Check for updates

OPEN ACCESS

EDITED BY Maria Emilia Eirin, CONICET Institute of Agrobiotechnology and Molecular Biology (IABIMO), Argentina

REVIEWED BY

Marian Louise Price-Carter, AgResearch Ltd, New Zealand Erica Lasek-Nesselquist, Wadsworth Center, United States Andries Van Tonder, University of Cambridge, United Kingdom

*CORRESPONDENCE Eleftheria Palkopoulou 🖂 eleftheria.palkopoulou@apha.gov.uk

¹PRESENT ADDRESS Stefan Berg, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

RECEIVED 23 October 2024 ACCEPTED 04 February 2025 PUBLISHED 25 March 2025

CITATION

Sandhu P, Nunez-Garcia J, Berg S, Wheeler J, Dale J, Upton P, Gibbens J, Hewinson RG, Downs SH, Ellis RJ and Palkopoulou E (2025) Enhanced analysis of the genomic diversity of *Mycobacterium bovis* in Great Britain to aid control of bovine tuberculosis. *Front. Microbiol.* 16:1515906. doi: 10.3389/fmicb.2025.1515906

COPYRIGHT

© 2025 Sandhu, Nunez-Garcia, Berg, Wheeler, Dale, Upton, Gibbens, Hewinson, Downs, Ellis and Palkopoulou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhanced analysis of the genomic diversity of *Mycobacterium bovis* in Great Britain to aid control of bovine tuberculosis

Prizam Sandhu¹, Javier Nunez-Garcia¹, Stefan Berg^{1†}, Jo Wheeler², James Dale¹, Paul Upton³, Jane Gibbens⁴, R. Glyn Hewinson⁵, Sara H. Downs³, Richard J. Ellis⁶ and Eleftheria Palkopoulou^{1*}

¹Department of Bacteriology, Animal and Plant Health Agency, Addlestone, United Kingdom, ²Field Epidemiology, Animal Health and Welfare Advice Team, Professional Advice and Standards Directorate, Animal and Plant Health Agency, Addlestone, United Kingdom, ³Department of Epidemiological Sciences, Animal and Plant Health Agency, Addlestone, United Kingdom, ⁴Consultant Veterinary Epidemiologist, London, United Kingdom, ⁵Sèr Cymru Centre of Excellence for Bovine TB, Aberystwyth University, Aberystwyth, United Kingdom, ⁶Department of Surveillance and Laboratory Services, Animal and Plant Health Agency, Addlestone, United Kingdom

Bovine tuberculosis (bTB) is an endemic disease in Great Britain (GB) that affects mainly cattle but also other livestock and wild mammal species, leading to significant economic and social impact. Traditional genotyping of Mycobacterium bovis (M. bovis) isolates, which cause bTB, had been used routinely since the late 1990s as the main resource of genetic information in GB to describe their population and to understand their epidemiology. Since 2017, whole-genome sequencing (WGS) has been implemented on *M. bovis* isolates collected during routine surveillance. In this study, we analysed genome sequences from 3,052 M. bovis isolates from across GB to characterise their diversity and population structure in more detail. Our findings show that the M. bovis population in GB, based on WGS, is more diverse than previously indicated by traditional genotyping and can be divided into seven major clades, with one of them subdivided further into 29 clades that differ from each other by at least 70 single-nucleotide polymorphisms (SNPs). Based on the observed phylogenetic structure, we present a SNP-based classification system that replaces the genotype scheme that had been used until recently in GB. The predicted function and associated processes of the genes harbouring these SNPs are discussed with potential implications for phenotypic/functional differences between the identified clades. At the local scale, we show that WGS provides greater discriminatory power and that it can reveal the origin of infection and associated risk pathways even in areas of high bTB prevalence. The difficulty in determining transmission pathways due to the limited discrimination of isolates by traditional typing methods has compromised bTB control, as without such information it is harder to determine the relative efficacy of potential intervention measures. This study demonstrates that the higher resolution provided by WGS data can improve determination of infection sources and transmission pathways, provide important insights that will inform and shape bTB control policies in GB, as well as improve farm specific advice on interventions that are likely to be effective.

KEYWORDS

Mycobacterium bovis, whole genome sequencing, bovine tuberculosis (bTB), epidemiology, phylogenetics, clade

1 Introduction

Bovine tuberculosis (bTB) is one of the most significant animal health challenges in Great Britain (GB) that impacts the livestock sector and results in considerable costs to the government and industry, and has major implications for farmers (Godfray et al., 2018). Despite long-standing control strategies including test-andslaughter regimes, slaughterhouse surveillance, pre-/post-movement tests and movement restrictions, the incidence and geographic spread of bTB increased steadily in parts of England and Wales from the mid-1980s up to 2010, when they reached a plateau (Godfray et al., 2018). An overall declining trend in prevalence rates has been observed in England in the past 5 years, although this decrease has been heterogeneous (Gov.uk, 2024). Whilst many countries and regions have successfully eradicated or significantly reduced the disease using similar schemes (e.g., Scotland gained Official TB Free status in 2009), control efforts are more challenging in areas in which reservoirs of infection exist in wildlife (Palmer, 2013; Allen et al., 2018), as for instance in populations of the European badger (Meles meles) in GB (Godfray et al., 2013) and in possum in New Zealand (Livingstone et al., 2015). Furthermore, bTB poses a public health risk due to its zoonotic nature (Olea-Popelka et al., 2017; WHO, FAO, and WOAH, 2017), although the risk of human infection remains low in most developed countries since the introduction of milk pasteurisation (Müller et al., 2013).

Bovine TB causes a chronic respiratory disease with a pathology characterised by the development of granulomas in affected tissues of the infected animal (Domingo et al., 2014). Mycobacterium bovis (M. bovis), which causes bTB, is a member of the Mycobacterium tuberculosis complex (MTBC), a group of closely related and highly clonal organisms with over 99.9% identity in their genomes (Brosch et al., 2000). Other members of the MTBC include Mycobacterium tuberculosis and M. africanum, which are mostly human-adapted, and several animal-adapted species or ecotypes such as M. microti, M. caprae, M. pinnipedii, M. mungi, M. orygis, and M. suricattae (Fabre et al., 2004; Smith et al., 2006; Alexander et al., 2010; van Ingen et al., 2012; Parsons et al., 2013; Rodriguez-Campos et al., 2014). Although M. bovis has been associated primarily with cattle, it exhibits a wide range of wild mammalian hosts such as badgers, brush-tail possums, deer, elk, wild boar and water buffalo, as well as farmed or park deer, goats, pigs, sheep and South American camelids (Delahay et al., 2002; Fitzgerald and Kaneene, 2013).

Molecular typing methods have been extensively used in many countries to characterise and monitor the spread of *M. bovis* (Collins et al., 1986; Skuce et al., 2005; Tsao et al., 2014; Hauer et al., 2015). Earlier studies used traditional genetic markers to define four major clonal complexes carrying distinct deletion, spoligotype and singlenucleotide polymorphism (SNP) signatures. The clonal complexes African 1 (Af1) and African 2 (Af2) were discovered in West and East Africa, respectively (Müller et al., 2009; Berg et al., 2011), while European 2 was found to be predominant in the Iberian Peninsula and Brazil (Rodriguez-Campos et al., 2012). The European 1 (Eu1) clonal complex revealed a global distribution with prevalence in the United Kingdom (UK) and Ireland, and former trading countries of the UK (Smith et al., 2011).

In GB, spoligotyping, the most common typing technique of the MTBC, had been routinely used by the Animal and Plant Health Agency (APHA) for over 25 years as part of the bTB surveillance

programme, complemented by multiple loci variable number tandem repeat (VNTR) analysis since 2005. Spoligotyping detects polymorphisms in the presence or absence of unique spacer sequences within the direct repeat (DR) region of the MTBC genome (Groenen et al., 1993) while VNTR measures variation in the number of tandem repeats across multiple repetitive regions of the genome (Frothingham and Meeker-O'Connell, 1998). The combined output of these two methods determines the genotype of an isolate, information that has supported epidemiological investigations at the individual incident level and allowed for large-scale analysis of the population structure and spread of bTB at regional and national levels (Waller et al., 2022).

Based on this type of data, it has been shown that the different genotypes found in Britain are geographically localised, leading to the hypothesis that the observed diversity of *M. bovis* is the result of a series of clonal expansions of specific genotypes following the substantial population reduction in the 1950s and 1960s due to the statutory eradication programme (Smith et al., 2003; Smith et al., 2006). The regional localisation of genotypes has been captured by the so-called home ranges; geographical areas where particular genotypes are commonly observed. In 2020, over 96% of isolates collected in Britain were represented by 27 genotypes, of which 53% were assigned to the three most frequent genotypes (GB genotype 25:a [which corresponds to a strain with a specific VNTR variant of spoligotype SB0129], GB genotype 17:a [as above but spoligotype SB0263] and GB genotype 11:a [as above but spoligotype SB0274]) displaying an extensive distribution in the South West and West of England and Wales (Gov.uk, 2021).

Although spoligotyping coupled with VNTR typing have proved to be invaluable in identifying groups of genetically related strains over larger geographical scales and in tracing the source of infection, particularly in cases in which long-distance cattle movement has been involved (Skuce et al., 2010), their discriminatory power is limited since they measure variation only at a small portion of the genome. This has been particularly problematic in areas of high endemicity, such as the high-risk area (HRA) of England, where certain genotypes are widespread, at the within-or between-herd level of transmission that involves much finer-scale genetic differences, as well as at multihost species systems, in which finer resolution could help infer the direction of transmission. Moreover, due to their homoplastic nature, identical spoligotype or VNTR patterns can arise by convergent evolution rather than by evolutionary descent and therefore phylogenetic relationships cannot be reliably inferred (Warren et al., 2002). A new spoligotype pattern arises from the loss of one or more spacers in the DR region of a replicated bacterial genome, which due to the clonal nature of M. bovis is passed on to all its descendants, making it a suitable phylogenetic marker (Smith et al., 2003). However, because the frequency of these loss of spacer(s) mutation events is not as common as other mutations (e.g., SNPs), spoligotyping has limited discriminatory power. On the other hand, VNTR typing offers additional resolution since it involves multiple loci but is not as reliable phylogenetically since the number of repeats in the VNTR loci can increase or decrease, making it highly homoplastic.

Technological advances and continuously decreasing costs have made it possible for whole-genome sequencing (WGS) to be used as a diagnostic tool for outbreak investigations both in human TB (Gardy et al., 2011; Roetzer et al., 2013) as well as in bTB (Biek et al., 2012; Trewby et al., 2016; Price-Carter et al., 2018). WGS offers higher resolution compared to conventional typing methods as it assesses variation across the entire bacterial genome, making it possible to distinguish between isolates with identical genotypes. The greater degree of discrimination provided by WGS allows for more accurate reconstruction of risk pathways underlying transmission, especially when combined with high-quality epidemiological data, promising to offer a better understanding of the persistence and spread of bTB. For this reason, APHA implemented WGS in 2017 and officially replaced genotyping in 2021, following the recommendation from the review of the UK Government's strategy to achieve officially bTB free status in England by 2038, that WGS should be used routinely as a surveillance tool (Godfray et al., 2018).

In this study, we analysed genome sequences from ~3,052 *M. bovis* isolates collected across GB to characterise their genomic diversity and population structure. Our phylogenetic reconstruction demonstrates that the *M. bovis* population in Britain is more diverse than previously thought and can be divided into seven major groups, with one of them subdivided further into 29 clades that differ from each other by at least 70 single-nucleotide polymorphisms (SNPs). Based on the obtained phylogenomic structure, we propose a SNP-based classification system to replace the traditional genotype scheme. We compare the two methods, discuss the benefits of WGS as well as provide a case study to showcase the insights that can be gained by higher resolution genome-wide data when combined with epidemiological information.

2 Materials and methods

2.1 Sampling and laboratory methods

As part of the bTB statutory surveillance programme, bTB culture-positive isolates collected during routine surveillance were characterised by spoligotyping since the late 1990s with VNTR-typing of six loci implemented in 2005. Whole-genome sequencing was implemented in 2017 in parallel with genotyping up until 2021, when WGS officially replaced genotyping. For our dataset, we selected sequences from 3,052 isolates collected between 2000 and 2023. The majority of sequences were routine surveillance isolates sampled across GB in 2022 (n = 2,615) (including bovines and non-bovines) capturing the diversity of the contemporaneous *M. bovis* population, while isolates pre-dating 2022 were added to ensure that most prevalent genotypes in GB were represented in our dataset.

Samples were inoculated on modified 7H11 slopes (Gallagher and Horwill, 1977) for 6-12 weeks and a single colony was harvested for molecular characterisation. These procedures were performed in Containment Level 3 (CL3) laboratories located at the Department of Bacteriology at APHA. Heat-inactivated isolates were spoligotyped and VNTR-typed following the methods described in Cadmus et al. (2011). Paired-end libraries were constructed directly from the heatinactivated material without a DNA extraction or DNA purification step using the Nextera XT DNA Library Preparation Kit (Illumina, Cambridge UK) and sequencing was performed at the Central Unit for Sequencing and PCR (CUSP) at APHA on an Illumina MiSeq or NextSeq 500/550 instrument generating 150 bp paired-end reads. Isolates sampled prior to 2017 (n = 251), were first retrieved from the frozen bTB archives at APHA and re-cultured for up to 12 weeks, followed by sequencing of the heat-inactivated material as described above. Sequencing reads were deposited in the European Nucleotide Archive (ENA; accession number: PRJEB81540).

2.2 Sequencing data processing

Raw sequencing reads were processed by APHA's in-house pipeline for *M. bovis* sequencing data, btb-seq¹ with a few modifications as described below. Duplicate reads were removed by FastUniq V1.1 (Xu et al., 2012) and adaptors and low quality bases (using a sliding window of 10 bp and minimum average quality per base of 20) were removed using trimmomatic 0.38 (Bolger et al., 2014). The filtered reads were then mapped onto the updated version of the AF2122/97 reference genome (NC_002945.4; Malone et al., 2017) using BWA mem v0.7.17 (Li, 2013) with default settings. Samples with <90% of reads mapped to the reference genome indicate contamination from other organisms and were therefore not included in further analysis. SAMtools v1.10 and BCFtools 1.10.2 (Danecek et al., 2021) were used to perform variant calling and normalisation. The following filters were applied: minimum mapping quality of 30, minimum base quality of 10, minimum read depth at variant sites of 10, and minimum support for alternative or reference allele of 0.9. Regions that are problematic for SNP calling, such as repeat regions, IS elements, PPE and PE genes (Ates, 2020) as well as poorly covered regions were masked (positions listed in Supplementary Table 1). Poorly covered regions were determined by comparing 89 highquality genome sequences and identifying regions of the genome that had insufficient read depth or insufficient support for the alternative or reference allele (Ns). An exponential decay function was applied to define the threshold of the density of Ns (number of genome sequences with an N at a given position) depending on the window size of Ns. This means that shorter N windows had a higher threshold (threshold of 0.08 for N window size = 1) while larger N windows had lower thresholds (threshold towards 0.04 for larger window sizes). Regions with Ns that exceeded these density thresholds based on their window size were masked. Moreover, SNPs that were found to be within 10 bp from each other were removed as they tend to be associated with repetitive regions (Gardy et al., 2011). The consensus genome sequences were combined in a multi-fasta alignment and snp-sites v2.5.1 (Page et al., 2016) was used to extract all polymorphic sites in our dataset.

2.3 Phylogenetic reconstruction

Maximum-likelihood (ML) phylogenies were constructed in IQ-Tree 2.2.0-beta (Minh et al., 2020) using the-fconst option to correct for ascertainment bias in our SNP-alignment. The K3Pu + F + R5 substitution model was used in our phylogenetic analyses, which was selected by IQ-Tree's extended ModelFinder (Kalyaanamoorthy et al., 2017) as the third most optimal model. UNREST+ +FO + R5 and TVM + F + R5 were the top two selected models but due to their added complexity, they failed to converge in a reasonable amount of time. To verify the validity of internal nodes, a total of 1,000 ultrafast bootstrap replicates (Hoang et al., 2017) were performed. The sequencing data from a *Mycobacterium caprae* (*M. caprae*) isolate (accession number SRR7617662) was processed as described above and included in the phylogenetic analysis to root the

¹ https://github.com/APHA-CSU/btb-seg

tree. The ML phylogeny was annotated in TreeViewer v2.2.0 (Bianchini and Sánchez-Baracaldo, 2024).

Given the large number of sequences in our alignment, Treemmer v0.3 (Menardo et al., 2018) was used to remove redundant branches and better visualise the phylogenetic tree without compromising its genetic diversity. To achieve this, the maximum number of samples per WGS clade (as defined below – see section "Results and Discussion") was limited to 50. This was done to downsize WGS clades with a large number of genetically highly similar sequences while maintaining the number of sequences in underrepresented WGS clades. The phylogenetic tree of the trimmed dataset was constructed in IQ-Tree as described above.

2.4 Clade-specific SNPs

Unique SNPs were identified that were exclusively shared by all isolates in a defined WGS clade but not present (i.e., they were not polymorphic) in any of the other clades. For this analysis, high-quality sequences per clade were selected (> 99% of reads aligned to the reference genome with >50-fold depth of coverage and > 99% of the reference genome covered; with a few exceptions for some of the clades; see below). The ancestral state reconstruction feature in Mesquite v3.81 (Maddison and Maddison, 2023) was employed to infer the sequence of the most recent common ancestor of each WGS clade, which were then used to detect unique SNPs per clade. To facilitate this analysis (Mesquite cannot handle very high numbers of sequences), we selected 10 high-quality sequences per clade from the original (i.e., non-trimmed) ML phylogeny (listed in Supplementary Table 2). For clades that showed higher diversity comprising relatively deeply diverged sub-clades (e.g., B5-11), all high-quality sequences within those clades were included. For clades that were underrepresented, all available sequences within those clades were included irrespective of the quality criteria above (B6-12, n = 10; B6-21, n = 8; B6-22, n = 10; B6-23, n = 9; B6-31, n = 10; B6-41, *n* = 10; B6-51, *n* = 10; B6-53, *n* = 10; B6-61, *n* = 2; B6-81, *n* = 3; B6-88, n = 4; B6-92, n = 9; B7-11, n = 3). These sequences still had an average genomic read depth of at least 23-fold, over 91% of the reads aligned to the reference genome, and over 99% of the genome covered. All polymorphic positions from this alignment (n = 4,803) were used to generate a maximum-parsimony (MP) phylogeny in MEGA v11.0.11 (Tamura et al., 2021) and the nodes representing the most recent common ancestor of each of the defined clades were selected on the output tree file, which was used as the input for Mesquite. The sequence of the selected ancestral nodes were inferred by Mesquite v3.81 under the parsimony model. Polymorphisms that were unique in each of the ancestral sequences were extracted using a presence/ absence matrix, representing unique SNPs in each of the defined clades. The list of clade-specific SNPs was tested against the entire dataset and the clade of each sequence determined from scanning these sites on the WGS data was compared to the phylogenetic clade they grouped within, in the original ML tree (Supplementary Figure 1).

2.5 In silico spoligotyping

In silico spoligotyping was performed on the raw sequencing reads using SpoTyping v2.0 (Xia et al., 2016) and the output was

converted to the respective SB number in the *M. bovis* spoligotype database.² Spoligotype patterns that had not been observed before were given a new unique SB number upon request from the database above. The genotypes of isolates that had been spoligotyped and VNTR typed in the laboratory were retrieved from APHA's cattle testing SAM database.

2.6 Pairwise SNP distances

Between-and within-clade pairwise SNP distances were estimated using snp-dists³ v 0.8.2. For the within-clade estimates, all pairs of isolates within a clade were compared. For the between-clade estimates, all sequences from one clade were compared against all sequences from the clade that appeared to be the closest neighbour in the original ML phylogeny (e.g., B1-11 against B3-11 and B6-81 against B6-82). Between-and within-clade pairwise genetic distance distributions were visualised with pyplot in the matplotlib module (Hunter, 2007).

2.7 Clade-specific SNPs functional analysis

To investigate the potential functional effect of the clade-specific SNPs, SNPeff v. 5.2c (Cingolani et al., 2012) was used with *M. bovis* AF2122/97 (gca_000195835) as the annotated reference database. This tool outputs variant and genetic information for all variants against the reference genome and predicts the effect of variants found on coding regions. Additional information about the genes on which variants with a moderate (non-disrupting) or high (disrupting) predicted impact were found, such as associated biological process, molecular function and cellular component, were obtained from Uniprot (Consortium, T.U, 2022), using annotation information from *M. bovis* in the first instance and prioritising reviewed entries. For genes where annotation information did not exist for *M. bovis*, the search was expanded to *M. tuberculosis* and then to all mycobacteria.

2.8 Case study

Our dataset included WGS data from three isolates collected from a farm in the southwest (SW) of England that had an outbreak in 2022. This data was analysed together with isolates in our database that were found to be within the same WGS clade as the three isolates and a MP tree was generated in MEGA v11.0.11 (Tamura et al., 2021). The phylogenetic tree was interrogated to identify closely related isolates and integrated with cattle location and testing data, stored in APHA's SAM database, as well as cattle movement data from the Department of Environment, Food and Rural Affairs' (DEFRA)'s cattle tracing system (CTS), to determine the origin of infection and potential transmission routes in this incident.

² https://www.mbovis.org/database.php

³ https://github.com/tseemann/snp-dists

3 Results and discussion

3.1 Assembling and processing the dataset

In this study, genome sequences from a total of $3,052 \ M.$ *bovis* isolates from across GB were analysed covering a time span from 2000 to 2023. A total of 2,615 isolates (85.7%) were sampled and sequenced in 2022 while 344 isolates (11.3%) were also traditionally spoligotyped and VNTR-typed as part of the routine surveillance scheme up to 2021. This dataset was selected to obtain a representative sample of the *M. bovis* population circulating in GB, including representatives of the most frequently observed genotypes (Gov.uk, 2021). The vast majority (n = 2,889; 94.7%) of isolates originated from cattle sourced from a total of 1,923 herds while other host-species, including badgers, deer, alpacas, pig, sheep and cats, were represented in our dataset by 150 samples in total (n = 13 had unknown host-species). Single samples per animal were analysed.

Samples with <90% of reads mapped to the reference genome (n = 128; 4.2%) were excluded from further analysis. The remaining dataset (n = 2,924) had an average depth of coverage of 107-fold (range: 13–578) with >99% of the genome covered (average = 99.8%). Sequences that had no calls (Ns) in more than 50 polymorphic positions were removed, resulting in a total of 2,823 sequences in our alignment, which was used to reconstruct a ML phylogeny (Supplementary Figure 1). Figure 1 shows the reduced ML phylogeny (938 sequences; 8,834 SNPs) generated by Treemmer that removes redundancy in our dataset, allowing for better visualisation.

3.2 Phylogenetic structure and estimates of genomic diversity

Seven major clades were identified, named B1 - B7, which are monophyletic and have diverged considerably from each other (Figure 1; Supplementary Figure 1). Major clades B1 - B5 and B7 represent deeply diverged lineages with pairwise SNP distances of at least 350 SNPs between them (Figure 2A; Supplementary Figure 2). B1 - B4 and B7 exhibit limited diversity, while B5 appears to be more diverse and can be split further as and when additional data becomes available (Figure 2B; Supplementary Figure 3). Major clade B6 appears to be highly diverse and was further subdivided into 29 minor monophyletic clades (B6-11 to B6-92; Figure 1) that are well-supported (bootstrap support values >95%) and separated by at least 70 SNPs (Figure 2A; Supplementary Figure 2). These clades were defined visually and were originally guided by our previous knowledge of the genotypes found in Britain. However, as inconsistencies between the phylogenomic and genotyping data were observed (discussed in section 3.3 below), we deviated from using the genotypes and applied a SNP distance threshold of 70 SNPs, which was the observed minimum distance between well-supported minor clades. The nomenclature presented above uses simple alphanumeric coding, where "B" stands for bovis, the first digit denotes the major clade and the next two digits denote the minor clade (groups within the major clade). This system is hierarchical, indicating relatedness between clades (e.g., B6-11 and B6-12 are more closely related to each other than to B6-81 and all three are more closely related to each other than to B4-11), and can be expanded to allow for potentially newly identified clades to be included and existing clades to be further subdivided.

The topology of the tree with most clades appearing "flat" and separated by deep branches suggests that they have gone through recent population expansions. Although we did not attempt to date these expansions, this observation is consistent with the theory of a series of clonal expansions in GB following the severe reduction in the population size of *M. bovis* in the 1950s and 1960s caused by the cattle testing and eradication programme (Smith et al., 2003; Smith et al., 2006). In future work, we aim to retrospectively sequence historical isolates from APHA's bTB collection going back to the late 1980s to enhance the temporal signal in our dataset, which will allow us to generate dated phylogenies and estimate the timings of expansions and divergence of the defined WGS clades.

The geographic distribution of each WGS clade was examined, by plotting the location of the source holdings of the cattle isolates using the online tool grid reference finder⁴. Similar to the geographic localisation previously observed for the different genotypes in GB (Smith et al., 2003), which led to the definition of home ranges, i.e., areas in which certain genotypes are expected, most WGS clades appear to have a defined geographical distribution. Geographical range does not appear to correlate with genomic diversity; clades with a limited geographical range (e.g., B6-16 and B6-41 in Figure 3) exhibit varying levels of genetic diversity. B6-41 exhibits low levels of diversity (pairwise SNP distances <15 SNPs) while B6-16 exhibits higher diversity (average: 18 SNPs and max: 35 SNPs). Clades that are geographically more widespread occupying large parts of the HRA, also exhibit varying levels of genomic diversity. For instance, B6-11, the most prevalent and widespread clade, displays high genomic diversity (up to 65 SNPs). On the other hand, B6-62, another clade with extensive geographic distribution covering parts of 12 counties, displays lower diversity (average: 11 SNPs and max: 32 SNPs), which is on average lower than that observed in B6-16. This needs to be further explored statistically in future studies, testing for isolation by distance, as well as comparing the phylogenomic data with cattle movement direction and frequency data (either specific data for the source holdings or typical patterns for farm types/locations) to better understand their role in geographical spread.

A few of the WGS clades have bi-or multi-modal distributions of genetic diversity with one or more of the observed peaks within the expected range of within-clade pairwise SNP distances (0-65 SNPs) and additional peaks at a higher range (Supplementary Figure 3; B5-11, B6-31, B6-51, B6-92). The latter peaks are mostly due to the presence of diverged isolates within these clades, which potentially represent under-sampled M. bovis variation within our dataset, most likely originating from imported infection. In fact, a few of our clades (e.g., B5-11, B6-89, B7-11) are associated with imported infected animals as evidenced from movement history and could therefore be considered as non-endemic. Despite their large genetic distances, these isolates were not separated out from their respective WGS clades since their numbers were limited. However, expanding our sampling spatially, including from countries where imported animals are likely sourced, as well as temporally could possibly fill in those gaps and allow us to further split some of these WGS clades into distinct groups. The nomenclature presented above uses simple alphanumeric coding that can be expanded to allow for potentially newly discovered clades to be included and existing clades to be further subdivided.

⁴ https://gridreferencefinder.com/



(Continued)

FIGURE 1 (Continued)

are labelled on the tree and minor clades (B6-11–B6-92) are shown next to the tips of the tree. WGS clade and *in silico* predicted spoligotype for each isolate are indicated in the bars next to the phylogeny (from left to right). Grey bars on the *in silico* predicted spoligotype annotations mean that a unique spoligotype pattern was identified that had not been observed for any other isolate (singletons). Bootstrap support values for major internal nodes including nodes that represent the most recent common ancestor of each of the WGS clades are over 95%.



FIGURE 2

Frequency distribution of pairwise SNP distances between isolates from select WGS clades (A) and between isolates within a WGS clade (B).



FIGURE 3

Frequency distributions of within-clade pairwise SNP distances (top) and geographical ranges (bottom) for clades B6-11 (**A**), B6-16 (**B**), B6-26 (**C**) and B6-41 (**D**). B6-11 and B6-62 (**A**, **D**) represent widespread clades with relatively high and low levels of genetic diversity, respectively, while B6-16 and B6-41 (**B**, **C**) represent clades with constrained geographical ranges and medium or low diversity, respectively. The y-axis in the frequency distribution plots shows the number of pairwise comparisons and the x-axis shows pairwise genetic distances in number of SNPs. Note that the scale of the x-axis differs in each of the clades.

3.3 Correlation of WGS clades and spoligotypes

Estimates of diversity based on the genome-wide data were compared against estimates based on *in silico* predicted spoligotyping. A detailed comparison between lab-based and *in silico* predicted spoligotypes is provided in the Supplementary Information. Overall, there appears to be agreement between the defined WGS clades and previously established spoligotypes. This is shown in Figure 1 as well as in Table 1 that maps the spoligotypes observed in each WGS clade. However, almost half of the WGS clades include more than one spoligotype with the majority of isolates coming from a single spoligotype (e.g., SB0140 [GB type 9] in clade B6-87) and a small proportion of isolates from different spoligotypes (SB0054 [GB type 65], SB0980 [GB type 103], and SB2830 in clade B6-87). By examining these spoligotype patterns, we can see that single mutation events in the most prevalent spoligotype within a clade have given rise to the less common spoligotypes that are present within the same clade via the loss of a single or multiple contiguous unique spacers within the DR region (see <u>Supplementary Figure 4</u> for example above), demonstrating that isolates with different spoligotypes can in fact be genetically closely related to each other, contrary to what their different spoligotypes might have suggested. There are also spoligotypes/genotypes that are polyphyletic and therefore split into multiple WGS clades across the phylogeny. For instance, SB0273 [GB type 13] is observed in clades B5-11, B6-51 and B6-71, which do not form a monophyletic group (Figure 1; Supplementary Figure 1 and Supplementary Table 1). Other spoligotypes observed in polyphyletic groups are SB0140 [GB type 9], SB0130 [GB type 21], and SB0263 [GB type 17]. This can be explained by a process by which an ancestral lineage with a prevalent spoligotype

B1-11	SB0134								
B2-11	SB0145								
B3-11	SB0129	SB1551							
B4-11	SB0130	SB1446	SB0914	SB2835					
B5-11	SB0273	SB0054	SB0140	SB2817					
B6-11	SB0263	SB0959	SB0957	SB1415	SB2059	SB2728	SB0661	SB0977	SB2032
	SB2828	SB2829							
B6-12	SB0140	SB1924							
B6-13	SB0140	SB1524							
B6-14	SB0140	SB0996	SB0139	SB1531	SB0054	SB0980	SB1924	SB0269	SB1524
	SB0673	SB2830	SB2826						
B6-15	SB0140								
B6-16	SB0140	SB2486							
B6-21	SB0140	SB1033							
B6-22	SB0140								
B6-23	SB0263								
B6-31	SB0140	SB0142	SB0980						
B6-41	SB1033								
B6-42	SB0275								
B6-51	SB0140	SB0273	SB0996	SB0269	SB2833				
B6-52	SB0271								
B6-53	SB0140								
B6-61	SB0140								
B6-62	SB0272	SB2032	SB2831	SB1952	SB2833				
B6-71	SB0273								
B6-81	SB0140								
B6-82	SB0140								
B6-83	SB0673	SB0140	SB0269	SB2486	SB2826				
B6-84	SB0140	SB0269							
B6-85	SB0274	SB1499	SB2827	SB2832	SB2834				
B6-86	SB0140								
B6-87	SB0140	SB0054	SB0980	SB2830					
B6-88	SB0140								
B6-89	SB0142								
B6-91	SB0140								
B6-92	SB0140								
B7-11	SB0130								

TABLE 1 WGS clades and their corresponding spoligotypes based on the lab-inferred and in silico predicted spoligotypes of the isolates in our dataset.

Italic font indicates less frequent spoligotypes while non-italic font indicates prevalent spoligotypes within each of the WGS clades. Spoligotypes that were observed only once (singletons) are not shown in this table.

08

10.3389/fmicb.2025.1515906

pattern gave rise to a large number of lineages (i.e., WGS clades), some of which share the same spoligotype pattern (i.e., no loss of spacer(s) events in their DR region during their evolution) while others evolved different spoligotype patterns due to the loss of a single or multiple spacers in their DR region. This process can also lead to homoplastic mutations in the direct repeat region, by which identical spoligotypes can emerge in phylogenetically unrelated lineages (Smith et al., 2006). Spoligotype SB0140 (GB type 9) is noteworthy; it is historically the most prevalent spoligotype in GB and is believed to have given rise to other common spoligotypes as well as diversified into many lineages (Smith et al., 2006). In our analysis, it is observed in 21 different WGS clades (Table 1; Figure 1; Supplementary Figure 1) across the phylogeny demonstrating that isolates of this type can be genetically quite different from one another despite their identical spoligotype. Interestingly, the majority of the home ranges of the WGS clades in which spoligotype SB0140 (GB type 9) is present, appear to be mostly confined and isolated from each other (Gov.uk, 2022), further supporting the evidence from the WGS data that these clades are genetically distinct.

These findings highlight that some of the assumptions based on conventional genotyping that were made in the past to inform epidemiological analyses might have been incorrect with potentially negative impact on case management, resulting in less effective control measures. Similar inconsistencies between WGS lineages and spoligotypes or genotypes have been reported previously (van Tonder et al., 2021) and show that WGS not only provides higher resolution, but also more accurate and reliable inference of genetic relatedness compared to conventional typing methods.

3.4 SNP markers for the *M. bovis* population in GB

Based on the WGS clades defined in Figure 1, a set of 2,058 polymorphic positions that are unique in each of the 35 clades, were identified from a subset of high-quality sequences (n = 358). Due to the strict clonality of the MTBC (Smith et al., 2006), these SNPs can be used as markers to assign isolates to the 35 different WGS clades that capture the genetic diversity of M. bovis in GB (Supplementary Table 3). The number of specific SNPs per clade varies, with B6-51 exhibiting the lowest number of unique SNPs (n = 9) and B1-11 the maximum number of unique SNPs (n = 294). When this set of SNPs was tested against the entire dataset, 7 SNPs that were originally found to be unique in clade B6-85, were found to be specific for a subgroup within this clade and not for the entire clade. This was due to the subsampling that was performed for this analysis and the isolates that were selected to represent clade B6-85, which did not capture its entire diversity. Similarly, two SNPs that were each originally found to be unique in clades B1-11 and B6-85 were found to have mutated back to the reference call in a subgroup of isolates and a single isolate within each clade, respectively, which was again due to the subsampling performed for this analysis. These sites were therefore removed from the SNP set since they were polymorphic and not fixed in the clades above. For the remaining set of SNPs (n = 2,049 sites), when tested against the entire dataset (n = 2,823 sequences), SNP-based clade assignment was consistent with the phylogenetic clade in which each sequence belonged with an average of 99.1% matching ratio across each clade's defining SNPs (range: 54-100%). Matching ratios lower than 100% were due to Ns in clade-defining SNP positions.

SNP-based tools or sets of lineage-specific SNPs have been published before to identify different members of the MTBC (Lipworth et al., 2019), to differentiate *M. bovis* from *M. caprae* and *M. orygis* (Zwyer et al., 2021), and to distinguish global lineages of *M. bovis* (Zimpel et al., 2020). Although these SNP sets are useful for characterisation of MTBC strains or *M. bovis* strains in a wider context, this present study identified unique SNPs that specifically characterise the diversity of *M. bovis* in GB. These SNP sets can further be used and expanded to include *M. bovis* strains isolated from neighbouring countries, as well as from other countries that may share similar genetic lineages due to cattle trade with the UK (Smith et al., 2011).

3.5 Annotation of the clade-specific SNPs

Annotation of the unique SNPs that define the 35 WGS clades identified in this study showed that 89% of the variants (n = 1,824) were located on protein-coding regions, of which 32.5% (n = 665) were synonymous and therefore assumed to be neutral. In addition, 12.2% (n = 250) were in intergenic regions and therefore non-coding but may have unknown regulatory functions. The remaining mutations were non-synonymous, meaning that the encoded amino acid is altered, a start or stop codon is lost, or a premature stop codon is introduced.

We focused on the mutations that were predicted to have high or moderate impact on the encoded protein by SNPeff (n = 1,133) and the genes they affect (Supplementary Table 3). High impact variants are assumed to be disruptive, causing protein truncation, loss of function or triggering nonsense mediated decay, while moderate impact variants are assumed to be non-disruptive, potentially changing the effectiveness of the protein (Cingolani et al., 2012). The highest number of such variants were found in genes that are involved in DNA synthesis and translation processes, as well as phosphorylation that plays an important role in the pathogen and host physiology, such as virulence, signalling and immune response (Wang and Xie, 2022). Other predicted genes that appear to be overrepresented among high and moderate impact variants are associated with processes such as cell wall organisation, the biosynthesis and transport of the cell wall lipid PDIM (phthiocerol dimycocerosates), secondary metabolite biosynthetic processes, and fatty acid and lipid biosynthetic processes. The latter have been shown to play an essential role in the formation of cell wall components, which are thought to be important in hostpathogen interactions, virulence and pathogenicity (Kinsella et al., 2003; Gago et al., 2018; Mondino et al., 2020). The cell wall lipid PDIM of *M. tuberculosis* has been shown to inhibit autophagy in mice, contributing to its virulence and immune evasion (Mittal et al., 2024).

Notably, 180 genes (of which 90 are of known function and 90 are of unknown function) were found to harbour more than one of these high or moderate impact mutations, largely defining different clades. One of these genes, which has eight clade-specific mutations (each specifying a different clade) is *whiA*, a gene encoding a DNA-binding protein that is involved in cell division and chromosome segregation in most Grampositive bacteria. This gene has been shown to be essential for sporulation in *Streptomyces coelicolor* A3 (Aínsa et al., 2000) and to affect fatty acid composition of the membrane in *Bacillus subtilis* (Suarez, 2018). Another gene containing six clade-specific mutation is *pks12*, which has been

10.3389/fmicb.2025.1515906

predicted to encode a novel polyketide that is presented to CD1cmediated T cells in *M. tuberculosis* and *M. bovis* Bacille-Calmette-Guerin and is therefore necessary for antigen production (Matsunaga et al., 2004). This gene has also been shown to be involved in PDIM production (mentioned above) and pathogenesis (Sirakova et al., 2003). Other genes containing four mutations are the following: *fhaA*, which plays a role in peptidoglycan biosynthesis and maintenance of the cell envelope integrity in mycobacteria (Viswanathan et al., 2017); *rsmH*, a member of the 16S RNA methyltransferase family, in which mutations are known to confer aminoglycoside resistance in other bacteria (Ledger et al., 2020); and *metY* that is involved in the biosynthesis of homocysteine, a sulphur metabolic pathway (trans-sulfuration) which is essential for survival and the expression of virulence in *M. tuberculosis* (Wheeler et al., 2008; Paritala and Carroll, 2013).

The observation that many of these genes are associated with virulence and pathogenicity in *M. bovis, M. tuberculosis* and/or other bacteria raises the hypothesis that the WGS clades identified in this study may exhibit phenotypic variation, which might have led to different evolutionary trajectories. However, experimental *in vitro* studies would be required to examine the functional effect of the clade-specific mutations as the above are merely computational predictions. Furthermore, genome-wide association studies and selection scans can in future be performed to investigate whether these mutations were driven to fixation owing to selection or random genetic drift.

3.6 Discriminatory power of WGS at the local scale

WGS data from a recent outbreak in 2022 on a farm (referred to as farm "A" from here on) in the HRA of England (SW) that had no history of TB incidents in the previous 10 years were analysed and integrated with location, cattle movement and testing data, in an attempt to determine the source of infection and risk pathway involved. Of the 10 animals that tested positive at the disclosing Single Intradermal Comparative Cervical skin test, three samples were collected for mycobacterial culture and sequencing, and subsequently three *M. bovis* isolates were all found to belong to clade B6-51. Based on the home range for clade B6-51, the incident was located within its expected geographic range, and the same clade had been identified in earlier outbreaks on nearby farms. Notably, badger culling had recently started in that area and a sick badger had been found on farm "A" a few months prior to the outbreak, leading to speculation that local spread via wildlife was a likely source attribution.

Using the genome sequences of the three isolates from farm "A" and all other sequences within our dataset that were assigned to clade B6-51, we built a MP tree to reconstruct their phylogenetic relationships. Figure 4A shows that the three isolates differed by 1–2 SNPs from each other. Isolate AF-21-000185-22 was identical to five other isolates from incidents dating from 2019 to 2022, which were geographically clustered to the north-west of farm "A" at ~35 miles distance. The remaining two isolates were each one SNP away from the identical isolates with which they share a common ancestor (Figure 4B).

This phylogenetic analysis instigated the inspection of movements of the animals from which the identical isolates were collected; two animals were found to have originated from the same farm (referred to as farm "B" from here on) and overlapped in time for approximately 2 years and 7 months. The animal from which the 2019 isolate (AF-21-08004-19) was collected, was born on farm "B" in 2017 and lived there until December 2019, when it tested positive and was sent for slaughter. The animal from which the 2022 isolate (AF-21-05767-22) was collected, was also born on farm "B" in 2012 and lived there for 8 years and 9 months before it was purchased by another farm (farm "C") via a local market (in September 2021). It was then disclosed as a reactor animal in April 2022 at a 12-month test on the purchase farm, following the conclusion of a previous outbreak on that farm, despite its clear



FIGURE 4

Maximum parsimony tree (A) and map (B) showing the phylogenetic relationships and relative location of the three isolates collected from a recent outbreak (farm A) in the SW of England. The tree (A) is a snippet from the phylogeny of WGS clade B6-51. Branch length indicates number of SNPs. Relative locations are shown on the map (B) to comply with data protection law in the UK. pre-movement test on farm "B" in August 2021. It should be noted that this animal had also tested negative through 3 previous whole-herd short interval tests and a 6-month check test on farm "B", after the disclosure of the 2019 isolate and two subsequent positive slaughterhouse cases.

Interestingly, by looking at the movement and testing data from farm "A," it was discovered that farm "A" had purchased two animals from farm "B" in August 2021, which both had a positive skin test following their movement, in August 2022 and November 2022 respectively, but for which no bacterial genetic data is available. The animal disclosed in August 2022 did not undergo post-mortem examination as it was within its drug withdrawal period and slaughtered on farm. However, according to its clinical history, it had been losing weight and was dyspnoeic as well as unresponsive to antibiotics. The second animal disclosed in November 2022, did not exhibit visible lesions and was not sampled. Both these animals were born on farm "B" in September 2019 and March 2018, respectively, and had skin-tested negative 7 and 8 times, respectively, on that farm prior to their movement. Hence, the genomic and epidemiological data analysed together suggest that one or both of these animals were infected on farm "B", and that at least one was an open case and source for other cattle on farm "A" upon their arrival.

By harnessing the higher resolution provided by WGS data and integrating it with epidemiological information, we were able to attribute the 2022 outbreak on farm "A" to cattle purchases in 2021 from "farm B." Our findings show that this farm had not cleared infection after its own outbreak starting in 2019, despite three rounds of subsequent whole-herd short-interval testing and a 6-month check test. Identifying farm "B" as the source of infection would not have been possible by only looking at the gross-level WGS clade characterisation of these isolates or at the previously used genotypes; the isolates would have been considered to be within home range and local spread via wildlife would have been assumed as the most likely source of infection, given the recent sighting of a sick badger on farm. In contrast, our findings based on the genomic and epidemiological data indicate that infected animals missed during testing along with cattle purchases, led to this outbreak on a farm with no TB history in the previous 10 years, which has implications for control policy measures. It should be noted that this is not an isolated case, our investigations have shown many examples where within-home range transmission can be explained by local cattle movement, adding to the growing evidence that within-species transmission is considerably more prevalent than between-species transmission (Crispell et al., 2019; van Tonder et al., 2021; Rossi et al., 2022; Akhmetova et al., 2023; Canini et al., 2023). Our analysis further demonstrates that genomic similarity can be used to highlight potential transmission links, narrowing down epidemiological investigations, but also in the absence of other information, which is one of the major benefits of WGS.

4 Conclusion

This study characterised the genomic diversity of *M. bovis* across GB from the analysis of ~3,000 genome sequences. A total of 35 WGS clades was identified, based on which, a SNP-based classification system was designed that differentiates bTB isolates from GB more accurately and with higher granularity compared to previously used typing schemes. The genes harbouring these mutations and the processes they are associated with were discussed with potential implications for functional differences between the identified clades. The utility and high resolution of WGS at the local scale was demonstrated by analysing genome

sequences from an outbreak in an endemic area of England, where source attribution remained challenging due to the limited discriminatory power of traditional typing methods. It is shown that WGS can be a powerful and informative tool when used alongside epidemiological data that can better determine transmission pathways, leading to more efficient case management policies and better-targeted control measures and interventions in the fight against bTB.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ebi.ac.uk/ena/browser/view/PRJEB81540.

Author contributions

PS: Data curation, Formal analysis, Writing - review & editing. JN-G: Conceptualization, Formal analysis, Methodology, Writing review & editing. SB: Conceptualization, Writing - review & editing. JW: Investigation, Writing - review & editing. JD: Investigation, Writing - review & editing. PU: Data curation, Writing - review & editing. JG: Writing - review & editing. GH: Conceptualization, Writing - review & editing. SD: Conceptualization, Writing - review & editing. RE: Conceptualization, Formal analysis, Methodology, Writing - review & editing. EP: Conceptualization, Formal analysis, Supervision, Writing - original draft, Writing - review & editing. PS performed all analyses with supervision from EP. JNG, SB, JG, GH, SD, RE and EP conceived and designed the study. JD performed experiments. JNG and RE developed the pipeline and devised the nomenclature. JW and PU contributed data to the study. EP wrote and prepared the original draft of the manuscript and all co-authors reviewed and edited it.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by the Department for Environment, Food and Rural Affairs (DEFRA) and Welsh Government and Scottish Government (devolved project SB4030).

Acknowledgments

We would like to thank Karen Gover and the Central Sequencing Unit at APHA, Weybridge for technical support. We would also like to thank Aaron Fishman for contributing code to the pipeline, Nick Pestell for contributing to the development of the masked regions and Hira Naveed for helpful comments on earlier drafts of the manuscript.

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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,

References

Aínsa, J. A., Ryding, N. J., Hartley, N., Findlay, K. C., Bruton, C. J., and Chater, K. F. (2000). WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 182, 5470–5478. doi: 10.1128/jb.182.19.5470-5478.2000

Akhmetova, A., Guerrero, J., McAdam, P., Salvador, L. C. M., Crispell, J., Lavery, J., et al. (2023). Genomic epidemiology of *Mycobacterium bovis* infection in sympatric badger and cattle populations in Northern Ireland. *Microb. Genom.* 9:1023. doi: 10.1099/mgen.0.001023

Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M., et al. (2010). Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi. Emerg. Inf. Dis.* 16, 1296–1299. doi: 10.3201/eid1608.100314

Allen, A. R., Skuce, R. A., and 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. doi: 10.3389/fvets.2018.00109

Ates, L. S. (2020). New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Mol. Microbiol.* 113, 4–21. doi: 10.1111/mmi.14409

Berg, S., Garcia-Pelayo, M. C., Müller, B., Hailu, E., Asiimwe, B., Kremer, K., et al. (2011). African 2, a clonal complex of *Mycobacterium bovis* epidemiologically important in East Africa. *J. Bacteriol.* 193, 670–678. doi: 10.1128/jb.00750-10

Bianchini, G., and Sánchez-Baracaldo, P. (2024). TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol. Evol.* 14:e10873. doi: 10.1002/ece3.10873

Biek, R., O'Hare, A., Wright, D., Mallon, T., McCormick, C., Orton, R. J., et al. (2012). Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog.* 8:e1003008. doi: 10.1371/journal.ppat.1003008

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brosch, R., Gordon, S. V., Pym, A., Eiglmeier, K., Garnier, T., and Cole, S. T. (2000). Comparative genomics of the mycobacteria. *Int. J. Med. Microbiol.* 290, 143–152. doi: 10.1016/s1438-4221(00)80083-1

Cadmus, S. I., Gordon, S. V., Hewinson, R. G., and Smith, N. H. (2011). Exploring the use of molecular epidemiology to track bovine tuberculosis in Nigeria: an overview from 2002 to 2004. *Vet. Microbiol.* 151, 133–138. doi: 10.1016/j.vetmic.2011.02.036

Canini, L., Modenesi, G., Courcoul, A., Boschiroli, M.-L., Durand, B., and Michelet, L. (2023). Deciphering the role of host species for two *Mycobacterium bovis* genotypes from the European 3 clonal complex circulation within a cattle-badger-wild boar multihost system. *MicrobiologyOpen* 12:e1331. doi: 10.1002/mbo3.1331

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

Collins, D. M., De Lisle, G. W., and Gabric, D. M. (1986). Geographic distribution of restriction types of *Mycobacterium bovis* isolates from brush-tailed possums (*Trichosurus vulpecula*) in New Zealand. *J. Hyg.* 96, 431–438. doi: 10.1017/s0022172400066201

Consortium, T.U (2022). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi: 10.1093/nar/gkac1052

Crispell, J., Benton, C. H., Balaz, D., De Maio, N., Ahkmetova, A., Allen, A., et al. (2019). Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *eLife* 8:e45833. doi: 10.7554/eLife.45833

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. doi: 10.1093/gigascience/giab008

Delahay, R. J., De Leeuw, A. N., Barlow, A. M., Clifton-hadley, R. S., and Cheeseman, C. L. (2002). The status of *Mycobacterium bovis* infection in UK wild mammals: a review. *Vet. J.* 164, 90–105. doi: 10.1053/tvjl.2001.0667

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2025.1515906/ full#supplementary-material

Domingo, M., Vidal, E., and Marco, A. (2014). Pathology of bovine tuberculosis. *Res. Vet. Sci.* 97, S20–S29. doi: 10.1016/j.rvsc.2014.03.017

Fabre, M., Koeck, J.-L., le, P., Simon, F., Hervé, V., Vergnaud, G., et al. (2004). High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "Mycobacterium canettii" strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "M. Canettii". *J. Clin. Microbiol.* 42, 3248–3255. doi: 10.1128/JCM.427.3248-3255.2004

Fitzgerald, S. D., and Kaneene, J. B. (2013). Wildlife reservoirs of bovine tuberculosis worldwide: hosts, pathology, surveillance, and control. *Vet. Pathol.* 50, 488–499. doi: 10.1177/0300985812467472

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

Gago, G., Diacovich, L., and Gramajo, H. (2018). Lipid metabolism and its implication in mycobacteria-host interaction. *Curr. Opin. Microbiol.* 41, 36–42. doi: 10.1016/j.mib.2017.11.020

Gallagher, J., and Horwill, D. M. (1977). A selective oleic acid albumin agar medium for the cultivation of *Mycobacterium bovis*. J. Hyg. 79, 155–160. doi: 10.1017/s0022172400052943

Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodkin, E., et al. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364, 730–739. doi: 10.1056/NEJMoa1003176

Godfray, C., Donnelly, C., Hewinson, G., Winter, M., and Wood, J. (2018). Bovine TB strategy review October 2018. Available at: https://assets.publishing.service.gov.uk/ media/5beed433e5274a2af111f622/tb-review-final-report-corrected.pdf (accessed Septemeber 6, 2024).

Godfray, H. C. J., Donnelly, C. A., Kao, R. R., Macdonald, D. W., McDonald, R. A., Petrokofsky, G., et al. (2013). A restatement of the natural science evidence base relevant to the control of bovine tuberculosis in Great Britain. *Proc. R. Soc. B Biol. Sci.* 280:20131634. doi: 10.1098/rspb.2013.1634

Gov.uk. (2021). Bovine tuberculosis in England in 2020. Epidemiological analysis of the 2020 data and historical trends. Available at: https://assets.publishing.service.gov.uk/ media/61715c7ad3bf7t5603ecf08e/tb-epidemiological-report-2020.pdf (accessed Septemeber 6, 2024).

Gov.uk. (2022). Bovine tuberculosis in Great Britain. Surveillance data for 2021 and historical trends. Available at: https://assets.publishing.service.gov.uk/ media/6398789c8fa8f552fe8711de/GB_data_report_2021_not_accessible.ods (accessed July 10, 2024).

Gov.uk. (2024). Latest accredited official statistics on tuberculosis (TB) in cattle in Great Britain - quarterly. Available at: https://www.gov.uk/government/statistics/ incidence-of-tuberculosis-tb-in-cattle-in-great-britain (accessed October 1, 2024).

Groenen, P. M., Bunschoten, A. E., van Soolingen, D., and 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. doi: 10.1111/j.1365-2958.1993.tb00976.x

Hauer, A., de, K., Cochard, T., Godreuil, S., Karoui, C., Henault, S., et al. (2015). Genetic evolution of *Mycobacterium bovis* causing tuberculosis in livestock and wildlife in France since 1978. *PLoS One* 10:e0117103. doi: 10.1371/journal.pone.0117103

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kinsella, R. J., Fitzpatrick, D. A., Creevey, C. J., and McInerney, J. O. (2003). Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. Proc. Natl. Acad. Sci. USA 100, 10320-10325. doi: 10.1073/pnas.1737230100

Ledger, L., Eidt, J., and Cai, H. Y. (2020). Identification of antimicrobial resistanceassociated genes through whole genome sequencing of *Mycoplasma bovis* isolates with different antimicrobial resistances. *Pathogens* 9:588. doi: 10.3390/pathogens9070588

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [preprint] arXiv:1303.3997v2 [q-bio.GN]. doi: 10.48550/arXiv.1303.3997

Lipworth, S., Jajou, R., de Neeling, A., Bradley, P., van der Hoek, W., Maphalala, G., et al. (2019). SNP-IT tool for identifying subspecies and associated lineages of *Mycobacterium tuberculosis* complex. *Emerg. Infect. Dis.* 25, 482–488. doi: 10.3201/eid2503.180894

Livingstone, P. G., Hancox, N., Nugent, G., and 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. *N. Z. Vet. J.* 63 Suppl 1, 4–18. doi: 10.1080/00480169.2014.971082

Maddison, W. P., and Maddison, D. R. (2023). Mesquite: a modular system for evolutionary analysis. Version 3.81. Available at: http://www.mesquiteproject.org (Accessed 15 May 2023).

Malone, K. M., Farrell, D., Stuber, T. P., Schubert, O. T., Aebersold, R., Robbe-Austerman, S., et al. (2017). Updated reference genome sequence and annotation of *Mycobacterium bovis* AF2122/97. *Genome Announc.* 5:17. doi: 10.1128/genomeA.00157-17

Matsunaga, I., Bhatt, A., Young, D. C., Cheng, T. Y., Eyles, S. J., Besra, G. S., et al. (2004). *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J. Exp. Med.* 200, 1559–1569. doi: 10.1084/jem.20041429

Menardo, F., Loiseau, C., Brites, D., Coscolla, M., Gygli, S. M., Rutaihwa, L. K., et al. (2018). Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinf*. 19:164. doi: 10.1186/s12859-018-2164-8

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Mittal, E., Prasad, G. V. R. K., Upadhyay, S., Sadadiwala, J., Olive, A. J., Yang, G., et al. (2024). *Mycobacterium tuberculosis* virulence lipid PDIM inhibits autophagy in mice. *Nat. Microbiol.* 9, 2970–2984. doi: 10.1038/s41564-024-01797-5

Mondino, S., Vázquez, C. L., Cabruja, M., Sala, C., Cazenave-Gassiot, A., Blanco, F. C., et al. (2020). FasR regulates fatty acid biosynthesis and is essential for virulence of *Mycobacterium tuberculosis. Front. Microbiol.* 11:586285. doi: 10.3389/fmicb.2020.586285

Müller, B., Dürr, S., Alonso, S., Hattendorf, J., Laisse, C. J., Parsons, S. D., et al. (2013). Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. *Emerg. Infect. Dis.* 19, 899–908. doi: 10.3201/eid1906.120543

Müller, B., Hilty, M., Berg, S., Garcia-Pelayo, M. C., Dale, J., Boschiroli, M. L., et al. (2009). African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. J. Bacteriol. 191, 1951–1960. doi: 10.1128/jb.01590-08

Olea-Popelka, F., Muwonge, A., Perera, A., Dean, A. S., Mumford, E., Erlacher-Vindel, E., et al. (2017). Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis*-a call for action. *Lancet Infect. Dis.* 17, e21–e25. doi: 10.1016/s1473-3099(16)30139-6

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial. Genom.* 2:e000056. doi: 10.1099/mgen.0.000056

Palmer, M. V. (2013). *Mycobacterium bovis*: characteristics of wildlife reservoir hosts. *Transbound. Emerg. Dis.* 60, 1–13. doi: 10.1111/tbed.12115

Paritala, H., and Carroll, K. S. (2013). New targets and inhibitors of mycobacterial sulfur metabolism. *Infect. Disord. Drug Targets* 13, 85–115. doi: 10.2174/18715265113139990022

Parsons, S. D., Drewe, J. A., Gey van Pittius, N. C., Warren, R. M., and van Helden, P. D. (2013). Novel cause of tuberculosis in meerkats, South Africa. *Emerg. Infect. Dis.* 19, 2004–2007. doi: 10.3201/eid1912.130268

Price-Carter, M., Brauning, R., de Lisle, G. W., Livingstone, P., Neill, M., Sinclair, J., et al. (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. doi: 10.3389/fvets.2018.00272

Rodriguez-Campos, S., Schürch, A. C., Dale, J., Lohan, A. J., Cunha, M. V., Botelho, A., et al. (2012). European 2--a clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infect. Genet. Evol.* 12, 866–872. doi: 10.1016/j.meegid.2011.09.004

Rodriguez-Campos, S., Smith, N. H., Boniotti, M. B., and Aranaz, A. (2014). Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. *Res. Vet. Sci.* 97, S5–s19. doi: 10.1016/j.rvsc.2014.02.009

Roetzer, A., Diel, R., Kohl, T. A., Rückert, C., Nübel, U., Blom, J., et al. (2013). Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med.* 10:e1001387. doi: 10.1371/journal.pmed.1001387

Rossi, G., Crispell, J., Brough, T., Lycett, S. J., White, P. C. L., Allen, A., et al. (2022). Phylodynamic analysis of an emergent *Mycobacterium bovis* outbreak in an area with no previously known wildlife infections. *J. Appl. Ecol.* 59, 210–222. doi: 10.1111/1365-2664.14046

Sirakova, T. D., Dubey, V. S., Kim, H.-J., Cynamon, M. H., and Kolattukudy, P. E. (2003). The largest open Reading frame *pks12* in the *Mycobacterium tuberculosis* genome is involved in pathogenesis and Dimycocerosyl Phthiocerol synthesis. *Infect. Immun.* 71, 3794–3801. doi: 10.1128/iai.71.7.3794-3801.2003

Skuce, R. A., Mallon, T. R., McCormick, C., McBride, S. H., Clarke, G., Thompson, A., et al. (2010). Bovine tuberculosis: herd-level surveillance of *Mycobacterium bovis* genotypes in Northern Ireland (2003-2008). *Adv. Anim. Biosci.* 1:112. doi: 10.1017/S2040470010002554

Skuce, R. A., McDowell, S. W., Mallon, T. R., Luke, B., Breadon, E. L., Lagan, P. L., et al. (2005). Discrimination of isolates of *Mycobacterium bovis* in Northern Ireland on the basis of variable numbers of tandem repeats (VNTRs). *Vet. Rec.* 157, 501–504. doi: 10.1136/vr.157.17.501

Smith, N. H., Berg, S., Dale, J., Allen, A., Rodriguez, S., Romero, B., et al. (2011). European 1: a globally important clonal complex of *Mycobacterium bovis*. *Infect. Genet. Evol.* 11, 1340–1351. doi: 10.1016/j.meegid.2011.04.027

Smith, N. H., Dale, J., Inwald, J., Palmer, S., Gordon, S. V., Hewinson, R. G., et al. (2003). The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc. Natl. Acad. Sci. USA* 100, 15271–15275. doi: 10.1073/pnas.2036554100

Smith, N. H., Gordon, S. V., de la Rua-Domenech, R., Clifton-Hadley, R. S., and Hewinson, R. G. (2006). Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis. Nat. Rev. Microbiol.* 4, 670–681. doi: 10.1038/nrmicro1472

Suarez, B. (2018). The dynamic role of the conserved WhiA and MinD proteins in chromosome segregation, fatty acid metabolism and cell division in *Bacillus subtilis*. Amsterdam: University of Amsterdam.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Trewby, H., Wright, D., Breadon, E. L., Lycett, S. J., Mallon, T. R., McCormick, C., et al. (2016). Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics* 14, 26–35. doi: 10.1016/j.epidem.2015.08.003

Tsao, K., Robbe-Austerman, S., Miller, R. S., Portacci, K., Grear, D. A., and Webb, C. (2014). Sources of bovine tuberculosis in the United States. *Infect. Genet. Evol.* 28, 137–143. doi: 10.1016/j.meegid.2014.09.025

van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeone, R., Brosch, R., et al. (2012). Characterization of Mycobacterium orygis as *M. tuberculosis* complex subspecies. *Emerg. Infect. Dis.* 18, 653–655. doi: 10.3201/eid1804.110888

van Tonder, A. J., Thornton, M. J., Conlan, A. J. K., Jolley, K. A., Goolding, L., Mitchell, A. P., et al. (2021). Inferring *Mycobacterium bovis* transmission between cattle and badgers using isolates from the randomised badger culling trial. *PLoS Pathog.* 17:e1010075. doi: 10.1371/journal.ppat.1010075

Viswanathan, G., Yadav, S., Joshi, S. V., and Raghunand, T. R. (2017). Insights into the function of FhaA, a cell division-associated protein in mycobacteria. *FEMS Microbiol. Lett.* 364:fnw294. doi: 10.1093/femsle/fnw294

Waller, E., Brouwer, A., Upton, P., Harris, K., Lawes, J., Duncan, D., et al. (2022). Bovine TB infection status in cattle in Great Britain in 2020. *Vet. Rec.* 191:e2513. doi: 10.1002/vetr.2513

Wang, Z., and Xie, J. (2022). Phosphoproteomics of Mycobacterium-host interaction and inspirations for novel measures against tuberculosis. *Cell. Signal.* 91:110238. doi: 10.1016/j.cellsig.2021.110238

Warren, R. M., Streicher, E. M., Sampson, S. L., van der Spuy, G. D., Richardson, M., Nguyen, D., et al. (2002). Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *J. Clin. Microbiol.* 40, 4457–4465. doi: 10.1128/jcm.40.12.4457-4465.2002

Wheeler, P. R., Brosch, R., Coldham, N. G., Inwald, J. K., Hewinson, R. G., and Gordon, S. V. (2008). Functional analysis of a clonal deletion in an epidemic strain of *Mycobacterium bovis* reveals a role in lipid metabolism. *Microbiology* 154, 3731–3742. doi: 10.1099/mic.0.2008/022269-0

WHO, FAO, and WOAH (2017). Roadmap for zoonotic tuberculosis. Available at: https://www.woah.org/app/uploads/2021/03/roadmap-zoonotic-tb.pdf (accessed September 6, 2024).

Xia, E., Teo, Y.-Y., and Ong, R. T.-H. (2016). SpoTyping: fast and accurate in silico Mycobacterium spoligotyping from sequence reads. *Genome Med.* 8:19. doi: 10.1186/s13073-016-0270-7

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., et al. (2012). FastUniq: A fast De novo duplicates removal tool for paired short reads. *PLoS One* 7:e52249. doi: 10.1371/journal.pone.0052249

Zimpel, C. K., Patané, J. S. L., Guedes, A. C. P., de Souza, R. F., Silva-Pereira, T. T., Camargo, N. C. S., et al. (2020). Global distribution and evolution of *Mycobacterium* bovis lineages. *Front. Microbiol.* 11:843. doi: 10.3389/fmicb.2020.00843

Zwyer, M., Çavusoglu, C., Ghielmetti, G., Pacciarini, M., Scaltriti, E., Van Soolingen, D., et al. (2021). A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics [version 2; peer review: 2 approved]. *Open Res. Europe* 1:14029. doi: 10.12688/openreseurope.14029.2