frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 05 January 2023
DOI 10.3389/fmicb.2022.982679

OPEN ACCESS

EDITED BY
Javier Pascual,
Darwin Bioprospecting Excellence, Spain

REVIEWED BY
Claudio Counoupas,
Royal Prince Alfred Hospital, Australia
Pushpendra Singh,
National Institute for Research in Tribal Health (ICMR), India

*CORRESPONDENCE
Mireia Coscolla
✉ mireia.coscolla@uv.es
Esther Julián
✉ Esther.Julian@uab.cat

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 30 June 2022
ACCEPTED 06 December 2022
PUBLISHED 05 January 2023

CITATION
Renau-Mínguez C, Herrero-Abadía P, Ruiz-Rodriguez P, Sentandreu V, Torrents E, Chiner-Oms Á, Torres-Puente M, Comas I, Julián E and Coscolla M (2023) Genomic analysis of *Mycobacterium brumae* sustains its nonpathogenic and immunogenic phenotype.
Front. Microbiol. 13:982679.
doi: 10.3389/fmicb.2022.982679

Genomic analysis of *Mycobacterium brumae* sustains its nonpathogenic and immunogenic phenotype

Chantal Renau-Mínguez[1], Paula Herrero-Abadía[2], Paula Ruiz-Rodriguez[1], Vicente Sentandreu[3], Eduard Torrents[4,5], Álvaro Chiner-Oms[6], Manuela Torres-Puente[6], Iñaki Comas[6], Esther Julián[2]* and Mireia Coscolla[1]*

[1]I2SysBio, University of Valencia-FISABIO Joint Unit, Paterna, Spain, [2]Genetics and Microbiology Department, Faculty of Biosciences, Autonomous University of Barcelona, Barcelona, Spain, [3]Genomics Unit, Central Service for Experimental Research (SCSIE), University of Valencia, Burjassot, Spain, [4]Bacterial Infections and Antimicrobial Therapies Group, Institute for Bioengineering of Catalonia (IBEC), Barcelona, Spain, [5]Microbiology Section, Department of Genetics, Microbiology, and Statistics, Biology Faculty, Universitat de Barcelona, Barcelona, Spain, [6]Instituto de Biomedicina de Valencia (IBV), CSIC, Valencia, Spain

Mycobacterium brumae is a rapid-growing, non-pathogenic *Mycobacterium* species, originally isolated from environmental and human samples in Barcelona, Spain. *Mycobacterium brumae* is not pathogenic and it's *in vitro* phenotype and immunogenic properties have been well characterized. However, the knowledge of its underlying genetic composition is still incomplete. In this study, we first describe the 4Mb genome of the *M. brumae* type strain ATCC 51384^T assembling PacBio reads, and second, we assess the low intraspecies variability by comparing the type strain with Illumina reads from three additional strains. *Mycobacterium brumae* genome is composed of a circular chromosome with a high GC content of 69.2% and containing 3,791 CDSs, 97 pseudogenes, one prophage and no CRISPR loci. *Mycobacterium brumae* has shown no pathogenic potential in *in vivo* experiments, and our genomic analysis confirms its phylogenetic position with other non-pathogenic and rapid growing mycobacteria. Accordingly, we determined the absence of virulence-related genes, such as ESX-1 locus and most PE/PPE genes, among others. Although the immunogenic potential of *M. brumae* was proved to be as high as *Mycobacterium bovis* BCG, the only mycobacteria licensed to treat cancer, the genomic content of *M. tuberculosis* T cell and B cell antigens in *M. brumae* genome is considerably lower than those antigens present in *M. bovis* BCG genome. Overall, this work provides relevant genomic data on one of the species of the mycobacterial genus with high therapeutic potential.

KEYWORDS

diversity, immunogenic, non-pathogenic, nontuberculous mycobacteria, therapeutic

Introduction

Mycobacterium brumae (synonym *Mycolicibacterium brumae*; Meehan et al., 2021) is a saprophytic bacterium originally isolated from environmental samples, soil, and water, and human sputum from asymptomatic individuals (Luquin et al., 1993). It is a rapid-growing *Mycobacterium* (RGM) obtaining colonies in approximately 5–7 days at 37°C when it is grown in a rich medium. In Middlebrook 7H10 medium, *M. brumae* forms irregular, rough, and non-pigmented colonies, while in Middlebrook 7H9 liquid medium can form pellicles and clumps with cording morphology (Brambilla et al., 2012).

Mycobacterium brumae has recently been described as a plausible immunotherapy agent for non-muscle invasive bladder cancer (NMIBC; Noguera-Ortega et al., 2016b). Currently, the standard treatment for avoiding recurrence and progression of NMIBC, after transurethral resection of tumors, consists of weekly intravesical instillations of viable *Mycobacterium bovis* BCG. Despite being efficacious, BCG has some limitations. Around 30 % of patients do not respond to BCG treatment, and a high percentage of BCG-treated patients show local and even serious systemic side effects such as pulmonary infections or sepsis (Guallar-Garrido and Julián, 2020) making adherence and continuity to treatment complicated. For that reason, safer alternatives are needed. In regard, *M. brumae* showed a safe profile *in vitro* and *in vivo* studies (summarized in Figure 1; Noguera-Ortega et al., 2016a; Bach-Griera et al., 2020).

Although the antitumor mechanism of action of BCG has not been fully elucidated, preclinical, and clinical data support sequential events that occur at bladder epithelium after BCG instillation. First, inhibition of tumor proliferation due to apoptosis and/or cell cycle arrest of the remaining tumor cells occur after their interaction with BCG. In addition, BCG triggers the production of a broad spectrum of cytokines and chemokines that leads to tumor infiltration of different immune populations that indirectly provide an antitumor role. Preclinical studies have shown that *M. brumae* inhibits bladder tumor cell proliferation being even more efficacious than BCG well-differentiated cells (low-grade tumors), and activated peripheral blood mononuclear cells triggering the production of cytokines and a cytotoxic profile against tumor cells (Noguera-Ortega et al., 2016a). In the orthotopic murine model of bladder cancer, the induction of local and systemic immune responses by *M. brumae* treatment leads to prolonging the survival of tumor-bearing mice with respect to non-treated tumor-bearing mice and by a similar ratio than BCG-treated mice (Noguera-Ortega et al., 2016a,b; Guallar-Garrido et al., 2022).

However, although biological, immunological, and virulence differences have been shown between *M. brumae* and BCG, it is still unknown which genomic differences might be responsible for the different biological properties of these two mycobacteria. The antigen/s responsible for the efficacious antitumor effect of BCG remains elusive. Given the similar ability of *M. brumae* and BCG to trigger an immunomodulatory and antitumor response, it can be hypothesized that both mycobacteria share immunostimulatory antigens. Thus, it is relevant to analyze the *M. brumae* genome since it is devoid of genes involved in pathogenicity but contains genes responsible for the immunomodulatory and antitumor mechanisms. The comparison of their genomes could provide clues for understanding their therapeutic capacities. In this work, we offer a detailed description of the complete genomic sequence of the reference strain of *M. brumae*, and we address the biological differences between *M. brumae* and BCG in the context of their genomic differences using a comparative genomics approach. Furthermore, we assess the intraspecies variability of *M. brumae* by comparing the genetic diversity between different isolates of the species.

Materials and methods

Bacterial strains

Four *M. brumae* strains [CR103, CR142, CR269, and CR270 (which corresponds to the type strain: ATCC 51384^T)] were provided by Prof. Luquin, obtained from the original isolates described in Luquin et al. (1993). Lyophilized cells were cultured on Middlebrook 7H10 agar medium (Difco Laboratories, Michigan, United States) supplemented with 10% oleic-albumin-dextrose-catalase (OADC) and incubated at 37°C for 1 week.

Genomes

We analyzed the phylogenetic relationships of published reference/complete genomes from different *Mycobacterium* species (Supplementary Table S1). Although the genome sequence is not closed, three additional previously published *M. brumae* sequences for the type strain were compared with our reference genome sequence: CIP1034565 (GCF_002553575.1), DSM44177 (GCF_004014795.1), and MBR1 (GCF_900073015.1).

The *Mycobacterium tuberculosis* H37Rv and *M. bovis* BCG Connaught genomes used for genome comparisons were downloaded from the GenBank database with accession numbers GCA_000195955.2 and GCA_001287325.1, respectively.

Drug resistance analysis

Sensititre RAPMYCOI and SLOMYCOI panels (Thermo Fisher Scientific, Massachusetts, United States) were used to determine the susceptibilities of *M. brumae* strains to different antibiotics. A wide variety of antibiotic groups were tested, including β-lactams, aminoglycosides, quinolones, macrolides, tetracyclines, oxazolidinones, sulfonamides, and various anti-tuberculosis drugs. A bacterial suspension adjusted to a McFarland 1 standard was prepared, and 50 µl of the suspension was transferred to a tube of Mueller-Hinton broth, obtaining an

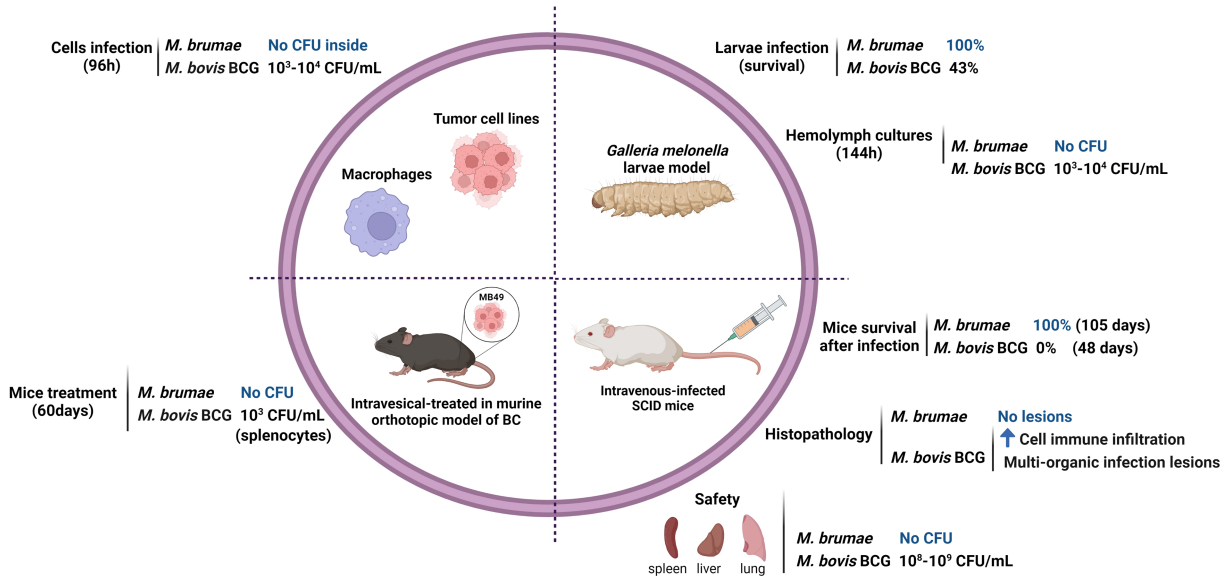


FIGURE 1
Safety profile of *Mycobacterium brumae*. Comparison of the pathogenicity and toxicity between *M. brumae* and *M. bovis* BCG Connaught. Summary of the results obtained in different *in vitro* studies infecting macrophages (Noguera-Ortega et al., 2016b) and bladder cancer cell lines (Noguera-Ortega et al., 2016a), and *in vivo* studies in different animal models: intrahemacoelical infection of *Galleria mellonella* (Bach-Griera et al., 2020), intravesical instillations in orthotopic murine model of bladder cancer (Noguera-Ortega et al., 2016a,b,c), and intravenous infection in SCID mice (Bach-Griera et al., 2020). Blue rising arrow (high cell immune infiltration).

inoculum of 5×10^5 CFU/ml. 100 µl of bacterial inoculum was added to each well of the Sensititre plates. The rest of the procedure was done according to the manufacturer's instructions. The method and guidelines for the interpretation of results were those of the Clinical and Laboratory Standards Institute. The antimicrobials *p*-aminosalicylic acid, capreomycin, cycloserine, and kanamycin susceptibility results were obtained from previously published studies (Luquin et al., 1993).

The reported mutations in *M. tuberculosis* H37Rv associated with drug resistance were obtained from the WHO 2021 annual catalog of mutations in *M. tuberculosis* (WHO, 2021) accessed in April 2022. The presence or absence of the mutations in the regions of interest (*rpoB* and *katG* genes and *inhA* regulation regions) was manually inspected in the gene sequence extracted from the *M. brumae* genome using MEGA X (Tamura et al., 2021).

DNA extraction

DNA was extracted from the four *M. brumae* strains using the UltraClean® Microbial DNA Isolation Kit following the manufacturer's instructions (MO BIO Laboratories, Inc. Carlsbad, California, United States) with slight modifications. Briefly, samples were heated to 65°C for 10 min in step 4 of the procedure, to improve production. The extracts obtained were electrophoresed on a 0.8% agarose gel using 6x NZYDNA Loading Dye (Nzytech, Lisboa, Portugal) and λ DNA/Hind III marker (Thermofisher, Massachusetts, United States). Samples were concentrated by evaporation at 45°C with Eppendorf® centrifugal vacuum concentrator 5301 (Sigma-Aldrich, Missouri, United States), and DNA quantification was performed using a NanoDrop™ 2000 spectrophotometer (NanoDrop Technologies, Inc. Wilmington, United States).

Library construction and genome sequencing

Whole-genome sequencing for the *M. brumae* ATCC 51384^T type strain was performed at the SCSIE Genomics Core Facility at the University of Valencia using the PacBio Sequel™ system (Pacific Biosciences, Menlo Park, CA, United States). A single non-multiplexed sequencing library was prepared following the manufacturer's protocol for 10 kb SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). A ready-to-sequence SMRT bell-Polymerase Complex was created using PacBio's Sequel binding kit 3.0 according to the manufacturer's instructions. The final library was sequenced on 1 Sequel™ SMRT® Cell 1 M v3, taking a 10 h movie using the Sequel Sequencing Kit 3.0. *M. brumae* strains CR103, CR142, and CR269 were subjected to whole-genome sequencing with Illumina. Genome libraries were constructed with a Nextera XT DNA library preparation kit (Illumina, San Diego, CA, United States), following the manufacturer's instructions. Sequencing was carried out at the Institute of Biomedicine of Valencia Ion an Illumina MiSeq platform (2 × 300 cycles paired-end reads).

Assembly and annotation of the completed genome

Sequence assembly was performed from 584,529,485 total bases (sequencing raw data used for the assembly is available in NCBI under the code PRJNA798885) with a read depth of 136 over the genome and read length N50 of 6,188 bases using SMRT Link v8.0.0 interface and Microbial Assembly analysis application (Pacific Biosciences). Genome assembly was conducted with genome size parameter set to 4 Mb and produced a single polished circular contig of 3,988,920 bases, and the quality of the assembly was evaluated using Inspector software (Chen et al., 2021), by mapping the long reads to the contig. Ori-Finder 2022 (https://tubic.org/Ori-Finder2022, accessed in August 2022) was used to identify the bacterial replication origin. The draft genome for the *M. brumae* ATCC 51384^T strain was annotated by the pipelines Rapid Prokaryotic Genome Annotation (Prokka) v 1.14.5 (Seemann, 2014), PGAP v 2022-04-14.build6021 (Tatusova et al., 2016), Bakta v1.4.2 (Cantalapiedra et al., 2021), and the Rapid Annotations using Subsystems Technology (RAST) server (Aziz et al., 2008). PseudoFinder (Syberg-Olsen et al., 2022) and PGAP were used for the detection of pseudogene candidates. Only those pseudogenes predicted by both approaches were kept as highly confident pseudogenes, while the rest were manually annotated as "putative pseudogene." PGAP annotation was curated by detailed inspection of virulence-related genes detailed in the next section. Additionally, all ESX regions were obtained from *M. tuberculosis* (Mycobrowser, accessed in April 2022) and BLASTp was used to identify and manually annotate those regions in our *M. brumae* genome. For the identification of other regions of interest in our genome, CRISPRFinder (Grissa et al., 2007) was used for the CRISPR-Cas system prediction. The internal conservation of the candidate direct repeats and the divergence of the candidate spacers were manually checked. PhiSpy v 4.2.21 was used to identify the prophages harbored in the *M. brumae* genome. CIRCOS software (Krzywinski et al., 2009) was used to represent the circular representation of the chromosome. To compare the functional categories between *M. brumae* and *M. tuberculosis* H37Rv, the COG assignment for each *M. brumae* protein was performed by EggNOG-mapper v 2.1.9 (Cantalapiedra et al., 2021) and *M. tuberculosis* COG assignments were obtained from the COG database.[1] The abundance of proteins in each category was compared. Row-wise Fisher's exact test was used within R v 4.2.1 to assess the statistical significance of differences between the *M. brumae* and *M. tuberculosis* H37Rv annotations for the gene abundance in each COG category.

Virulence genes analysis

To obtain a list of virulence genes, we looked for virulence-related genes from *M. tuberculosis* H37Rv in *M. brumae* and *M. bovis* BCG Connaught. The list of genes was compiled by combining 249 genes included in the Virulence Factors Database (VFDB; http://www.mgc.ac.cn/VFs/, accessed in July 2022) and 268 additional genes reported to be involved in virulence *in vivo* or *in vitro* by reviewing the bibliography. Because 193 genes were retrieved by both approaches, a final list of 324 genes was analyzed. The bibliographic search was done looking for in PubMed the following terms "((virulence[Title/Abstract]) AND (*Mycobacterium tuberculosis*[Title/Abstract])) AND ((*in vivo*[Title/Abstract]) OR (*in vitro*[Title/Abstract]))." We retrieved 594 manuscripts and included those which studied a virulence phenotype with *M. tuberculosis*, obtaining a list of 199 articles. The complete list of references consulted for the virulence-related genes is indicated in Supplementary Table S2.

To assess the presence or absence of these *M. tuberculosis* genes in other genomes, we first retrieved the nucleotide sequences from H37Rv genomes (accession number GCA_000195955.2) by indexing the nucleotide position with Python module Biopython version 1.79 (Cock et al., 2009), and retrieved the coordinates and orientation of the genes from Mycobrowser.[2] Next, we translated and obtained the amino acid sequence. For genes in the negative strand, we performed the complementary and reverse translation by using the Biopython module. tBLASTn was then used to find the amino acid similarity between *M. tuberculosis* proteins and both *M. bovis* BCG Connaught and *M. brumae* proteins using BLAST+ v 2.10.0 (Camacho et al., 2009) by a custom Python script v 3.6.0 (GitHub repository: https://github.com/PathoGenOmics/mbrumae_closedgenome). After exploring different percentages of identity, a gene was considered present when 70% identity and 70% coverage were reached. Additionally, the similarity between Mmpl, Mce, and PE/PPE proteins from our *M. brumae* genome (CP104302) and previously published *M. brumae* contigs (GCF_002553575.1) and *Mycobacterium fallax* (GCA_010726955.1) was assessed using the same procedure explained above.

Immunogenic genes analysis

For the analysis of immunogenic genes, all the experimentally verified T and B cell epitopes were obtained from the Immune Epitope Database for *M. tuberculosis* H37Rv (IEDB, 2020: Free epitope database and prediction resource, accessed in April 2020) studied in humans in the context of

1 https://www.ncbi.nlm.nih.gov/research/cog

2 https://mycobrowser.epfl.ch/

infectious diseases, allergy, autoimmunity, and transplantation (Supplementary Table S2). Then, the immunogenic epitope-associated genes were obtained from *M. tuberculosis* H37Rv using a custom pipeline.[3] The procedure followed to assess the presence of a gene was as explained above for the virulence-related genes.

Cell wall biosynthesis genes analysis

To obtain a list of *M. tuberculosis* genes involved in the cell wall biosynthesis the following bibliographic search was done in PubMed: "(*Mycobacterium tuberculosis*[Title/Abstract]) AND (biosynthesis[Title/Abstract]) AND ((mycolic acids[Title/Abstract]) OR (PDIM[Title/Abstract]) OR (PGL[Title/Abstract]) OR (PIM[Title/Abstract]) OR (TDM[Title/Abstract]) OR (TMM[Title/Abstract]))." We obtained a list of 221 articles, and we filtered for studies that described genes with cell-wall related function in *M. tuberculosis*, obtaining a final list of 95 articles and 179 genes. The complete list of references consulted for the cell wall genes is indicated in Supplementary Table S2. The procedure followed to assess the presence of a gene was the same as explained above for the virulence-related genes.

Mapping and variant calling in illumina sequences

We used two pipelines previously described (Chiner-Oms et al., 2019; Coscolla et al., 2021) for mapping and variant calling of three different *M. brumae* strains respect to the reference genome reported here. Only SNPs that reached fixation within an isolate were considered homozygous SNPs (within-host frequency, i.e., SNP frequency within the reads of the same sample, higher than 90%), being classified as heterozygous SNPs otherwise. All SNPs were annotated using SnpEff v 4.11 (Cingolani et al., 2012), and the *M. brumae* ATCC 51384^T strain annotation (CP104302). Repeated regions in the *M. brumae* genome were obtained using run-mummer3 and a custom pipeline using Python to identify repeated genes with a minimum threshold of 300 bp of identity. The custom script used and the list of repeated genes were uploaded to GitHub (see footnote 3). SNPs located in those regions were classified as "low confidence SNPs." For the deletions analysis, mean coverages per gene, corrected by the size of the gene were calculated (see footnote 3). Only those genes with at least 10% of the gene length with less than five reads-per-site were considered deleted. As repetitive regions could compromise the accuracy of the analysis, genes located in repeated regions were also excluded.

3 https://github.com/PathoGenOmics/mbrumae_closedgenome

Phylogenetic reconstruction

For the main phylogenetic construction, we selected a manageable subset of genomes from the *Mycobacterium* genus that would represent different parts of the genus, especially the closest taxa to *M. brumae*. We only used closed and well-described genomes (Fedrizzi et al., 2017). Our dataset included 21 downloaded genome assemblies (Supplementary Table S1), our newly sequenced *M. brumae* assembly, and the CR103, CR142, and CR269 *M. brumae* strains. Roary software v 3.11.2 (Page et al., 2015) was used to obtain the aligned core-genome, defined as the subset of genes found in at least 99% of the samples, and included 177 conserved genes (Supplementary Table S3). The alignment was used to infer a maximum-likelihood phylogenetic tree using IQ-TREE v 1.6.12 (Nguyen et al., 2015) with GTR model and the tree was re-examined using the bootstrap (1,000 replicates) resampling method. *Hoyosella subflava* (family Mycobacteriaceae) was used to root the tree as it is the sequenced bacterium closest to the *Mycobacterium* genus (Gupta et al., 2018). Virulence potential and growth rate (Fedrizzi et al., 2017) were annotated on the phylogeny using the iTOL software v 4 (Letunic and Bork, 2019). Possible recombination events in the tree were checked using Gubbins v 3.2 (Croucher et al., 2015).

An additional phylogenetic analysis was performed to confirm the consistency of our genomic sequence of *M. brumae* reference genome with other genomic sequences of the same strain. We included *M. brumae* ATCC 51384^T sequence, the three Illumina strains (CR103, CR142, and CR269) and three published *M. brumae* sequences for the type strain which are not closed (named in Supplementary Figure S1 as CIP1034565, DSM44177, and MBR1 with accession numbers GCF_002553575.1, GCF_004014795.1, and GCF_900073015.1 respectively). Panaroo software v 1.2.9 (Tonkin-Hill et al., 2020) was used to obtain the core-genome, using the following parameters "panaroo -i *.gff -o results_panaroo_core --clean-mode strict -a core," and the alignment was constructed using IQ-TREE software v 1.6.10 with GTR model and the tree was re-examined using the bootstrap (1,000 replicates) resampling method.

Results

Genome composition of the reference strain

To gain further insights into its genome composition and intraspecific variation, the complete *M. brumae* reference genome sequence was determined by PacBio. The complete genome from *M. brumae* and the manually curated annotation has been deposited in the GenBank database under the BioProject code PRJNA798885 with accession number CP104302. The *M. brumae* genome consisted of a single circular chromosome of 3,988,920 bp with an average GC content of 69% (Table 1). The quality of the assembly was assessed by a quality value score of 63 using

TABLE 1 Genome features.

Features	*M. brumae* PGAP	*M. brumae* Bakta	*M. brumae* RAST	*M. brumae* Prokka	*M. tuberculosis* H37Rv	*M. bovis* BCG Connaught	*M. fallax*
Genome size	4	4	4	4	4.41	4.29	4.16
GC content (%)	69.2	69.2	69.2	69.2	65.6	65.6	70.4
Contigs number	1	1	1	1	1	1	1
CDSs	3,791	3,794	3,827	3,781	3,906	3,804	3,995
Pseudogenes	97	-	-	-	30	190	118
Total RNAs	57	68	52	89	70	50	55
tRNAs	48	49	46	58	45	45	47
rRNAs	6	6	6	6	3	3	6
other RNAs	3	25	-	25	22	2	2
Hypothetical proteins	529	568	1,587	1,542	1,051	473	641
Downloaded accession number (NCBI)	-	-	-	-	GCA_000195955.2	GCA_001287325.1	GCA_010726955.1

Genome overview of *M. brumae* based on PGAP, Bakta, RAST, and Prokka annotations and compared with *M. tuberculosis* H37Rv, *M. bovis BCG* Connaught, and *M. fallax* JCM 6405.

Inspector software (Chen et al., 2021). The *oriC* was predicted next to the *rnpA-rpmH-dnaA-dnaN-recF-gyrB-gyrA* region (Supplementary Table S4) and shown in Figure 2. Conversely, the GC skew did not accurately indicate the *oriC* (Figure 2). Different annotations of *M. brumae* resulted in an estimate of 3,791, 3,794, 3,827, and 3,781 protein-coding sequences (CDS) using PGAP, Bakta, RAST, and Prokka, respectively (Supplementary Table S4). Because PGAP delivered fewer CDS as hypothetical proteins (Table 1), we kept the PGAP annotation for manual curation (Supplementary Table S4). Additionally, we found 48 genes encoding tRNA and two rRNA operons, each one including three rRNAs.

Pseudogene annotation was performed by combining the results from two approaches. PGAP predicted 97 and PseudoFinder predicted 326 pseudogenes, of which 47 were predicted by both. To be conservative, we annotated those detected by both approaches as pseudogenes and those detected just by PGAP as a probable pseudogene (Supplementary Table S4). The number of pseudogenes found with both approaches was similar to the number of pseudogenes in *M. tuberculosis* (GCA_000195955.2) and lower than those found in *M. bovis* (GCA_001287325.1) or *M. fallax* (GCA_010726955.1; Table 1). Another interesting feature in the genome refers to the bacterial adaptive systems against viral infections. We analyzed the clustered regularly interspaced short palindromic repeats (CRISPRs) and prophages. The CRISPRfinder software predicted that there were no CRISPR sequences throughout the *M. brumae* genome. Next, we used the PhiSpy software to identify the prophages present in the genome, and one insert was detected between the positions 83,324–110,205 within the *M. brumae* genome, including a total of 235 CDS (Supplementary Table S5).

For a general genome content description, CDS predicted by PGAP annotation were subclassified into 26 different COG categories (Figure 3) and compared to *M. tuberculosis*. We found 2,617 (69%) *M. brumae* CDSs associated with a COG category, including 661 (25%) assigned to less informative categories such as "General function prediction only" and "Function unknown" (Figure 3). Compared to the *M. tuberculosis* annotation, *M. brumae* had a significantly bigger representation in L ("Replication and repair") and M ("Cell wall/membrane/envelope biogenesis") categories with 6.57 and 7.87% of the genes in *M. brumae* compared to 3.35 and 4.97% in *M. tuberculosis* (value of ps of 4.70e−07 and 1.86e−04 respectively). *Mycobacterium brumae* was also more abundant (8.06%) than *M. tuberculosis* (6.18%). On the contrary, *M. tuberculosis* showed a statistically higher abundance of genes in six categories compared to *M. brumae*, including lipid metabolism (8.75 and 4.47% for *M. tuberculosis* and *M. brumae* respectively, value of p 3.08e−09), coenzyme metabolism (6.63 and 3.59%, value of p of 2.80e−06), mobilome (2.67 and 0%, value of p of 3.13e−21), secondary structure (5.61 and 3.52%, value of p of 2.03e−04), defense mechanisms (3.28 and 1.91%, value of p of 1.42e−03), and signal transduction (3.51 and 2.18%, value of p of 2.97e−03).

Phylogenetic position of *Mycobacterium brumae* in the genus

To investigate the *M. brumae* position in the evolutionary history of mycobacteria, we constructed a phylogeny using a representative sample of species from the genus (Figure 4). The alignment was constructed using all the core genes among the

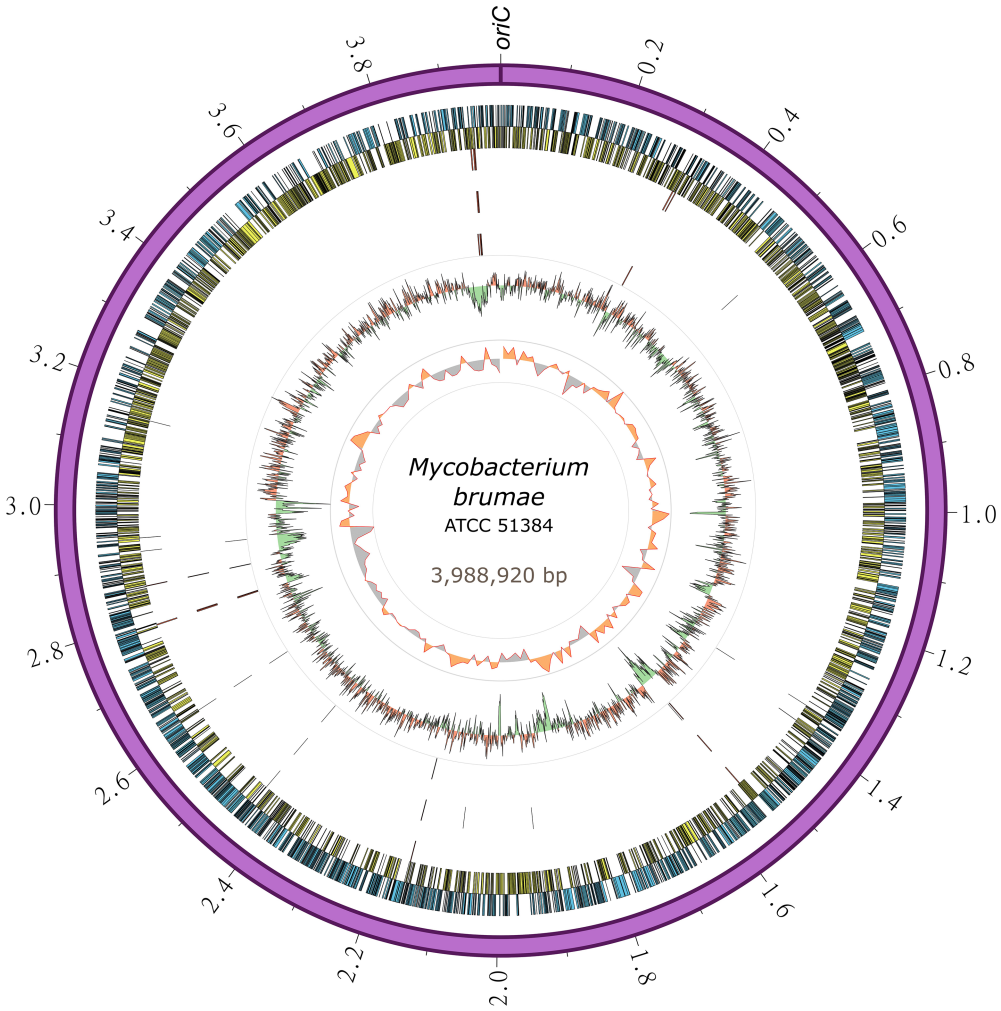


FIGURE 2
Circular representation of the *Mycobacterium brumae* chromosome. The outermost ring shows the chromosome length in megabases, followed by the forward (blue) and reverse genes (yellow). SNPs in *M. brumae* strains (CR269, CR142, and CR103) with respect to the ATCC 51384^T reference genome is indicated in the next three rings, respectively. Moving inward, the next ring shows the GC content, with values over the mean filled in red and values below the mean in green. The last ring shows the GC skew (G–C)/(G+C) using a 20-kb window.

different species to provide a higher resolution of the relationships among them. Recombination events were eliminated from the alignment and the structure of the tree remained unaltered, which supports the robustness of the analysis. *Mycobacterium brumae* clusters with other non-pathogenic rapidly growing mycobacteria. *Mycobacterium fallax* was the closest mycobacteria to *M. brumae*, followed by *Mycobacterium insubricum*, and all three clustering with the monophyletic group formed by *Mycobacterium confluentis* and *Mycobacterium chitae*.

Although the genome presented here is the first *M. brumae* closed genome, the same strain has been sequenced previously. We assessed the consistency of our closed genome sequence with previously published *M. brumae*-type strain sequences. Phylogenetic analysis shows that our *M. brumae* sequence clustered with the monophyletic group formed by MBR1 and CIP1034565, all of them corresponding to the type strain (Supplementary Figure S1). Another monophyletic group was formed by the other three strains (CR103, CR142, and CR269) and DSM44177 which also corresponds to the type strain. This result indicates the consistency of our results with two out of three genomic sequences of the same strain.

Drug resistance phenotypic and genotypic profile

The minimum inhibitory concentrations (MICs) of different antimicrobial drugs for the four *M. brumae* strains were determined using broth microdilution panels (Sensititre RAPMYCOI and SLOMYCOI; Supplementary Table S6). The test results were identical for the four strains and demonstrated resistance to *p*-aminosalicylic acid, cefepime, isoniazid (≤ 10 μg/

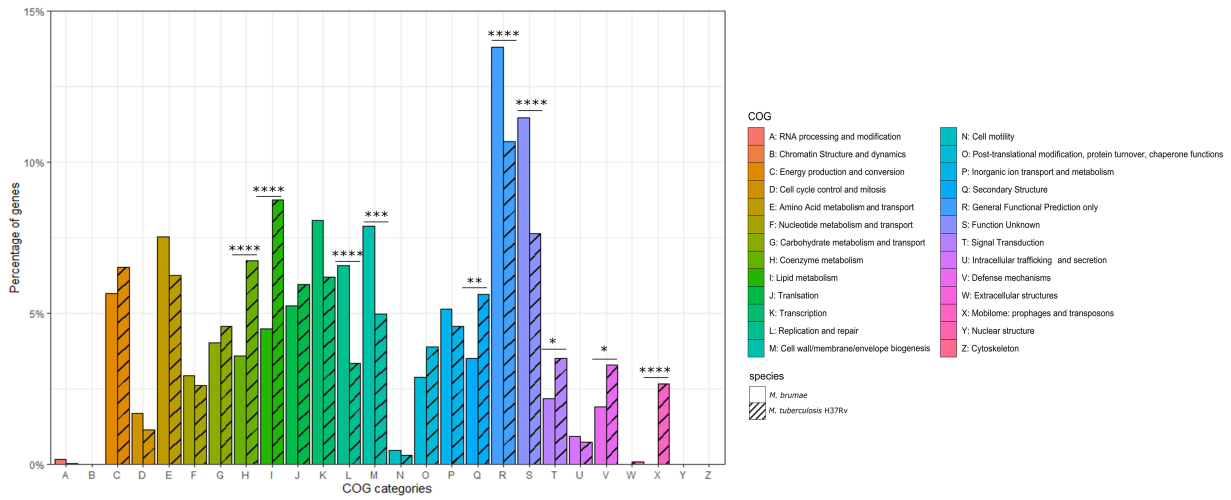


FIGURE 3
Gene annotation comparison of COG distributions shared in *Mycobacterium brumae* and in *Mycobacterium tuberculosis* H37Rv. COG refers to Clusters of Orthologous Groups. The vertical axis shows the percentage of genes in each category. The different categories are represented on the horizontal axis and the legend indicates the correspondence with each COG category. The corresponding strain for each bar is indicated in stripe pattern for *M. tuberculosis* H37Rv and in plain pattern for *M. brumae*. Fisher test *p* values for each category indicated as following: *p<0.05; **p<0.01; ***p<0.001; and ****p<0.0001.

ml antimicrobial concentrations), and rifampin and intermediate susceptibility to ceftriaxone. The four *M. brumae* strains proved susceptibility to amikacin, amoxicillin/clavulanic acid 2:1, capreomycin, cefoxitin, ciprofloxacin, clarithromycin, cycloserine, doxycycline, ethambutol, ethionamide, imipenem, isoniazid (≥ 10 μg/ml antimicrobial concentration), kanamycin, linezolid, minocycline, moxifloxacin, rifabutin, streptomycin, sulfamethoxazole, tigecycline, and tobramycin.

To explain the antibiotic resistance profile of *M. brumae*, drug-resistance-related mutations for rifampicin (RIF) and isoniazid (see Methods) were analyzed. Regarding RIF resistance, we did not find any of the 90 reported mutations in the *rpoB* gene associated with resistance in *M. tuberculosis*. Similarly, for isoniazid resistance, nine reported mutations that included the *katG* gene and promoter mutations upstream of the *fabG1-inhA* operon were not found in *M. brumae*. However, the *M. brumae* sequence held several mutations in these regions compared to the sequence of *M. tuberculosis* susceptible strains (Supplementary Figure S2), indicating that the drug-resistance associated mutations could be potentially located in other parts of these genes.

Virulence related genes

For the presence or absence of *M. tuberculosis* virulence genetic factors in *M. brumae and M. bovis* BCG Connaught, we focused on a list of *M. tuberculosis* virulence-related genes, which include PE/PPE proteins, ESX export systems, Mce proteins, MmpL proteins, and proteins from the Two-Component System (TCS) among others (Supplementary Table S2). We found that *M. brumae* only held 57 out of 324 *M. tuberculosis* virulence-associated genes with a range of protein identity from 71.0 to 98.7% covering between 92.7 and 100% of the *M. tuberculosis* gene sequence (Table 2). From these, 56 genes were also shared with the *M. bovis* BCG strain (Figure 5A). Only the phoT protein, involved in the import of inorganic phosphate across the membrane, was found in *M. brumae* but not in *M. bovis* BCG Connaught, due to a frameshift mutation in BCG that alters the aminoacid sequence from position 156 and likely alters phoT function (Collins et al., 2003). We did not detect genes from PE/PPE, Mce, or MmpL families with more than 70% protein identity to *M. tuberculosis* or *M. bovis* BCG in *M. brumae*, which was consist with the lack of these genes in the annotation performed with Prokka. However, *M. brumae* PGAP annotation predicted the presence of 9 MmpL, 15 Mce, and 11 PE-PPE family genes. MmpL and Mce genes showed high protein identity to the ones annotated in previously published *M. brumae* contigs and the closely related species *M. fallax*. However, the 11 PE-PPE family genes showed very low protein similarity or coverage to both. However, because the synteny of PE/PPE within ESX regions was consistent with the gene organization known for other mycobacteria, we kept them in the annotation. Although not all ESX clusters include *M. tuberculosis* virulence genes, we next analyzed in detail the presence or absence of the five ESX clusters in *M. brumae* and we found that *M. brumae* only had ESX-3 and ESX-4 gene clusters (Supplementary Figure S3), although one gene in ESX-3 is annotated as a putative pseudogene (*eccD4*), indicated in Supplementary Table S4. *M. brumae* does not contain the virulence-related ESX-1, which is consistent with its absence in all the species clustering together in the phylogeny: *M. fallax*, *M. chitae*, and *M. confluentis* (Figure 4). However, although ESX-2 was not complete in *M. brumae*, we found 6 genes that could

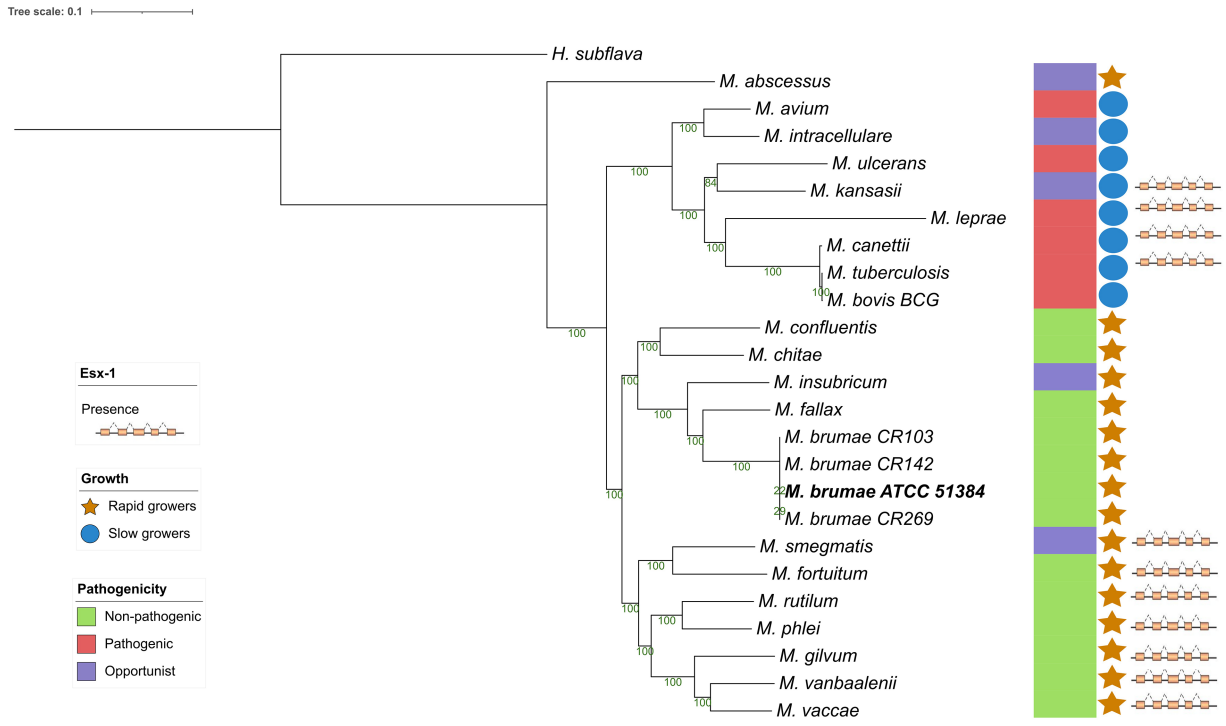


FIGURE 4
Phylogenetic relationship of *Mycobacterium brumae* within the *Mycobacterium* genus. Maximum likelihood phylogenetic tree of type strains from the *Mycobacterium* genus based on an alignment with 177 core genes, highlighting the position of *M. brumae* ATCC 51384^T type strain. GTR model was used with bootstrap confidence of 1,000 replicates. Branch lengths are proportional to nucleotide substitutions and the topology is rooted with *Hoyosella subflava*. The phylogenetic tree was annotated with the pathogenicity, the ESX-1 presence or absence, and the growth rate (see Methods).

be homologous to half of the genes in the ESX-2 *M. tuberculosis* H37Rv cluster (Rv3884c to Rv3889c).

Comparison of the immunogenic capacity

We next analyzed the immunogenic capacity of *M. brumae* as a potential immunotherapy agent. Immunity to infections caused by mycobacteria species such as *M. tuberculosis* depends mainly on T lymphocytes. For the *in-silico* analysis of the antigens in *M. brumae*, experimentally proved human T cell epitopes were obtained and the corresponding gene for each epitope was assessed (see Materials and methods), obtaining a list of 442 non-redundant antigens, including the ESAT-6 gene cluster as potent immunogenic and virulence-associated regions. The presence or absence of each gene was evaluated and compared in the *M. bovis* BCG Connaught and *M. brumae* ATCC 51384^T strains. As a result, we found that *M. bovis* BCG strain shared 374 out of the 441 immune-related genes with *M. tuberculosis*, while the *M. brumae* genome only had 85 genes (Supplementary Table S2), from which 83 genes were shared with the *M. bovis* BCG and the genes Rv2074 (a possible pyridoxamine 5′-phosphate oxidase) and Rv3427c, a possible transposase, were only found in *M. brumae* (Figure 5B). Interestingly, there were no ESAT-6 antigens among the genes found in *M. brumae*, which correlates with its non-virulent phenotype and shows that the immunogenic activity in *M. brumae* could be produced by other genes. B cells and humoral immunity can modulate the immune response to different pathogens, including *M. tuberculosis*. To explore other mechanisms for immunity that could modulate the immune response, we also analyzed the presence of B cell antigens in *M. brumae*. From a list of 127 non-redundant genes related to the immunogenic activity mediated by B cells, the *M. bovis* BCG strain shared 81 genes with *M. tuberculosis*, while the *M. brumae* genome only contained 17 genes (Supplementary Table S2), all of them shared with *M. bovis* BCG (Figure 5C).

Cell wall composition

To compare the cell wall composition at the genetic level, we assessed the presence in *M. brumae* of 179 genes related to the mycobacterial cell wall (see Methods) and associated with the synthesis and transport of mycolic acids, PDIM, PGL, PIM, PL, and the TDM and TMM glycolipids. As expected, the *M. brumae* genome contained genes for the synthesis and transport of α-mycolic acids and for the PIM, TDM, and TMM production, and lacked PDIM and PGL associated genes (Supplementary Table S2).

TABLE 2 *M. tuberculosis* virulence genes with high similarity in *M. brumae*.

Locus	Gene name	Start	End	Strand	Function	Protein identity (%)	Query coverage (%)
Rv0126	*treS*	152,324	154,129	Forward	Lipid Synthesis	82.7	94.7
Rv0129c	*fbpC*	156,578	157,600	Reverse	Secreted antigen	71	96.2
Rv0198c	*zmp1*	234,516	236,507	Reverse	Metallo-protease	75.6	100
Rv0148	Rv0148	174,833	175,693	Forward	Other virulence factors	82.9	100
Rv0211	*pckA*	251,782	253,602	Forward	Other virulence factors	83.5	99.8
Rv0244c	*fadE5*	293,798	295,633	Reverse	Other virulence factors	78.6	100
Rv0334	*rmlA*	398,658	399,524	Forward	Other virulence factors	80.3	98.6
Rv0353	*hspR*	423,639	424,019	Forward	Other transcriptional regulators	74	95.3
Rv0409	*ackA*	493,851	495,008	Forward	Other virulence factors	40.5	100
Rv0410c	*pknG*	495,062	497,314	Reverse	Protein kinase	76.7	100
Rv0467	*icl1*	557,527	558,813	Forward	Other genes related in lipid synthesis	89.7	100
Rv0491	*senX3-regX3*	580,809	581,492	Forward	Two-Component System (TCS)	90.8	100
Rv0500A	Rv0500A	591,111	591,347	Forward	Other virulence factors	98.7	100
Rv0667	rpoB	759,807	763,325	Forward	Other virulence factors	91.6	100
Rv0757	*phoP*	851,608	852,351	Forward	Gene Expression Regulators	83.9	92.7
Rv0780	*purC*	873,343	874,236	Forward	Other virulence factors	76.5	100
Rv0820	*phoT*	912,726	913,502	Forward	Cell wall protein	84.2	100
Rv0821c	*phoY2*	913,558	914,199	Reverse	Other transcriptional regulators	84.6	99.5
Rv0903c	*prrA*	1,005,852	1,006,562	Reverse	Two-Component System (TCS)	91.1	100
Rv0981	*mprAB*	1,096,822	1,097,508	Forward	Two-Component System (TCS)	89.1	100
Rv0982	*mprAB*	1,097,508	1,099,022	Forward	Two-Component System (TCS)	76.2	100
Rv1002c	Rv1002c	1,118,428	1,119,939	Reverse	Other virulence factors	72.4	100
Rv1236	*sugA*	1,378,927	1,379,850	Forward	Protein ABC transporter	73.1	93.8
Rv1237	*sugB*	1,379,855	1,380,679	Forward	Protein ABC transporter	81.1	98.9

(Continued)

TABLE 2 (Continued)

Locus	Gene name	Start	End	Strand	Function	Protein identity (%)	Query coverage (%)
Rv1238	*sugC*	1,380,684	1,381,865	Forward	Protein ABC transporter	78.7	100
Rv1293	*lysA*	1,448,028	1,449,371	Forward	Other virulence factors	76.1	99.6
Rv1346	*mbtN*	1,510,846	1,512,006	Forward	Other virulence factors	75.7	99
Rv1540	Rv1540	1,742,845	1,743,771	Forward	Other virulence factors	81.6	100
Rv1821	*secA2*	2,066,457	2,068,883	Forward	Secretion system	81.2	100
Rv1854c	*ndh*	2,101,651	2,103,042	Reverse	Other virulence factors	71.1	99.8
Rv2097c	*pafA*	2,355,319	2,356,677	Reverse	Proteasome-associated protein	92.7	100
Rv2115c	*mpa*	2,374,461	2,376,290	Reverse	Proteasome-associated protein	87.7	100
Rv2136c	Rv2136c	2,396,008	2,396,838	Reverse	Oxidative and nitrosative stresses	74.7	100
Rv2192c	*trpD*	2,455,631	2,456,743	Reverse	Other virulence factors	72.8	92.7
Rv2220	*glnA1*	2,487,615	2,489,051	Forward	Other virulence factors	83.1	100
Rv2246	*kasB*	2,519,396	2,520,712	Forward	Lipids and Fatty Acids Metabolism	74	98.4
Rv2428	*ahpC*	2,726,193	2,726,780	Forward	Oxidative and nitrosative stresses	78.6	97.4
Rv2445c	*ndk*	2,745,314	2,745,724	Reverse	Oxidative and nitrosative stresses	73.7	100
Rv2583c	*relA*	2,907,826	2,910,198	Reverse	Other virulence factors	87.6	100
Rv2610c	*pimA*	2,937,865	2,939,001	Reverse	Other virulence factors	77.8	100
Rv2794c	*pptT*	3,103,257	3,103,940	Reverse	Other virulence factors	74.6	100
Rv2869c	*rip1*	3,180,548	3,181,762	Reverse	Metallo-protease	71.6	100
Rv2987c	*leuD*	3,344,033	3,344,629	Reverse	Other virulence factors	81.4	100
Rv3003c	*ilvB1*	3,361,130	3,362,986	Reverse	Other virulence factors	83	99.7
Rv3223c	*sigH*	3,598,901	3,599,551	Reverse	Sigma Factors	82	100
Rv3245c	*mtrB*	3,624,910	3,626,613	Reverse	Other virulence factors	75.2	98.1
Rv3370c	*dnaE2*	3,781,501	3,784,740	Reverse	Other virulence factors	80.9	100
Rv3409c	*choD*	3,826,991	3,828,727	Reverse	Catabolism of cholesterol	79.8	99.7

(Continued)

TABLE 2 (Continued)

Locus	Gene name	Start	End	Strand	Function	Protein identity (%)	Query coverage (%)
Rv3414c	*sigD*	3,833,038	3,833,676	Reverse	Sigma Factors	71.4	93.4
Rv3601c	*panD*	4,043,862	4,044,281	Reverse	Other virulence factors	82.9	97.9
Rv3681c	*whiB4*	4,121,198	4,121,554	Reverse	Other virulence factors	73.1	100
Rv3682	*ponA2*	4,121,916	4,124,348	Forward	Oxidative and nitrosative stresses	72.9	94.2
Rv3722c	Rv3722c	4,166,821	4,168,128	Reverse	Other virulence factors	71.8	95.2
Rv3810	*erp*	4,273,739	4,274,593	Forward	Cell envelope protein	73	100
Rv3846	*sodA*	4,320,704	4,321,327	Forward	Other virulence factors	79.8	100
Rv3862c	*whiB6*	4,338,171	4,338,521	Reverse	Secretion system	74.5	86
Rv3876	*espI*	4,353,010	4,355,010	Forward	Secretion system	71.6	92

List of the virulence-associated genes from *M. tuberculosis* H37Rv found with high similarity in the newly sequenced *M. brumae* ATCC 51384^T strain. The respective *M. tuberculosis* Rv code, gene name, orientation, position within the genome, predicted function, protein identity compared to *M. tuberculosis* and query coverage are shown.

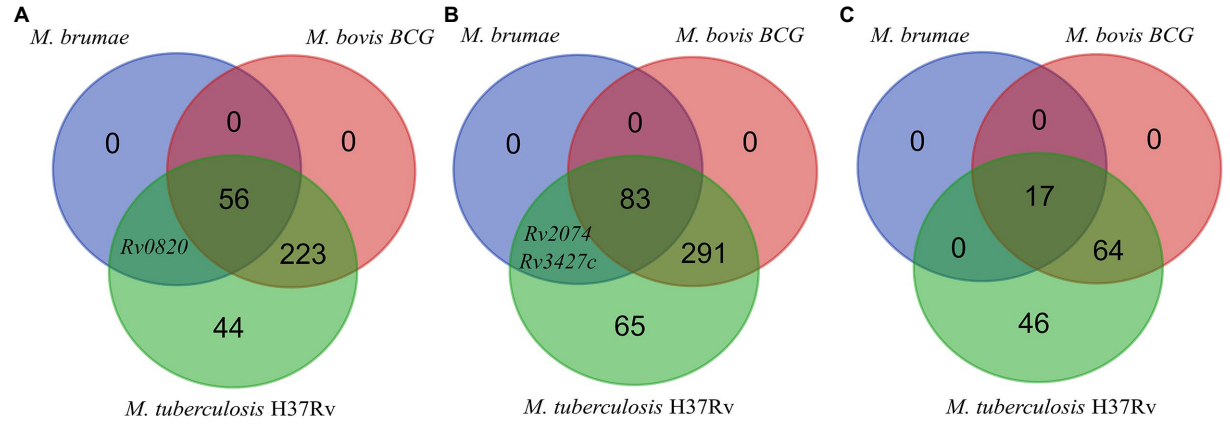


FIGURE 5
Genes presence or absence analysis. Venn diagram showing the comparison of the presence/absence and abundance of *M. tuberculosis* virulence-associated genes **(A)**, antigens recognized by T cells **(B)**, and antigens recognized by B cells **(C)** in the *M. bovis* BCG Connaught and *M. brumae* genomes.

Interestingly, among the 96 genes related to mycolic acid production, we only found in *M. brumae* 31 mycolic-acids-associated genes, associated with general fatty acid biosynthesis functions. From the 83 total assessed genes related to the PIM, TDM, TMM, and other PL production, we detected 13 genes in *M. brumae* (Rv0126, Rv0129c, Rv0982, Rv1166, Rv1236-Rv1238, Rv1564c, Rv2188, Rv2610c, Rv2869c, Rv3264c, and Rv3793).

Intraspecific variability

To assess the intraspecific variation among the different *M. brumae* strains, we analyzed the genomes of three isolates (CR103, CR142, and CR269 strains). The sequences from the three different *M. brumae* strains were uploaded to the ENA-EBI database under the BioProject code PRJEB52012 and with accession numbers ERR9463983 (CR103), ERR9463984 (CR142), and ERR9463985 (CR269). Intraspecies variability was assessed by calling SNPs and deletions compared to our newly sequenced *M. brumae* ATCC 51384^T complete genome using two different pipelines (see Materials and methods). Mapping qualities were sufficient to study variability in SNPs and regions of difference (Supplementary Table S7). Both pipelines predicted a similar number of SNPs before filtering, and the same number of SNPs after filtering for the SNPs found in non-repetitive

regions. All variable SNPs located in repeated regions that were discarded are shown in Supplementary Table S8. Strains CR103, CR142, and CR269 differed from the reference genome in five, three, and six high-confidence SNPs, respectively. The majority (five of eight) of the SNPs among *M. brumae* strains were present in coding regions (Table 3). Compared to *M. brumae* ATCC 51384^T reference, the three *M. brumae* isolates presented a missense variant related to the aminopeptidase N and another SNP located in a pseudogene. The CR103 strain presented only one unique SNP in a hypothetical protein (missense variant). The CR142 strain showed one unique SNP related to the propionyl-CoA--succinate CoA transferase (synonymous variant). The CR269 harbored two unique SNPs in intergenic regions which could potentially modify genes related to the glycerol-3-phosphate dehydrogenase/oxidase and the aquaporin Z.

We also examined if genes from the reference ATCC 51384^T strain were absent in any of the three *M. brumae* genomes, by analyzing the coverage of Illumina reads in each position (see Materials and methods). We did not find any gene completely deleted in any of the three additional strains. However, we found 14 genes for which at least 10% of their length was not covered by sequence data in at least one of the strains. Only three of them contained regions not covered in any of the three additional strains: L2Z93_000058, L2Z93_000604, and L2Z93_001659, coding for a nucleoid-structuring protein H-NS, a transglycosylase family protein, and a C40 family peptidase respect (Table 4).

Discussion

In this study, we performed the assembly and compared the genome of the *M. brumae* type strain with *M. bovis* BCG and *M. tuberculosis* to precisely unravel the peculiar characteristics shown phenotypically by this species: its non-pathogenicity and high immunostimulatory ability. We also compared some genomic features of *M. brumae* with *M. fallax*, a fast-growing mycobacterial species that was initially isolated from water samples. *Mycobacterium fallax* shares phenotypic characteristics with *M. brumae* and in initial phylogenetic comparisons appeared as the closest species to *M. brumae* (Luquin et al., 1993; Noguera-Ortega et al., 2016a). The phylogenetic position of *M. brumae* is congruent with the microbiological features observed (Bach-Griera et al., 2020) as non-pathogenic rapidly-growing mycobacteria. Genome-based phylogenetic reconstruction confirms the position of *M. brumae* within Mycobacteriaceae, and closely related to *M. fallax*, which was shown in previous analysis with limited data based on 16S rRNA sequences (Noguera-Ortega et al., 2016b).

We confirmed the small size of *M. brumae* genome. Although it is accepted that the size of the genome of nonpathogenic mycobacteria is longer than those of opportunistic and obligated pathogenic mycobacteria (Rahman et al., 2014; Jia et al., 2021), *M. brumae* genome size is similar to *M. tuberculosis* genome and comparable only to their related *Mycobacterium* species such as *M. fallax*. *Mycobacterium brumae* genome is among the smallest

TABLE 3 Predicted impact and characteristics of the SNPs present in the *M. brumae* strains compared to the *M. brumae* reference genome.

Position	Ref	Alt	Start	End	CR103	CR142	CR269	SNP Effect	Variant	Function
254,422	A	G	Intergenic region	Intergenic region	–	–	Yes	Modifier	Modifier	Aquaporin Z
1,357,198	T	C	1,357,157	1,357,465	Yes	–	Yes	Low	Synonymous	Hypothetical protein
1,541,659	A	G	1,541,571	1,542,494	Yes	–	–	Moderate	Missense	Hypothetical protein
1,929,203	C	T	1,928,301	1,929,584	–	Yes	–	Low	Synonymous	Propionyl-CoA succinate CoA transferase
2,440,841	C	T	2,440,345	2,441,145	Yes	Yes	Yes	Moderate	Missense	Pseudogene
2,620,902	A	G	2,618,705	2,621,296	Yes	Yes	Yes	Moderate	Missense	Aminopeptidase N
2,919,681	A	G	Intergenic region	Intergenic region	Yes	–	Yes	Modifier	Modifier	Hypothetical protein
3,150,050	A	G	Intergenic region	Intergenic region	–	–	Yes	Modifier	Modifier	Glycerol-3-phosphate dehydrogenase/oxidase

SNPs found with both variant calling pipelines and outside of repetitive regions (see Methods). The table shows the position of the SNP within the reference genome, the reference base and the respective SNP found, the CDS start and end positions, the presence ("yes") or absence ("-") of the SNP in each *M. brumae* strain and the predicted impact of the mutation by the SnpEff software. The protein function is decided for the involved CDS or the following gene in intergenic regions.

TABLE 4 Genes affected by deletions in the different *M. brumae* strains compared to the *M. brumae* reference genome.

Gene annotation		Percentage deleted (%)		
CDS ID	Product	CR103	CR142	CR269
L2Z93_000058	Nucleoid-structuring protein H-NS	21	13	20
L2Z93_000315	IS256 family transposase	21	0	0
L2Z93_000322	Hypothetical protein	0	0	19
L2Z93_000342	Hypothetical protein	18	26	0
L2Z93_000604	Transglycosylase family protein	37	37	32
L2Z93_001423	Hypothetical protein	0	20	23
L2Z93_001447	Hypothetical protein	29	29	23
L2Z93_001544	HNH endonuclease	10	0	0
L2Z93_001659	C40 family peptidase	37	32	38
L2Z93_001702	Alanine and proline-rich secreted protein Apa	0	10	25
L2Z93_001714	Hypothetical protein	16	0	0
L2Z93_001739	Hypothetical protein	0	11	0
L2Z93_002425	Translation initiation factor IF-2	15	15	0
L2Z93_003782	IS3 family transposase	24	0	0

CDS ID indicates the PGAP identity code for genes affected by deletions. The product obtained for each gene was predicted by PGAP.

mycobacterial genomes, after *Mycobacterium uberis* (Benjak et al., 2018), *Mycobacterium lepromatosis* (Silva et al., 2022), and *Mycobacterium leprae* (Rahman et al., 2014; Matsumoto et al., 2019). Our results are highly congruent with the draft genome obtained with an Illumina MiSeq, which had a length of 4,026,006 bp and a mean GC ratio of 69.1% (D'Auria et al., 2016). The percentage of GC content in *M. brumae* genome is in accordance with other RGMs genomes (between 66.0 and 69.0%) and higher than for slow-growing mycobacteria; only *M. chelonae-M. abscessus* group contains a percentage between 63.9 and 64.1% of GC contents among RGMs (Sassi and Drancourt, 2014). Despite sharing a genome size with pathogenic mycobacteria, the pattern of gene categories is different between *M. brumae* and *M. tuberculosis* or *M. bovis* BCG. A lower percentage of genes related to defense mechanisms and lipid metabolism, among others, is found in *M. brumae* in comparison to *M. tuberculosis* genome, which correlates with its non-pathogenicity (Gago et al., 2018).

The mycobacterial Type VII secretion system (Type VII; ESX1-5) is specialized for the secretion of various protein substrates across the complex cell envelope, and they have evolved through duplication leading to a variety of ESX clusters across the mycobacterial genus (Newton-Foot et al., 2016). *Mycobacterium tuberculosis* and some nonpathogenic *Mycobacterium* such as *Mycobacterium paragordonae* (Kim et al., 2019) harbor five ESX clusters, other non-pathogenic mycobacteria, such as *Mycobacterium smegmatis*, *Mycobacterium flavescens*, *M. phlei*, *Mycobacterium gilvum*, *Mycobacterium vaccae*, or *Mycobacterium gordonae* only three (Gcebe et al., 2016; Newton-Foot et al., 2016). Here we show that, like *M. abcessus* and *M. massiliensse* (Gcebe et al., 2016), *M. brumae* genome harbors only ESX-4 and ESX-3 and lacks other ESX clusters. The ESX-1 system has been implicated in virulence, and its deletion has been related to the attenuation of the pathogenic mycobacteria as in the case of *M. bovis* BCG (Hsu et al., 2003; Pym et al., 2003). However, its presence throughout the *Mycobacterium* genus, including non-pathogenic and fast-growing organisms (Gcebe et al., 2016) suggests that the primary function of this gene cluster is not virulence and that the virulence-associated function could have evolved more recently in pathogenic organisms. PE/PPE family proteins represent about 10% of the total *M. tuberculosis* genome and they are expressed by *M. tuberculosis* upon infection of macrophages and play critical roles in virulence, antigenic diversity, and modulation of the host immune response (Mukhopadhyay and Balaji, 2011; Fishbein et al., 2015). Similar to what is found for *M. fallax*, very few PE/PPE family genes are annotated in *M. brumae* genome, and sequence identity from the few annotated PE/PPE genes is very low compared to *M. tuberculosis* and *M. fallax*. The distant nature and the low number of these genes in *M. brumae* and *M. fallax* are consistent with the fact that these genes are more abundant in slow-growing mycobacteria (Gey van Pittius et al., 2006; Qian et al., 2020).

Mce proteins are lipid/sterol transporters (Cantrell et al., 2013) that play an important role in interfering with the host cell signaling modulation (Fenn et al., 2019) and they are implicated in the entry and survival inside macrophages (Saini et al., 2008; Fenn et al., 2019). We could only detect three *mce* clusters *M. brumae* genome but showed very low sequence similarity with *M. tuberculosis mce* genes, which impedes accurately hypothesizing its orthology with specific *mce* clusters in *M. tuberculosis*. The number of *mce* operons present in the genome has been related to pathogenicity in actinomycetes: both in mycobacteria and nocardia. *Mycobacterium abscessus* contains seven *mce* operons (*mce1-7*) while *M. smegmatis* contains only four (Ripoll et al., 2009). There are six *mce* operons in *Nocardia farcinica*, one of the agents causing nocardiosis, whereas *Streptomyces avermitilis and*

Streptomyces coelicolor, both nonpathogenic soil bacteria, each have only one copy of the *mce* operon (Ishikawa et al., 2004). In a recent study, Bachmann et al. (2020) found that the *mce1* operon is highly variable across all five mycobacterial sub-genera. The study indicates that the sequence similarity of proteins encoded by the *mce1* genes between *M. tuberculosis* and *M. brumae* was very low, which confirms our results. Although there are yet no described genetic or phenotypic markers whose appearance correlates with nontuberculous mycobacterial virulence (Falkinham, 2009), the absence of virulence genes, such as ESX-1, and most PE/PPE can explain the results previously obtained *in vitro* and *in vivo* models, demonstrating the safety and non-toxicity of *M. brumae* (summarized in Figure 1; Noguera-Ortega et al., 2016b, 2020; Bach-Griera et al., 2020). *Mycobacterium brumae*, as most of the more than 190 mycobacterial species that have been identified, have not been linked with human, animal, or plant disease. Furthermore, the absence of ESX-1 secretion system in *M. brumae* is interesting since this chromosomal region is involved in a unique conjugation mechanism to some RGM, as has been demonstrated for *M. smegmatis*, potentially enabling saprophytic mycobacteria to change niche and evolve to an opportunistic or specialized persistent pathogen (Coros et al., 2008; Gray and Derbyshire, 2018).

Previous studies related to resistance to the main anti-tuberculosis drugs in NTMs indicate that it is mainly due to specific mutations in genes that encode their target or the enzymes that activate them, as well as their promoter regions (Alcaide et al., 2017; Bakuła et al., 2018). Resistance to rifampin in *M. tuberculosis* is conferred by mutations mainly in the *rpoB* gene, and similar mutations in the same cluster have been described for *M. leprae* and some NTM species such as *Mycobacterium kansasii*, *M. smegmatis*, *M. avium* complex, and *Mycobacterium ulcerans* (Brown-Elliott et al., 2012; van Ingen et al., 2012; Alcaide et al., 2017). *Mycobacterium brumae* lacks point mutations within RIF-resistance determining region (RRDR) of the *rpoB* gene but has mutations in adjoining areas of the region. Similarly, in some resistant isolates of *M. avium*, no internal mutations were found in *rpoB* but in a downstream region of this gene. This fact together with the mycobacterial cell wall composition, which has a role in the intrinsic resistance to rifampin (Brown-Elliott et al., 2012; Huh et al., 2019), could explain the resistance to rifampin observed in *M. brumae*. Regarding isoniazid resistance, often associated with a loss of catalase-peroxidase activity, point mutations in the sequence of *katG* gene (responsible for activating the drug) have been described in *M. tuberculosis* and some NTMs such as *M. smegmatis* or *M. kansasii* (Brown-Elliott et al., 2012; Alcaide et al., 2017; Iwao and Nakata, 2018). In addition, resistance to isoniazid has been described in *M. smegmatis*, *M. bovis* BCG, and *M. tuberculosis* due to overexpression or mutation of the *inhA* structural gene or its promoter, which are involved in the synthesis of the bacterial wall (Larsen et al., 2002; Brown-Elliott et al., 2012; Alcaide et al., 2017). *Mycobacterium brumae* strains lack point mutations in the most relevant positions described for isoniazid. However, the low cell wall permeability of mycobacteria and the different efflux pump systems or porin channels could also play a crucial role in intrinsic drug resistance in NTMs such as *M. fortuitum* or *M. smegmatis* (van Ingen et al., 2012; Bakuła et al., 2018) and could explain the intrinsic resistance to isoniazid found in *M. brumae*. Finally, similar to isolates of *M. fortuitum* and *M. avium* complex (Cheng et al., 2017), *M. brumae* is also resistant to *p*-aminosalicylic acid, although the mutations responsible for resistance have not been specifically identified (Yang et al., 2022). Thus, due to the differences in drug resistance genetic determinants among *Mycobacterium* species, additional efforts are needed to determine drug susceptibility by using molecular approaches in NTM such as *M. brumae*.

The comparative analysis of immunogenic T cell and B cell epitopes highlights that *M. brumae* is sharing very few antigens with BCG and *M. tuberculosis* and it can still trigger an immune response to similar levels than BCG in preclinical studies related to bladder cancer immunotherapy (summarized in Figure 6). *Mycobacterium brumae* might have other immunogenic epitopes that are not described in *M. tuberculosis*, and finding those would provide a better knowledge of *M. brumae* immunogenic profile. Interestingly, some of the shared genes are critical immunogenic heat-shock proteins (*groEL*, *groES*, or *dnaK*) involved in the immunostimulatory ability of BCG (Qazi et al., 2005). Heat-shock proteins have been the focus of numerous studies as potential immunotherapy tools in different types of cancer (Das et al., 2019). Purified heat-shock proteins from mycobacteria species or BCG overexpressing diverse heat-shock proteins have been studied as adjuvants for the treatment of different cancers with some success (reviewed in Noguera-Ortega et al., 2020). Therefore, the high *M. brumae* immune potential could be related to a higher expression of antigens or the presence of alternative immunogenic molecules. Another source of immunogenic molecules is the mycobacterial cell wall. The cell wall biosynthesis genes found in the *M. brumae* genome are in accordance with the cell wall composition described previously for this species (Guallar-Garrido et al., 2021; Supplementary Figure S4). The hallmark of mycobacteria is their unique abundance of lipids and glycolipids in their cell wall (Chiaradia et al., 2017; Minnikin and Brennan, 2020), which include the exceptionally long chain fatty acids (mycolic acids, MA). Mycolic acids are useful taxonomic markers, and they are related to host-mycobacteria interaction. *Mycobacterium brumae* is one of the few species that only contains α-mycolic acids (type I; Luquin et al., 1993; Supplementary Figure S4). Like most mycobacteria species, *M. brumae* contains TMM, TDM, PIM, and PL (Guallar-Garrido et al., 2021). However, hydrophilic glycopeptidolipids (GPLs) or lipooligosaccharides (LOS), present in other NTM are absent in *M. brumae*. Unlike *M. bovis* BCG, *M. brumae* does not contain lipids related to virulence like highly hydrophobic PDIM or their related PGLs, or sulfolipid (SL) and di-acyl- and poly-acyl trehaloses (DAT, PAT) present in the *M. tuberculosis* cell wall (Minnikin et al., 2002; Supplementary Figure S4).

Apart from the reference strain, other three *M. brumae* strains are available. The type *M. brumae* strain (ATCC 51384^T initially

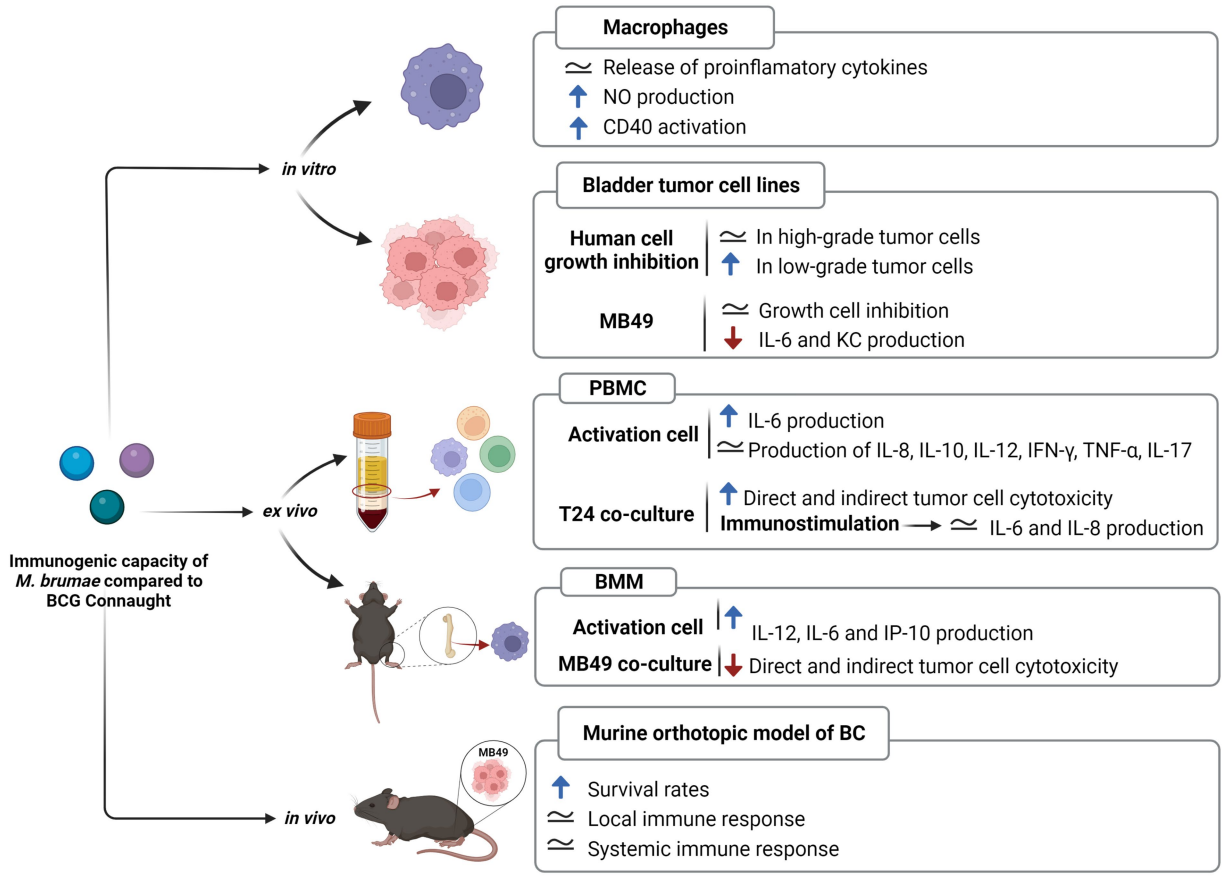


FIGURE 6
Immunogenic capacity of *Mycobacterium brumae* compared to *Mycobacterium bovis* BCG Connaught. Summary of the results obtained in studies treating with both mycobacteria J774 macrophages cell line (Noguera-Ortega et al., 2016b), bladder cancer cell lines (Noguera-Ortega et al., 2016a), human peripheral mononuclear cells (Noguera-Ortega et al., 2016b), and murine bone-marrow macrophages (unpublished); and after treating intravesically with each mycobacteria bladder cancer tumor-bearing mice (Noguera-Ortega et al., 2016a,b, 2018). Blue rising arrow indicates higher than *M. bovis* BCG Connaught, red falling arrow indicates lower than *M. bovis* BCG Connaught and the equals symbol indicates similar to *M. bovis* BCG Connaught.

named CR270), and CR103 and CR142 strains were isolated from water samples taken from the Llobregat River in Spain between 1983 and 1987, and CR269 from the soil between 1989 and 1991 (Luquin et al., 1993). Despite coming from slightly different environmental sources due to a different timeframe, all four strains showed similar *in vitro* genotypes and high genomic similarity. Phenotypically, they showed the same colony morphology when grown in solid media (Supplementary Figure S5), a similar pattern of lipids and glycolipids on the cell wall, and the same susceptibility pattern to a wide range of antibiotics. At the genome level, they showed a maximum of six SNPs and seven regions of difference, which is a very low genomic diversity for different environmental strains. Other mycobacteria, such as *M. abscessus* have shown very low genetic diversity across wide temporal and geographical scales too. Person to person transmission has been refuted to be the reason for the low genetic diversity in *M. abscessus*, and a lower mutation rate due to differences in a repair mechanism has been proposed as the most probable explanation (Commins et al., 2022).

In summary, the *M. brumae* diversity analysis demonstrated that *M. brumae* is genetically less diverse than expected from the diverse origin of the strains compared. The gene composition reported reveals that *M. brumae* interacts differently with the host immune system compared to host-specialized mycobacterial groups, such as *M. bovis* BCG or *M. tuberculosis*. Our research not only reveals the variety of genomic features among nontuberculous mycobacteria species, but it also provides an insight into finding critical antigenic molecules that could be the basis for designing bacteria-based therapies to combat cancer or other immune-based disorders.

Data availability statement

The datasets presented in this study can be found in online repositories. *M. brumae* reference genome and annotation is under the BioProject code PRJNA798885 with accession number CP104302. Illumina reads are included in the BioProject code

PRJEB52012 with accession numbers ERR9463983 (CR103), ERR9463984 (CR142), and ERR9463985 (CR269). The names of the repository/repositories and accession number(s) of previously published genomes can be found in the article/Supplementary material.

Author contributions

CR-M: formal analysis, investigation, visualization, and writing original draft. PH-A: investigation, visualization, and writing original draft. VS: data curation and review and edit draft. PR-R: data curation, formal analysis, investigation, visualization, software, and review and edit draft. ET: resources and review and edit draft. ÁC-O and MT-P: investigation and review and edit draft. IC: resources, project administration, and review and edit draft. EJ and MC: resources, funding, conceptualization, project administration, supervision, and writing original draft. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Spanish Ministry of Science, Innovation, and Universities (MICIN) and co-funded by the FEDER Funds (RTI2018-098777-B-I00, RTI2018-094399-A-I00, PID2021-125801OB-I00, PID2021-122331OB-I00, and PID2021-123443OB-I00), the AGAUR-Generalitat of Catalunya (2017SGR-229 and 2017SGR-1079), Ramón y Cajal fellowship (RYC-2015-18213 MICIN), FPI fellowship (PRE-2019-088141 MICIN), Generalitat Valenciana (SEJI/2019/011), MycoNET (RED2018-102677-T), and the European Commission—NextGenerationEU (Regulation EU 2020/2094), through CSIC's Global Health Program (PTI+ Salud Global).

Acknowledgments

We acknowledge the access to Centro de Investigaciones Príncipe Felipe (CIPF) scientific computing core facility in Valencia. Figures 1, 6, and Supplementary Figure S4 were created with Biorender.com.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.982679/full#supplementary-material

References

Alcaide, F., Esteban, J., González-Martin, J., and Palacios, J.-J. (2017). Methods for determining the antimicrobial susceptibility of mycobacteria. *Enfermedades Infecc. Microbiol. Clin. Engl. Ed* 35, 527–533. doi: 10.1016/j.eimce.2017.08.005

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bach-Griera, M., Campo-Pérez, V., Barbosa, S., Traserra, S., Guallar-Garrido, S., Moya-Andérico, L., et al. (2020). Mycolicibacterium brumae is a safe and non-toxic immunomodulatory agent for cancer treatment. *Vaccine* 8:198. doi: 10.3390/vaccines8020198

Bachmann, N. L., Salamzade, R., Manson, A. L., Whittington, R., Sintchenko, V., Earl, A. M., et al. (2020). Key transitions in the evolution of rapid and slow growing mycobacteria identified by comparative genomics. *Front. Microbiol.* 10:3019. doi: 10.3389/fmicb.2019.03019

Bakuła, Z., Kościuch, J., Safianowska, A., Proboszcz, M., Bielecki, J., van Ingen, J., et al. (2018). Clinical, radiological and molecular features of mycobacterium kansasii pulmonary disease. *Respir. Med.* 139, 91–100. doi: 10.1016/j.rmed.2018.05.007

Benjak, A., Avanzi, C., Benito, Y., Breysse, F., Chartier, C., Boschiroli, M. L., et al. (2018). Highly reduced genome of the new species *mycobacterium uberis*, the causative agent of nodular Thelitis and Tuberculoid Scrotitis in livestock and a close relative of the leprosy bacilli. *mSphere* 3, e00405–e00418. doi: 10.1128/mSphere.00405-18

Brambilla, C., Sánchez-Chardi, A., Pérez-Trujillo, M., Julián, E., and Luquin, M. (2012). (2012). Cyclopropanation of α-mycolic acids is not required for cording in mycobacterium brumae and Mycobacterium fallax. *Microbiology* 158, 1615–1621. doi: 10.1099/mic.0.057919-0

Brown-Elliott, B. A., Nash, K. A., and Wallace, R. J. (2012). Antimicrobial susceptibility testing, drug resistance mechanisms, and therapy of infections with nontuberculous mycobacteria. *Clin. Microbiol. Rev.* 25, 545–582. doi: 10.1128/CMR.05030-11

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). Egg NOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Cantrell, S. A., Leavell, M. D., Marjanovic, O., Iavarone, A. T., Leary, J. A., and Riley, L. W. (2013). Free mycolic acid accumulation in the cell wall of the mce 1 operon mutant strain of mycobacterium tuberculosis. *J. Microbiol. Seoul Korea* 51, 619–626. doi: 10.1007/s12275-013-3092-y

Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., and Chong, Z. (2021). Accurate long-read de novo assembly evaluation with inspector. *Genome Biol.* 22:312. doi: 10.1186/s13059-021-02527-4

Cheng, G., Xu, D., Wang, J., Liu, C., Zhou, Y., Cui, Y., et al. (2017). Isolation and identification of multiple drug resistant nontuberculous mycobacteria from organs of cattle produced typical granuloma lesions. *Microb. Pathog.* 107, 313–316. doi: 10.1016/j.micpath.2017.03.047

Chiaradia, L., Lefebvre, C., Parra, J., Marcoux, J., Burlet-Schiltz, O., Etienne, G., et al. (2017). Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. *Sci. Rep.* 7:12807. doi: 10.1038/s41598-017-12718-4

Chiner-Oms, Á., Berney, M., Boinett, C., González-Candelas, F., Young, D. B., Gagneux, S., et al. (2019). Genome-wide mutational biases fuel transcriptional diversity in the mycobacterium tuberculosis complex. *Nat. Commun.* 10:3994. doi: 10.1038/s41467-019-11948-6

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, Snp eff. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Collins, D. M., Kawakami, R. P., Buddle, B. M., Wards, B. J., and de Lisle, G. W. (2003). Different susceptibility of two animal species infected with isogenic mutants of Mycobacterium bovis identifies pho T as having roles in tuberculosis virulence and phosphate transport. *Microbiol. Read. Engl.* 149, 3203–3212. doi: 10.1099/mic.0.26469-0

Commins, N., Sullivan, M. R., McGowen, K., Koch, E., Rubin, E. J., and Farhat, M. (2022). Mutation rates and adaptive variation among the clinically dominant clusters of *Mycobacterium abscessus*. bioRxiv [Preprint]. doi: 10.1101/2022.04.27.489597

Coros, A., Callahan, B., Battaglioli, E., and Derbyshire, K. M. (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in mycobacterium smegmatis. *Mol. Microbiol.* 69, 794–808. doi: 10.1111/j.1365-2958.2008.06299.x

Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., et al. (2021). Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. *Microb. Genomics* 7:477. doi: 10.1099/mgen.0.000477

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 43:e15. doi: 10.1093/nar/gku1196

D'Auria, G., Torrents, E., Luquin, M., Comas, I., and Julián, E. (2016). Draft genome sequence of mycobacterium brumae ATCC 51384. *Genome Announc.* 4, e00237–e00246. doi: 10.1128/genomeA.00237-16

Das, J. K., Xiong, X., Ren, X., Yang, J.-M., and Song, J. (2019). Heat shock proteins in cancer immunotherapy. *J. Oncol.* 2019:e3267207. doi: 10.1155/2019/3267207

Falkinham, J. O. (2009). Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *J. Appl. Microbiol.* 107, 356–367. doi: 10.1111/j.1365-2672.2009.04161.x

Fedrizzi, T., Meehan, C. J., Grottola, A., Giacobazzi, E., Fregni Serpini, G., Tagliazucchi, S., et al. (2017). Genomic characterization of nontuberculous mycobacteria. *Sci. Rep.* 7:45258. doi: 10.1038/srep45258

Fenn, K., Wong, C. T., and Darbari, V. C. (2019). Mycobacterium tuberculosis uses Mce proteins to interfere with host cell signaling. *Front. Mol. Biosci.* 6:149. doi: 10.3389/fmolb.2019.00149

Fishbein, S., van Wyk, N., Warren, R. M., and Sampson, S. L. (2015). Phylogeny to function: PE/PPE protein evolution and impact on mycobacterium tuberculosis pathogenicity. *Mol. Microbiol.* 96, 901–916. doi: 10.1111/mmi.12981

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

Gcebe, N., Michel, A., Gey van Pittius, N. C., and Rutten, V. (2016). Comparative genomics and proteomic analysis of four non-tuberculous mycobacterium species and mycobacterium tuberculosis complex: occurrence of shared immunogenic proteins. *Front. Microbiol.* 7:795. doi: 10.3389/fmicb.2016.00795

Gey van Pittius, N. C., Sampson, S. L., Lee, H., Kim, Y., van Helden, P. D., and 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. doi: 10.1186/1471-2148-6-95

Gray, T. A., and Derbyshire, K. M. (2018). Blending genomes: distributive conjugal transfer in mycobacteria, a sexier form of HGT. *Mol. Microbiol.* 108, 601–613. doi: 10.1111/mmi.13971

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Guallar-Garrido, S., Campo-Pérez, V., Pérez-Trujillo, M., Cabrera, C., Senserrich, J., Sánchez-Chardi, A., et al. (2022). Mycobacterial surface characters remodeled by growth conditions drive different tumor-infiltrating cells and systemic IFN-γ/IL-17 release in bladder cancer treatment. *Onco Immunol.* 11:2051845. doi: 10.1080/2162402X.2022.2051845

Guallar-Garrido, S., and Julián, E. (2020). Bacillus Calmette-Guerin (BCG) therapy for bladder cancer: an update. *Immuno Targets Ther.* 9, 1–11. doi: 10.2147/ITT.S202006

Guallar-Garrido, S., Luquin, M., and Julián, E. (2021). Analysis of the lipid composition of mycobacteria by thin layer chromatography. *J. Vis. Exp.* 170:e62368. doi: 10.3791/62368

Gupta, R. S., Lo, B., and 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. doi: 10.3389/fmicb.2018.00067

Hsu, T., Hingley-Wilson, S. M., Chen, B., Chen, M., Dai, A. Z., Morin, P. M., et al. (2003). The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl. Acad. Sci.* 100, 12420–12425. doi: 10.1073/pnas.1635213100

Huh, H. J., Kim, S.-Y., Jhun, B. W., Shin, S. J., and Koh, W.-J. (2019). Recent advances in molecular diagnostics and understanding mechanisms of drug resistance in nontuberculous mycobacterial diseases. *Infect. Genet. Evol.* 72, 169–182. doi: 10.1016/j.meegid.2018.10.003

IEDB (2020). Free epitope database and prediction resource. Available at: http://www.iedb.org (Accessed June 28, 2022).

Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., et al. (2004). The complete genomic sequence of Nocardia farcinica IFM 10152. *Proc. Natl. Acad. Sci.* 101, 14925–14930. doi: 10.1073/pnas.0406410101

Iwao, Y., and Nakata, N. (2018). Roles of the three mycobacterium smegmatis kat G genes for peroxide detoxification and isoniazid susceptibility. *Microbiol. Immunol.* 62, 158–167. doi: 10.1111/1348-0421.12574

Jia, X., Yang, L., Li, C., Xu, Y., Yang, Q., and Chen, F. (2021). Combining comparative genomic analysis with machine learning reveals some promising diagnostic markers to identify five common pathogenic non-tuberculous mycobacteria. *Microb. Biotechnol.* 14, 1539–1549. doi: 10.1111/1751-7915.13815

Kim, B.-J., Cha, G.-Y., Kim, B.-R., Kook, Y.-H., and Kim, B.-J. (2019). Insights from the genome sequence of mycobacterium paragordonae, a potential novel live vaccine for preventing mycobacterial infections: the putative role of type VII secretion Systems for an Intracellular Lifestyle within Free-Living Environmental Predators. *Front. Microbiol.* 10:1524. doi: 10.3389/fmicb.2019.01524

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Larsen, M. H., Vilchèze, C., Kremer, L., Besra, G. S., Parsons, L., Salfinger, M., et al. (2002). Overexpression of inh a, but not kas a, confers resistance to isoniazid and ethionamide in mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. *Mol. Microbiol.* 46, 453–466. doi: 10.1046/j.1365-2958.2002.03162.x

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Luquin, M., Ausina, V., Vincent-Lévy-Frébault, V., Lanéelle, M. A., Belda, F., Garcí Abarceló, M., et al. (1993). Mycobacterium brumae sp. nov., a rapidly growing, nonphotochromogenic mycobacterium. *Int. J. Syst. Evol. Microbiol.* 43, 405–413. doi: 10.1099/00207713-43-3-405

Matsumoto, Y., Kinjo, T., Motooka, D., Nabeya, D., Jung, N., Uechi, K., et al. (2019). Comprehensive subspecies identification of 175 nontuberculous mycobacteria species based on 7547 genomic profiles. *Emerg. Microbes Infect.* 8, 1043–1053. doi: 10.1080/22221751.2019.1637702

Meehan, C. J., Barco, R. A., Loh, Y.-H. E., Cogneau, S., and Rigouts, L. (2021). Reconstituting the genus mycobacterium. *Int. J. Syst. Evol. Microbiol.* 71:004922. doi: 10.1099/ijsem.0.004922

Minnikin, D. E., and Brennan, P. J. (2020). "Lipids of clinically significant mycobacteria" in *Health Consequences of Microbial Interactions With Hydrocarbons, Oils, and Lipids Handbook of Hydrocarbon and Lipid Microbiology*. ed. H. Goldfine (Cham: Springer International Publishing), 1–76.

Minnikin, D. E., Kremer, L., Dover, L. G., and Besra, G. S. (2002). The methyl-branched fortifications of mycobacterium tuberculosis. *Chem. Biol.* 9, 545–553. doi: 10.1016/S1074-5521(02)00142-4

Mukhopadhyay, S., and Balaji, K. N. (2011). The PE and PPE proteins of mycobacterium tuberculosis. *Tuberculosis* 91, 441–447. doi: 10.1016/j.tube.2011.04.004

Newton-Foot, M., Warren, R. M., Sampson, S. L., van Helden, P. D., and Gey van Pittius, N. C. (2016). The plasmid-mediated evolution of the mycobacterial ESX (type VII) secretion systems. *BMC Evol. Biol.* 16:62. doi: 10.1186/s12862-016-0631-2

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Noguera-Ortega, E., Blanco-Cabra, N., Rabanal, R. M., Sánchez-Chardi, A., Roldán, M., Guallar-Garrido, S., et al. (2016a). Mycobacteria emulsified in olive oil-in-water trigger a robust immune response in bladder cancer treatment. *Sci. Rep.* 6:27232. doi: 10.1038/srep27232

Noguera-Ortega, E., Guallar-Garrido, S., and Julián, E. (2020). Mycobacteria-based vaccines as immunotherapy for non-urological cancers. *Cancer* 12:1802. doi: 10.3390/cancers12071802

Noguera-Ortega, E., Rabanal, R. M., Gómez-Mora, E., Cabrera, C., Luquin, M., and Julián, E. (2018). Intravesical mycobacterium brumae triggers both local and systemic immunotherapeutic responses against bladder cancer in mice. *Sci. Rep.* 8:15102. doi: 10.1038/s41598-018-33253-w

Noguera-Ortega, E., Rabanal, R. M., Secanella-Fandos, S., Torrents, E., Luquin, M., and Julián, E. (2016c). γ irradiated mycobacteria enhance survival in bladder tumor bearing mice although less efficaciously than live mycobacteria. *J. Urol.* 195, 198–205. doi: 10.1016/j.juro.2015.07.011

Noguera-Ortega, E., Secanella-Fandos, S., Eraña, H., Gasión, J., Rabanal, R. M., Luquin, M., et al. (2016b). Nonpathogenic mycobacterium brumae inhibits bladder cancer growth in vitro, ex vivo, and in vivo. *Eur. Urol. Focus* 2, 67–76. doi: 10.1016/j.euf.2015.03.003

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pym, A. S., Brodin, P., Majlessi, L., Brosch, R., Demangel, C., Williams, A., et al. (2003). Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat. Med.* 9, 533–539. doi: 10.1038/nm859

Qazi, K. R., Qazi, M. R., Julián, E., Singh, M., Abedi-Valugerdi, M., and Fernández, C. (2005). Exposure to mycobacteria primes the immune system for evolutionarily diverse heat shock proteins. *Infect. Immun.* 73, 7687–7696. doi: 10.1128/IAI.73.11.7687-7696.2005

Qian, J., Chen, R., Wang, H., and Zhang, X. (2020). Role of the PE/PPE family in host–pathogen interactions and prospects for anti-tuberculosis vaccine and diagnostic tool design. *Front. Cell. Infect. Microbiol.* 10:594288. doi: 10.3389/fcimb.2020.594288

Rahman, S. A., Singh, Y., Kohli, S., Ahmad, J., Ehtesham, N. Z., Tyagi, A. K., et al. (2014). Comparative analyses of nonpathogenic, opportunistic, and totally pathogenic mycobacteria reveal genomic and biochemical variabilities and highlight the survival attributes of mycobacterium tuberculosis. *MBio* 5, e02020–e02014. doi: 10.1128/mBio.02020-14

Ripoll, F., Pasek, S., Schenowitz, C., Dossat, C., Barbe, V., Rottman, M., et al. (2009). Non mycobacterial virulence genes in the genome of the emerging pathogen mycobacterium abscessus. *PLoS One* 4:e5660. doi: 10.1371/journal.pone.0005660

Saini, N. K., Sharma, M., Chandolia, A., Pasricha, R., Brahmachari, V., and Bose, M. (2008). Characterization of Mce 4A protein of mycobacterium tuberculosis: role in invasion and survival. *BMC Microbiol.* 8:200. doi: 10.1186/1471-2180-8-200

Sassi, M., and Drancourt, M. (2014). Genome analysis reveals three genomospecies in mycobacterium abscessus. *BMC Genomics* 15:359. doi: 10.1186/1471-2164-15-359

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Silva, F. J., Santos-Garcia, D., Zheng, X., Zhang, L., and Han, X. Y. (2022). Construction and analysis of the complete genome sequence of leprosy agent *mycobacterium lepromatosis*. *Microbiol. Spectr.* 10:e0169221. doi: 10.1128/spectrum.01692-21

Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., and Husnik, F. (2022). Pseudofinder: detection of pseudogenes in prokaryotic genomes. *Mol. Biol. Evol.* 39:msac153. doi: 10.1093/molbev/msac153

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

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Tonkin-Hill, G., Mac Alasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4

van Ingen, J., Boeree, M. J., van Soolingen, D., and Mouton, J. W. (2012). Resistance mechanisms and drug susceptibility testing of nontuberculous mycobacteria. *Drug Resist. Updat.* 15, 149–161. doi: 10.1016/j.drup.2012.04.001

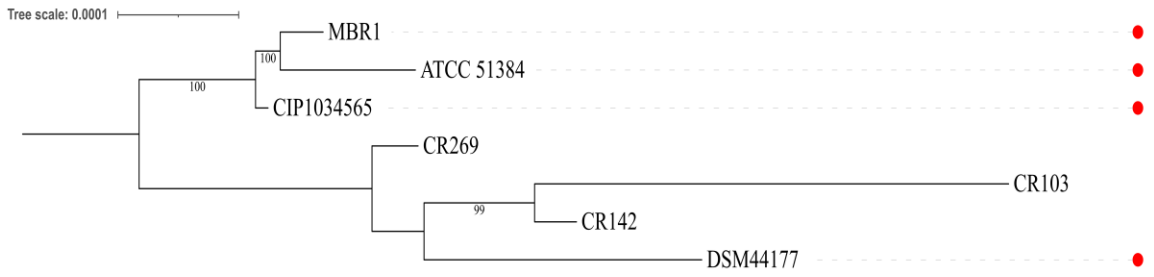
WHO Catalogue of mutations in mycobacterium tuberculosis complex and their association with drug resistance. (2021). Available at: https://www.who.int/publications-detail-redirect/9789240028173 (Accessed June 28, 2022).

Yang, T., Gan, M., Liu, Q., Liang, W., Tang, Q., Luo, G., et al. (2022). SAM-TB: a whole genome sequencing data analysis website for detection of mycobacterium tuberculosis drug resistance and transmission. *Brief. Bioinform.* 23:bbac030. doi: 10.1093/bib/bbac030

Supplementary Figure S1. Phylogenetic relationship of the available *M. brumae* genomic sequences. A) Maximum likelihood phylogenetic tree based on an alignment of the core-genome with a length of 293,792 bp and including *M. fallax* as outgroup. GTR model was used with bootstrap confidence of 1000 replicates. B) Zoom-in to the *M. brumae* cluster of the previous tree showed in A. Red dots indicate genomic sequences of the type strain (same as ATCC 51384).

A)

B)



Supplementary Figure S2. Protein alignments for the analysis of drug-resistance mutations. Query represents *M. tuberculosis* H37Rv and Sbjct represents *M. brumae*. All *M. brumae* mutations are indicated in the alignment. **A)** RpoB protein alignment for rifampicin resistance analysis. **B)** KatG protein alignment for isoniazid resistance analysis.

A)

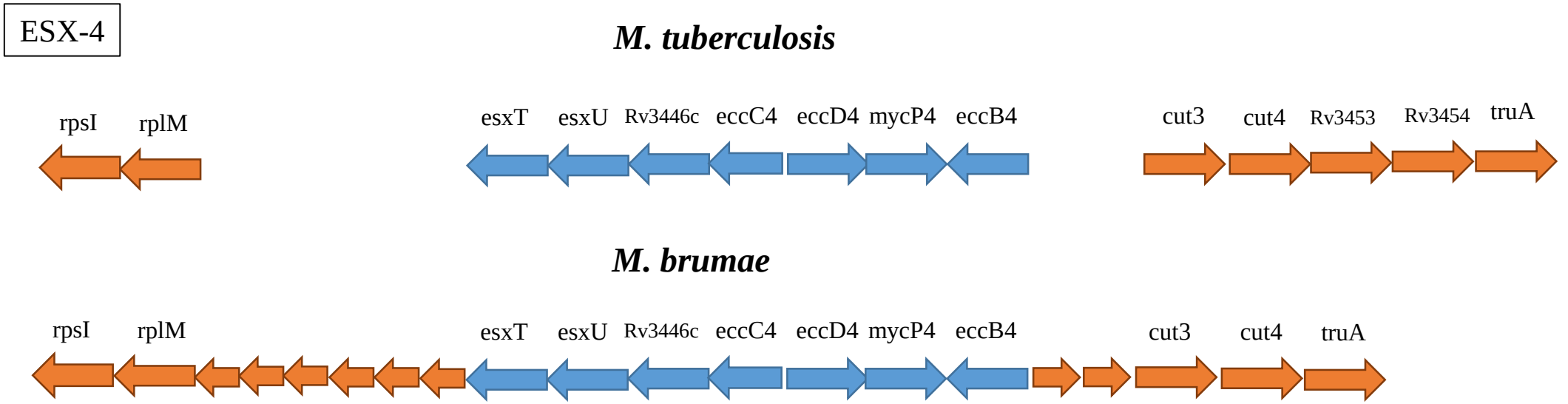
Score	Expect	Method	Identities	Positives	Gaps
2151 bits(5574)	0.0	Compositional matrix adjust.	1059/1179(90%)	1113/1179(94%)	18/1179(1%)

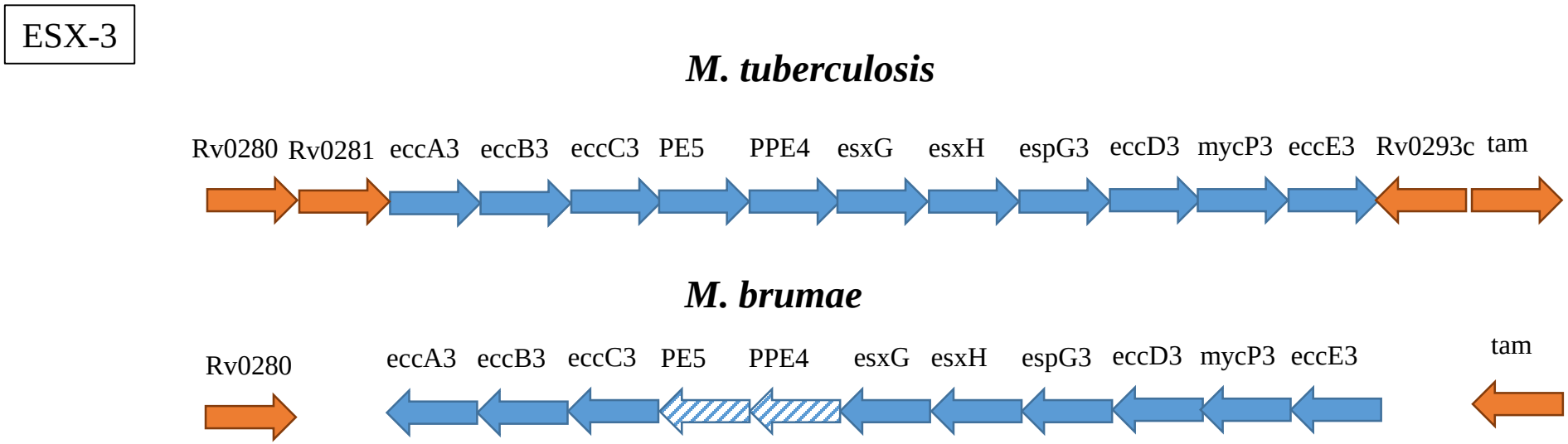
```
Query  1     LADSRQSKTAASPSPSRPQSSSNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIG  60
             LA SRQSK+ A+           NSVPGAPNR+SFAKLREPLEVPGLLDVQTDSFEWLIG
Sbjct  7     LAVSRQSKSVAT-----------NSVPGAPNRISFAKLREPLEVPGLLDVQTDSFEWLIG  55

Query  61    SPRWRESAAERGDV-NPVGGLEEVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDK  119
             S +WR SA+ RGD+ NP GGLEEVL ELSPIEDFSGSMSLSFSDPRFD+VKAPV ECKDK
Sbjct  56    SDKWRASASARGDIANPKGGLEEVLDELSPIEDFSGSMSLSFSDPRFDEVKAPVAECKDK  115

Query  120   DMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGV  179
             DMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGV
Sbjct  116   DMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGV  175

Query  180   YFDETIDKSTDKTLHSVKVIPSRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTS  239
             YFDE+IDKST+KTLHSVKVIP RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTS
Sbjct  176   YFDESIDKSTEKTLHSVKVIPGRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTS  235

Query  240   EQIVERFGFSEIMRSTLEKDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKR  299
             EQI ERFGFSEIM  TLEKD T G DEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKR
Sbjct  236   EQIRERFGFSEIMMGTLEKDPTAGPDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKR  295

Query  300   YDLARVGRYKVNKKLGLHVGEPI--TSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEV  357
             YDLARVGRYKVNKKLGL++ +P+  +S+TLTEED+VATIEYLVRLHEGQTTMT PGG EV
Sbjct  296   YDLARVGRYKVNKKLGLNLDKPVNTSSNTLTEEDIVATIEYLVRLHEGQTTMTAPGGSEV  355

Query  358   PVETDDIDHFGNRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPV  417
             PVE DDIDHFGNRRLRTVGELIQNQIRVG+SRMERVVRERMTTQDVEAITPQTLINIRPV
Sbjct  356   PVEVDDIDHFGNRRLRTVGELIQNQIRVGLSRMERVVRERMTTQDVEAITPQTLINIRPV  415

Query  418   VAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRM  477
             VAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVH SHYGRM
Sbjct  416   VAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHSSHYGRM  475

Query  478   CPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQ  537
             CPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKV+DGV++D+I YLTADEEDRHVVAQ
Sbjct  476   CPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVIDGVITDQIDYLTADEEDRHVVAQ  535

Query  538   ANSPI-DADGR--FVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHD  594
             ANSP  + DG   F + RVLVRRK GEVE+V ++EVDYMDVSPRQMVSVATAMIPFLEHD
Sbjct  536   ANSPTAEGDGGLVFTDDRVLVRRKGGEVEFVSAAEVDYMDVSPRQMVSVATAMIPFLEHD  595

Query  595   DANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYITV  654
             DANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVV+E++GV+EEVSADYITV
Sbjct  596   DANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVSEKAGVVEEVSADYITV  655

Query  655   MHDNGTRRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLL  714
             M D+GTR+TYRMRKFARSNHGTCANQ PIVDAG RVE+GQV+ADGPCT++GEMALGKNLL
Sbjct  656   MADDGTRQTYRMRKFARSNHGTCANQRPIVDAGQRVESGQVVADGPCTENGEMALGKNLL  715

Query  715   VAIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNISDEV  774
             VA+MPWEGHNYEDAIILS RLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPN+SDEV
Sbjct  716   VAVMPWEGHNYEDAIILSQRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNVSDEV  775

Query  775   LADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVP  834
             LADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVP
Sbjct  776   LADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVP  835

Query  835   HGESGKVIGIRVFSREDEDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKILP  894
             HGESGKVIGIRVFSRED+DELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKILP
Sbjct  836   HGESGKVIGIRVFSREDDDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKILP  895

Query  895   VEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKG-VPDWAARL  953
             +EDMPFL DGTPVDIILNTHGVPRRMNIGQILETHLGW A +GW V  A G +P+WA  L
Sbjct  896   IEDMPFLPDGTPVDIILNTHGVPRRMNIGQILETHLGWVAKAGWNVQVAGGEIPEWAGNL  955

Query  954   PDELLEAQPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLFDGRSGEPF  1013
             P++LL A  ++IVSTPVFDGA+E ELQGLLS TLPNRDG+V+VD DGKA LFDGRSGEPF
Sbjct  956   PEDLLSAPSDSIVSTPVFDGAREEELQGLLSSTLPNRDGEVMVDGDGKARLFDGRSGEPF  1015

Query  1014  PYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQAY  1073
             PYPVTVGYMYI+KLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQAY
Sbjct  1016  PYPVTVGYMYILKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQAY  1075

Query  1074  GAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEVLS  1133
             GAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEVLS
Sbjct  1076  GAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEVLS  1135

Query  1134  SDGAAIELREGEDEDLERAAANLGINLSRNESASVEDLA  1172
             SDGAAIE+R+G+DEDLERAAANLGINLSRNESASVEDLA
Sbjct  1136  SDGAAIEMRDGDDEDLERAAANLGINLSRNESASVEDLA  1174
```

B)

Score	Expect	Method	Identities	Positives	Gaps
972 bits(2513)	0.0	Compositional matrix adjust.	488/745(66%)	576/745(77%)	18/745(2%)

```
Query  1    VPEQHPPITETTTGAASNGCPVVGHMKYPVEGGGNQDWWPNRLNLKVLHQNPAVADPMGA  60
            VPE+ PPI E  T    +GCP+   +K PVEGG N+DWWPN++NLK+L +NP + DP
Sbjct  1    VPEETPPIGEAQT---ESGCPM--RIKPPVEGGSNRDWWPNQVNLKILQKNPDIIDPEDE  55

Query  61   AFDYAAEVATIDVDALTRDIEEVMTTSQPWWPADYGHYGPLFIRMAWHAAGTYRIHDGRG  120
             +DY   V T+D +    D + ++T SQ WWPAD+GHYGPLF+RM+WHAAGTYR+ DGRG
Sbjct  56   GYDYRQVVQTLDFEEFQADFDALLTDSQSWWPADFGHYGPLFVRMSWHAAGTYRVQDGRG  115

Query  121  GAGGGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKKLSWADLIVFAGNCALESMGFKT  180
            GAG GMQRF PLNSWPDN  LD+ARRLLWP+KKKYG K+SWADLI +AGN A+E MGFKT
Sbjct  116  GAGRGMQRFEPLNSWPDNVLLDQARRLLWPLKKKYGNKISWADLIAYAGNRAMEHMGFKT  175

Query  181  FGFGFGRVDQWEPDE-VYWGKEATWLGDE-RYSG--KRDLENPLAAVQMGLIYVNPEGPN  236
             GF FGR D WEP+E V+WG EA WLG + RY G  +  L+NPLAA  MGLIYVNPEGP
Sbjct  176  AGFAFGRADCWEPEEDVFWGAEAEWLGSQDRYQGSDRTKLDNPLAATMMGLIYVNPEGPE  235

Query  237  GNPDPMAAAVDIRETFRRMAMNDVETAALIVGGHTFGKTHGAGPADLVGPEPEAAPLEQM  296
            G PDP+AAA+DIRETF RMAMNDVETAALIVGGHTFGKTHG G A+ +GPEP AAPL++M
Sbjct  236  GVPDPLAAAIDIRETFGRMAMNDVETAALIVGGHTFGKTHGNGDAEALGPEPAAAPLQEM  295

Query  297  GLGWKSSYGTGTGKDAITSGIEVVWTNTPTKWDNSFLEILYGYEWELTKSPAGAWQYTAK  356
            GLGWK+   TG   D + SG+EV+WT+TPTKWDNSFLEILY  EWEL KS AGA Q+  K
Sbjct  296  GLGWKNPNDTGNPNDRVGSGLEVIWTHTPTKWDNSFLEILYSNEWELFKSKAGAQQWRPK  355

Query  357  DGAGAGTIPDPFGGPGRSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFAKAWYKLI  416
            D   A ++P P       P ML TDLS+R DPIY +ITRRWL+HP+ELA+EFAKAW+KL+
Sbjct  356  DNGWANSVPTPDLKGRTHPAMLTTDLSMREDPIYGKITRRWLDHPDELAEEFAKAWFKLM  415

Query  417  HRDMGPVARYLGPLVPKQTLLWQDPVPAVSHDLVGEAEIASLKSQIRASGLTVSQLVSTA  476
            HRDMGP  RYLGP VPK T +WQDPVPA + +L  +A++A+LK+ I  SGLTV QLVSTA
Sbjct  416  HRDMGPAVRYLGPFVPKDTWVWQDPVPAGNANL-SDADVAALKAAIADSGLTVPQLVSTA  474

Query  477  WAAASSFRGSDKRGGANGGRIRLQPQVGWEVNDPDGDLRKVIRTLEEIQESFNSAAPGNI  536
            W AA+S+R SD RGGANGGRIRLQPQ+GWE N+PD +L +VIR LEEIQ S        +
Sbjct  475  WKAAASYRNSDMRGGANGGRIRLQPQLGWESNEPD-ELAQVIRKLEEIQAS------SGV  527

Query  537  KVSFADLVVLGGCAAIEKAAKAAGHNITVPFTPGRTDASQEQTDVESFAVLEPKADGFRN  596
             VSFAD+VVL G   +EKAAKAAG +I VPFTPGR DA+QE TD +SF+ LEPKADGFRN
Sbjct  528  NVSFADVVVLAGNVGVEKAAKAAGFDIDVPFTPGRGDATQEMTDADSFSYLEPKADGFRN  587

Query  597  YLGKGNPLPAEYMLLDKANLLTLSAPEMTVLVGGLRVLGANYKRLPLGVFTEASESLTND  656
            Y GKG  LPAEY L+D+AN L LS PEMTVLVGGLR L AN+    LGV TE   +LT D
Sbjct  588  YAGKGLNLPAEYHLIDRANQLNLSGPEMTVLVGGLRALEANHGGSKLGVLTERPGALTTD  647

Query  657  FFVNLLDMGITWEPSPADDGTYQGKD-GSGKVKWTGSRVDLVFGSNSELRALVEVYGADD  715
            FFV++ DMG+ W PS ADDGTY G D  +G+ K+T SRVDL+FGSNS+LRAL EVY ADD
Sbjct  648  FFVSICDMGLKWAPSSADDGTYVGSDRATGEPKYTASRVDLLFGSNSQLRALAEVYAADD  707

Query  716  AQPKFVQDFVAAWDKVMNLDRFDVR  740
            A+ KFV+DFVAAW KVM+ DRFDV+
Sbjct  708  AKEKFVRDFVAAWTKVMDADRFDVK  732
```





Supplementary Figure S3. Comparison of the ESX-3 and ESX-4 clusters in *M. tuberculosis* H37Rv and *M. brumae*. Genes from the cluster are colored in blue and flanquing genes are colored in orange. Stripped arrows indicate the genes found with low similitude (under 70% protein identity).