterium bovis* in bovine and bubaline tissues using nested-PCR for TbD1. PLoS One 9, e91023.

Araújo, C.P., Osório, A.L., Jorge, K.S., Ramos, C.A., Souza Filho, A.F., Vidal, C.E., Vargas, A.P., Roxo, E., Rocha, A.S., Suffys, P.N., Fonseca, A.A., Jr., Silva, M.R., Barbosa Neto, J.D., Cerqueira, V.D., Araújo, F.R., 2014b. Direct detection of *Mycobacterium tuberculosis* complex in bovine and bubaline tissues through nested-PCR. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology] 45, 633-640.

Araújo, C.P., Osório, A.L.A.R., Jorge, K.S.G., Ramos, C.A.N., Filho, A.F.S., Vidal, C.E.S., Roxo, E., Nishibe, C., Almeida, N.F., António A. F. Júnior, M.R.S., José Diomedes B. Neto, Cerqueira, V.D., Martín J. Zumárraga9, F.R.A., 2014a. Detection of *Mycobacterium bovis* in Bovine and Bubaline Tissues Using Nested-PCR for TbD1. PLOS One 9.

Araújo, C.P., Osório, A.L.A.R., Jorge, K.S.G., Ramos, C.A.N., Filho, A.F.S., Vidal, C.E.S., Vargas, A.P.C., Roxo, E., Rocha, A.S., Suffys, P.N., Júnior, A.A.F., Silva, M.R., Neto, J.D.B., Cerqueira, V.D., Araújo, F.R., 2014b. Direct detection of *Mycobacterium tuberculosis* complex in bovine and bubaline tissues through nested-PCR. Brazilian Journal of Microbiology 45, 633-640.

Armstrong, D.T., Parrish, N., 2021. Current Updates on Mycobacterial Taxonomy, 2018 to 2019. J Clin Microbiol 59, e0152820.

Arnot, L.F., Michel, A., 2020. Challenges for controlling bovine tuberculosis in South Africa. The Onderstepoort journal of veterinary research 87, e1-e8.

Asante-Poku, A., Aning, K.G., Boi-Kikimoto, B., Yeboah-Manu, D., 2014. Prevalence of bovine tuberculosis in a dairy cattle farm and a research farm in Ghana. Onderstepoort J Vet Res 81, E1-6.

Atkins, P.J., 2016. 'Your enemy the cow': the construction of early medical and veterinary knowledge about bovine tuberculosis, In: A history of uncertainty: bovine tuberculosis in Britain, 1850 to the present. Winchester University Press.

Bacanelli, G., Olarte, L.C., Silva, M.R., Rodrigues, R.A., Carneiro, P.A.M., Kaneene, J.B., Pasquatti, T.N., Takatani, H., Zumárraga, M.J., Etges, R.N., Araújo, F.R., Verbisck, N.V., 2019. Matrix Assisted Laser Desorption Ionization-Time-of-Flight mass

spectrometry identification of *Mycobacterium bovis* in *Bovinae*. The Journal of veterinary medical science 81, 1400-1408.

Bailey, S.S., Crawshaw, T.R., Smith, N.H., Palgrave, C.J., 2013. *Mycobacterium bovis* infection in domestic pigs in Great Britain. Veterinary journal (London, England : 1997) 198, 391-397.

Bajaj, A.O., Saraswat, S., Knuuttila, J.E.A., Freeke, J., Stielow, J.B., Barker, A.P., 2021. Accurate Identification of Closely Related *Mycobacterium tuberculosis* Complex Species by High Resolution Tandem Mass Spectrometry. Frontiers in cellular and infection microbiology 11, 656880.

Bajaj, A.O., Slechta, E.S., Barker, A.P., 2022. Rapid and Accurate Differentiation of *Mycobacteroides abscessus* Complex Species by Liquid Chromatography-Ultra-High-Resolution Orbitrap Mass Spectrometry. Frontiers in cellular and infection microbiology 12, 809348.

Balada-Llasat, J.M., Kamboj, K., Pancholi, P., 2013. Identification of mycobacteria from solid and liquid media by matrix-assisted laser desorption ionization-time of flight mass spectrometry in the clinical laboratory. J Clin Microbiol 51, 2875-2879.

Barbier, E., Boschiroli, M.L., Gueneau, E., Rochelet, M., Payne, A., de Cruz, K., Blieux, A.L., Fossot, C., Hartmann, A., 2016. First molecular detection of *Mycobacterium bovis* in environmental samples from a French region with endemic bovine tuberculosis. J Appl Microbiol 120, 1193-1207.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Bayraktar, B., Togay, A., 2011. *Mycobacterium caprae* Causing Lymphadenitis in a Child. The Pediatric infectious disease journal 30, 2.

Berg, S., Garcia-Pelayo, M.C., Muller, B., Hailu, E., Asiimwe, B., Kremer, K., Dale, J., Boniotti, M.B., Rodriguez, S., Hilty, M., Rigouts, L., Firdessa, R., Machado, A., Mucavele, C., Ngandolo, B.N., Bruchfeld, J., Boschiroli, L., Muller, A., Sahraoui, N., Pacciarini, M., Cadmus, S., Joloba, M., van Soolingen, D., Michel, A.L., Djonne, B., Aranaz, A., Zinsstag, J., van Helden, P., Portaels, F., Kazwala, R., Kallenius, G., Hewinson, R.G., Aseffa, A., Gordon, S.V., Smith, N.H., 2011. African 2, a clonal complex of *Mycobacterium bovis* epidemiologically important in East Africa. J Bacteriol 193, 670-678.

Bermúdez, H.R., Rentería, E.T., Medina, B.G., Hori-Oshima, S., De la Mora Valle, A., López, V.G., 2012. Evaluation of a lateral flow assay for the diagnosis of *Mycobacterium bovis* infection in dairy cattle. Journal of immunoassay & immunochemistry 33, 59-65.

Bezos, J., Alvarez, J., de Juan, L., Romero, B., Rodríguez, S., Fernández-de-Mera, I.G., Hewinson, R.G., Vordermeier, M., Mateos, A., Domínguez, L., Aranaz, A., 2011a. Assessment of in vivo and in vitro tuberculosis diagnostic tests in *Mycobacterium caprae* naturally infected caprine flocks. Prev Vet Med 100, 187-192.

Bezos, J., Alvarez, J., Juan, L., Romero, B., Rodriguez, S., Castellanos, E., Saez-Llorente, J.L., Mateos, A., Dominguez, L., Aranaz, A., 2011b. Factors influencing the performance of an interferon-gamma assay for the diagnosis of tuberculosis in goats. Veterinary journal (London, England : 1997) 190, 131-135.

Bezos, J., Alvarez, J., Minguez, O., Marques, S., Martin, O., Vigo, V., Pieltain, C., Romero, B., Rodriguez, S., Casal, C., Mateos, A., Dominguez, L., de Juan, L., 2012a. Evaluation of specificity of tuberculosis diagnostic assays in caprine flocks under different epidemiological situations. Res Vet Sci 93, 636-640.

Bezos, J., Alvarez, J., Romero, B., Aranaz, A., Juan, L., 2012b. Tuberculosis in goats: assessment of current in vivo cell-mediated and antibody-based diagnostic assays. Veterinary journal (London, England : 1997) 191, 161-165.

Bezos, J., Casal, C., Romero, B., Schroeder, B., Hardegger, R., Raeber, A.J., Lopez, L., Rueda, P., Dominguez, L., 2014. Current *ante-mortem* techniques for diagnosis of bovine tuberculosis. Res Vet Sci 97 Suppl, S44-52.

Bhalla, G.S., Sarao, M.S., Kalra, D., Bandyopadhyay, K., John, A.R., 2018. Methods of phenotypic identification of non-tuberculous mycobacteria. Pract Lab Med 12, e00107.

Bibb, L.A., Hatfull, G.F., 2002. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. Mol Microbiol 45, 1515-1526.

Biek, R., O'Hare, A., Wright, D., Mallon, T., McCormick, C., Orton, R.J., McDowell, S., Trewby, H., Skuce, R.A., Kao, R.R., 2012. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. PLoS Pathog 8, e1003008.

Biet, F., Boschiroli, M.L., 2014. Non-tuberculous mycobacterial infections of veterinary relevance. Res Vet Sci 97 Suppl, S69-77.

Biet, F., Boschiroli, M.L., Thorel, M.F., Guilloteau, L.A., 2005. Zoonotic aspects of *Mycobacterium bovis* and *Mycobacterium avium-intracellulare* complex (MAC). Vet Res 36, 411-436.

Blackwood, K.S., He, C., Gunton, J., Turenne, C.Y., Wolfe, J., Kabani, A.M., 2000. Evaluation of *recA* sequences for identification of *Mycobacterium* species. J Clin Microbiol 38, 2846-2852.

Blas, I.d., Ruiz-Zarzuela, I., Vallejo, A., 2006. WinEpi: Working In Epidemiology. An Online Epidemiological Tool, In: Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, Australia.

Blouin, Y., Cazajous, G., Dehan, C., Soler, C., Vong, R., Hassan, M.O., Hauck, Y., Boulais, C., Andriamanantena, D., Martinaud, C., Martin, E., Pourcel, C., Vergnaud, G., 2014. Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg Infect Dis 20, 21-28.

Bogaerts, B., Delcourt, T., Soetaert, K., Boarbi, S., Ceyssens, P.J., Winand, R., Van Braekel, J., De Keersmaecker, S.C.J., Roosens, N.H.C., Marchal, K., Mathys, V., Vanneste, K., 2021. A Bioinformatics Whole-Genome Sequencing Workflow for Clinical *Mycobacterium tuberculosis* Complex Isolate Analysis, Validated Using a Reference Collection Extensively Characterized with Conventional Methods and In Silico Approaches. J Clin Microbiol 59.

Bolado-Martinez, E., Benavides-Davila, I., Candia-Plata Mdel, C., Navarro-Navarro, M., Aviles-Acosta, M., Alvarez-Hernandez, G., 2015. Proposal of a Screening MIRU-VNTR Panel for the Preliminary Genotyping of *Mycobacterium bovis* in Mexico. BioMed research international 2015, 416479.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Boniotti, M.B., Gaffuri, A., Gelmetti, D., Tagliabue, S., Chiari, M., Mangeli, A., Spisani, M., Nassuato, C., Gibelli, L., Sacchi, C., Zanoni, M., Pacciarini, M.L., 2014. Detection and molecular characterization of *Mycobacterium microti* isolates in wild boar from northern Italy. J Clin Microbiol 52, 2834-2843.

Boniotti, M.B., Goria, M., Loda, D., Garrone, A., Benedetto, A., Mondo, A., Tisato, E., Zanoni, M., Zoppi, S., Dondo, A., Tagliabue, S., Bonora, S., Zanardi, G., Pacciarini, M.L., 2009. Molecular typing of *Mycobacterium bovis* strains isolated in Italy from 2000 to 2006 and evaluation of variable-number tandem repeats for geographically optimized genotyping. J Clin Microbiol 47, 636-644.

Boritsch, E.C., Frigui, W., Cascioferro, A., Malaga, W., Etienne, G., Laval, F., Pawlik, A., Le Chevalier, F., Orgeur, M., Ma, L., Bouchier, C., Stinear, T.P., Supply, P., Majlessi, L., Daffe, M., Guilhot, C., Brosch, R., 2016a. *pks5*-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. Nat Microbiol 1, 15019.

Boritsch, E.C., Khanna, V., Pawlik, A., Honore, N., Navas, V.H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T.P., Brosch, R., 2016b. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc Natl Acad Sci U S A 113, 9876-9881.

Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514, 494-497.

Bottai, D., Stinear, T.P., Supply, P., Brosch, R., 2014. Mycobacterial Pathogenomics and Evolution. Microbiol Spectr 2, MGM2-0025-2013.

Bouzinbi, N., Marcy, O., Bertolotti, T., Chiron, R., Bemer, P., Pestel-Caron, M., Peuchant, O., Guet-Revillet, H., Fangous, M.S., Hery-Arnaud, G., Ouedraogo, A.S., Banuls, A.L., Godreuil, S., 2020. Evaluation of the GenoType NTM-DR assay performance for the identification and molecular detection of antibiotic resistance in *Mycobacterium abscessus* complex. PLoS One 15, e0239146.

Bovine Tuberculosis Subgroup of the Task Force, 2013. Working Document on Eradication of Bovine Tuberculosis in the EU. European Commission, Health & Consumer protection Directorate. General. Veterinary and International Affairs. Unit G5 -Veterinary Programmes. Brussels.

Branger, M., Loux, V., Cochard, T., Boschiroli, M.L., Biet, F., Michelet, L., 2020. The complete genome sequence of *Mycobacterium bovis* Mb3601, a SB0120 spoligotype strain representative of a new clonal group. Infect Genet Evol 82, 104309.

Brites, D., Gagneux, S., 2017. The Nature and Evolution of Genomic Diversity in the *Mycobacterium tuberculosis* Complex, In: Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control.

Brites, D., Loiseau, C., Menardo, F., Borrell, S., Boniotti, M.B., Warren, R., Dippenaar, A., Parsons, S.D.C., Beisel, C., Behr, M.A., Fyfe, J.A., Coscolla, M., Gagneux, S., 2018. A New Phylogenetic Framework for the Animal-Adapted *Mycobacterium tuberculosis* Complex. Front Microbiol 9, 2820.

Brodin, P., Majlessi, L., Brosch, R., Smith, D., Bancroft, G., Clark, S., Williams, A., Leclerc, C., Cole, S.T., 2004a. Enhanced protection against tuberculosis by vaccination with recombinant *Mycobacterium microti* vaccine that induces T cell immunity against region of difference 1 antigens. J Infect Dis 190, 115-122.

Brodin, P., Rosenkrands, I., Andersen, P., Cole, S.T., Brosch, R., 2004b. ESAT-6 proteins: protective antigens and virulence factors? Trends Microbiol 12, 500-508.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T., 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A 99, 3684-3689.

Bruning-Fann, C.S., Robbe-Austerman, S., Kaneene, J.B., Thomsen, B.V., Tilden, J.D., Ray, J.S., Smith, R.W., Fitzgerald, S.D., Bolin, S.R., O'Brien, D.J., Mullaney, T.P., Stuber, T.P., Averill, J.J., Marks, D., 2017. Use of whole-genome sequencing and evaluation of the apparent sensitivity and specificity of *antemortem* tuberculosis tests in the investigation of an unusual outbreak of *Mycobacterium bovis* infection in a Michigan dairy herd. Javma-J Am Vet Med A 251, 206-216.

Bruning-Fann, C.S., Schmitt, S.M., Fitzgerald, S.D., Fierke, J.S., Friedrich, P.D., Kaneene, J.B., Clarke, K.A., Butler, K.L., Payeur, J.B., Whipple, D.L., Cooley, T.M., Miller, J.M.,

Muzo, D.P., 2001. Bovine tuberculosis in free-ranging carnivores from Michigan. Journal of wildlife diseases 37, 58-64.
Buchan, B.W., Ledeboer, N.A., 2013. Advances in identification of clinical yeast isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 51, 1359-1366.
Buddle, B.M., Vordermeier, H.M., Chambers, M.A., de Klerk-Lorist, L.M., 2018. Efficacy and Safety of BCG Vaccine for Control of Tuberculosis in Domestic Livestock and Wildlife. Front Vet Sci 5, 259.
Burthe, S., Bennett, M., Kipar, A., Lambin, X., Smith, A., Telfer, S., Begon, M., 2008. Tuberculosis (*Mycobacterium microti*) in wild field vole populations. Parasitology 135, 309-317.
Bushnell, B., 2014. BBMap, Joint Genome Institute.
Butler, W.R., Guthertz, L.S., 2001. Mycolic acid analysis by high-performance liquid chromatography for identification of *Mycobacterium* species. Clin Microbiol Rev 14, 704-726, table of contents.
Cadmus, S., Palmer, S., Okker, M., Dale, J., Gover, K., Smith, N., Jahans, K., Hewinson, R.G., Gordon, S.V., 2006. Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. J Clin Microbiol 44, 29-34.
Cadmus, S.I., Diarra, B., Traore, B., Maiga, M., Siddiqui, S., Tounkara, A., Falodun, O., Lawal, W., Adewole, I.F., Murphy, R., van Soolingen, D., Taiwo, B., 2016. Nontuberculous Mycobacteria Isolated from Tuberculosis Suspects in Ibadan, Nigeria. Journal of pathogens 2016, 6547363.
Cadmus, S.I., Yakubu, M.K., Magaji, A.A., Jenkins, A.O., van Soolingen, D., 2010. *Mycobacterium bovis*, but also *M. africanum* present in raw milk of pastoral cattle in north-central Nigeria. Tropical animal health and production 42, 1047-1048.
Caffrey, J.P., 1994. Status of bovine tuberculosis eradication programmes in Europe. Veterinary Microbiology 40, 1-4.
Caimi, K., Romano, M.I., Alito, A., Zumarraga, M., Bigi, F., Cataldi, A., 2001. Sequence analysis of the direct repeat region in *Mycobacterium bovis*. J Clin Microbiol 39, 1067-1072.
Cambau, E., Drancourt, M., 2014. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 20, 196-201.
Canchaya, C., Proux, C., Fournous, G., Bruttin, A., Brussow, H., 2003. Prophage genomics. Microbiol Mol Biol Rev 67, 238-276, table of contents.
Casal, C., Bezos, J., Diez-Guerrier, A., Alvarez, J., Romero, B., de Juan, L., Rodriguez-Campos, S., Vordermeier, M., Whelan, A., Hewinson, R.G., Mateos, A., Dominguez, L., Aranaz, A., 2012. Evaluation of two cocktails containing ESAT-6, CFP-10 and Rv-3615c in the intradermal test and the interferon-gamma assay for diagnosis of bovine tuberculosis. Prev Vet Med 105, 149-154.
Casal, C., Diez-Guerrier, A., Alvarez, J., Rodriguez-Campos, S., Mateos, A., Linscott, R., Martel, E., Lawrence, J.C., Whelan, C., Clarke, J., O'Brien, A., Dominguez, L., Aranaz, A., 2014. Strategic use of serology for the diagnosis of bovine tuberculosis after intradermal skin testing. Vet Microbiol 170, 342-351.
Casal, C., Infantes, J.A., Risalde, M.A., Diez-Guerrier, A., Dominguez, M., Moreno, I., Romero, B., de Juan, L., Saez, J.L., Juste, R., Gortazar, C., Dominguez, L., Bezos, J., 2017. Antibody detection tests improve the sensitivity of tuberculosis diagnosis in cattle. Res Vet Sci 112, 214-221.
Cassidy, J.P., 2006. The pathogenesis and pathology of bovine tuberculosis with insights from studies of tuberculosis in humans and laboratory animal models. Vet Microbiol 112, 151-161.

Castets, M., Boisvert, H., Grumbach, F., Brunel, M., Rist, N., 1968. [Tuberculosis bacilli of the African type: preliminary note]. Rev Tuberc Pneumol (Paris) 32, 179-184.

Cavanagh, R., Begon, M., Bennett, M., Ergon, T., Graham, I.M., De Haas, P.E., Hart, C.A., Koedam, M., Kremer, K., Lambin, X., Roholl, P., Soolingen Dv, D., 2002. *Mycobacterium microti* infection (vole tuberculosis) in wild rodent populations. J Clin Microbiol 40, 3281-3285.

Chen, C.Y., Weng, J.Y., Huang, H.H., Yen, W.C., Tsai, Y.H., Cheng, T.C., Jou, R., 2019. A new oligonucleotide array for the detection of multidrug and extensively drug-resistance tuberculosis. Sci Rep 9, 4425.

Chen, H., 2022. Package 'VennDiagram', 1.7.3 ed.

Chen, X., Yang, Q., Kong, H., Chen, Y., 2012. Real-time PCR and Amplified MTD(R) for rapid detection of *Mycobacterium tuberculosis* in pulmonary specimens. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease 16, 235-239.

Chiari, M., Ferrari, N., Giardiello, D., Avisani, D., Pacciarini, M.L., Alborali, L., Zanoni, M., Boniotti, M.B., 2016. Spatiotemporal and Ecological Patterns of *Mycobacterium microti* Infection in Wild Boar (*Sus scrofa*). Transbound Emerg Dis 63, e381-388.

Ciaravino, G., Vidal, E., Cortey, M., Martin, M., Sanz, A., Mercader, I., Perea, C., Robbe-Austerman, S., Allepuz, A., Perez de Val, B., 2021. Phylogenetic relationships investigation of *Mycobacterium caprae* strains from sympatric wild boar and goats based on whole genome sequencing. Transbound Emerg Dis 68, 1476-1486.

Clark, A.E., Kaleta, E.J., Arora, A., Wolk, D.M., 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev 26, 547-603.

Claydon, M.A., Davey, S.N., Edwards-Jones, V., Gordon, D.B., 1996. The rapid identification of intact microorganisms using mass spectrometry. Nat Biotechnol 14, 1584-1586.

Clegg, T.A., Duignan, A., Whelan, C., Gormley, E., Good, M., Clarke, J., Toft, N., More, S.J., 2011. Using latent class analysis to estimate the test characteristics of the gamma-interferon test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions. Vet Microbiol 151, 68-76.

Coad, M., Hewinson, R.G., Clifford, D., Vordermeier, H.M., Whelan, A.O., 2007. Influence of skin testing and blood storage on interferon-gamma production in cattle affected naturally with *Mycobacterium bovis*. The Veterinary record 160, 660-662.

Coetsier, C., Vannuffel, P., Blondeel, N., Denef, J.F., Cocito, C., Gala, J.L., 2000. Duplex PCR for differential identification of *Mycobacterium bovis*, *M. avium*, and *M. avium* subsp. *paratuberculosis* in formalin- fixed paraffin-embedded tissues from cattle. J Clin Microbiol 38, 3048-3054.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537-544.

Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigao, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nature communications 5, 4812.

Collins, D.M., Radford, A.J., Delisle, G.W., Billmanjacobe, H., 1994. Diagnosis and Epidemiology of Bovine Tuberculosis Using Molecular Biological Approaches. Veterinary Microbiology 40, 83-94.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet 45, 1176-1182.

Comas, I., Homolka, S., Niemann, S., Gagneux, S., 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4, e7815.

Corner, L.A., 1994. *Post mortem* diagnosis of *Mycobacterium bovis* infection in cattle. Veterinary Microbiology 40, 53-63.

Corner, L.A., Gormley, E., Pfeiffer, D.U., 2012. Primary isolation of *Mycobacterium bovis* from bovine tissues: conditions for maximising the number of positive cultures. Vet Microbiol 156, 162-171.

Corner, L.A., Murphy, D., Gormley, E., 2011. *Mycobacterium bovis* infection in the Eurasian badger (*Meles meles*): the disease, pathogenesis, epidemiology and control. Journal of comparative pathology 144, 1-24.

Corner, L.A., Pearson, C.W., 1978. Response of cattle to inoculation with atypical mycobacteria of bovine origin. Aust Vet J 54, 379-382.

Corner, L.A., Trajstman, A.C., 1988. An Evaluation of 1-Hexadecylpyridinium Chloride as a Decontaminant in the Primary Isolation of *Mycobacterium-Bovis* from Bovine Lesions. Veterinary Microbiology 18, 127-134.

Corner, L.A., Trajstman, A.C., Lund, K., 1995. Determination of the optimum concentration of decontaminants for the primary isolation of *Mycobacterium bovis*. N Z Vet J 43, 129-133.

Coros, A., DeConno, E., Derbyshire, K.M., 2008. IS*6110*, a *Mycobacterium tuberculosis* complex-specific insertion sequence, is also present in the genome of *Mycobacterium smegmatis*, suggestive of lateral gene transfer among mycobacterial species. J Bacteriol 190, 3408-3410.

Coscolla, M., Lewin, A., Metzger, S., Maetz-Rennsing, K., Calvignac-Spencer, S., Nitsche, A., Dabrowski, P.W., Radonic, A., Niemann, S., Parkhill, J., Couacy-Hymann, E., Feldman, J., Comas, I., Boesch, C., Gagneux, S., Leendertz, F.H., 2013. Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. Emerg Infect Dis 19, 969-976.

Costa-Alcalde, J.J., Barbeito-Castineiras, G., Gonzalez-Alba, J.M., Aguilera, A., Galan, J.C., Perez-Del-Molino, M.L., 2019. Comparative evaluation of the identification of rapidly growing non-tuberculous mycobacteria by mass spectrometry (MALDI-TOF MS), GenoType *Mycobacterium* CM/AS assay and partial sequencing of the *rpob* gene with phylogenetic analysis as a reference method. Enferm Infecc Microbiol Clin (Engl Ed) 37, 160-166.

Costa, P., Botelho, A., Couto, I., Viveiros, M., Inacio, J., 2014. Standing of nucleic acid testing strategies in veterinary diagnosis laboratories to uncover *Mycobacterium tuberculosis* complex members. Front Mol Biosci 1, 16.

Costa, P., Ferreira, A.S., Amaro, A., Albuquerque, T., Botelho, A., Couto, I., Cunha, M.V., Viveiros, M., Inacio, J., 2013. Enhanced detection of tuberculous mycobacteria in animal tissues using a semi-nested probe-based real-time PCR. PLoS One 8, e81337.

Costello, E., O'Grady, D., Flynn, O., O'Brien, R., Rogers, M., Quigley, F., Egan, J., Griffin, J., 1999. Study of restriction fragment length polymorphism analysis and

spoligotyping for epidemiological investigation of *Mycobacterium bovis* infection. J Clin Microbiol 37, 3217-3222.

Courcoul, A., Moyen, J.L., Brugère, L., Faye, S., Hénault, S., Gares, H., Boschiroli, M.L., 2014. Estimation of sensitivity and specificity of bacteriology, histopathology and PCR for the confirmatory diagnosis of bovine tuberculosis using latent class analysis. PLoS One 9, e90334.

Cousins, D., 2001. *Mycobacterium bovis* infection and control in domestic livestock. Revue Scientifique et Technique 20, 15.

Cousins, D., Williams, S., Liebana, E., Aranaz, A., Bunschoten, A., Van Embden, J., Ellis, T., 1998. Evaluation of four DNA typing techniques in epidemiological investigations of bovine tuberculosis. J Clin Microbiol 36, 168-178.

Cousins, D.V., Bastida, R., Cataldi, A., Quse, V., Redrobe, S., Dow, S., Duignan, P., Murray, A., Dupont, C., Ahmed, N., Collins, D.M., Butler, W.R., Dawson, D., Rodriguez, D., Loureiro, J., Romano, M.I., Alito, A., Zumarraga, M., Bernardelli, A., 2003. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. Int J Syst Evol Microbiol 53, 1305-1314.

Cousins, D.V., Peet, R.L., Gaynor, W.T., Williams, S.N., Gow, B.L., 1994. Tuberculosis in imported hyrax (*Procavia capensis*) caused by an ususual variant belonging to the *Mycobacterium tuberculosis* complex. Veterinary Microbiology 42, 135-145.

Cousins, D.V., Roberts, J.L., 2001. Australia's campaign to eradicate bovine tuberculosis: the battle for freedom and beyond. Tuberculosis (Edinburgh, Scotland) 81, 5-15.

Cousins, D.V., Williams, S.N., Reuter, R., Forshaw, D., Chadwick, B., Coughran, D., Collins, P., Gales, N., 1993. Tuberculosis in wild seals and characterisation of the seal bacillus. Aust Vet J 70, 92-97.

Cousins, D.V., Wilton, S.D., Francis, B.R., 1991. Use of DNA amplification for the rapid identification of *Mycobacterium bovis*. Veterinary Microbiology 27, 187-195.

Cowan, L.S., Diem, L., Brake, M.C., Crawford, J.T., 2004. Transfer of a *Mycobacterium tuberculosis* genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. J Clin Microbiol 42, 474-477.

Crawshaw, T., Daniel, R., Clifton-Hadley, R., Clark, J., Evans, H., Rolfe, S., de la Rua-Domenech, R., 2008. TB in goats caused by *Mycobacterium bovis*. The Veterinary record 163, 127.

Crawshaw, T.R., Chanter, J.I., McGoldrick, A., Line, K., 2014. A proof of concept study to assess the potential of PCR testing to detect natural *Mycobacterium bovis* infection in South American camelids. Irish veterinary journal 67, 5.

Crispell, J., Balaz, D., Gordon, S.V., 2019a. HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. Microb Genom 5.

Crispell, J., Benton, C.H., Balaz, D., De Maio, N., Ahkmetova, A., Allen, A., Biek, R., Presho, E.L., Dale, J., Hewinson, G., Lycett, S.J., Nunez-Garcia, J., Skuce, R.A., Trewby, H., Wilson, D.J., Zadoks, R.N., Delahay, R.J., Kao, R.R., 2019b. Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. Elife 8.

Crispell, J., Cassidy, S., Kenny, K., McGrath, G., Warde, S., Cameron, H., Rossi, G., MacWhite, T., White, P.C.L., Lycett, S., Kao, R.R., Moriarty, J., Gordon, S.V., 2020. *Mycobacterium bovis* genomics reveals transmission of infection between cattle and deer in Ireland. Microb Genom 6.

Crispell, J., Zadoks, R.N., Harris, S.R., Paterson, B., Collins, D.M., de-Lisle, G.W., Livingstone, P., Neill, M.A., Biek, R., Lycett, S.J., Kao, R.R., Price-Carter, M., 2017.

Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. BMC genomics 18, 180.
Croxatto, A., Prod'hom, G., Greub, G., 2012. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev 36, 380-407.
da Silva Rabello, M.C., de Oliveira, R.S., Silva, R.M., Leao, S.C., 2010. Natural occurrence of horizontal transfer of *Mycobacterium avium*-specific insertion sequence IS*1245* to *Mycobacterium kansasii*. J Clin Microbiol 48, 2257-2259.
Dai, J., Chen, Y., Lauzardo, M., 2011. Web-accessible database of *hsp65* sequences from *Mycobacterium* reference strains. J Clin Microbiol 49, 2296-2303.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158.
David, H.L., Jahan, M.T., Jumin, A., Grandry, J., Lehman, E.H., 1978. Numerical Taxonomy Analysis of *Mycobacterium africanum*. International Journal of Systematic Bacteriology 28, 464-472.
Davidson, J.A., Loutet, M.G., O'Connor, C., Kearns, C., Smith, R.M., Lalor, M.K., Thomas, H.L., Abubakar, I., Zenner, D., 2017. Epidemiology of *Mycobacterium bovis* Disease in Humans in England, Wales, and Northern Ireland, 2002-2014. Emerg Infect Dis 23, 377-386.
Dawson, K.L., Bell, A., Kawakami, R.P., Coley, K., Yates, G., Collins, D.M., 2012. Transmission of *Mycobacterium orygis* (*M. tuberculosis* complex species) from a tuberculosis patient to a dairy cow in New Zealand. J Clin Microbiol 50, 3136-3138.
de Azevedo Issa, M., Martins Soares Filho, P., Fonseca Junior, A.A., Arrais Hodon, M., Cristian Dos Santos, L., Karlisson Pimenta Dos Reis, J., Cerqueira Leite, R., 2017. Comparative study of *Mycobacterium bovis* primary isolation methods. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology] 48, 139-144.
de Jong, B.C., Antonio, M., Gagneux, S., 2010. *Mycobacterium africanum*--review of an important cause of human tuberculosis in West Africa. PLoS neglected tropical diseases 4, e744.
de la Cruz, M.L., Branscum, A.J., Nacar, J., Pages, E., Pozo, P., Perez, A., Grau, A., Saez, J.L., de Juan, L., Diaz, R., Minguez, O., Alvarez, J., 2018. Evaluation of the Performance of the IDvet IFN-Gamma Test for Diagnosis of Bovine Tuberculosis in Spain. Front Vet Sci 5, 229.
de la Rua-Domenech, R., Goodchild, A.T., Vordermeier, H.M., Hewinson, R.G., Christiansen, K.H., Clifton-Hadley, R.S., 2006. *Ante mortem* diagnosis of tuberculosis in cattle: a review of the tuberculin tests, gamma-interferon assay and other ancillary diagnostic techniques. Res Vet Sci 81, 190-210.
de Lisle, G.W., Havill, P.F., 1985. Mycobacteria isolated from deer in New Zealand from 1970-1983. N Z Vet J 33, 138-140.
De Maio, F., Berisio, R., Manganelli, R., Delogu, G., 2020. PE_PGRS proteins of *Mycobacterium tuberculosis*: A specialized molecular task force at the forefront of host-pathogen interaction. Virulence 11, 898-915.
de Val, B.P., Sanz, A., Soler, M., Allepuz, A., Michelet, L., Boschiroli, M.L., Vidal, E., 2019. *Mycobacterium microti* Infection in Free-Ranging Wild Boar, Spain, 2017-2019. Emerg Infect Dis 25, 2152-2154.
Deforges, L., Boulouis, H.J., Thibaud, J.L., Boulouha, L., Sougakoff, W., Blot, S., Hewinson, G., Truffot-Pernot, C., Haddad, N., 2004. First isolation of *Mycobacterium microti* (Llama-type) from a dog. Vet Microbiol 103, 249-253.
Delahay, R.J., Smith, G.C., Barlow, A.M., Walker, N., Harris, A., Clifton-Hadley, R.S., Cheeseman, C.L., 2007. Bovine tuberculosis infection in wild mammals in the

South-West region of England: a survey of prevalence and a semi-quantitative assessment of the relative risks to cattle. Veterinary journal (London, England : 1997) 173, 287-301.

Delibes-Mateos, M., Farfan, M.A., Olivero, J., Marquez, A.L., Vargas, J.M., 2009. Long-term changes in game species over a long period of transformation in the Iberian Mediterranean landscape. Environ Manage 43, 1256-1268.

Didkowska, A., Krajewska-Wedzina, M., Bielecki, W., Brzezinska, S., Augustynowicz-Kopec, E., Olech, W., Anusz, K., Sridhara, A.A., Johnathan-Lee, A., Elahi, R., Miller, M.A., Ray Waters, W., Lyashchenko, K.P., 2021. Antibody responses in European bison (*Bison bonasus*) naturally infected with *Mycobacterium caprae*. Vet Microbiol 253, 108952.

Dippenaar, A., Parsons, S.D.C., Miller, M.A., Hlokwe, T., Gey van Pittius, N.C., Adroub, S.A., Abdallah, A.M., Pain, A., Warren, R.M., Michel, A.L., van Helden, P.D., 2017. Progenitor strain introduction of *Mycobacterium bovis* at the wildlife-livestock interface can lead to clonal expansion of the disease in a single ecosystem. Infect Genet Evol 51, 235-238.

Dippenaar, A., Parsons, S.D.C., Sampson, S.L., van der Merwe, R.G., Drewe, J.A., Abdallah, A.M., Siame, K.K., Gey van Pittius, N.C., van Helden, P.D., Pain, A., Warren, R.M., 2015. Whole genome sequence analysis of *Mycobacterium suricattae*. Tuberculosis (Edinburgh, Scotland) 95, 682-688.

Djemal, S.E., Siala, M., Smaoui, S., Kammoun, S., Marouane, C., Bezos, J., Messadi-Akrout, F., Romero, B., Gdoura, R., 2017. Genetic diversity assessment of Tunisian *Mycobacterium bovis* population isolated from cattle. BMC Vet Res 13, 393.

Domingo, M., Vidal, E., Marco, A., 2014. Pathology of bovine tuberculosis. Res Vet Sci 97 Suppl, S20-29.

Domogalla, J., Prodinger, W.M., Blum, H., Krebs, S., Gellert, S., Muller, M., Neuendorf, E., Sedlmaier, F., Buttner, M., 2013. Region of difference 4 in alpine *Mycobacterium caprae* isolates indicates three variants. J Clin Microbiol 51, 1381-1388.

Donnelly, C.A., Wei, G., Johnston, W.T., Cox, D.R., Woodroffe, R., Bourne, F.J., Cheeseman, C.L., Clifton-Hadley, R.S., Gettinby, G., Gilks, P., Jenkins, H.E., Le Fevre, A.M., McInerney, J.P., Morrison, W.I., 2007. Impacts of widespread badger culling on cattle tuberculosis: concluding analyses from a large-scale field trial. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases 11, 300-308.

Doyle, L.P., Courcier, E.A., Gordon, A.W., O'Hagan, M.J.H., Menzies, F.D., 2016. Bovine tuberculosis in Northern Ireland: Risk factors associated with duration and recurrence of chronic herd breakdowns. Prev Vet Med 131, 1-7.

Doyle, R.M., Burgess, C., Williams, R., Gorton, R., Booth, H., Brown, J., Bryant, J.M., Chan, J., Creer, D., Holdstock, J., Kunst, H., Lozewicz, S., Platt, G., Romero, E.Y., Speight, G., Tiberi, S., Abubakar, I., Lipman, M., McHugh, T.D., Breuer, J., 2018. Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant *Mycobacterium tuberculosis* Faster than MGIT Culture Sequencing. J Clin Microbiol 56.

Duarte, E.L., Domingos, M., Amado, A., Botelho, A., 2008. Spoligotype diversity of *Mycobacterium bovis* and *Mycobacterium caprae* animal isolates. Vet Microbiol 130, 415-421.

Duarte, E.L., Domingos, M., Amado, A., Cunha, M.V., Botelho, A., 2010. MIRU-VNTR typing adds discriminatory value to groups of *Mycobacterium bovis* and *Mycobacterium caprae* strains defined by spoligotyping. Vet Microbiol 143, 299-306.

Duault, H., Michelet, L., Boschiroli, M.L., Durand, B., Canini, L., 2022. A Bayesian evolutionary model towards understanding wildlife contribution to F4-family *Mycobacterium bovis* transmission in the South-West of France. Vet Res 53, 28.

Duffy, S.C., Srinivasan, S., Schilling, M.A., Stuber, T., Danchuk, S.N., Michael, J.S., Venkatesan, M., Bansal, N., Maan, S., Jindal, N., Chaudhary, D., Dandapat, P., Katani, R., Chothe, S., Veerasami, M., Robbe-Austerman, S., Juleff, N., Kapur, V., Behr, M.A., 2020. Reconsidering *Mycobacterium bovis* as a proxy for zoonotic tuberculosis: a molecular epidemiological surveillance study. The Lancet Microbe 1, e66-e73.

Dunn, J.R., Kaneene, J.B., Grooms, D.L., Bolin, S.R., Bolin, C.A., Bruning-Fann, C.S., 2005. Effects of positive results for *Mycobacterium avium* subsp *paratuberculosis* as determined by microbial culture of feces or antibody ELISA on results of caudal fold tuberculin test and interferon-gamma assay for tuberculosis in cattle. J Am Vet Med Assoc 226, 429-435.

Durr, P.A., Hewinson, R.G., Clifton-Hadley, R.S., 2000. Molecular epidemiology of bovine tuberculosis. I. *Mycobacterium bovis* genotyping. Revue scientifique et technique (International Office of Epizootics) 19, 675-688.

Durr, S., Muller, B., Alonso, S., Hattendorf, J., Laisse, C.J., van Helden, P.D., Zinsstag, J., 2013. Differences in primary sites of infection between zoonotic and human tuberculosis: results from a worldwide systematic review. PLoS neglected tropical diseases 7, e2399.

Dvorska, L., Matlova, L., Bartos, M., Parmova, I., Bartl, J., Svastova, P., Bull, T.J., Pavlik, I., 2004. Study of *Mycobacterium avium* complex strains isolated from cattle in the Czech Republic between 1996 and 2000. Vet Microbiol 99, 239-250.

ECDC, E., 2021. The European Union One Health 2020 Zoonoses Report. EFSA Journal 19.

Egbe, N.F., Muwonge, A., Ndip, L., Kelly, R.F., Sander, M., Tanya, V., Ngwa, V.N., Handel, I.G., Novak, A., Ngandalo, R., Mazeri, S., Morgan, K.L., Asuquo, A., de, C.B.B.M., 2017. Molecular epidemiology of *Mycobacterium bovis* in Cameroon. Sci Rep 7, 4652.

El Khechine, A., Henry, M., Raoult, D., Drancourt, M., 2009. Detection of *Mycobacterium tuberculosis* complex organisms in the stools of patients with pulmonary tuberculosis. Microbiology (Reading) 155, 2384-2389.

Emmanuel, F.X., Seagar, A.L., Doig, C., Rayner, A., Claxton, P., Laurenson, I., 2007. Human and animal infections with *Mycobacterium microti*, Scotland. Emerg Infect Dis 13, 1924-1927.

Erler, W., Martin, G., Sachse, K., Naumann, L., Kahlau, D., Beer, J., Bartos, M., Nagy, G., Cvetnic, Z., Zolnir-Dovc, M., Pavlik, I., 2004. Molecular fingerprinting of *Mycobacterium bovis* subsp. *caprae* isolates from central Europe. J Clin Microbiol 42, 2234-2238.

Erwin, P.C., Bemis, D.A., McCombs, S.B., Sheeler, L.L., Himelright, I.M., Halford, S.K., Diem, L., Metchock, B., Jones, T.F., Schilling, M.G., Thomsen, B.V., 2004. *Mycobacterium tuberculosis* transmission from human to canine. Emerg Infect Dis 10, 2258-2210.

European Food Safety, A., European Centre for Disease, P., Control, 2021. The European Union One Health 2019 Zoonoses Report. EFSA J 19, e06406.

Faksri, K., Xia, E., Tan, J.H., Teo, Y.Y., Ong, R.T., 2016. *In silico* region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. BMC genomics 17, 847.

Fan, X., Xie, L., Li, W., Xie, J., 2014. Prophage-like elements present in *Mycobacterium* genomes. BMC genomics 15, 243.

Fedrizzi, T., Meehan, C.J., Grottola, A., Giacobazzi, E., Fregni Serpini, G., Tagliazucchi, S., Fabio, A., Bettua, C., Bertorelli, R., De Sanctis, V., Rumpianesi, F., Pecorari, M., Jousson, O., Tortoli, E., Segata, N., 2017. Genomic characterization of Nontuberculous Mycobacteria. Sci Rep 7, 45258.

Feizabadi, M.M., Robertson, I.D., Cousins, D.V., Hampson, D.J., 1996. Genomic analysis of *Mycobacterium bovis* and other members of the *Mycobacterium tuberculosis* complex by isoenzyme analysis and pulsed-field gel electrophoresis. J Clin Microbiol 34, 1136-1142.

Fell, S., Bröckl, S., Büttner, M., Rettinger, A., Zimmermann, P., Straubinger, R.K., 2016. Two alternative DNA extraction methods to improve the detection of *Mycobacterium-tuberculosis*-complex members in cattle and red deer tissue samples. BMC microbiology 16, 213.

Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbon, M.H., Bobadilla del Valle, M., Fyfe, J., Garcia-Garcia, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., Leon, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendon, A., Sifuentes-Osornio, J., Ponce de Leon, A., Cave, M.D., Fleischmann, R., Whittam, T.S., Alland, D., 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188, 759-772.

Forbes, B.A., Hall, G.S., Miller, M.B., Novak, S.M., Rowlinson, M.C., Salfinger, M., Somoskovi, A., Warshauer, D.M., Wilson, M.L., 2018. Practical Guidance for Clinical Microbiology Laboratories: Mycobacteria. Clin Microbiol Rev 31.

Frampton, M., Houlston, R., 2012. Generation of artificial FASTQ files to evaluate the performance of next-generation sequencing pipelines. PLoS One 7, e49110.

Franco, M.M., Paes, A.C., Ribeiro, M.G., de Figueiredo Pantoja, J.C., Santos, A.C., Miyata, M., Leite, C.Q., Motta, R.G., Listoni, F.J., 2013. Occurrence of mycobacteria in bovine milk samples from both individual and collective bulk tanks at farms and informal markets in the southeast region of Sao Paulo, Brazil. BMC Vet Res 9, 85.

Frothingham, R., 1995. Differentiation of strains in *Mycobacterium tuberculosis* complex by DNA sequence polymorphisms, including rapid identification of *M. bovis* BCG. J Clin Microbiol 33, 840-844.

Frothingham, R., Meeker-O'connell, W.A., 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 144, 1189-1196.

Gagneux, S., 2018. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol 16, 202-213.

García-Bocanegra, I., Barranco, I., Rodríguez-Gómez, I.M., Pérez, B., Gómez-Laguna, J., Rodríguez, S., Ruiz-Villamayor, E., Perea, A., 2010. Tuberculosis in alpacas (*Lama pacos*) caused by *Mycobacterium bovis*. J Clin Microbiol 48, 1960-1964.

Garcia-Jimenez, W.L., Fernandez-Llario, P., Benitez-Medina, J.M., Cerrato, R., Cuesta, J., Garcia-Sanchez, A., Goncalves, P., Martinez, R., Risco, D., Salguero, F.J., Serrano, E., Gomez, L., Hermoso-de-Mendoza, J., 2013. Reducing Eurasian wild boar (*Sus scrofa*) population density as a measure for bovine tuberculosis control: effects in wild boar and a sympatric fallow deer (*Dama dama*) population in Central Spain. Prev Vet Med 110, 435-446.

Garrido, J.M., Vazquez, P., Molina, E., Plazaola, J.M., Sevilla, I.A., Geijo, M.V., Alonso-Hearn, M., Juste, R.A., 2013. Paratuberculosis vaccination causes only limited cross-reactivity in the skin test for diagnosis of bovine tuberculosis. PLoS One 8, e80985.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint.

Gcebe, N., Hlokwe, T.M., 2017. Non-tuberculous Mycobacteria in South African Wildlife: Neglected Pathogens and Potential Impediments for Bovine Tuberculosis Diagnosis. Frontiers in cellular and infection microbiology 7, 15.

Gcebe, N., Rutten, V., Pittius, N.G.V., Naicker, B., Michel, A., 2017. *Mycobacterium malmesburyense* sp. nov., a non-tuberculous species of the genus *Mycobacterium* revealed by multiple gene sequence characterization. Int J Syst Evol Microbiol 67, 832-838.

Getahun, H., Matteelli, A., Chaisson, R.E., Raviglione, M., 2015. Latent *Mycobacterium tuberculosis* infection. N Engl J Med 372, 2127-2135.

Gey Van Pittius, N.C., Gamieldien, J., Hide, W., Brown, G.D., Siezen, R.J., Beyers, A.D., 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol 2, RESEARCH0044.

Gey van Pittius, N.C., Perrett, K.D., Michel, A.L., Keet, D.F., Hlokwe, T., Streicher, E.M., Warren, R.M., van Helden, P.D., 2012. Infection of African buffalo (*Syncerus caffer*) by oryx bacillus, a rare member of the antelope clade of the *Mycobacterium tuberculosis* complex. Journal of wildlife diseases 48, 849-857.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., van Helden, P.D., Warren, R.M., 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. BMC Evol Biol 6, 95.

Ghielmetti, G., Friedel, U., Scherrer, S., Sarno, E., Landolt, P., Dietz, O., Hilbe, M., Zweifel, C., Stephan, R., 2018. Non-tuberculous Mycobacteria isolated from lymph nodes and faecal samples of healthy slaughtered cattle and the abattoir environment. Transbound Emerg Dis 65, 711-718.

Ghielmetti, G., Hilbe, M., Friedel, U., Menegatti, C., Bacciarini, L., Stephan, R., Bloemberg, G., 2021a. Mycobacterial infections in wild boars (*Sus scrofa*) from Southern Switzerland: Diagnostic improvements, epidemiological situation and zoonotic potential. Transbound Emerg Dis 68, 573-586.

Ghielmetti, G., Kupca, A.M., Hanczaruk, M., Friedel, U., Weinberger, H., Revilla-Fernandez, S., Hofer, E., Riehm, J.M., Stephan, R., Glawischnig, W., 2021b. *Mycobacterium microti* Infections in Free-Ranging Red Deer (Cervus elaphus). Emerg Infect Dis 27, 2025-2032.

Ghielmetti, G., Scherrer, S., Friedel, U., Frei, D., Suter, D., Perler, L., Wittenbrink, M.M., 2017. Epidemiological tracing of bovine tuberculosis in Switzerland, multilocus variable number of tandem repeat analysis of *Mycobacterium bovis* and *Mycobacterium caprae*. PLoS One 12, e0172474.

Gingeras, T.R., Higuchi, R., Kricka, L.J., Lo, Y.M., Wittwer, C.T., 2005. Fifty years of molecular (DNA/RNA) diagnostics. Clin Chem 51, 661-671.

Glaser, L., Carstensen, M., Shaw, S., Robbe-Austerman, S., Wunschmann, A., Grear, D., Stuber, T., Thomsen, B., 2016. Descriptive Epidemiology and Whole Genome Sequencing Analysis for an Outbreak of Bovine Tuberculosis in Beef Cattle and White-Tailed Deer in Northwestern Minnesota. PLoS One 11, e0145735.

Glickman, C., Kammlade, S.M., Hasan, N.A., Epperson, L.E., Davidson, R.M., Strong, M., 2020. Characterization of integrated prophages within diverse species of clinical nontuberculous mycobacteria. Virol J 17, 124.

Goig, G.A., Cancino-Munoz, I., Torres-Puente, M., Villamayor, L.M., Navarro, D., Borras, R., Comas, I., 2020. Whole-genome sequencing of *Mycobacterium tuberculosis* directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study. Lancet Microbe 1, e175-e183.

Gomez-Buendia, A., Romero, B., Bezos, J., Lozano, F., de Juan, L., Alvarez, J., 2021. Spoligotype-specific risk of finding lesions in tissues from cattle infected by *Mycobacterium bovis*. BMC Vet Res 17, 148.

Gonzalo-Asensio, J., Perez, I., Aguilo, N., Uranga, S., Pico, A., Lampreave, C., Cebollada, A., Otal, I., Samper, S., Martin, C., 2018. New insights into the transposition mechanisms of IS*6110* and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. PLoS Genet 14, e1007282.

Good, M., Bakker, D., Duignan, A., Collins, D.M., 2018. The History of *In Vivo* Tuberculin Testing in Bovines: Tuberculosis, a "One Health" Issue. Front Vet Sci 5, 59.

Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M.E., Suzuki, K.I., Ludwig, W., Whitman, W.B., Michael Goodfellow, P.K., Hans-Jürgen Busse,, Martha E. Trujillo, K.-i.S., Wolfgang Ludwig, Whitman, a.W.B., 2012. Bergey's Manual of Systematic Bacteriology, volume 5: The *Actinobacteria*, 2 ed. Springer.

Gordon, S.V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., Cole, S.T., 1999a. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32, 643-655.

Gordon, S.V., Heym, B., Parkhill, J., Barrell, B., Cole, S.T., 1999b. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. Microbiology (Reading) 145 (Pt 4), 881-892.

Gormley, E., Corner, L.A., Costello, E., Rodriguez-Campos, S., 2014. Bacteriological diagnosis and molecular strain typing of *Mycobacterium bovis* and *Mycobacterium caprae*. Res Vet Sci 97 Suppl, S30-43.

Gormley, E., Doyle, M.B., Fitzsimons, T., McGill, K., Collins, J.D., 2006. Diagnosis of *Mycobacterium bovis* infection in cattle by use of the gamma-interferon (Bovigam) assay. Vet Microbiol 112, 171-179.

Gormley, E., Doyle, M.B., McGill, K., Costello, E., Good, M., Collins, J.D., 2004. The effect of the tuberculin test and the consequences of a delay in blood culture on the sensitivity of a gamma-interferon assay for the detection of *Mycobacterium bovis* infection in cattle. Veterinary immunology and immunopathology 102, 413-420.

Gortazar, C., Torres, M.J., Vicente, J., Acevedo, P., Reglero, M., de la Fuente, J., Negro, J.J., Aznar-Martin, J., 2008. Bovine tuberculosis in Donana Biosphere Reserve: the role of wild ungulates as disease reservoirs in the last Iberian lynx strongholds. PLoS One 3, e2776.

Gortazar, C., Vicente, J., Samper, S., Garrido, J.M., Fernandez-De-Mera, I.G., Gavin, P., Juste, R.A., Martin, C., Acevedo, P., De La Puente, M., Hofle, U., 2005. Molecular characterization of *Mycobacterium tuberculosis* complex isolates from wild ungulates in south-central Spain. Vet Res 36, 43-52.

Goyal, M., Young, D., Zhang, Y., Jenkins, P.A., Shaw, R.J., 1994. PCR amplification of variable sequence upstream of *katG* gene to subdivide strains of *Mycobacterium tuberculosis* complex. J Clin Microbiol 32, 3070-3071.

Gray, T.A., Derbyshire, K.M., 2018. Blending genomes: distributive conjugal transfer in mycobacteria, a sexier form of HGT. Mol Microbiol 108, 601-613.

Gray, T.A., Krywy, J.A., Harold, J., Palumbo, M.J., Derbyshire, K.M., 2013. Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS Biol 11, e1001602.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D., van Embden, J.D., 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol 10, 1057-1065.

Gudan, A., Artukovic, B., Cvetnic, Z., Spicic, S., Beck, A., Hohsteter, M., Naglic, T., Bata, I., Grabarevic, Z., 2008. Disseminated tuberculosis in hyrax (*Procavia capensis*) caused by *Mycobacterium africanum*. Journal of zoo and wildlife medicine : official publication of the American Association of Zoo Veterinarians 39, 386-391.

Guerin-Faublee, V., Flandrois, J.P., Pichat, C., Boschiroli, M.L., Lamy, B., 2013. *Mycobacterium bourgelatii* sp. nov., a rapidly growing, non-chromogenic species isolated from the lymph nodes of cattle. Int J Syst Evol Microbiol 63, 4669-4674.

Guimaraes, A.M.S., Zimpel, C.K., 2020. *Mycobacterium bovis*: From Genotyping to Genome Sequencing. Microorganisms 8.

Gupta, R.S., Lo, B., Son, J., 2018. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. Front Microbiol 9, 67.

Guta, S., Casal, J., Garcia-Saenz, A., Saez, J.L., Pacios, A., Garcia, P., Napp, S., Allepuz, A., 2014. Risk factors for bovine tuberculosis persistence in beef herds of Southern and Central Spain. Prev Vet Med 115, 173-180.

Gutierrez Reyes, J.A., Garcia Casanova, L., Romero Torres, C., Sosa Gallegos, S.L., Canto Alarcon, G.J., Mercado Pezzat, M., Pizano Martinez, O., Estrada Chavez, C., Milian Suazo, F., 2012. Population structure of *Mycobacterium bovis* isolates from cattle in Mexico. Prev Vet Med 106, 1-8.

Haddad, N., Ostyn, A., Karoui, C., Masselot, M., Thorel, M.F., Hughes, S.L., Inwald, J., Hewinson, R.G., Durand, B., 2001. Spoligotype diversity of *Mycobacterium bovis* strains isolated in France from 1979 to 2000. J Clin Microbiol 39, 3623-3632.

Hadgu, A., 1996. The discrepancy in discrepant analysis. Lancet 348, 592-593.

Halse, T.A., Escuyer, V.E., Musser, K.A., 2011. Evaluation of a single-tube multiplex real-time PCR for differentiation of members of the *Mycobacterium tuberculosis* complex in clinical specimens. J Clin Microbiol 49, 2562-2567.

Harmsen, D., Dostal, S., Roth, A., Niemann, S., Rothganger, J., Sammeth, M., Albert, J., Frosch, M., Richter, E., 2003. RIDOM: comprehensive and public sequence database for identification of *Mycobacterium* species. BMC infectious diseases 3, 26.

Harrington, N., Howden, K., Turcotte, C., 2014. Canada's bovine tuberculosis eradication program, In: Zoonotic Tuberculosis: *Mycobacterium bovis* and Other Pathogenic Mycobacteria, 3rd ed. John Wiley & Sons.

Harris, N.B., Payeur, J., Bravo, D., Osorio, R., Stuber, T., Farrell, D., Paulson, D., Treviso, S., Mikolon, A., Rodriguez-Lainz, A., Cernek-Hoskins, S., Rast, R., Ginsberg, M., Kinde, H., 2007. Recovery of *Mycobacterium bovis* from soft fresh cheese originating in Mexico. Appl Environ Microbiol 73, 1025-1028.

Hauer, A., Michelet, L., Cochard, T., Branger, M., Nunez, J., Boschiroli, M.L., Biet, F., 2019. Accurate Phylogenetic Relationships Among *Mycobacterium bovis* Strains Circulating in France Based on Whole Genome Sequencing and Single Nucleotide Polymorphism Analysis. Front Microbiol 10, 955.

Hauer, A., Michelet, L., De Cruz, K., Cochard, T., Branger, M., Karoui, C., Henault, S., Biet, F., Boschiroli, M.L., 2016. MIRU-VNTR allelic variability depends on *Mycobacterium bovis* clonal group identity. Infect Genet Evol 45, 165-169.

Hayden, H.S., Joshi, S., Radey, M.C., Vo, A.T., Forsberg, C., Morgan, S.J., Waalkes, A., Holmes, E.A., Klee, S.M., Emond, M.J., Singh, P.K., Salipante, S.J., 2022. Genome Capture Sequencing Selectively Enriches Bacterial DNA and Enables Genome-Wide Measurement of Intrastrain Genetic Diversity in Human Infections. mBio, e0142422.

Helke, K.L., Mankowski, J.L., Manabe, Y.C., 2006. Animal models of cavitation in pulmonary tuberculosis. Tuberculosis (Edinburgh, Scotland) 86, 337-348.

Hermans, P.W., van Soolingen, D., Bik, E.M., de Haas, P.E., Dale, J.W., van Embden, J.D., 1991. Insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infection and immunity 59, 2695-2705.

Hermans, P.W., van Soolingen, D., Dale, J.W., Schuitema, A.R., McAdam, R.A., Catty, D., van Embden, J.D., 1990. Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J Clin Microbiol 28, 2051-2058.

Hermans, P.W., van Soolingen, D., van Embden, J.D., 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. J Bacteriol 174, 4157-4165.

Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. PLoS One 3, e3426.

Hewinson, R.G., Vordermeier, H.M., Smith, N.H., Gordon, S.V., 2006. Recent advances in our knowledge of *Mycobacterium bovis*: a feeling for the organism. Vet Microbiol 112, 127-139.

Hijmans, C.K., Williams, E., Vennes, C., 2021. Geosphere, 1.5-14 ed.

Hines, N., Payeur, J.B., Hoffman, L.J., 2006. Comparison of the recovery of *Mycobacterium bovis* isolates using the BACTEC MGIT 960 system, BACTEC 460 system, and Middlebrook 7H10 and 7H11 solid media. Journal of Veterinary Diagnostic Investigation 18, 243-250.

Hlavsa, M.C., Moonan, P.K., Cowan, L.S., Navin, T.R., Kammerer, J.S., Morlock, G.P., Crawford, J.T., Lobue, P.A., 2008. Human tuberculosis due to *Mycobacterium bovis* in the United States, 1995-2005. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47, 168-175.

Hlokwe, T.M., van Helden, P., Michel, A.L., 2014. Evidence of increasing intra and inter-species transmission of *Mycobacterium bovis* in South Africa: are we losing the battle? Prev Vet Med 115, 10-17.

Holland, R.D., Wilkes, J.G., Rafii, F., Sutherland, J.B., Persons, C.C., Voorhees, K.J., Lay, J.O., 1996. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 10, 1227-1232.

Homolka, S., Projahn, M., Feuerriegel, S., Ubben, T., Diel, R., Nubel, U., Niemann, S., 2012. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. PLoS One 7, e39855.

Hope, J.C., Thom, M.L., Villarreal-Ramos, B., Vordermeier, H.M., Hewinson, R.G., Howard, C.J., 2005. Exposure to *Mycobacterium avium* induces low-level protection from *Mycobacterium bovis* infection but compromises diagnosis of disease in cattle. Clinical and experimental immunology 141, 432-439.

Howard, S.T., Oughton, M.T., Haddad, A., Johnson, W.M., 1998. Absence of the genetic marker IS*6110* from a strain of *Mycobacterium tuberculosis* isolated in Ontario. Can J Infect Dis 9, 48-53.

Huard, R.C., Fabre, M., de Haas, P., Lazzarini, L.C., van Soolingen, D., Cousins, D., Ho, J.L., 2006. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. J Bacteriol 188, 4271-4287.

Huard, R.C., Lazzarini, L.C., Butler, W.R., van Soolingen, D., Ho, J.L., 2003. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. J Clin Microbiol 41, 1637-1650.

Hughes, M.S., Ball, N.W., McCarroll, J., Erskine, M., Taylor, M.J., Pollock, J.M., Skuce, R.A., Neill, S.D., 2005. Molecular analyses of mycobacteria other than the *M. tuberculosis* complex isolated from Northern Ireland cattle. Vet Microbiol 108, 101-112.

Infantes-Lorenzo, J.A., Moreno, I., Risalde, M.L.A., Roy, A., Villar, M., Romero, B., Ibarrola, N., de la Fuente, J., Puentes, E., de Juan, L., Gortazar, C., Bezos, J., Dominguez, L., Dominguez, M., 2017. Proteomic characterisation of bovine and avian purified protein derivatives and identification of specific antigens for serodiagnosis of bovine tuberculosis. Clin Proteomics 14, 36.

Infantes-Lorenzo, J.A., Moreno, I., Roy, A., Risalde, M.A., Balseiro, A., de Juan, L., Romero, B., Bezos, J., Puentes, E., Akerstedt, J., Tessema, G.T., Gortazar, C., Dominguez, L., Dominguez, M., 2019. Specificity of serological test for detection of tuberculosis in cattle, goats, sheep and pigs under different epidemiological situations. BMC Vet Res 15, 70.

Infantes-Lorenzo, J.A., Romero, B., Rodriguez-Bertos, A., Roy, A., Ortega, J., de Juan, L., Moreno, I., Dominguez, M., Dominguez, L., Bezos, J., 2020. Tuberculosis caused by *Mycobacterium caprae* in a camel (*Camelus dromedarius*). BMC Vet Res 16, 435.

Iwai, H., Kato-Miyazawa, M., Kirikae, T., Miyoshi-Akiyama, T., 2015. CASTB (the comprehensive analysis server for the *Mycobacterium tuberculosis* complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. Tuberculosis (Edinburgh, Scotland) 95, 843-844.

Jahans, K., Palmer, S., Inwald, J., Brown, J., Abayakoon, S., 2004. Isolation of *Mycobacterium microti* from a male Charolais-Hereford cross. The Veterinary record 155, 2.

Jajou, R., Kohl, T.A., Walker, T., Norman, A., Cirillo, D.M., Tagliani, E., Niemann, S., de Neeling, A., Lillebaek, T., Anthony, R.M., van Soolingen, D., 2019. Towards standardisation: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. Euro Surveill 24.

Javed, M.T., Aranaz, A., de Juan, L., Bezos, J., Romero, B., Alvarez, J., Lozano, C., Mateos, A., Dominguez, L., 2007. Improvement of spoligotyping with additional spacer sequences for characterization of *Mycobacterium bovis* and *M. caprae* isolates from Spain. Tuberculosis (Edinburgh, Scotland) 87, 437-445.

Jia Khor, M., Broda, A., Kostrzewa, M., Drobniewski, F., Larrouy-Maumus, G., 2021. An Improved Method for Rapid Detection of *Mycobacterium abscessus* Complex Based on Species-Specific Lipid Fingerprint by Routine MALDI-TOF. Front Chem 9, 715890.

Jolley, M.E., Nasir, M.S., Surujballi, O.P., Romanowska, A., Renteria, T.B., De la Mora, A., Lim, A., Bolin, S.R., Michel, A.L., Kostovic, M., Corrigan, E.C., 2007. Fluorescence polarization assay for the detection of antibodies to *Mycobacterium bovis* in bovine sera. Vet Microbiol 120, 113-121.

Jombart, T., Kendall, M., Almagro-Garcia, J., Colijn, C., 2017. treespace: Statistical exploration of landscapes of phylogenetic trees. Mol Ecol Resour 17, 1385-1392.

Jones, G.J., Whelan, A., Clifford, D., Coad, M., Vordermeier, H.M., 2012. Improved skin test for differential diagnosis of bovine tuberculosis by the addition of Rv3020c-derived peptides. Clinical and vaccine immunology : CVI 19, 620-622.

Jones, R.C., Harris, L.G., Morgan, S., Ruddy, M.C., Perry, M., Williams, R., Humphrey, T., Temple, M., Davies, A.P., 2019. Phylogenetic Analysis of *Mycobacterium tuberculosis* Strains in Wales by Use of Core Genome Multilocus Sequence Typing To Analyze Whole-Genome Sequencing Data. J Clin Microbiol 57.

Kam, K.M., Yip, C.W., Tse, L.W., Leung, K.L., Wong, K.L., Ko, W.M., Wong, W.S., 2006. Optimization of variable number tandem repeat typing set for differentiating *Mycobacterium tuberculosis* strains in the Beijing family. FEMS microbiology letters 256, 258-265.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35, 907-914.

Kaminski, M.M., Abudayyeh, O.O., Gootenberg, J.S., Zhang, F., Collins, J.J., 2021. CRISPR-based diagnostics. Nat Biomed Eng 5, 643-656.

Kapalamula, T.F., Thapa, J., Akapelwa, M.L., Hayashida, K., Gordon, S.V., Hang' Ombe, B.M., Munyeme, M., Solo, E.S., Bwalya, P., Nyenje, M.E., Tamaru, A., Suzuki, Y., Nakajima, C., 2021. Development of a loop-mediated isothermal amplification (LAMP) method for specific detection of *Mycobacterium bovis*. PLoS neglected tropical diseases 15, e0008996.

Karamanou, M., Androutsos, G., 2012. The masterful description of pulmonary tuberculosis by Soranus of Ephesus (c. 98-138 A.D.). Am J Respir Crit Care Med 186, 571.

Karlsson, A.G., Lessel, E.F., 1970. *Mycobacterium bovis* nom. nov. International Journal of Systematic Bacteriology 20.

Karolemeas, K., McKinley, T.J., Clifton-Hadley, R.S., Goodchild, A.V., Mitchell, A., Johnston, W.T., Conlan, A.J., Donnelly, C.A., Wood, J.L., 2011. Recurrence of bovine tuberculosis breakdowns in Great Britain: risk factors and prediction. Prev Vet Med 102, 22-29.

Kasai, H., Ezaki, T., Harayama, S., 2000. Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. J Clin Microbiol 38, 301-308.

Kassa, G.M., Abebe, F., Worku, Y., Legesse, M., Medhin, G., Bjune, G., Ameni, G., 2012. Tuberculosis in Goats and Sheep in Afar Pastoral Region of Ethiopia and Isolation of *Mycobacterium tuberculosis* from Goat. Veterinary medicine international 2012, 869146.

Keating, L.A., Wheeler, P.R., Mansoor, H., Inwald, J.K., Dale, J., Hewinson, R.G., Gordon, S.V., 2005. The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: implications for *in vivo* growth. Mol Microbiol 56, 163-174.

Kiers, A., Klarenbeek, A., Mendelts, B., van Soolingen, D., Koëter, G., 2008. Transmission of *Mycobacterium pinnipedii* to humans in a zoo with marine mammals. Int J Tuberc Lung Dis 12, 5.

Kim, J.H., Sohn, H.J., Kang, K.I., Kim, W.I., An, J.S., Jean, Y.H., 2002. *Mycobacterium bovis* infection in a farmed elk in Korea. Journal of veterinary science 3, 163-166.

Kinne, J., Johnson, B., Jahans, K.L., Smith, N.H., Ul-Haq, A., Wernery, U., 2006. Camel tuberculosis--a case report. Tropical animal health and production 38, 207-213.

Kipar, A., Burthe, S.J., Hetzel, U., Rokia, M.A., Telfer, S., Lambin, X., Birtles, R.J., Begon, M., Bennett, M., 2014. *Mycobacterium microti* tuberculosis in its maintenance host, the field vole (*Microtus agrestis*): characterization of the disease and possible routes of transmission. Veterinary pathology 51, 903-914.

Koch, R., 1882. Die aetiologie der tuberculose. Berliner klinische Wochenschrift 19, 9.

Koh, W.J., 2017. Nontuberculous Mycobacteria-Overview. Microbiol Spectr 5.

Kohl, T.A., Diel, R., Harmsen, D., Rothganger, J., Walter, K.M., Merker, M., Weniger, T., Niemann, S., 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 52, 2479-2486.

Kohl, T.A., Harmsen, D., Rothganger, J., Walker, T., Diel, R., Niemann, S., 2018a. Harmonized Genome Wide Typing of Tubercle Bacilli Using a Web-Based Gene-By-Gene Nomenclature System. EBioMedicine 34, 131-138.

Kohl, T.A., Kranzer, K., Andres, S., Wirth, T., Niemann, S., Moser, I., 2020. Population Structure of *Mycobacterium bovis* in Germany: a Long-Term Study Using Whole-Genome Sequencing Combined with Conventional Molecular Typing Methods. J Clin Microbiol 58.

Kohl, T.A., Utpatel, C., Niemann, S., Moser, I., 2018b. *Mycobacterium bovis* Persistence in Two Different Captive Wild Animal Populations in Germany: a Longitudinal Molecular Epidemiological Study Revealing Pathogen Transmission by Whole-Genome Sequencing. J Clin Microbiol 56.

Kohl, T.A., Utpatel, C., Schleusener, V., De Filippo, M.R., Beckert, P., Cirillo, D.M., Niemann, S., 2018c. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. PeerJ 6, e5895.

Konjek, J., Souded, S., Guerardel, Y., Trivelli, X., Bernut, A., Kremer, L., Welte, B., Joyeux, M., Dubrou, S., Euzeby, J.P., Gaillard, J.L., Sapriel, G., Heym, B., 2016. *Mycobacterium lutetiense* sp. nov., *Mycobacterium montmartrense* sp. nov. and *Mycobacterium arcueilense* sp. nov., members of a novel group of non-pigmented rapidly growing mycobacteria recovered from a water distribution system. Int J Syst Evol Microbiol 66, 3694-3702.

Krajewska-Wedzina, M., Kozinska, M., Orlowska, B., Weiner, M., Szulowski, K., Augustynowicz-Kopec, E., Anusz, K., Smith, N.H., 2018. Molecular characterisation of *Mycobacterium caprae* strains isolated in Poland. The Veterinary record 182, 292.

Kremer, K., Au, B.K., Yip, P.C., Skuce, R., Supply, P., Kam, K.M., van Soolingen, D., 2005. Use of variable-number tandem-repeat typing to differentiate *Mycobacterium tuberculosis* Beijing family isolates from Hong Kong and comparison with IS*6110* restriction fragment length polymorphism typing and spoligotyping. J Clin Microbiol 43, 314-320.

Kremer, K., van Soolingen, D., Frothingham, R., Haas, W.H., Hermans, P.W.M., Martín, C., Palittapongarnpim, P., Plikaytis, B.B., Riley, L.W., Yakurs, M.A., Musser, J.M., van Embden, J.D.A., 1999. Comparison of Methods Based on Different Molecular Epidemiological Markers for Typing of *Mycobacterium tuberculosis* Complex Strains: Interlaboratory Study of Discriminatory Power and Reproducibility. Journal of Clinical Microbiology 37, 12.

Kriz, P., Kralik, P., Slany, M., Slana, I., Svobodova, J., Parmova, I., Barnet, V., Jurek, V., Pavlik, I., 2011. *Mycobacterium pinnipedii* in a captive Southern sea lion (*Otaria flavescens*): a case report. Veterinarni Medicina 56, 7.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35, 1547-1549.

Kuypers, J., Jerome, K.R., 2017. Applications of Digital PCR for Clinical Microbiology. J Clin Microbiol 55, 1621-1628.

Lahlou, O., Millet, J., Chaoui, I., Sabouni, R., Filali-Maltouf, A., Akrim, M., El Mzibri, M., Rastogi, N., El Aouad, R., 2012. The genotypic population structure of *Mycobacterium tuberculosis* complex from Moroccan patients reveals a predominance of Euro-American lineages. PLoS One 7, e47113.

Lahuerta-Marin, A., Milne, M.G., McNair, J., Skuce, R.A., McBride, S.H., Menzies, F.D., McDowell, S.J.W., Byrne, A.W., Handel, I.G., de, C.B.B.M., 2018. Bayesian latent class estimation of sensitivity and specificity parameters of diagnostic tests for bovine tuberculosis in chronically infected herds in Northern Ireland. Veterinary journal (London, England : 1997) 238, 15-21.

Lamine-Khemiri, H., Martinez, R., Garcia-Jimenez, W.L., Benitez-Medina, J.M., Cortes, M., Hurtado, I., Abassi, M.S., Khazri, I., Benzarti, M., Hermoso-de-Mendoza, J., 2014. Genotypic characterization by spoligotyping and VNTR typing of *Mycobacterium bovis* and *Mycobacterium caprae* isolates from cattle of Tunisia. Tropical animal health and production 46, 305-311.

Lantos, A., Niemann, S., Mezosi, L., Sos, E., Erdelyi, K., David, S., Parsons, L.M., Kubica, T., Rusch-Gerdes, S., Somoskovi, A., 2003. Pulmonary tuberculosis due to *Mycobacterium bovis* subsp. *caprae* in captive Siberian tiger. Emerg Infect Dis 9, 1462-1464.

Larenas-Muñoz, F., Sánchez-Carvajal, J.M., Galán-Relaño, Á., Ruedas-Torres, I., Vera-Salmoral, E., Gómez-Gascón, L., Maldonado, A., Carrasco, L., Tarradas, C., Luque, I., Rodríguez-Gómez, I.M., Gómez-Laguna, J., 2022. The Role of Histopathology as a Complementary Diagnostic Tool in the Monitoring of Bovine Tuberculosis. Frontiers in Veterinary Science 9.

Lasserre, M., Fresia, P., Greif, G., Iraola, G., Castro-Ramos, M., Juambeltz, A., Nunez, A., Naya, H., Robello, C., Berna, L., 2018. Whole genome sequencing of the monomorphic pathogen *Mycobacterium bovis* reveals local differentiation of cattle clinical isolates. BMC genomics 19, 2.

Le Fleche, P., Fabre, M., Denoeud, F., Koeck, J.L., Vergnaud, G., 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC microbiology 2, 37.

Lee, R.S., Behr, M.A., 2016. Does Choice Matter? Reference-Based Alignment for Molecular Epidemiology of Tuberculosis. J Clin Microbiol 54, 1891-1895.

Lesellier, S., Boschiroli, M.L., Barrat, J., Wanke, C., Salguero, F.J., Garcia-Jimenez, W.L., Nunez, A., Godinho, A., Spiropoulos, J., Palmer, S., Dave, D., Anderson, P., Boucher, J.M., de Cruz, K., Henault, S., Michelet, L., Gowtage, S., Williams, G.A., Nadian, A.K., Monchatre-Leroy, E., Boue, F., Chambers, M.A., Richomme, C., 2019. Detection of live *M. bovis* BCG in tissues and IFN-gamma responses in European badgers (*Meles meles*) vaccinated by oropharyngeal instillation or directly in the ileum. BMC Vet Res 15, 445.

Lessel, E.F., 1960. Status of the generic name *Mycobacterium* Lehmann and Neumann. International Bulletin of Bacteriological Nomenclature and Taxonomy 10, 8.

Leth, C., Varadharajan, A., Mester, P., Fischaleck, M., Rossmanith, P., Schmoll, F., Fink, M., 2017. Matrixlysis, an improved sample preparation method for recovery of Mycobacteria from animal tissue material. PLoS One 12, e0181157.

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49, W293-W296.

Leung, K.S., Siu, G.K., Tam, K.K., Ho, P.L., Wong, S.S., Leung, E.K., Yu, S.H., Ma, O.C., Yam, W.C., 2018. Diagnostic evaluation of an in-house developed single-tube, duplex, nested IS*6110* real-time PCR assay for rapid pulmonary tuberculosis diagnosis. Tuberculosis (Edinburgh, Scotland) 112, 120-125.

Levy-Frebault, V.V., Portaels, F., 1992. Proposed minimal standards for the genus *Mycobacterium* and for description of new slowly growing *Mycobacterium* species. Int J Syst Bacteriol 42, 315-323.

Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., Sherman, D.R., 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. J Infect Dis 187, 117-123.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

Liébana, E., Aranaz, A., Dominguez, L., Mateos, A., González-Llamazares, O., Rodriguez-Ferri, E.F., Domingo, M., Vidal, D., Cousins, D., 1997. The insertion element IS*6110* is a useful tool for DNA fingerprinting of *Mycobacterium bovis* isolates from cattle and goats in Spain. Veterinary Microbiology 54, 223-233.

Liebana, E., Aranaz, A., Francis, B., Cousins, D., 1996. Assessment of genetic markers for species differentiation within the *Mycobacterium tuberculosis* complex. J Clin Microbiol 34, 933-938.

Liébana, E., Aranaz, A., Mateos, A., Vilafranca, M., Gomez-Mampaso, E., Tercero, J.C., Alemany, J., Suárez, G., Domingo, M., Dominguez, L., 1995. Simple and Rapid Detection of *Mycobacterium tuberculosis* Complex Organisms in Bovine Tissue Samples by PCR. Journal of Clinical Microbiology 33, 33-36.

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of naturally occurring bovine tuberculosis in England and Wales. Veterinary journal (London, England : 1997) 176, 354-360.

Livingstone, P.G., Hancox, N., Nugent, G., de Lisle, G.W., 2015. Toward eradication: the effect of *Mycobacterium bovis* infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. New Zealand Veterinary Journal 63, 4-18.

Loeffler, S.H., de Lisle, G.W., Neill, M.A., Collins, D.M., Price-Carter, M., Paterson, B., Crews, K.B., 2014. The seal tuberculosis agent, *Mycobacterium pinnipedii*, infects domestic cattle in New Zealand: epidemiologic factors and DNA strain typing. Journal of wildlife diseases 50, 180-187.

Loiseau, C., Brites, D., Moser, I., Coll, F., Pourcel, C., Robbe-Austerman, S., Escuyer, V., Musser, K.A., Peacock, S.J., Feuerriegel, S., Kohl, T.A., Niemann, S., Gagneux, S., Koser, C.U., 2019. Revised Interpretation of the Hain Lifescience GenoType MTBC To Differentiate *Mycobacterium canettii* and Members of the *Mycobacterium tuberculosis* Complex. Antimicrobial agents and chemotherapy 63.

Loiseau, C., Menardo, F., Aseffa, A., Hailu, E., Gumi, B., Ameni, G., Berg, S., Rigouts, L., Robbe-Austerman, S., Zinsstag, J., Gagneux, S., Brites, D., 2020. An African origin for *Mycobacterium bovis*. Evol Med Public Health 2020, 49-59.

Lombard, J.E., Patton, E.A., Gibbons-Burgener, S.N., Klos, R.F., Tans-Kersten, J.L., Carlson, B.W., Keller, S.J., Pritschet, D.J., Rollo, S., Dutcher, T.V., Young, C.A., Hench, W.C., Thacker, T.C., Perea, C., Lehmkuhl, A.D., Robbe-Austerman, S., 2021. Human-to-Cattle *Mycobacterium tuberculosis* Complex Transmission in the United States. Front Vet Sci 8, 691192.

Lomme, J.R., Thoen, C.O., Himes, E.M., Vinson, J.W., King, R.E., 1976. *Mycobacterium tuberculosis* infection in two East African oryxes. 169.

Lopes, B.C., Dos Reis, E.M., de Bitencourt, F.B.R., Loiko, M.R., Bezerra, A.V.A., Bueno, T.S., Lape, I.T., Cerva, C., Coppola, M.M., Rodrigues, R.O., Vargas, J.E., Bertagnolli, A.C., Mayer, F.Q., 2020. A molecular strategy to optimize bovine tuberculosis *post-mortem* diagnosis and the exposure to *Mycobacterium tuberculosis* variant bovis. Molecular biology reports 47, 7291-7296.

Lorente-Leal, V., Farrell, D., Romero, B., Alvarez, J., de Juan, L., Gordon, S.V., 2021. Performance and Agreement Between WGS Variant Calling Pipelines Used for Bovine Tuberculosis Control: Toward International Standardization. Front Vet Sci 8, 780018.

Lumb, R., Lanser, J.A., Lim, I.S., 1993. Rapid identification of mycobacteria by the Gen-Probe Accuprobe system. Pathology 25, 313-315.

Luquin, M., Laneelle, M.A., Ausina, V., Garcia Barcelo, M., Belda, F., Alonso, C., Prats, G., 1991. Distribution of a novel mycolic acid in species of the genus *Mycobacterium*. Int J Syst Bacteriol 41, 390-394.

Macedo, R., Isidro, J., Gomes, M.C., Botelho, A., Albuquerque, T., Sogorb, A., Bernardino, R., Fernandes, T.L., Mourato, T., Durval, M., Gomes, J.P., 2020. Animal-to-human transmission of *Mycobacterium pinnipedii*. The European respiratory journal 56.

Machen, A., Kobayashi, M., Connelly, M.R., Wang, Y.F., 2013. Comparison of heat inactivation and cell disruption protocols for identification of mycobacteria from solid culture media by use of vitek matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 51, 4226-4229.

Magnani, R., Cavalca, M., Pierantoni, M., Luppi, A., Cantoni, A.M., Prosperi, A., Pacciarini, M., Zanoni, M., Tamba, M., Santi, A., Bonardi, S., 2020. Infection by *Mycobacterium caprae* in three cattle herds in Emilia-Romagna Region, Northern Italy. Ital J Food Saf 9, 8467.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., Stover, C.K., 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol 178, 1274-1282.

Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol Mol Biol Rev 62, 725-774.

Majoor, C.J., Magis-Escurra, C., van Ingen, J., Boeree, M.J., van Soolingen, D., 2011. Epidemiology of *Mycobacterium bovis* disease in humans, The Netherlands, 1993-2007. Emerg Infect Dis 17, 457-463.

Malone, K.M., Gordon, S.V., 2017. *Mycobacterium tuberculosis* Complex Members Adapted to Wild and Domestic Animals, In: Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control, 2017/11/09 ed, pp. 135-154.

MAPA, 2021. Manual para la realización de estudios histopatológicos, inmunohistoquímicos y de PCR directa de tejidos para el diagnóstico rápido de la tuberculosis bovina por el Complejo *Mycobacterium tuberculosis* (CMT).

MAPA, 2022. Programa Nacional de Erradicación de Tuberculosis Bovina 2022 (Infección por el complejo *Mycobacterium tuberculosis*). p. 31.

MAPA, VISAVET, 2017. Manual de procedimiento para la toma y envío de muestras para el cultivo microbiológico y diagnóstico de tuberculosis, https://www.mapa.gob.es/es/ganaderia/temas/sanidad-animal-higiene-ganadera/.

MAPAMA, AECOSAN, 2017. Plan de Actuación sobre Tuberculosis en Especies Silvestres

Marcos, L.A., Spitzer, E.D., Mahapatra, R., Ma, Y., Halse, T.A., Shea, J., Isabelle, M., Lapierre, P., Escuyer, V.E., 2017. *Mycobacterium orygis* Lymphadenitis in New York, USA. Emerg Infect Dis 23, 1749-1751.

Marin, M., Vargas, R., Harris, M., Jeffrey, B., Epperson, L.E., Durbin, D., Strong, M., Salfinger, M., Iqbal, Z., Akhundova, I., Vashakidze, S., Crudu, V., Rosenthal, A., Farhat, M.R., 2021. Genomic sequence characteristics and the empiric accuracy of short-read sequencing. bioRxiv, 2021.2004.2008.438862.

Marrakchi, H., Laneelle, M.A., Daffe, M., 2014. Mycolic acids: structures, biosynthesis, and beyond. Chem Biol 21, 67-85.

Martinez-Guijosa, J., Lima-Barbero, J.F., Acevedo, P., Cano-Terriza, D., Jimenez-Ruiz, S., Barasona, J.A., Boadella, M., Garcia-Bocanegra, I., Gortazar, C., Vicente, J., 2021. Description and implementation of an On-farm Wildlife Risk Mitigation

Protocol at the wildlife-livestock interface: Tuberculosis in Mediterranean environments. Prev Vet Med 191, 105346.

Martinez-Lopez, B., Barasona, J.A., Gortazar, C., Rodriguez-Prieto, V., Sanchez-Vizcaino, J.M., Vicente, J., 2014. Farm-level risk factors for the occurrence, new infection or persistence of tuberculosis in cattle herds from South-Central Spain. Prev Vet Med 116, 268-278.

Martinez, L.R., Harris, B., Black, W.C.t., Meyer, R.M., Brennan, P.J., Vissa, V.D., Jones, R.L., 2008. Genotyping North American animal *Mycobacterium bovis* isolates using multilocus variable number tandem repeat analysis. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 20, 707-715.

Martiny, D., Busson, L., Wybo, I., El Haj, R.A., Dediste, A., Vandenberg, O., 2012. Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 50, 1313-1325.

Mather, C.A., Rivera, S.F., Butler-Wu, S.M., 2014. Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. J Clin Microbiol 52, 130-138.

Matsumoto, C.K., Bispo, P.J., Santin, K., Nogueira, C.L., Leao, S.C., 2014. Demonstration of plasmid-mediated drug resistance in *Mycobacterium abscessus*. J Clin Microbiol 52, 1727-1729.

McCallan, L., Brooks, C., Couzens, C., Young, F., McNair, J., Byrne, A.W., 2017. Assessment of serological tests for diagnosis of bovine tuberculosis. The Veterinary record 181, 90.

McElvania Tekippe, E., Shuey, S., Winkler, D.W., Butler, M.A., Burnham, C.A., 2013. Optimizing identification of clinically relevant Gram-positive organisms by use of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system. J Clin Microbiol 51, 1421-1427.

McEvoy, C.R., Cloete, R., Muller, B., Schurch, A.C., van Helden, P.D., Gagneux, S., Warren, R.M., Gey van Pittius, N.C., 2012. Comparative analysis of *Mycobacterium tuberculosis pe* and *ppe* genes reveals high sequence variation and an apparent absence of selective constraints. PLoS One 7, e30593.

McEvoy, C.R., Falmer, A.A., Gey van Pittius, N.C., Victor, T.C., van Helden, P.D., Warren, R.M., 2007. The role of IS*6110* in the evolution of *Mycobacterium tuberculosis*. Tuberculosis (Edinburgh, Scotland) 87, 393-404.

McEvoy, C.R., van Helden, P.D., Warren, R.M., Gey van Pittius, N.C., 2009a. Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis PPE38* gene region. BMC Evol Biol 9, 237.

McEvoy, C.R., Warren, R.M., van Helden, P.D., Gey van Pittius, N.C., 2009b. Multiple, independent, identical IS*6110* insertions in *Mycobacterium tuberculosis PPE* genes. Tuberculosis (Edinburgh, Scotland) 89, 439-442.

McHugh, T.D., Newport, L.E., Gillespie, S.H., 1997. IS*6110* homologs are present in multiple copies in mycobacteria other than tuberculosis-causing mycobacteria. J Clin Microbiol 35, 1769-1771.

McIlroy, S.G., Neill, S.D., McCracken, R.M., 1986. Pulmonary lesions and *Mycobacterium bovis* excretion from the respiratory tract of tuberculin reacting cattle. The Veterinary record 118, 718-721.

McLernon, J., Costello, E., Flynn, O., Madigan, G., Ryan, F., 2010. Evaluation of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat analysis and spoligotyping for genotyping of *Mycobacterium bovis* isolates and

a comparison with restriction fragment length polymorphism typing. J Clin Microbiol 48, 4541-4545.

Medeiros, L., Marassi, C.D., Duarte, R.S., da Silva, M.G., Lilenbaum, W., 2012. Comparison of decontamination methods for primary isolation of *Mycobacterium bovis* in paucibacillary bovine tissues. Letters in applied microbiology 54, 182-186.

Meehan, C.J., Barco, R.A., Loh, Y.E., Cogneau, S., Rigouts, L., 2021. Reconstituting the genus *Mycobacterium*. Int J Syst Evol Microbiol 71.

Meehan, C.J., Goig, G.A., Kohl, T.A., Verboven, L., Dippenaar, A., Ezewudo, M., Farhat, M.R., Guthrie, J.L., Laukens, K., Miotto, P., Ofori-Anyinam, B., Dreyer, V., Supply, P., Suresh, A., Utpatel, C., van Soolingen, D., Zhou, Y., Ashton, P.M., Brites, D., Cabibbe, A.M., de Jong, B.C., de Vos, M., Menardo, F., Gagneux, S., Gao, Q., Heupink, T.H., Liu, Q., Loiseau, C., Rigouts, L., Rodwell, T.C., Tagliani, E., Walker, T.M., Warren, R.M., Zhao, Y., Zignol, M., Schito, M., Gardy, J., Cirillo, D.M., Niemann, S., Comas, I., Van Rie, A., 2019. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. Nat Rev Microbiol 17, 533-545.

Menin, Á., Fleith, R., Reck, C., Marlow, M., Fernandes, P., Pilati, C., Báfica, A., 2013. Asymptomatic Cattle Naturally Infected with *Mycobacterium bovis* Present Exacerbated Tissue Pathology and Bacterial Dissemination. PLOS One 8.

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat Rev Genet 11, 31-46.

Michelet, L., Cruz, K.d., Karoui, C., Hénault, S., Boschiroli, M.L., 2017a. *Mycobacterium microti*, un agent tuberculoux méconnu. Épidémiologie et santé animale 71, 129-138.

Michelet, L., De Cruz, K., Henault, S., Tambosco, J., Richomme, C., Reveillaud, E., Gares, H., Moyen, J.L., Boschiroli, M.L., 2018a. *Mycobacterium bovis* Infection of Red Fox, France. Emerg Infect Dis 24, 1150-1153.

Michelet, L., de Cruz, K., Karoui, C., Hénault, S., Boschiroli, M.L., 2017b. *Mycobacterium microti* infection in a cow in France. The Veterinary record 180, 429.

Michelet, L., de Cruz, K., Karoui, C., Tambosco, J., Moyen, J.L., Henault, S., Boschiroli, M.L., 2018b. Second line molecular diagnosis for bovine tuberculosis to improve diagnostic schemes. PLoS One 13, e0207614.

Michelet, L., de Cruz, K., Phalente, Y., Karoui, C., Henault, S., Beral, M., Boschiroli, M.L., 2016. *Mycobacterium microti* Infection in Dairy Goats, France. Emerg Infect Dis 22, 569-570.

Michelet, L., de Cruz, K., Phalente, Y., Karoui, C., Hénault, S., Boschiroli, M.L., 2015a. *Mycobacterium microti* detection in French wildlife. The Veterinary record 177, 1.

Michelet, L., de Cruz, K., Tambosco, J., Henault, S., Boschiroli, M.L., 2020. *Mycobacterium microti* Interferes with Bovine Tuberculosis Surveillance. Microorganisms 8.

Michelet, L., de Cruz, K., Zanella, G., Aaziz, R., Bulach, T., Karoui, C., Henault, S., Joncour, G., Boschiroli, M.L., 2015b. Infection with *Mycobacterium microti* in animals in France. J Clin Microbiol 53, 981-985.

Middleton, S., Steinbach, S., Coad, M., McGill, K., Brady, C., Duignan, A., Wiseman, J., Gormley, E., Jones, G.J., Vordermeier, H.M., 2021. A molecularly defined skin test reagent for the diagnosis of bovine tuberculosis compatible with vaccination against Johne's Disease. Sci Rep 11, 2929.

Miller, J., Jenny, A., Rhyan, J., Saari, D., Suarez, D., 1997. Detection of *Mycobacterium bovis* in formalin-fixed, paraffin-embedded tissues of cattle and elk by PCR amplification of an IS*6110* sequence specific for *Mycobacterium tuberculosis* complex organisms. Journal of veterinary diagnostic investigation : official

publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 9, 244-249.
Miller, J.M., Jenny, A.L., Payeur, J.B., 2002. Polymerase chain reaction detection of *Mycobacterium tuberculosis* complex and *Mycobacterium avium* organisms in formalin-fixed tissues from culture-negative ruminants. Vet Microbiol 87, 15-23.
Miller, R.S., Sweeney, S.J., 2013. *Mycobacterium bovis* (bovine tuberculosis) infection in North American wildlife: current status and opportunities for mitigation of risks of further infection in wildlife populations. Epidemiology and infection 141, 1357-1370.
Mishra, A., Singhal, A., Chauhan, D.S., Katoch, V.M., Srivastava, K., Thakral, S.S., Bharadwaj, S.S., Sreenivas, V., Prasad, H.K., 2005. Direct Detection and Identification of *Mycobacterium tuberculosis* and *Mycobacterium bovis* in Bovine Samples by a Novel Nested PCR Assay: Correlation with Conventional Techniques. Journal of Clinical Microbiology 43, 5670-5678.
Mittal, M., Chakravarti, S., Sharma, V., Sanjeeth, B.S., Churamani, C.P., Kanwar, N.S., 2014. Evidence of presence of *Mycobacterium tuberculosis* in bovine tissue samples by multiplex PCR: possible relevance to reverse zoonosis. Transbound Emerg Dis 61, 97-104.
Modlin, S.J., Robinhold, C., Morrissey, C., Mitchell, S.N., Ramirez-Busby, S.M., Shmaya, T., Valafar, F., 2021. Exact mapping of Illumina blind spots in the *Mycobacterium tuberculosis* genome reveals platform-wide and workflow-specific biases. Microb Genom 7.
Monaghan, M.L., Doherty, M.L., Collins, J.D., Kazda, J.F., Quinn, P.J., 1994. The tuberculin test. Veterinary Microbiology 40, 111-124.
Montali, R.J., Mikota, S.K., Cheng, L.I., 2001. *Mycobacterium tuberculosis* in zoo and wildlife species. Revue scientifique et technique (International Office of Epizootics) 20, 291-303.
Moser, I., Prodinger, W.M., Hotzel, H., Greenwald, R., Lyashchenko, K.P., Bakker, D., Gomis, D., Seidler, T., Ellenberger, C., Hetzel, U., Wuennemann, K., Moisson, P., 2008. *Mycobacterium pinnipedii*: transmission from South American sea lion (*Otaria byronia*) to Bactrian camel (*Camelus bactrianus bactrianus*) and Malayan tapirs (*Tapirus indicus*). Vet Microbiol 127, 399-406.
Mostowy, S., Onipede, A., Gagneux, S., Niemann, S., Kremer, K., Desmond, E.P., Kato-Maeda, M., Behr, M., 2004. Genomic analysis distinguishes *Mycobacterium africanum*. J Clin Microbiol 42, 3594-3599.
Mukundan, H., Chambers, M.A., Waters, W.R., Larsen, M.H., 2015. Tuberculosis, Leprosy and Mycobacterial Diseases of Man and Animals: The Many Hosts of Mycobacteria. CABI.
Muller, B., Durr, S., Alonso, S., Hattendorf, J., Laisse, C.J., Parsons, S.D., van Helden, P.D., Zinsstag, J., 2013. Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. Emerg Infect Dis 19, 899-908.
Muller, B., Hilty, M., Berg, S., Garcia-Pelayo, M.C., Dale, J., Boschiroli, M.L., Cadmus, S., Ngandolo, B.N., Godreuil, S., Diguimbaye-Djaibe, C., Kazwala, R., Bonfoh, B., Njanpop-Lafourcade, B.M., Sahraoui, N., Guetarni, D., Aseffa, A., Mekonnen, M.H., Razanamparany, V.R., Ramarokoto, H., Djonne, B., Oloya, J., Machado, A., Mucavele, C., Skjerve, E., Portaels, F., Rigouts, L., Michel, A., Muller, A., Kallenius, G., van Helden, P.D., Hewinson, R.G., Zinsstag, J., Gordon, S.V., Smith, N.H., 2009. African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. J Bacteriol 191, 1951-1960.
Munoz Mendoza, M., Juan, L., Menendez, S., Ocampo, A., Mourelo, J., Saez, J.L., Dominguez, L., Gortazar, C., Garcia Marin, J.F., Balseiro, A., 2012. Tuberculosis due

to *Mycobacterium bovis* and *Mycobacterium caprae* in sheep. Veterinary journal (London, England : 1997) 191, 267-269.
Murai, K., Tizzani, P., Awada, L., Mur, L., Mapitse, N.J., Caceres, P., 2019. Controlling bovine tuberculosis: a One Health challenge, In: Panorama. OIE, OIE bulletin, p. 85.
Murcia, M.I., Tortoli, E., Menendez, M.C., Palenque, E., Garcia, M.J., 2006. *Mycobacterium colombiense* sp. nov., a novel member of the *Mycobacterium avium* complex and description of MAC-X as a new ITS genetic variant. Int J Syst Evol Microbiol 56, 2049-2054.
Muskens, J., van Zijderveld, F., Eger, A., Bakker, D., 2002. Evaluation of the long-term immune response in cattle after vaccination against paratuberculosis in two Dutch dairy herds. Veterinary Microbiology 86, 269-278.
Mustafa, M.I., Makhawi, A.M., 2021. SHERLOCK and DETECTR: CRISPR-Cas Systems as Potential Rapid Diagnostic Tools for Emerging Infectious Diseases. J Clin Microbiol 59.
Muwonge, A., Johansen, T.B., Vigdis, E., Godfroid, J., Olea-Popelka, F., Biffa, D., Skjerve, E., Djonne, B., 2012. *Mycobacterium bovis* infections in slaughter pigs in Mubende district, Uganda: a public health concern. BMC Vet Res 8, 168.
Naaktgeboren, C.A., Bertens, L.C., van Smeden, M., de Groot, J.A., Moons, K.G., Reitsma, J.B., 2013. Value of composite reference standards in diagnostic research. BMJ 347, f5605.
Nakanishi, N., Wada, T., Arikawa, K., Millet, J., Rastogi, N., Iwamoto, T., 2013. Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages. Infect Genet Evol 16, 174-177.
Naranjo, V., Gortazar, C., Vicente, J., de la Fuente, J., 2008. Evidence of the role of European wild boar as a reservoir of *Mycobacterium tuberculosis* complex. Vet Microbiol 127, 1-9.
Neill, S.D., Bryson, D.G., Pollock, J.M., 2001. Pathogenesis of tuberculosis in cattle. Tuberculosis (Edinburgh, Scotland) 81, 79-86.
Niemann, S., Richter, E., Dalugge-Tamm, H., Schlesinger, H., Graupner, D., Konigstein, B., Gurath, G., Greinert, U., Rusch-Gerdes, S., 2000. Two cases of *Mycobacterium microti* derived tuberculosis in HIV-negative immunocompetent patients. Emerg Infect Dis 6, 539-542.
Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., Hase, T., 2000. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res 28, E63.
Nugent, G., Buddle, B.M., Knowles, G., 2015a. Epidemiology and control of *Mycobacterium bovis* infection in brushtail possums (*Trichosurus vulpecula*), the primary wildlife host of bovine tuberculosis in New Zealand. N Z Vet J 63 Suppl 1, 28-41.
Nugent, G., Gortazar, C., Knowles, G., 2015b. The epidemiology of *Mycobacterium bovis* in wild deer and feral pigs and their roles in the establishment and spread of bovine tuberculosis in New Zealand wildlife. N Z Vet J 63 Suppl 1, 54-67.
Nuñez-Garcia, J., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R., Cook, A.J., de la Rua-Domenech, R., Goodchild, A.V., Gunn, J., More, S.J., Rhodes, S., Rolfe, S., Sharp, M., Upton, P.A., Vordermeier, H.M., Watson, E., Welsh, M., Whelan, A.O., Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., 2018. Meta-analyses of the sensitivity and specificity of *ante-mortem* and *post-mortem* diagnostic tests for bovine tuberculosis in the UK and Ireland. Prev Vet Med 153, 94-107.
Nyaruaba, R., Mwaliko, C., Kering, K.K., Wei, H., 2019. Droplet digital PCR applications in the tuberculosis world. Tuberculosis (Edinburgh, Scotland) 117, 85-92.

O'Hare, A., Balaz, D., Wright, D.M., McCormick, C., McDowell, S., Trewby, H., Skuce, R.A., Kao, R.R., 2021. A new phylodynamic model of *Mycobacterium bovis* transmission in a multi-host system uncovers the role of the unobserved reservoir. PLoS Comput Biol 17, e1009005.

O'Reilly, L.M., Daborn, C.J., 1995. The epidemiology of *Mycobacterium bovis* infections in animals and man: a review. Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease 76 Suppl 1, 1-46.

Oevermann, A., Pfyffer, G.E., Zanolari, P., Meylan, M., Robert, N., 2004. Generalized tuberculosis in llamas (*Lama glama*) due to *Mycobacterium microti*. J Clin Microbiol 42, 1818-1821.

OIE, 2018. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. World Organisation of Animal Health.

OIE, 2021. Infection with *Mycobacterium tuberculosis* complex, In: OIE - Terrestrial Animal Health Code. OIE, p. 6.

Olea-Popelka, F., Muwonge, A., Perera, A., Dean, A.S., Mumford, E., Erlacher-Vindel, E., Forcella, S., Silk, B.J., Ditiu, L., El Idrissi, A., Raviglione, M., Cosivi, O., LoBue, P., Fujiwara, P.I., 2017. Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis* —a call for action. The Lancet Infectious Diseases 17, e21-e25.

Olsen, I., Boysen, P., Kulberg, S., Hope, J.C., Jungersen, G., Storset, A.K., 2005. Bovine NK cells can produce gamma interferon in response to the secreted mycobacterial proteins ESAT-6 and MPP14 but not in response to MPB70. Infection and immunity 73, 5628-5635.

Orloski, K., Robbe-Austerman, S., Stuber, T., Hench, B., Schoenbaum, M., 2018. Whole Genome Sequencing of *Mycobacterium bovis* Isolated From Livestock in the United States, 1989-2018. Front Vet Sci 5, 253.

Orlowska, B., Krajewska-Wedzina, M., Augustynowicz-Kopec, E., Kozinska, M., Brzezinska, S., Zabost, A., Didkowska, A., Welz, M., Kaczor, S., Zmuda, P., Anusz, K., 2020. Epidemiological characterization of *Mycobacterium caprae* strains isolated from wildlife in the Bieszczady Mountains, on the border of Southeast Poland. BMC Vet Res 16, 362.

Ortiz, A.P., Perea, C., Davalos, E., Velazquez, E.F., Gonzalez, K.S., Camacho, E.R., Garcia Latorre, E.A., Lara, C.S., Salazar, R.M., Bravo, D.M., Stuber, T.P., Thacker, T.C., Robbe-Austerman, S., 2021. Whole Genome Sequencing Links *Mycobacterium bovis* From Cattle, Cheese and Humans in Baja California, Mexico. Front Vet Sci 8, 674307.

Ozturk-Gurgen, H., Rieseberg, B., Leipig-Rudolph, M., Straubinger, R.K., Hermanns, W., 2020. Morphology of Naturally-Occurring Tuberculosis in Cattle Caused by *Mycobacterium caprae*. Journal of comparative pathology 174, 120-139.

Palaniyandi, K., Kumar, N., Veerasamy, M., Kabir Refaya, A., Dolla, C., Balaji, S., Baskaran, D., Thiruvengadam, K., Rajendran, A., Narayanan, S., Raj, D., Swaminathan, S., Peacock, S.J., 2019. Isolation and comparative genomics of *Mycobacterium tuberculosis* isolates from cattle and their attendants in South India. Sci Rep 9, 17892.

Palgrave, C.J., Benato, L., Eatwell, K., Laurenson, I.F., Smith, N.H., 2012. *Mycobacterium microti* infection in two meerkats (*Suricata suricatta*). Journal of comparative pathology 146, 278-282.

Palmer, M.V., Thacker, T.C., Waters, W.R., Robbe-Austerman, S., Aldwell, F.E., 2014. Persistence of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) Danish in white-tailed deer (*Odocoileus virginianus*) vaccinated with a lipid-formulated oral vaccine. Transbound Emerg Dis 61, 266-272.

Palmer, M.V., Waters, W.R., 2011. Bovine tuberculosis and the establishment of an eradication program in the United States: role of veterinarians. Veterinary medicine international 2011, 816345.

Palmer, M.V., Waters, W.R., Thacker, T.C., 2007. Lesion development and immunohistochemical changes in granulomas from cattle experimentally infected with *Mycobacterium bovis*. Veterinary pathology 44, 863-874.

Palmer, M.V., Waters, W.R., Thacker, T.C., Greenwald, R., Esfandiari, J., Lyashchenko, K.P., 2006. Effects of different tuberculin skin-testing regimens on gamma interferon and antibody responses in cattle experimentally infected with *Mycobacterium bovis*. Clinical and vaccine immunology : CVI 13, 387-394.

Panteix, G., Gutierrez, M.C., Boschiroli, M.L., Rouviere, M., Plaidy, A., Pressac, D., Porcheret, H., Chyderiotis, G., Ponsada, M., Van Oortegem, K., Salloum, S., Cabuzel, S., Banuls, A.L., Van de Perre, P., Godreuil, S., 2010. Pulmonary tuberculosis due to *Mycobacterium microti*: a study of six recent cases in France. Journal of medical microbiology 59, 984-989.

Papaventsis, D., Dougas, G., Kalkouni, O., Karabela, S., Manika, K., 2021. Occupational Exposure to Zoonotic Tuberculosis Caused by *Mycobacterium caprae*, Northern Greece, 2019. Emerg Infect Dis 27, 1997-1999.

Parra, A., García, N., García, A., Lacombe, A., Moreno, F., Freire, F., Moran, J., Hermoso de Mendoza, J., 2008a. Development of a molecular diagnostic test applied to experimental abattoir surveillance on bovine tuberculosis. Vet Microbiol 127, 315-324.

Parra, A., García, N., García, A., Lacombe, A., Moreno, F., Freire, F., Moran, J., Mendoza, J.H.d., 2008b. Development of a molecular diagnostic test applied to experimental abattoir surveillance on bovine tuberculosis. Veterinary Microbiology 127, 315-324.

Parsons, L.M., Brosch, R., Cole, S.T., Somoskovi, A., Loder, A., Bretzel, G., Van Soolingen, D., Hale, Y.M., Salfinger, M., 2002. Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. J Clin Microbiol 40, 2339-2345.

Parsons, S., Smith, S.G., Martins, Q., Horsnell, W.G., Gous, T.A., Streicher, E.M., Warren, R.M., van Helden, P.D., Gey van Pittius, N.C., 2008. Pulmonary infection due to the dassie bacillus (*Mycobacterium tuberculosis* complex sp.) in a free-living dassie (rock hyrax-*Procavia capensis*) from South Africa. Tuberculosis (Edinburgh, Scotland) 88, 80-83.

Parsons, S.D., Drewe, J.A., Gey van Pittius, N.C., Warren, R.M., van Helden, P.D., 2013. Novel cause of tuberculosis in meerkats, South Africa. Emerg Infect Dis 19, 2004-2007.

Parsons, S.D., Warren, R.M., Ottenhoff, T.H., Gey van Pittius, N.C., van Helden, P.D., 2012. Detection of *Mycobacterium tuberculosis* infection in dogs in a high-risk setting. Res Vet Sci 92, 414-419.

Parte, A.C., Sarda Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C., Goker, M., 2020. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70, 5607-5612.

Patane, J.S., Martins, J., Beatriz Castelao, A., Nishibe, C., Montera, L., Bigi, F., Zumarraga, M.J., Cataldi, A.A., Fonseca Junior, A., Roxo, E., Luiza, A., Osorio, A.R., Jorge Ufms, K.S., Thacker, T.C., Almeida, N.F., Araujo, F.R., Setubal, J.C., 2017. Patterns and processes of *Mycobacterium bovis* evolution revealed by phylogenomic analyses. Genome Biol Evol.

Pate, M., Svara, T., Gombac, M., Paller, T., Zolnir-Dovc, M., Emersic, I., Prodinger, W.M., Bartos, M., Zdovc, I., Krt, B., Pavlik, I., Cvetnic, Z., Pogacnik, M., Ocepek, M., 2006. Outbreak of tuberculosis caused by *Mycobacterium caprae* in a zoological

garden. Journal of veterinary medicine. B, Infectious diseases and veterinary public health 53, 387-392.
Paudel, S., Sreevatsan, S., 2020. Tuberculosis in elephants: Origins and evidence of interspecies transmission. Tuberculosis (Edinburgh, Scotland) 123, 101962.
Pearson, C.W., Corner, L.A., Lepper, A.W., 1977. Tuberculin sensitivity of cattle inoculated with atypical mycobacteria isolated from cattle, feral pigs and trought water. Aust Vet J 53, 67-71.
Peker, N., Schuele, L., Kok, N., Terrazos, M., Neuenschwander, S.M., de Beer, J., Akkerman, O., Peter, S., Ramette, A., Merker, M., Niemann, S., Couto, N., Sinha, B., Rossen, J.W., 2021. Evaluation of whole-genome sequence data analysis approaches for short- and long-read sequencing of *Mycobacterium tuberculosis*. Microb Genom 7.
Perea, C., Ciaravino, G., Stuber, T., Thacker, T.C., Robbe-Austerman, S., Allepuz, A., Val, B.P., 2021. Whole-Genome SNP Analysis Identifies Putative *Mycobacterium bovis* Transmission Clusters in Livestock and Wildlife in Catalonia, Spain. Microorganisms 9.
Pérez, A., Reniero, A., Forteis, A., Meregalli, S., López, B., Ritacco, V., 2002. [Study of *Mycobacterium bovis* in milk using bacteriological methods and the polymerase chain reaction]. Revista Argentina de microbiologia 34, 45-51.
Pérez de Val, B., Sanz, A., Soler, M., Allepuz, A., Michelet, L., Boschiroli, M.L., Vidal, E., 2019. *Mycobacterium microti* Infection in Free-Ranging Wild Boar, Spain, 2017-2019. Emerg Infect Dis 25, 2152-2154.
Perez de Val, B., Vidal, E., Lopez-Soria, S., Marco, A., Cervera, Z., Martin, M., Mercader, I., Singh, M., Raeber, A., Domingo, M., 2016. Assessment of safety and interferon gamma responses of *Mycobacterium bovis* BCG vaccine in goat kids and milking goats. Vaccine 34, 881-886.
Perumaalla, V.S., Adams, L.G., Payeur, J.B., Jarnagin, J.L., Baca, D.R., Suarez Guemes, F., Ficht, T.A., 1996. Molecular epidemiology of *Mycobacterium bovis* in Texas and Mexico. J Clin Microbiol 34, 2066-2071.
Peterhans, S., Landolt, P., Friedel, U., Oberhansli, F., Dennler, M., Willi, B., Senn, M., Hinden, S., Kull, K., Kipar, A., Stephan, R., Ghielmetti, G., 2020. *Mycobacterium microti*: Not Just a Coincidental Pathogen for Cats. Front Vet Sci 7, 590037.
Phelan, J.E., Coll, F., Bergval, I., Anthony, R.M., Warren, R., Sampson, S.L., Gey van Pittius, N.C., Glynn, J.R., Crampin, A.C., Alves, A., Bessa, T.B., Campino, S., Dheda, K., Grandjean, L., Hasan, R., Hasan, Z., Miranda, A., Moore, D., Panaiotov, S., Perdigao, J., Portugal, I., Sheen, P., de Oliveira Sousa, E., Streicher, E.M., van Helden, P.D., Viveiros, M., Hibberd, M.L., Pain, A., McNerney, R., Clark, T.G., 2016. Recombination in *pe/ppe* genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. BMC genomics 17.
Pinsky, B.A., Banaei, N., 2008. Multiplex real-time PCR assay for rapid identification of *Mycobacterium tuberculosis* complex members to the species level. J Clin Microbiol 46, 2241-2246.
Pollock, J.M., Welsh, M.D., McNair, J., 2005. Immune responses in bovine tuberculosis: towards new strategies for the diagnosis and control of disease. Veterinary immunology and immunopathology 108, 37-43.
Poulet, S., Cole, S.T., 1995. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. Arch Microbiol 163, 87-95.
Pozo, P. 2020. Use of Epidemiological Analytic Tools and Molecular Characterization Techniques to Inform Bovine Tuberculosis Eradication Programs. Doctoral thesis, Universidad Complutense de Madrid.

Pozo, P., Cardenas, N.C., Bezos, J., Romero, B., Grau, A., Nacar, J., Saez, J.L., Minguez, O., Alvarez, J., 2021. Evaluation of the performance of slaughterhouse surveillance for bovine tuberculosis detection in Castilla y Leon, Spain. Prev Vet Med 189, 105307.

Pozo, P., Romero, B., Bezos, J., Grau, A., Nacar, J., Saez, J.L., Minguez, O., Alvarez, J., 2020. Evaluation of Risk Factors Associated With Herds With an Increased Duration of Bovine Tuberculosis Breakdowns in Castilla y Leon, Spain (2010-2017). Front Vet Sci 7, 545328.

Pranada, A.B., Witt, E., Bienia, M., Kostrzewa, M., Timke, M., 2017. Accurate differentiation of *Mycobacterium chimaera* from *Mycobacterium intracellulare* by MALDI-TOF MS analysis. Journal of medical microbiology 66, 670-677.

Praud, A., Boschiroli, M.L., Meyer, L., Garin-Bastuji, B., Dufour, B., 2015. Assessment of the sensitivity of the gamma-interferon test and the single intradermal comparative cervical test for the diagnosis of bovine tuberculosis under field conditions. Epidemiology and infection 143, 157-166.

Price-Carter, M., Brauning, R., de Lisle, G.W., Livingstone, P., Neill, M., Sinclair, J., Paterson, B., Atkinson, G., Knowles, G., Crews, K., Crispell, J., Kao, R., Robbe-Austerman, S., Stuber, T., Parkhill, J., Wood, J., Harris, S., Collins, D.M., 2018. Whole Genome Sequencing for Determining the Source of *Mycobacterium bovis* Infections in Livestock Herds and Wildlife in New Zealand. Front Vet Sci 5, 272.

Prodinger, W.M., Brandstatter, A., Naumann, L., Pacciarini, M., Kubica, T., Boschiroli, M.L., Aranaz, A., Nagy, G., Cvetnic, Z., Ocepek, M., Skrypnyk, A., Erler, W., Niemann, S., Pavlik, I., Moser, I., 2005. Characterization of *Mycobacterium caprae* isolates from Europe by mycobacterial interspersed repetitive unit genotyping. J Clin Microbiol 43, 4984-4992.

Prodinger, W.M., Eigentler, A., Allerberger, F., Schonbauer, M., Glawischnig, W., 2002. Infection of red deer, cattle, and humans with *Mycobacterium bovis* subsp. *caprae* in western Austria. J Clin Microbiol 40, 2270-2272.

Pucken, V.B., Knubben-Schweizer, G., Döpfer, D., Groll, A., Hafner-Marx, A., Hörmansdorfer, S., Sauter-Louis, C., Straubinger, R.K., Zimmermann, P., Hartnack, S., 2017. Evaluating diagnostic tests for bovine tuberculosis in the southern part of Germany: A latent class analysis. PLoS One 12, e0179847.

Puk, K., Banach, T., Wawrzyniak, A., Adaszek, L., Zietek, J., Winiarczyk, S., Guz, L., 2018. Detection of *Mycobacterium marinum, M. peregrinum, M. fortuitum* and *M. abscessus* in aquarium fish. J Fish Dis 41, 153-156.

Rabello, M.C., Matsumoto, C.K., Almeida, L.G., Menendez, M.C., Oliveira, R.S., Silva, R.M., Garcia, M.J., Leao, S.C., 2012. First description of natural and experimental conjugation between Mycobacteria mediated by a linear plasmid. PLoS One 7, e29884.

Radomski, N., Kreitmann, L., McIntosh, F., Behr, M.A., 2013. The Critical Role of DNA Extraction for Detection of Mycobacteria in Tissues. PLOS One 8.

Rahim, M.R., Möllers, M., te Koppele-Vije, A., de Beer, J., Zaman, K., Matin, M.A., Kamal, M., Raquib, R., van Soolingen, D., Baqi, M.A., Heilman, F.G.C., van der Zanden, A.G.M., 2007. Characterization of *Mycobacterium africanum* subtype I among cows in a dairy farm in Bangladesh using spoligotyping. Southeast Asian J Trop Med Public Health 38, 7.

Rahim, Z., Thapa, J., Fukushima, Y., van der Zanden, A.G.M., Gordon, S.V., Suzuki, Y., Nakajima, C., 2017. Tuberculosis Caused by *Mycobacterium orygis* in Dairy Cattle and Captured Monkeys in Bangladesh: a New Scenario of Tuberculosis in South Asia. Transbound Emerg Dis 64, 1965-1969.

Rajwani, R., Shehzad, S., Siu, G.K.H., 2018. MIRU-profiler: a rapid tool for determination of 24-loci MIRU-VNTR profiles from assembled genomes of *Mycobacterium tuberculosis*. PeerJ 6, e5090.

Randall, L.P., Lemma, F., Koylass, M., Rogers, J., Ayling, R.D., Worth, D., Klita, M., Steventon, A., Line, K., Wragg, P., Muchowski, J., Kostrzewa, M., Whatmore, A.M., 2015. Evaluation of MALDI-ToF as a method for the identification of bacteria in the veterinary diagnostic laboratory. Res Vet Sci 101, 42-49.

Raphaka, K., Matika, O., Sanchez-Molano, E., Mrode, R., Coffey, M.P., Riggio, V., Glass, E.J., Woolliams, J.A., Bishop, S.C., Banos, G., 2017. Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. BMC Genet 18, 27.

Raphaka, K., Sanchez-Molano, E., Tsairidou, S., Anacleto, O., Glass, E.J., Woolliams, J.A., Doeschl-Wilson, A., Banos, G., 2018. Impact of Genetic Selection for Increased Cattle Resistance to Bovine Tuberculosis on Disease Transmission Dynamics. Front Vet Sci 5, 237.

Rasolofo Razanamparany, V., Quirin, R., Rapaoliarijaona, A., Rakotoaritahina, H., Vololonirina, E.J., Rasolonavalona, T., Ferdinand, S., Sola, C., Rastogi, N., Ramarokoto, H., Chanteau, S., 2006. Usefulness of restriction fragment length polymorphism and spoligotyping for epidemiological studies of *Mycobacterium bovis* in Madagascar: description of new genotypes. Vet Microbiol 114, 115-122.

Ravva, S.V., Harden, L.A., Sarreal, C.Z., 2017. Characterization and Differentiation of *Mycobacterium avium* subsp. *paratuberculosis* from Other Mycobacteria Using Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. Frontiers in cellular and infection microbiology 7, 297.

Reddington, K., O'Grady, J., Dorai-Raj, S., Maher, M., van Soolingen, D., Barry, T., 2011a. Novel multiplex real-time PCR diagnostic assay for identification and differentiation of *Mycobacterium tuberculosis*, *Mycobacterium canettii*, and *Mycobacterium tuberculosis* complex strains. J Clin Microbiol 49, 651-657.

Reddington, K., O'Grady, J., Dorai-Raj, S., Niemann, S., van Soolingen, D., Barry, T., 2011b. A novel multiplex real-time PCR for the identification of mycobacteria associated with zoonotic tuberculosis. PLoS One 6, e23481.

Reddington, K., Zumla, A., Bates, M., van Soolingen, D., Niemann, S., Barry, T., O'Grady, J., 2012. SeekTB, a two-stage multiplex real-time-PCR-based method for differentiation of the *Mycobacterium tuberculosis* complex. J Clin Microbiol 50, 2203-2206.

Refaya, A.K., Kumar, N., Raj, D., Veerasamy, M., Balaji, S., Shanmugam, S., Rajendran, A., Tripathy, S.P., Swaminathan, S., Peacock, S.J., Palaniyandi, K., 2019. Whole-Genome Sequencing of a *Mycobacterium orygis* Strain Isolated from Cattle in Chennai, India. Microbiol Resour Announc 8.

Refregier, G., Sola, C., Guyeux, C., 2020. Unexpected diversity of CRISPR unveils some evolutionary patterns of repeated sequences in *Mycobacterium tuberculosis*. BMC genomics 21, 841.

Reis, A.C., Cunha, M.V., 2021. Genome-wide estimation of recombination, mutation and positive selection enlightens diversification drivers of *Mycobacterium bovis*. Sci Rep 11, 18789.

Reis, A.C., Salvador, L.C.M., Robbe-Austerman, S., Tenreiro, R., Botelho, A., Albuquerque, T., Cunha, M.V., 2021. Whole Genome Sequencing Refines Knowledge on the Population Structure of *Mycobacterium bovis* from a Multi-Host Tuberculosis System. Microorganisms 9.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217-223.

Reviriego Gordejo, F.J., Vermeersch, J.P., 2006. Towards eradication of bovine tuberculosis in the European Union. Vet Microbiol 112, 101-109.

Reynolds, J., Moyes, R.B., Breakwell, D.P., 2009. Differential staining of bacteria: acid fast stain. Curr Protoc Microbiol Appendix 3, Appendix 3H.

Ricchi, M., Mazzarelli, A., Piscini, A., Di Caro, A., Cannas, A., Leo, S., Russo, S., Arrigoni, N., 2017. Exploring MALDI-TOF MS approach for a rapid identification of *Mycobacterium avium* ssp. *paratuberculosis* field isolates. J Appl Microbiol 122, 568-577.

Richter, E., Niemann, S., Rusch-Gerdes, S., Hoffner, S., 1999. Identification of *Mycobacterium kansasii* by using a DNA probe (AccuProbe) and molecular techniques. J Clin Microbiol 37, 964-970.

Richter, E., Weizenegger, M., Rusch-Gerdes, S., Niemann, S., 2003. Evaluation of genotype MTBC assay for differentiation of clinical *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 41, 2672-2675.

Riojas, M.A., McGough, K.J., Rider-Riojas, C.J., Rastogi, N., Hazbon, M.H., 2018. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*. Int J Syst Evol Microbiol 68, 324-332.

Ritz, N., Hanekom, W.A., Robins-Browne, R., Britton, W.J., Curtis, N., 2008. Influence of BCG vaccine strain on the immune response and protection against tuberculosis. FEMS Microbiol Rev 32, 821-841.

Robbe-Austerman, S., Bravo, D.M., Harris, B., 2013. Comparison of the MGIT 960, BACTEC 460 TB and solid media for isolation of *Mycobacterium bovis* in United States veterinary specimens. BMC Veterinary Research 9:74.

Rodrigues, R.A., Ribeiro Araujo, F., Rivera Davila, A.M., Etges, R.N., Parkhill, J., van Tonder, A.J., 2021. Genomic and temporal analyses of *Mycobacterium bovis* in southern Brazil. Microb Genom 7.

Rodriguez-Campos, S., Gonzalez, S., de Juan, L., Romero, B., Bezos, J., Casal, C., Alvarez, J., Fernandez-de-Mera, I.G., Castellanos, E., Mateos, A., Saez-Llorente, J.L., Dominguez, L., Aranaz, A., Spanish Network on Surveillance Monitoring of Animal, T., 2012a. A database for animal tuberculosis (mycoDB.es) within the context of the Spanish national programme for eradication of bovine tuberculosis. Infect Genet Evol 12, 877-882.

Rodriguez-Campos, S., Navarro, Y., Romero, B., de Juan, L., Bezos, J., Mateos, A., Golby, P., Smith, N.H., Hewinson, G.R., Dominguez, L., Garcia-de-Viedma, D., Aranaz, A., 2013. Splitting of a prevalent *Mycobacterium bovis* spoligotype by variable-number tandem-repeat typing reveals high heterogeneity in an evolving clonal group. J Clin Microbiol 51, 3658-3665.

Rodriguez-Campos, S., Schurch, A.C., Dale, J., Lohan, A.J., Cunha, M.V., Botelho, A., De Cruz, K., Boschiroli, M.L., Boniotti, M.B., Pacciarini, M., Garcia-Pelayo, M.C., Romero, B., de Juan, L., Dominguez, L., Gordon, S.V., van Soolingen, D., Loftus, B., Berg, S., Hewinson, R.G., Aranaz, A., Smith, N.H., 2012b. European 2--a clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. Infect Genet Evol 12, 866-872.

Rodriguez-Campos, S., Smith, N.H., Boniotti, M.B., Aranaz, A., 2014. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. Res Vet Sci 97 Suppl, S5-S19.

Rodriguez-Sanchez, B., Ruiz-Serrano, M.J., Ruiz, A., Timke, M., Kostrzewa, M., Bouza, E., 2016. Evaluation of MALDI Biotyper Mycobacteria Library v3.0 for Identification of Nontuberculous Mycobacteria. J Clin Microbiol 54, 1144-1147.

Rodriguez-Temporal, D., Alcaide, F., Marekovic, I., O'Connor, J.A., Gorton, R., van Ingen, J., Van den Bossche, A., Hery-Arnaud, G., Beauruelle, C., Orth-Holler, D., Palacios-Gutierrez, J.J., Tudo, G., Bou, G., Ceyssens, P.J., Garrigo, M., Gonzalez-Martin, J., Greub, G., Hrabak, J., Ingebretsen, A., Mediavilla-Gradolph, M.C., Oviano, M., Palop, B., Pranada, A.B., Quiroga, L., Ruiz-Serrano, M.J., Rodriguez-Sanchez, B., 2022. Multicentre study on the reproducibility of MALDI-TOF MS for nontuberculous mycobacteria identification. Sci Rep 12, 1237.

Rodriguez-Temporal, D., Perez-Risco, D., Struzka, E.A., Mas, M., Alcaide, F., 2017. Impact of updating the MALDI-TOF MS database on the identification of nontuberculous mycobacteria. J Mass Spectrom 52, 597-602.

Rodriguez-Temporal, D., Perez-Risco, D., Struzka, E.A., Mas, M., Alcaide, F., 2018. Evaluation of Two Protein Extraction Protocols Based on Freezing and Mechanical Disruption for Identifying Nontuberculous Mycobacteria by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry from Liquid and Solid Cultures. J Clin Microbiol 56.

Rodriguez-Temporal, D., Rodriguez-Sanchez, B., Alcaide, F., 2020. Evaluation of MALDI Biotyper Interpretation Criteria for Accurate Identification of Nontuberculous Mycobacteria. J Clin Microbiol 58.

Rodriguez, S., Bezos, J., Romero, B., de Juan, L., Alvarez, J., Castellanos, E., Moya, N., Lozano, F., Javed, M.T., Saez-Llorente, J.L., Liebana, E., Mateos, A., Dominguez, L., Aranaz, A., Spanish Network on, S., Monitoring of Animal, T., 2011. *Mycobacterium caprae* infection in livestock and wildlife, Spain. Emerg Infect Dis 17, 532-535.

Rodriguez, S., Romero, B., Bezos, J., de Juan, L., Alvarez, J., Castellanos, E., Moya, N., Lozano, F., Gonzalez, S., Saez-Llorente, J.L., Mateos, A., Dominguez, L., Aranaz, A., Spanish Network on, S., Monitoring of Animal, T., 2010. High spoligotype diversity within a *Mycobacterium bovis* population: clues to understanding the demography of the pathogen in Europe. Vet Microbiol 141, 89-95.

Roe, W.D., Lenting, B., Kokosinska, A., Hunter, S., Duignan, P.J., Gartrell, B., Rogers, L., Collins, D.M., de Lisle, G.W., Gedye, K., Price-Carter, M., 2019. Pathology and molecular epidemiology of *Mycobacterium pinnipedii* tuberculosis in native New Zealand marine mammals. PLoS One 14, e0212363.

Romano, M.I., Alito, A., Fisanotti, J.C., Bigi, F., Kantor, I., Cicuta, M.E., Cataldi, A., 1996. Comparison of different genetic markers for molecular epidemiology of bovine tuberculosis. Veterinary Microbiology 50, 59-71.

Romero, B., Aranaz, A., Sandoval, A., Alvarez, J., de Juan, L., Bezos, J., Sanchez, C., Galka, M., Fernandez, P., Mateos, A., Dominguez, L., 2008. Persistence and molecular evolution of *Mycobacterium bovis* population from cattle and wildlife in Doñana National Park revealed by genotype variation. Vet Microbiol 132, 87-95.

Romero, B., Rodriguez, S., Bezos, J., Diaz, R., Copano, M.F., Merediz, I., Minguez, O., Marques, S., Palacios, J.J., Garcia de Viedma, D., Saez, J.L., Mateos, A., Aranaz, A., Dominguez, L., de Juan, L., 2011. Humans as source of *Mycobacterium tuberculosis* infection in cattle, Spain. Emerg Infect Dis 17, 2393-2395.

Roring, S., Scott, A., Brittain, D., Walker, I., Hewinson, G., Neill, S., Skuce, R., 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J Clin Microbiol 40, 2126-2133.

Ross, B.C., Raios, K., Jackson, K., Dwyer, B., 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J Clin Microbiol 30, 942-946.

Rossi, G., Crispell, J., Brough, T., Lycett, S.J., White, P.C.L., Allen, A., Ellis, R.J., Gordon, S.V., Harwood, R., Palkopoulou, E., Presho, E.L., Skuce, R., Smith, G.C., Kao, R.R.,

2021. Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections. Journal of Applied Ecology.

Roth, A., Fischer, M., Hamid, M.E., Michalke, S., Ludwig, W., Mauch, H., 1998. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. J Clin Microbiol 36, 139-147.

Rothel, J.S., Jones, S.L., Corner, L.A., Cox, J.C., Wood, P.R., 1990. A sandwich enzyme immunoassay for bovine interferon-gamma and its use for the detection of tuberculosis in cattle. Aust Vet J 67, 134-137.

Rothel, J.S., Jones, S.L., Corner, L.A., Cox, J.C., Wood, P.R., 1992. The gamma-interferon assay for diagnosis of bovine tuberculosis in cattle: conditions affecting the production of gamma-interferon in whole blood culture. Aust Vet J 69, 1-4.

Roupie, V., Alonso-Velasco, E., Van Der Heyden, S., Holbert, S., Duytschaever, L., Berthon, P., Van Dosselaer, I., Van Campe, W., Mostin, L., Biet, F., Roels, S., Huygen, K., Fretin, D., 2018. Evaluation of mycobacteria-specific gamma interferon and antibody responses before and after a single intradermal skin test in cattle naturally exposed to *M. avium* subsp. *paratuberculosis* and experimentally infected with *M. bovis*. Veterinary immunology and immunopathology 196, 35-47.

Roy, A., Infantes-Lorenzo, J.A., Blazquez, J.C., Venteo, A., Mayoral, F.J., Dominguez, M., Moreno, I., Romero, B., de Juan, L., Grau, A., Dominguez, L., Bezos, J., 2018. Temporal analysis of the interference caused by paratuberculosis vaccination on the tuberculosis diagnostic tests in goats. Prev Vet Med 156, 68-75.

Roy, A., Tome, I., Romero, B., Lorente-Leal, V., Infantes-Lorenzo, J.A., Dominguez, M., Martin, C., Aguilo, N., Puentes, E., Rodriguez, E., de Juan, L., Risalde, M.A., Gortazar, C., Dominguez, L., Bezos, J., 2019. Evaluation of the immunogenicity and efficacy of BCG and MTBVAC vaccines using a natural transmission model of tuberculosis. Vet Res 50, 82.

Roychowdhury, T., Mandal, S., Bhattacharya, A., 2015. Analysis of IS*6110* insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. Sci Rep 5, 12567.

Rüfenacht, S., Bogli-Stuber, K., Bodmer, T., Jaunin, V.F., Jmaa, D.C., Gunn-Moore, D.A., 2011. *Mycobacterium microti* infection in the cat: a case report, literature review and recent clinical experience. Journal of feline medicine and surgery 13, 195-204.

Runyon, E.H., 1959. Anonymous Mycobacteria in Pulmonary Disease. Medical Clinics of North America 43, 273-290.

Ryan, T.J., Buddle, B.M., De Lisle, G.W., 2000. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. Res Vet Sci 69, 57-61.

Sabio, Y.G.J., Bigi, M.M., Klepp, L.I., Garcia, E.A., Blanco, F.C., Bigi, F., 2020. Does *Mycobacterium bovis* persist in cattle in a non-replicative latent state as *Mycobacterium tuberculosis* in human beings? Vet Microbiol 247, 108758.

Sahraoui, N., Muller, B., Guetarni, D., Boulahbal, F., Yala, D., Ouzrout, R., Berg, S., Smith, N.H., Zinsstag, J., 2009. Molecular characterization of *Mycobacterium bovis* strains isolated from cattle slaughtered at two abattoirs in Algeria. BMC Vet Res 5, 4.

Sakula, A., 1983. BCG: who were Calmette and Guérin? Thorax 38, 6.

Saleeb, P.G., Drake, S.K., Murray, P.R., Zelazny, A.M., 2011. Identification of mycobacteria in solid-culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 49, 1790-1794.

Saltykova, A., Mattheus, W., Bertrand, S., Roosens, N.H.C., Marchal, K., De Keersmaecker, S.C.J., 2019. Detailed Evaluation of Data Analysis Tools for Subtyping of Bacterial Isolates Based on Whole Genome Sequencing: *Neisseria meningitidis* as a Proof of Concept. Front Microbiol 10, 2897.

Salvador, L.C.M., O'Brien, D.J., Cosgrove, M.K., Stuber, T.P., Schooley, A.M., Crispell, J., Church, S.V., Grohn, Y.T., Robbe-Austerman, S., Kao, R.R., 2019. Disease management at the wildlife-livestock interface: Using whole-genome sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan, USA. Mol Ecol 28, 2192-2205.

Sampson, S.L., 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clin Dev Immunol 2011, 497203.

Sanchez-Carvajal, J.M., Galan-Relano, A., Ruedas-Torres, I., Jurado-Martos, F., Larenas-Munoz, F., Vera, E., Gomez-Gascon, L., Cardoso-Toset, F., Rodriguez-Gomez, I.M., Maldonado, A., Carrasco, L., Tarradas, C., Gomez-Laguna, J., Luque, I., 2021. Real-Time PCR Validation for *Mycobacterium tuberculosis* Complex Detection Targeting IS*6110* Directly From Bovine Lymph Nodes. Front Vet Sci 8, 643111.

Savine, E., Warren, R.M., van der Spuy, G.D., Beyers, N., van Helden, P.D., Locht, C., Supply, P., 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 *loci* in serial isolates of *Mycobacterium tuberculosis*. J Clin Microbiol 40, 4561-4566.

Scherrer, S., Landolt, P., Carroli, N., Stephan, R., 2018. Molecular Characterization of *Mycobacterium avium* subsp. *hominissuis* of Two Groups of Lymph Nodes, Being Intradermal Tuberculin or Interferon-Gamma Test Positive and Negative, Isolated from Swiss Cattle at Slaughter. Front Vet Sci 5, 32.

Schiller, I., Waters, W.R., Vordermeier, H.M., Jemmi, T., Welsh, M., Keck, N., Whelan, A., Gormley, E., Boschiroli, M.L., Moyen, J.L., Vela, C., Cagiola, M., Buddle, B.M., Palmer, M., Thacker, T., Oesch, B., 2011. Bovine tuberculosis in Europe from the perspective of an officially tuberculosis free country: trade, surveillance and diagnostics. Vet Microbiol 151, 153-159.

Schiller, I., Waters, W.R., Vordermeier, H.M., Nonnecke, B., Welsh, M., Keck, N., Whelan, A., Sigafoose, T., Stamm, C., Palmer, M., Thacker, T., Hardegger, R., Marg-Haufe, B., Raeber, A., Oesch, B., 2009. Optimization of a whole-blood gamma interferon assay for detection of *Mycobacterium bovis*-infected cattle. Clinical and vaccine immunology : CVI 16, 1196-1202.

Scorpio, A., Zhang, Y., 1996. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat Med 2, 662-667.

Sergeant, E., Happold, J., Langstaff, I., 2017. Evaluation of Australian surveillance for freedom from bovine tuberculosis. Aust Vet J 95, 474-479.

Serrano, M., Sevilla, I.A., Fuertes, M., Geijo, M., Risalde, M.A., Ruiz-Fons, J.F., Gortazar, C., Juste, R.A., Dominguez, L., Elguezabal, N., Garrido, J.M., 2018. Different lesion distribution in calves orally or intratracheally challenged with *Mycobacterium bovis*: implications for diagnosis. Vet Res 49, 74.

Seva, J., Sanes, J.M., Ramis, G., Mas, A., Quereda, J.J., Villarreal-Ramos, B., Villar, D., Pallares, F.J., 2014. Evaluation of the single cervical skin test and interferon gamma responses to detect *Mycobacterium bovis* infected cattle in a herd co-infected with *Mycobacterium avium* subsp. *paratuberculosis*. Vet Microbiol 171, 139-146.

Sevilla, I.A., Arnal, M.C., Fuertes, M., Martin, E., Comenge, J., Elguezabal, N., Fernandez de Luco, D., Garrido, J.M., 2020. Tuberculosis outbreak caused by

Mycobacterium caprae in a rabbit farm in Spain. Transbound Emerg Dis 67, 431-441.

Sevilla, I.A., Molina, E., Elguezabal, N., Pérez, V., Garrido, J.M., Juste, R.A., 2015. Detection of mycobacteria, *Mycobacterium avium* subspecies, and *Mycobacterium tuberculosis* complex by a novel tetraplex real-time PCR assay. J Clin Microbiol 53, 930-940.

Shin, S.J., Lee, B.S., Koh, W.J., Manning, E.J., Anklam, K., Sreevatsan, S., Lambrecht, R.S., Collins, M.T., 2010. Efficient differentiation of *Mycobacterium avium* complex species and subspecies by use of five-target multiplex PCR. J Clin Microbiol 48, 4057-4062.

Shinnick, T.M., Good, R.C., 1995. Diagnostic mycobacteriology laboratory practices. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 21, 291-299.

Shury, T.K., Bergeson, D., 2011. Lesion Distribution and Epidemiology of *Mycobacterium bovis* in Elk and White-Tailed Deer in South-Western Manitoba, Canada. Veterinary medicine international 2011, 591980.

Siddiqui, S.H., Rüsh-Gerdes, S., 2006. MGIT™ Procedure Manual: for BACTEC™ MGIT 960™ TB System. Foundation for Innovative New Diagnostics.

Silva-Pereira, T.T., Ikuta, C.Y., Zimpel, C.K., Camargo, N.C.S., de Souza Filho, A.F., Ferreira Neto, J.S., Heinemann, M.B., Guimaraes, A.M.S., 2019. Genome sequencing of *Mycobacterium pinnipedii* strains: genetic characterization and evidence of superinfection in a South American sea lion (*Otaria flavescens*). BMC genomics 20, 1030.

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 3940-3941.

Sintchenko, V., Jelfs, P., Dally, M., Crighton, T., Gilbert, G.L., 2006. A case of urinary tuberculosis due to *Mycobacterium bovis* subspecies *caprae*. Pathology 38, 376-378.

Skuce, R., McCorry, T.P., McCarrol, J.F., Roring, S.M.M., Scott, A.N., Brittain, D., Hughes, S.L., Hewinson, G.R., Neill, S.D., 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology 148, 519-528.

Skuce, R.A., Brittain, D., Hughes, M.S., Beck, L.A., Neill, S.D., 1994. Genomic fingerprinting of *Mycobacterium bovis* from cattle by restriction fragment length polymorphism analysis. J Clin Microbiol 32, 2387-2392.

Skuce, R.A., Brittain, D., Hughes, M.S., Neill, S.D., 1996. Differentiation of *Mycobacterium bovis* isolates from animals by DNA typing. J Clin Microbiol 34, 2469-2474.

Skuce, R.A., Mallon, T.R., McCormick, C.M., McBride, S.H., Clarke, G., Thompson, A., Couzens, C., Gordon, A.W., McDowell, S.W., 2010. *Mycobacterium bovis* genotypes in Northern Ireland: herd-level surveillance (2003 to 2008). The Veterinary record 167, 684-689.

Smith, N., 1960. The 'Dassie' bacillus. Tubercle 41, 203-212.

Smith, N., 1965. Animal pathogenicity of the 'dassie bacillus'. Tubercle 46, 58-64.

Smith, N.H., 2012. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. Infect Genet Evol 12, 857-865.

Smith, N.H., Berg, S., Dale, J., Allen, A., Rodriguez, S., Romero, B., Matos, F., Ghebremichael, S., Karoui, C., Donati, C., Machado Ada, C., Mucavele, C., Kazwala, R.R., Hilty, M., Cadmus, S., Ngandolo, B.N., Habtamu, M., Oloya, J., Muller, A., Milian-Suazo, F., Andrievskaia, O., Projahn, M., Barandiaran, S., Macias, A., Muller, B., Zanini, M.S., Ikuta, C.Y., Rodriguez, C.A., Pinheiro, S.R., Figueroa, A., Cho, S.N., Mosavari, N., Chuang, P.C., Jou, R., Zinsstag, J., van Soolingen, D.,

Costello, E., Aseffa, A., Proano-Perez, F., Portaels, F., Rigouts, L., Cataldi, A.A., Collins, D.M., Boschiroli, M.L., Hewinson, R.G., Ferreira Neto, J.S., Surujballi, O., Tadyon, K., Botelho, A., Zarraga, A.M., Buller, N., Skuce, R., Michel, A., Aranaz, A., Gordon, S.V., Jeon, B.Y., Kallenius, G., Niemann, S., Boniotti, M.B., van Helden, P.D., Harris, B., Zumarraga, M.J., Kremer, K., 2011. European 1: a globally important clonal complex of *Mycobacterium bovis*. Infect Genet Evol 11, 1340-1351.

Smith, N.H., Dale, J., Inwald, J., Palmer, S., Gordon, S.V., Hewinson, R.G., Smith, J.M., 2003. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. Proc Natl Acad Sci U S A 100, 15271-15275.

Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006a. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat Rev Microbiol 4, 670-681.

Smith, N.H., Kremer, K., Inwald, J., Dale, J., Driscoll, J.R., Gordon, S.V., van Soolingen, D., Hewinson, R.G., Smith, J.M., 2006b. Ecotypes of the *Mycobacterium tuberculosis* complex. J Theor Biol 239, 220-225.

Smith, T., 1898. A Comparative Study of Bovine Tubercle Bacilli and of Human Bacilli from Sputum. J Exp Med 3, 451-511.

Smittipat, N., Palittapongarnpim, P., 2000. Identification of possible *loci* of variable number of tandem repeats in *Mycobacterium tuberculosis*. Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease 80, 69-74.

Sohn, J.I., Nam, J.W., 2018. The present and future of de novo whole-genome assembly. Brief Bioinform 19, 23-40.

Sorensen, A.L., Nagai, S., Houen, G., Andersen, P., Andersen, A.B., 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infection and immunity 63, 1710-1717.

Soto, C.Y., Menendez, M.C., Perez, E., Samper, S., Gomez, A.B., Garcia, M.J., Martin, C., 2004. IS*6110* mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol 42, 212-219.

Spurgiesz, R.S., Quitugua, T.N., Smith, K.L., Schupp, J., Palmer, E.G., Cox, R.A., Keim, P., 2003. Molecular typing of *Mycobacterium tuberculosis* by using nine novel variable-number tandem repeats across the Beijing family and low-copy-number IS*6110* isolates. J Clin Microbiol 41, 4224-4230.

Srilohasin, P., Chaiprasert, A., Tokunaga, K., Nao, N., Prammananan, T., 2014. Novel DNA chip based on a modified DigiTag2 assay for high-throughput species identification and genotyping of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol 52, 1962-1968.

Srinivasan, S., Jones, G., Veerasami, M., Steinbach, S., Holder, T., Zewude, A., Fromsa, A., Ameni, G., Easterling, L., Bakker, D., Juleff, N., Gifford, G., Hewinson, R.G., Vordermeier, H.M., Kapur, V., 2019. A defined antigen skin test for the diagnosis of bovine tuberculosis. Sci Adv 5, eaax4899.

Srinivasan, S., Subramanian, S., Shankar Balakrishnan, S., Ramaiyan Selvaraju, K., Manomohan, V., Selladurai, S., Jothivelu, M., Kandasamy, S., Gopal, D.R., Kathaperumal, K., Conlan, A.J.K., Veerasami, M., Bakker, D., Vordermeier, M., Kapur, V., 2020. A Defined Antigen Skin Test That Enables Implementation of BCG Vaccination for Control of Bovine Tuberculosis: Proof of Concept. Front Vet Sci 7, 391.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Steensels, D., Fauville-Dufaux, M., Boie, J., De Beenhouwer, H., 2013. Failure of PCR-Based IS*6110* analysis to detect vertebral spondylodiscitis caused by *Mycobacterium bovis*. J Clin Microbiol 51, 366-368.

Steinparzer, R., Stanclova, G., Bago, Z., Revilla-Fernandez, S., Leth, C., Hofer, E., Pohl, B., Schmoll, F., 2020. Generalized Tuberculosis Due to *Mycobacterium caprae* in a Red Fox (*Vulpes vulpes*) in Austria. Journal of wildlife diseases 56, 956-958.

Stephens, N., Vogelnest, L., Lowbridge, C., Christensen, A., Marks, G.B., Sintchenko, V., McAnulty, J., 2013. Transmission of *Mycobacterium tuberculosis* from an Asian elephant (*Elephas maximus*) to a chimpanzee (*Pan troglodytes*) and humans in an Australian zoo. Epidemiology and infection 141, 1488-1497.

Stewart, E.J., 2012. Growing unculturable bacteria. J Bacteriol 194, 4151-4160.

Stewart, L.D., McNair, J., McCallan, L., Gordon, A., Grant, I.R., 2013. Improved detection of *Mycobacterium bovis* infection in bovine lymph node tissue using immunomagnetic separation (IMS)-based methods. PLoS One 8, e58374.

Stucki, D., Malla, B., Hostettler, S., Huna, T., Feldmann, J., Yeboah-Manu, D., Borrell, S., Fenner, L., Comas, I., Coscolla, M., Gagneux, S., 2012. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. PLoS One 7, e41253.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H., Roring, S., Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M.C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C., van Soolingen, D., 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44, 4498-4510.

Supply, P., Brosch, R., 2017. The Biology and Epidemiology of *Mycobacterium canettii*, In: Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control. Springer International Publishing AG, pp. 27 - 38.

Supply, P., Magdalena, J., Himpens, S., Locht, C., 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Molecular Microbiology 26, 991-1003.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A., Fiette, L., Orgeur, M., Fabre, M., Parmentier, C., Frigui, W., Simeone, R., Boritsch, E.C., Debrie, A.-S., Willery, E., Walker, D., Quail, M.A., Ma, L., Bouchier, C., Salvignol, G., Sayes, F., Cascioferro, A., Seemann, T., Barbe, V., Locht, C., Gutierrez, M.-C., Leclerc, C., Bentley, S.D., Stinear, T.P., Brisse, S., Médigue, C., Parkhill, J., Cruveiller, S., Brosch, R., 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nature Genetics 45, 172.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B., Locht, C., 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol 36, 762-771.

Sweeney, F.P., Courtenay, O., Hibberd, V., Hewinson, R.G., Reilly, L.A., Gaze, W.H., Wellington, E.M., 2007. Environmental monitoring of *Mycobacterium bovis* in badger feces and badger sett soil by real-time PCR, as confirmed by immunofluorescence, immunocapture, and cultivation. Appl Environ Microbiol 73, 7471-7473.

Tafess, K., Ng, T.T.L., Lao, H.Y., Leung, K.S.S., Tam, K.K.G., Rajwani, R., Tam, S.T.Y., Ho, L.P.K., Chu, C.M.K., Gonzalez, D., Sayada, C., Ma, O.C.K., Nega, B.H., Ameni, G., Yam, W.C., Siu, G.K.H., 2020. Targeted-Sequencing Workflows for Comprehensive Drug Resistance Profiling of *Mycobacterium tuberculosis* Cultures Using Two

Commercial Sequencing Platforms: Comparison of Analytical and Diagnostic Performance, Turnaround Time, and Cost. Clin Chem 66, 809-820.
Tagliapietra, V., Boniotti, M.B., Mangeli, A., Karaman, I., Alborali, G., Chiari, M., D'Incau, M., Zanoni, M., Rizzoli, A., Pacciarini, M.L., 2021. *Mycobacterium microti* at the Environment and Wildlife Interface. Microorganisms 9.
Tantivitayakul, P., Panapruksachat, S., Billamas, P., Palittapongarnpim, P., 2010. Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. Tuberculosis (Edinburgh, Scotland) 90, 311-318.
Tao, Y., Yun, J., Wang, J., Xu, P., Li, C., Liu, H., Lan, Y., Pan, J., Du, W., 2020. High-performance detection of *Mycobacterium bovis* in milk using digital LAMP. Food Chem 327, 126945.
Tar, S.Y., Bozdemir, B., Gurel, M.S., Bilgin, F.U., Baran, M.F., Demirkesen, C., 2009. *Mycobacterium bovis caprae*: a rare agent of *lupus vulgaris*. Clin Exp Dermatol 34, 532-533.
Taylor, C., Jahans, K., Palmer, S., Okker, M., Brown, J., Steer, K., 2006. *Mycobacterium microti* isolated from two pigs. The Veterinary record 159, 59-60.
Taylor, G.M., Murphy, E., Hopkins, R., Rutland, P., Chistov, Y., 2007a. First report of *Mycobacterium bovis* DNA in human remains from the Iron Age. Microbiology 153, 1243-1249.
Taylor, G.M., Worth, D.R., Palmer, S., Jahans, K., Hewinson, R.G., 2007b. Rapid detection of *Mycobacterium bovis* DNA in cattle lymph nodes with visible lesions using PCR. BMC Vet Res 3, 12.
Taylor, M.J., Hughes, M.S., Skuce, R.A., Neill, S.D., 2001. Detection of *Mycobacterium bovis* in bovine clinical specimens using real-time fluorescence and fluorescence resonance energy transfer probe rapid-cycle PCR. J Clin Microbiol 39, 1272-1278.
Telenti, A., Marchesi, F., Balz, M., Bally, F., Bottger, E.C., Bodmer, T., 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J Clin Microbiol 31, 175-178.
Thacker, T.C., Harris, B., Palmer, M.V., Waters, W.R., 2011. Improved specificity for detection of *Mycobacterium bovis* in fresh tissues using IS*6110* real-time PCR. BMC Vet Res 7, 50.
Thapa, J., Nakajima, C., Maharjan, B., Poudel, A., Suzuki, Y., 2015. Molecular characterization of *Mycobacterium orygis* isolates from wild animals of Nepal. Japanese Journal of Veterinary Research 63, 8.
Thapa, J., Paudel, S., Sadaula, A., Shah, Y., Maharjan, B., Kaufman, G.E., McCauley, D., Gairhe, K.P., Tsubota, T., Suzuki, Y., Nakajima, C., 2016. *Mycobacterium orygis*-Associated Tuberculosis in Free-Ranging Rhinoceros, Nepal, 2015. Emerg Infect Dis 22, 570-572.
Thierry, D., Cave, M.D., Eisenach, K.D., Crawford, J.T., Bates, J.H., Gicquel, B., Guesdon, J.L., 1990. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Research 18.
Thoen, C., Lobue, P., de Kantor, I., 2006. The importance of *Mycobacterium bovis* as a zoonosis. Vet Microbiol 112, 339-345.
Thomas, J., Balseiro, A., Gortazar, C., Risalde, M.A., 2021. Diagnosis of tuberculosis in wildlife: a systematic review. Vet Res 52, 31.
Thompson, P.J., Cousins, D.V., Gow, B.L., Collins, D.M., Williamson, B.H., Dagnia, H.T., 1993. Seals, seal trainers, and mycobacterial infection. Am Rev Respir Dis 147, 164-167.
Thorel, M.F., 1980. Isolation of *Mycobacterium africanum* from monkeys. Tubercle 61, 101-104.

Tomita, N., Mori, Y., Kanda, H., Notomi, T., 2008. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. Nat Protoc 3, 877-882.

Tompkins, D.M., Ramsey, D.S., Cross, M.L., Aldwell, F.E., de Lisle, G.W., Buddle, B.M., 2009. Oral vaccination reduces the incidence of tuberculosis in free-living brushtail possums. Proc Biol Sci 276, 2987-2995.

Tortoli, E., 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin Microbiol Rev 16, 319-354.

Tortoli, E., 2010. Standard operating procedure for optimal identification of mycobacteria using 16S rRNA gene sequences. Stand Genomic Sci 3, 145-152.

Tortoli, E., Brown-Elliott, B.A., Chalmers, J.D., Cirillo, D.M., Daley, C.L., Emler, S., Floto, R.A., Garcia, M.J., Hoefsloot, W., Koh, W.J., Lange, C., Loebinger, M., Maurer, F.P., Morimoto, K., Niemann, S., Richter, E., Turenne, C.Y., Vasireddy, R., Vasireddy, S., Wagner, D., Wallace, R.J., Jr., Wengenack, N., van Ingen, J., 2019. Same meat, different gravy: ignore the new names of mycobacteria. The European respiratory journal 54.

Tortoli, E., Fedrizzi, T., Meehan, C.J., Trovato, A., Grottola, A., Giacobazzi, E., Serpini, G.F., Tagliazucchi, S., Fabio, A., Bettua, C., Bertorelli, R., Frascaro, F., De Sanctis, V., Pecorari, M., Jousson, O., Segata, N., Cirillo, D.M., 2017. The new phylogeny of the genus *Mycobacterium*: The old and the news. Infect Genet Evol 56, 19-25.

Tortoli, E., Mariottini, A., Mazzarelli, G., 2003. Evaluation of INNO-LiPA MYCOBACTERIA v2: improved reverse hybridization multiple DNA probe assay for mycobacterial identification. J Clin Microbiol 41, 4418-4420.

Tortoli, E., Pecorari, M., Fabio, G., Messino, M., Fabio, A., 2010. Commercial DNA probes for mycobacteria incorrectly identify a number of less frequently encountered species. J Clin Microbiol 48, 307-310.

Tortoli, E., Rindi, L., Garcia, M.J., Chiaradonna, P., Dei, R., Garzelli, C., Kroppenstedt, R.M., Lari, N., Mattei, R., Mariottini, A., Mazzarelli, G., Murcia, M.I., Nanetti, A., Piccoli, P., Scarparo, C., 2004. Proposal to elevate the genetic variant MAC-A, included in the *Mycobacterium avium* complex, to species rank as *Mycobacterium chimaera* sp. nov. Int J Syst Evol Microbiol 54, 1277-1285.

Tran, A., Alby, K., Kerr, A., Jones, M., Gilligan, P.H., 2015. Cost Savings Realized by Implementation of Routine Microbiological Identification by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol 53, 2473-2479.

Tran, V., Liu, J., Behr, M.A., 2014. BCG Vaccines. Microbiol Spectr 2, MGM2-0028-2013.

Trewby, H., Wright, D., Breadon, E.L., Lycett, S.J., Mallon, T.R., McCormick, C., Johnson, P., Orton, R.J., Allen, A.R., Galbraith, J., Herzyk, P., Skuce, R.A., Biek, R., Kao, R.R., 2016. Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. Epidemics 14, 26-35.

Tsuchida, S., Umemura, H., Nakayama, T., 2020. Current Status of Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry (MALDI-TOF MS) in Clinical Diagnostic Microbiology. Molecules 25.

Turenne, C.Y., Semret, M., Cousins, D.V., Collins, D.M., Behr, M.A., 2006. Sequencing of *hsp65* distinguishes among subsets of the *Mycobacterium avium* complex. J Clin Microbiol 44, 433-440.

Turenne, C.Y., Wallace, R., Jr., Behr, M.A., 2007. *Mycobacterium avium* in the postgenomic era. Clin Microbiol Rev 20, 205-229.

Twomey, D.F., Crawshaw, T.R., Anscombe, J.E., Farrant, L., Evans, L.J., Mc Elligot, W.S., Higgins, R.J., Dean, G., Vordermeier, M., Jahans, K., de la Rua-Domenech, R., 2007. TB in llamas caused by *Mycobacterium bovis*. The Veterinary record 160, 1.

Uchiya, K., Takahashi, H., Nakagawa, T., Yagi, T., Moriyama, M., Inagaki, T., Ichikawa, K., Nikai, T., Ogawa, K., 2015. Characterization of a novel plasmid, pMAH135, from *Mycobacterium avium* subsp. *hominissuis*. PLoS One 10, e0117797.

Vågene, A.J., Honap, T.P., Harkins, K.M., Rosenberg, M.S., Giffin, K., Cardenas-Arroyo, F., Leguizamon, L.P., Arnett, J., Buikstra, J.E., Herbig, A., Krause, J., Stone, A.C., Bos, K.I., 2022. Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. Nature communications 13, 1195.

Valcheva, V., Perea, C., Savova-Lalkovska, T., Dimitrova, A., Radulski, L., Mokrousov, I., Marinov, K., Najdenski, H., Bonovska, M., 2022. *Mycobacterium bovis* and *M. caprae* in Bulgaria: insight into transmission and phylogeography gained through whole-genome sequencing. BMC Vet Res 18, 148.

Valcheva, V., Savova-Lalkovska, T., Vyazovaya, A., Dimitrova, A., Bonovska, M., Najdenski, H., 2020. First insight into phylogeography of *Mycobacterium bovis* and *M. caprae* from cattle in Bulgaria. Infect Genet Evol 81, 104240.

van den Heever, L.W., 1984. Tuberculosis in milch goats. J S Afr Vet Assoc. 55, 2.

van der Zanden, A.G., Kremer, K., Schouls, L.M., Caimi, K., Cataldi, A., Hulleman, A., Nagelkerke, N.J., van Soolingen, D., 2002. Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* complex isolates by introduction of new spacer oligonucleotides. J Clin Microbiol 40, 4628-4639.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., Shinnick, T.M., et al., 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31, 406-409.

van Embden, J.D., van Gorkom, T., Kremer, K., Jansen, R., van Der Zeijst, B.A., Schouls, L.M., 2000. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J Bacteriol 182, 2393-2401.

van Heusden, P., Mashologu, Z., Lose, T., Warren, R., Christoffels, A., 2022. The COMBAT-TB Workbench: Making Powerful *Mycobacterium tuberculosis* Bioinformatics Accessible. mSphere, e0099121.

van Ingen, J., Boeree, M.J., Kosters, K., Wieland, A., Tortoli, E., Dekhuijzen, P.N., van Soolingen, D., 2009. Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. Int J Syst Evol Microbiol 59, 2277-2282.

van Ingen, J., Rahim, Z., Mulder, A., Boeree, M.J., Simeone, R., Brosch, R., van Soolingen, D., 2012. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. Emerg Infect Dis 18, 653-655.

van Soolingen, D., de Haas, P.E.W., Haagsma, J., Eger, T., Hermans, P.W.M., Ritacco, V., Alito, A., van Embden, J.D.A., 1994. Use of Various Genetic Markers in Differentiation of *Mycobacterium bovis* Strains from Animals and Humans and for Studying Epidemiology of Bovine Tuberculosis.

van Soolingen, D., Hermans, P.W., de Haas, P.E., Soll, D.R., van Embden, J.D., 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol 29, 2578-2586.

van Soolingen, D., Hermans, P.W., de Haas, P.E., van Embden, J.D., 1992. Insertion element IS*1081*-associated restriction fragment length polymorphisms in *Mycobacterium tuberculosis* complex species: a reliable tool for recognizing *Mycobacterium bovis* BCG. J Clin Microbiol 30, 1772-1777.

van Soolingen, D., van der Zanden, A.G.M., de Haas, P.E.W., Noordhoek, A., Foudraine, N.A., Portaels, F., Kolk, A.H.J., Kremer, K., van Embden, J.D.A., 1998.

Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. Journal of Clinical Microbiology 36, 5.

van Tonder, A.J., Thornton, M.J., Conlan, A.J.K., Jolley, K.A., Goolding, L., Mitchell, A.P., Dale, J., Palkopoulou, E., Hogarth, P.J., Hewinson, R.G., Wood, J.L.N., Parkhill, J., 2021. Inferring *Mycobacterium bovis* transmission between cattle and badgers using isolates from the Randomised Badger Culling Trial. PLoS Pathog 17, e1010075.

Varello, K., Pezzolato, M., Mascarino, D., Ingravalle, F., Caramelli, M., Bozzetta, E., 2008. Comparison of histologic techniques for the diagnosis of bovine tuberculosis in the framework of eradication programs. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 20, 164-169.

Vayr, F., Martin-Blondel, G., Savall, F., Soulat, J.M., Deffontaines, G., Herin, F., 2018. Occupational exposure to human *Mycobacterium bovis* infection: A systematic review. PLoS neglected tropical diseases 12, e0006208.

VENoMYC, 2006. Protocol and guidelines for Multilocus Variable Number Tandem Repeat Genotyping of *Mycobacterium bovis* (Annex), In: VNTR/MIRUs and DVR spoligotyping for *M. bovis* typing, Toledo, Spain.

Veyrier, F., Saïd-Salim, B., Behr, M.A., 2008. Evolution of the Mycobacterial *SigK* Regulon. Journal of Bacteriology 190, 1891-1899.

Vicente, J., Barasona, J.A., Acevedo, P., Ruiz-Fons, J.F., Boadella, M., Diez-Delgado, I., Beltran-Beck, B., Gonzalez-Barrio, D., Queiros, J., Montoro, V., de la Fuente, J., Gortazar, C., 2013. Temporal trend of tuberculosis in wild ungulates from Mediterranean Spain. Transbound Emerg Dis 60 Suppl 1, 92-103.

Vidal, E., Grasa, M., Perálvarez, T., Martín, M., Mercader, I., Pérez de Val, B., 2018. Transmission of tuberculosis caused by *Mycobacterium caprae* between dairy sheep and goats. Small Ruminant Research 158, 22-25.

Viljoen, I.M., van Helden, P.D., Millar, R.P., 2015. *Mycobacterium bovis* infection in the lion (*Panthera leo*): Current knowledge, conundrums and research challenges. Vet Microbiol 177, 252-260.

Votintseva, A.A., Bradley, P., Pankhurst, L., Del Ojo Elias, C., Loose, M., Nilgiriwala, K., Chatterjee, A., Smith, E.G., Sanderson, N., Walker, T.M., Morgan, M.R., Wyllie, D.H., Walker, A.S., Peto, T.E.A., Crook, D.W., Iqbal, Z., 2017. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. J Clin Microbiol 55, 1285-1298.

Wagner, J.C., Buchanan, G., Bokkenheuser, V., Leviseur, S., 1958. An acid-fast bacillus isolated from the lungs of the Cape hyrax, *Procavia capensis* (Pallas). Nature 181, 284-285.

Walker, T.M., Ip, C.L.C., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E.A., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. The Lancet Infectious Diseases 13, 137-146.

Walsh, H.E., Kidd, M.G., Moum, T., Friesen, V.L., 1999. Polytomies and the Power of Phylogenetic Inference. Evolution 53, 932-937.

Walter, K.S., Colijn, C., Cohen, T., Mathema, B., Liu, Q., Bowers, J., Engelthaler, D.M., Narechania, A., Lemmer, D., Croda, J., Andrews, J.R., 2020. Genomic variant-identification methods may alter *Mycobacterium tuberculosis* transmission inferences. Microb Genom 6.

Wangoo, A., Johnson, L., Gough, J., Ackbar, R., Inglut, S., Hicks, D., Spencer, Y., Hewinson, G., Vordermeier, M., 2005. Advanced Granulomatous Lesions in *Mycobacterium bovis*-infected Cattle are Associated with Increased Expression

of Type I Procollagen, γδ (WC1+) T Cells and CD 68+ Cells. Journal of comparative pathology 133, 223-234.

Wards, B.J., Collins, D.M., Lisle, G.W., 1995. Detection of *Mycobacterium bovis* in tissues by polymerase chain reaction. Veterinary Microbiology 43, 227-240.

Warren, R.M., Streicher, E.M., Sampson, S.L., van der Spuy, G.D., Richardson, M., Nguyen, D., Behr, M.A., Victor, T.C., van Helden, P.D., 2002. Microevolution of the Direct Repeat Region of *Mycobacterium tuberculosis*: Implications for Interpretation of Spoligotyping Data. Journal of Clinical Microbiology 40, 4457-4465.

Waters, W.R., Buddle, B.M., Vordermeier, H.M., Gormley, E., Palmer, M.V., Thacker, T.C., Bannantine, J.P., Stabel, J.R., Linscott, R., Martel, E., Milian, F., Foshaug, W., Lawrence, J.C., 2011. Development and evaluation of an enzyme-linked immunosorbent assay for use in the detection of bovine tuberculosis in cattle. Clinical and vaccine immunology : CVI 18, 1882-1888.

Wei, W., Cao, J., Wu, X.C., Cheng, L.P., Shen, X.N., Sha, W., Sun, Q., 2022. Diagnostic performance of metagenomic next-generation sequencing in non-tuberculous mycobacterial pulmonary disease when applied to clinical practice. Infection.

Wells, A.Q., 1937. Tuberculosis in Wild Voles. The Lancet 229.

Wells, A.Q., Robb-Smith, A.H.T., 1946. The murine type of tubercle bacillus (the vole acid-fast bacillus). Sir William Dunn School of Pathology, University of Oxford, Spec. Rep. Ser. Med., p. 47.

Welsh, M.D., Cunningham, R.T., Corbett, D.M., Girvin, R.M., McNair, J., Skuce, R.A., Bryson, D.G., Pollock, J.M., 2005. Influence of pathological progression on the balance between cellular and humoral immune responses in bovine tuberculosis. Immunology 114, 101-111.

Wernery, U., Kinne, J., Jahans, K.L., Vordermeier, H.M., Esfandiari, J., Greenwald, R., Johnson, B., Ul-Haq, A., Lyashchenko, K.P., 2007. Tuberculosis outbreak in a dromedary racing herd and rapid serological detection of infected camels. Vet Microbiol 122, 108-115.

Whelan, A.O., Coad, M., Cockle, P.J., Hewinson, G., Vordermeier, M., Gordon, S.V., 2010a. Revisiting host preference in the *Mycobacterium tuberculosis* complex: experimental infection shows *M. tuberculosis* H37Rv to be avirulent in cattle. PLoS One 5, e8527.

Whelan, C., Whelan, A.O., Shuralev, E., Kwok, H.F., Hewinson, G., Clarke, J., Vordermeier, H.M., 2010b. Performance of the Enferplex TB assay with cattle in Great Britain and assessment of its suitability as a test to distinguish infected and vaccinated animals. Clinical and vaccine immunology : CVI 17, 813-817.

Whipple, D.L., Bolin, C.A., Miller, J.M., 1996. Distribution of lesions in cattle infected with *Mycobacterium bovis*. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 8, 351-354.

WHO, 2020. Global Tuberculosis Report, Geneva.

WHO, FAO, OIE, 2017. Roadmap for zoonotic tuberculosis.

Wilton, S., Cousins, D., 1992. Detection and Identification of Multiple Mycobacterial Pathogens by DNA Amplification in a Single Tube. Genome Research 1, 269-273.

Woolford, A.J., Hewinson, R.G., Woodward, M., Dale, J.W., 1997. Sequence heterogeneity of an *mpb70* gene analogue in *Mycobacterium kansasii*. FEMS microbiology letters 148, 43-48.

Xia, E., Teo, Y.Y., Ong, R.T., 2016. SpoTyping: fast and accurate *in silico Mycobacterium* spoligotyping from sequence reads. Genome Med 8, 19.

Xin, T., Jia, H., Ding, J., Li, P., Yang, H., Hou, S., Yuan, W., Guo, X., Wang, H., Liang, Q., Li, M., Wang, B., Zhu, H., 2013. Assessment of a protein cocktail-based skin test for

bovine tuberculosis in a double-blind field test in cattle. Clinical and vaccine immunology : CVI 20, 482-490.

Xu, P., Yang, K., Yang, L., Wang, Z., Jin, F., Wang, Y., Feng, J., 2022. Next-Generation Metagenome Sequencing Shows Superior Diagnostic Performance in Acid-Fast Staining Sputum Smear-Negative Pulmonary Tuberculosis and Non-tuberculous Mycobacterial Pulmonary Disease. Front Microbiol 13, 898195.

Yates, G.F., Price-Carter, M., Bland, K., Joyce, M.A., Khan, F., Surrey, M., de Lisle, G.W., 2017. Comparison of the BBL mycobacteria growth indicator tube, the BACTEC 12B, and solid media for the isolation of *Mycobacterium bovis*. Journal of Veterinary Diagnostic Investigation 29, 508-512.

Yeboah-Manu, D., de Jong, B.C., Gehre, F., 2017. The Biology and Epidemiology of *Mycobacterium africanum*, In: Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control,. Springer.

Yeh, R.W., Ponce de Leon, A., Agasino, C.B., Hahn, J.A., Daley, C.L., Hopewell, P.C., Small, P.M., 1998. Stability of *Mycobacterium tuberculosis* DNA genotypes. J Infect Dis 177, 1107-1111.

Yesilkaya, H., Dale, J.W., Strachan, N.J., Forbes, K.J., 2005. Natural transposon mutagenesis of clinical isolates of *Mycobacterium tuberculosis*: how many genes does a pathogen need? J Bacteriol 187, 6726-6732.

Yue, J.X., Liti, G., 2019. simuG: a general-purpose genome simulator. Bioinformatics 35, 4442-4444.

Zanardi, G., Boniotti, M.B., Gaffuri, A., Casto, B., Zanoni, M., Pacciarini, M.L., 2013. Tuberculosis transmission by *Mycobacterium bovis* in a mixed cattle and goat herd. Res Vet Sci 95, 430-433.

Zanini, M.S., Moreira, E.C., Lopes, M.T., Oliveira, R.S., Leão, S.C., Fioravanti, R.L., Roxo, E., Zumarraga, M., Romano, M.I., Cataldi, A., Salas, C.E., 2001. *Mycobacterium bovis*: polymerase chain reaction identification in bovine lymphonode biopsies and genotyping in isolates from Southeast Brazil by spolygotyping and restriction fragment length polymorphism. Memorias do Instituto Oswaldo Cruz 96, 809-813.

Zanolari, P., Robert, N., Lyashchenko, K.P., Pfyffer, G.E., Greenwald, R., Esfandiari, J., Meylan, M., 2009. Tuberculosis caused by *Mycobacterium microti* in South American camelids. Journal of veterinary internal medicine 23, 1266-1272.

Zeng, W., Zhang, Y., Zhao, X., Huang, G., Jiang, Y., Dong, H., Li, X., Wan, K., He, C., 2013. Occurrence of non-tuberculous mycobacteria species in livestock from northern China and first isolation of *Mycobacterium caprae*. Epidemiology and infection 141, 1545-1551.

Zhu, R.Y., Zhang, K.X., Zhao, M.Q., Liu, Y.H., Xu, Y.Y., Ju, C.M., Li, B., Chen, J.D., 2009. Use of visual loop-mediated isotheral amplification of *rimM* sequence for rapid detection of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. J Microbiol Methods 78, 339-343.

Zimpel, C.K., Patane, J.S.L., Guedes, A.C.P., de Souza, R.F., Silva-Pereira, T.T., Camargo, N.C.S., de Souza Filho, A.F., Ikuta, C.Y., Neto, J.S.F., Setubal, J.C., Heinemann, M.B., Guimaraes, A.M.S., 2020. Global Distribution and Evolution of *Mycobacterium bovis* Lineages. Front Microbiol 11, 843.

Zmak, L., Obrovac, M., Jankovic Makek, M., Perko, G., Trkanjec, J.T., 2019. From Peruvian mummies to living humans: first case of pulmonary tuberculosis caused by *Mycobacterium pinnipedii*. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease 23, 1283-1285.

Zolg, J.W., Philippi-Schulz, S., 1994. The superoxide dismutase gene, a target for detection and identification of mycobacteria by PCR. J Clin Microbiol 32, 2801-2812.

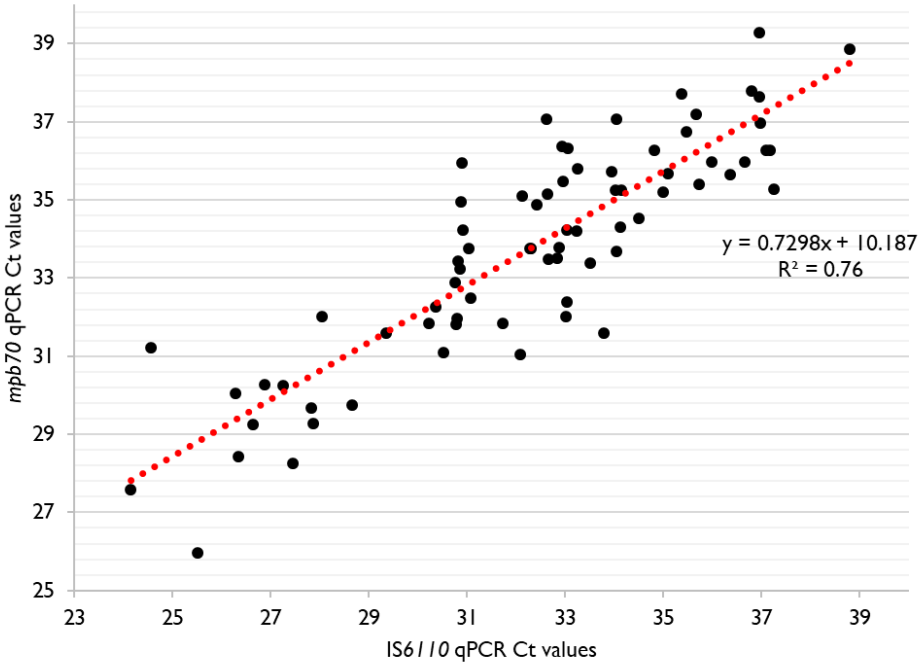
Zumarraga, M.J., Arriaga, C., Barandiaran, S., Cobos-Marin, L., de Waard, J., Estrada-Garcia, I., Figueiredo, T., Figueroa, A., Gimenez, F., Gomes, H.M., Gonzalez, Y.M.J.A., Macias, A., Milian-Suazo, F., Rodriguez, C.A., Santillan, M.A., Suffys, P.N., Trangoni, M.D., Zarraga, A.M., Cataldi, A., 2013. Understanding the relationship between *Mycobacterium bovis* spoligotypes from cattle in Latin American countries. Res Vet Sci 94, 9-21.

Zumarraga, M.J., Bernardell, A., Bastida, R., Quse, V., Loureiro, J., Cataldi, A., Bigi, F., Alito, A., Castro Ramos, M., Samper, S., Otal, I., Martin, C., Romano, M.I., 1999a. Molecular characterization of mycobacteria isolated from seals. Microbiology (Reading) 145 (Pt 9), 2519-2526.

Zumarraga, M.J., Martin, C., Samper, S., Alito, A., Latini, O., Bigi, F., Roxo, E., Cicuta, M.E., Errico, F., Ramos, M.C., Cataldi, A., van Soolingen, D., Romano, M.I., 1999b. Usefulness of spoligotyping in molecular epidemiology of *Mycobacterium bovis*-related infections in South America. J Clin Microbiol 37, 296-303.

Zwyer, M., Çavusoglu, C., Ghielmetti, G., Pacciarini, M.L., Scaltriti, E., Van Soolingen, D., Dötsch, A., Reinhard, M., Gagneux, S., Brites, D., 2021. A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics. Open Research Europe 1.

Appendix



Appendix 1. Pairwise representation of Ct values obtained with the *mpb70* and IS*6110* qPCRs in a subsample of 200 tissue samples from the IS*6110* study.