Polish Journal of Veterinary Sciences Vol. 28, No. 4 (2025), 583–595

DOI 10.24425/pjvs.2025.156086

Original article

In vitro validation of *in silico*-selected targets for PCR detection of genus *Mycobacterium* and species *Mycobacterium marinum*

D. Zygała-Pytlos[1,2], M. Kuzioła[1,2], N. Wrońska[3], A. Zabost[4], M. Kozińska[4], E. Augustynowicz-Kopeć[4], J. Dziadek[1], A. Minias[1]

[1] Institute of Medical Biology, Polish Academy of Sciences, ul. Lodowa 106, 93-232, Lodz, Poland
[2] The Bio-Med-Chem Doctoral School of the University of Lodz and Lodz Institutes of the Polish Academy of Sciences, ul. Matejki 21/23, 90-237, Lodz, Poland
[3] Department of Industrial Microbiology and Biotechnology, Faculty of Biology and Environmental Protection, University of Lodz, ul. Banacha 12/16, 90-237, Lodz, Poland
[4] Department of Microbiology, National Tuberculosis and Lung Diseases Research Institute, ul. Płocka 26, Warsaw, Poland

Correspondence to: A. Minias, e-mail: aminias@cbm.pan.pl

Abstract

Mycobacterial infections pose significant diagnostic challenges due to the genetic diversity of species, limitations of current detection methods, and the need for rapid and accurate identification tools. In this study, we developed and validated a novel molecular approach for the specific detection of the *Mycobacterium* genus and the *Mycobacterium marinum* species based on the identification of unique DNA sequences. Using comparative genomic alignments and *in silico* screening of curated genomic databases, we identified a 391 bp region of the *mmpl* gene specific to the *Mycobacterium* genus, and a 202 bp region of the *espE*_2 gene specific to *M. marinum*. Primers were designed for both targets and validated for specificity using *in silico* BLAST analysis and in vitro PCR and qPCR assays. Experimental validation involved DNA from 52 bacterial isolates, including 44 *Mycobacterium* species and 6 *M. marinum* strains. The *mmpl* target showed a sensitivity of 95% and specificity of 100% for *Mycobacterium*, while the *espE_2* target achieved 100% sensitivity and specificity for *M. marinum*. We further demonstrated the applicability of our method using mock clinical samples spiked with bacteria and subjected to standard diagnostic workflows. Although qPCR sensitivity was reduced in complex matrices like sputum, likely due to DNA degradation and eukaryotic DNA interference, our method showed strong performance in buccal swabs and saliva. The assay offers a rapid, cost-effective, and adaptable alternative for the detection of mycobacteria, particularly in laboratories with limited resources. Future work will expand validation across a broader panel of strains and clinical specimens to enhance diagnostic confidence.

Keywords: *Mycobacterium, Mycobacterium marinum, mmpl gene, espE-2* gene, PCR

Introduction

Mycobacteria are Gram-positive, acid-fast aerobic bacilli that are diverse in terms of diseases, pathogenesis, and species they infect. Phylogenetically the genus contains over 190 species, divided into complexes and species (Meehan et al. 2021). The group consists of obligatory pathogens, opportunistic pathogens, and saprophytic bacteria. Mycobacterial infections are marked by a severe disease course and long-lasting treatment (Kanabalan et al. 2021, Bagcchi 2023). Mycobacterial infections in animals are usually not treated, and all fish or other flocks of animals infected by mycobacteria are euthanized (Biet and Boschiroli 2014, Francis-Floyd 2011).

The *Mycobacterium marinum*, belongs to *M. marinum complex (M. marinum, Mycobacterium shottsii, Mycobacterium pseudoshottsii, Mycobacterium ulcerans, Mycobacterium liflandii)* and is often detected in stagnant water fish tanks or non-chlorinated swimming pools. At least 20 non-tuberculous mycobacteria species are responsible for causing diseases in fish, however, the most common are *M. marinum, Mycobacterium chelonae*, and *Mycobacterium fortuitum* (Hashish et al. 2018). *M. marinum* can also cause human pulmonary, skin, and soft tissue diseases. About half of *M. marinum* infections in humans were reported to be aquarium-related, of which less than 30% were related to injuries caused by fish or shellfish and about 10% to injuries and contact with contaminated saltwater or brackish water (Strobel et al. 2022, Akram and Aboobacker 2023).

Molecular methods have revolutionized the identification of non-tuberculous mycobacteria (NTM) by targeting specific gene markers. These methods rely on the amplification and analysis of nucleic acid sequences. Molecular methods offer higher specificity and can differentiate closely related species. The PCR technique is widely used to amplify specific DNA region, such as 16S rRNA gene, internal transcribed spacer, *hsp65* gene, or *rpoB* gene. Following amplification, DNA sequencing or hybridization techniques are used to compare the obtained sequences with reference databases for species identification. The GenoType *Mycobacterium* common mycobacteria/additional species (CM/AS) assay (Hain Lifescience, Nehren Germany) is a commercial kit developed to differentiate and identify different species of NTM from bacteria cultures. It involves DNA amplification targeting the 23S rRNA gene region, followed by reverse hybridization to specific oligonucleotide probes immobilized on membrane strips (Singh et al. 2013). However, the GenoType *Mycobacterium* CM/AS assay makes rapid and accurate identification of NTM species, it requires a specialized set up and trained laboratory personnel (Chihota et al. 2010). Current methods for identifying *M. marinum* and other *Mycobacterium* species are predominantly time-consuming, labor-consuming, and expensive, and their results may be ambiguous. There are several molecular methods for identifying *M. marinum* species, but due to the genetic diversity of strains, they are not 100% sensitive. One of the methods is based on the amplification of the *hsp65* fragment sequence. In published studies, this assay confirmed 88% of infections of laboratory zebrafish (*Danio rerio*) caused by *M. marinum* (Peterson et al. 2013). There are also commercial tests based on PCR and hybridization of the 23S rRNA fragment or the ribosomal gene spacer (16S-23S), but they do not allow differentiation between the *M. marinum* and *M. ulcerans* species (Aubry et al. 2017). Published loop-mediated isothermal amplification (LAMP) assay for the *mrsA* fragment gene detection allows the identification of species belonging to the *M. marinum complex*. Other data indicate the identification of *Mycobacterium* genus in 11/24 tested cases (46%) by 16S rRNA amplification. The identification of the exact species in each probe was impossible because of the amplified fragment of gene, which was 100% identical for species from this group (Sia et al. 2016). Therefore, improving the laboratory diagnosis of mycobacterial infections is still necessary by using simple, quick, and sensitive methods.

In the current article, we present a new method of identification of the *Mycobacterium* genus and *M. marinum* species (Fig. 1). The method is based on the identification of specific sequences and employs infrastructure common in diagnostic laboratories around the world, PCR and qPCR. We identify sequences specific for mycobacteria and *M. marinum* and test the specificity of these sequences *in silico* and *in vitro* in experiments by applying the PCR technique for cultured bacteria and qPCR for mock samples.

Materials and Methods

Construction of virtual databases

Virtual databases of genomic sequences of isolates belonging to *M. marinum* species (n=13), *M. chelonae* (n=9), *Mycobacterium* genus (n=166), *Actinobacteria* phylum (a higher phylogenetic level to *Mycobacterium* genus) excluding mycobacterial species (n=340), and bacteria (n=6263) were constructed (Supplemental databases). To construct the databases, genomic sequences were downloaded from the The National Center for Biotechnology Information (NCBI) Genome Database via the Geneious Prime software (Biomatters, Auckland, New Zealand). The species identification was assigned *a priori* to the sequences deposited in the database. The quality control of the sequences, except

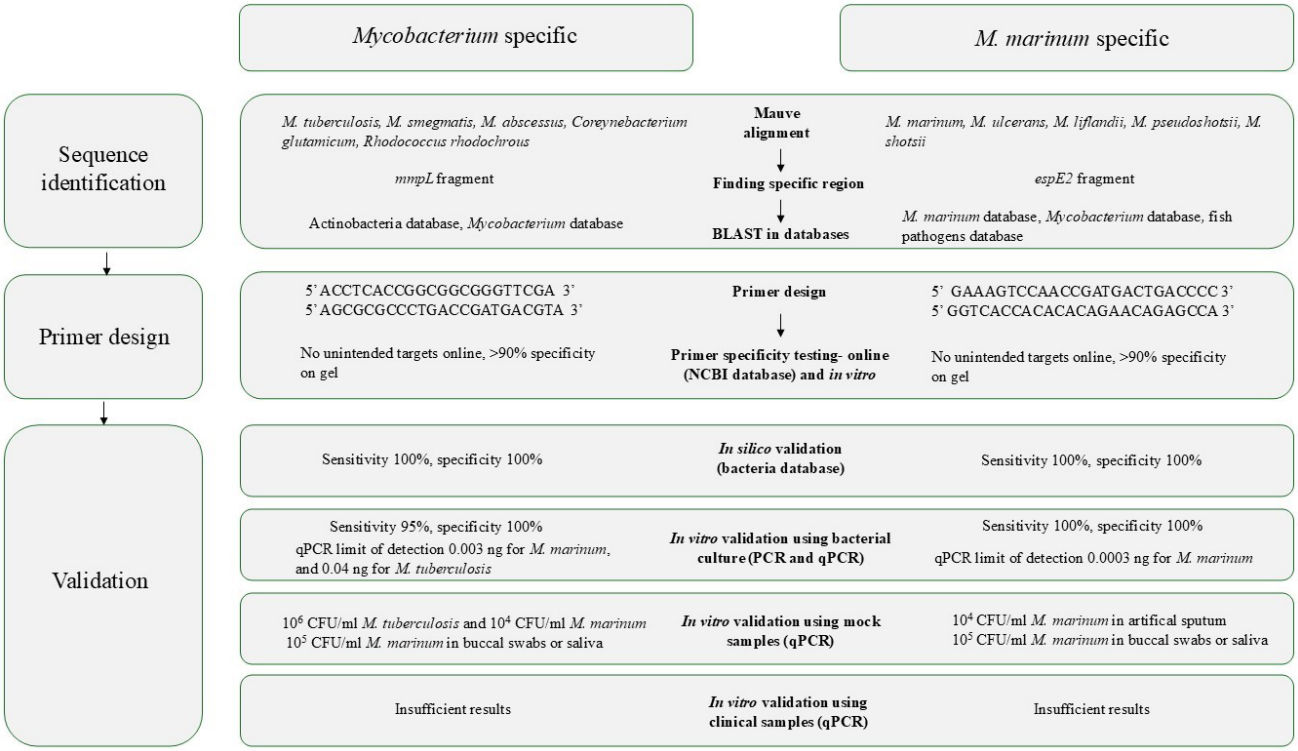

Fig. 1. Visual representation of the organization of work in this manuscript.

Table 1. Primers designed and used in this study.

Target	Forward primer	Reverse primer
espE_2 gene fragment	5' GAAAGTCCAACCG ATGACTGACCCC 3'	5' GGTCACCACACACA GAACAGAGCCA 3'
mmpl gene fragment	5' ACCTCACCGGCGGC GGGTTCGA 3'	5' AGCGCGCCCTGACC GATGACGTA 3'
16S rRNA gene fragment	5' CCTACGGGNGGCW GCAG 3'	5' GACTACHVGGGTAT CTAATCC 3'
YWHAZ gene fragment	5' TGCTTGCATCCCACA GACTA 3'	5' AGGCAGACAATGA CAGACCA 3'
B2M gene fragment	5' TGGGTTTCATCCATC CGACA 3'	5' TCAGTGGGGGTGAA TTCAGTG 3'

for the bacteria database, was based on the N50 value, which measures the number of contigs compared to genome size, and the sequences reached the 28 000 bp value acceptable in the scientific literature (Jauhal and Newcomb 2021). For the bacteria database, we used sequences marked as Reference Sequences.

Identification of *Mycobacterium* genus-specific sequence and *M. marinum*-specific sequence

The Geneious Prime was the primary bioinformatic software used in this study. For identification of *Mycobacterium*-specific sequence, we performed the Geneious in-built Mauve alignment of five species: *Mycobacterium tuberculosis* (NC_000962), *Mycobacterium smegmatis* (NC_008596), *Mycobacterium abscessus* (NC_010397), *Corynebacterium glutamicum* (NC_003450), *Rhodococcus rhodochrous* (NZ_CP032675). The alignment was manually curated, and we found a specific fragment for three *Mycobacterium* species – the fragment of the *mmpl* gene (Supplementary_sequences).

We used the BLAST tool from Geneious Prime (blastn, max e-value 1e-10) to search two databases with the sequence of the *mmpl* gene: 1) the *Actinobacteria* database composed of 340 genome sequences to confirm the lack of the sequence, and 2) the *Mycobacterium* database comprised of 166 genome sequences to confirm the presence of the sequence in *Mycobacterium* species (Table 1).

For *M. marinum* specific sequence, we ran the Mauve alignment of five closely phylogenetically related mycobacterial species: *M. marinum* (AP018496, CP000854), *M. ulcerans* (CP000325, CP092429), *M. liflandii* (CP003899, CP023138), *M. pseudoshotsii* (NZ_AP018410, NZ_AP026367), *M. shotsii* (NZ_AP022572, NZ_CP014860) (Coimbra et al. 2020), two strains of each species. We found the putative specific region for two *M. marinum* strains in the *espE_2* gene (Supplementary_sequences). The *espE_2* gene encodes

the secretion system ESX-1-associated protein EspE_2. The *M. marinum* ESX-1 secretion system is required for virulence and causes phagosomal lysis within macrophages (Chirakos et al. 2020). The system mediates sliding motility and is crucial for biofilm formation (Lai et al. 2018).

For *M. marinum* specific sequence, we used the BLAST tool from Geneious Prime (blastn, max e-value 1e-2), to search the constructed database of 13 *M. marinum* genome sequences for the presence of the specific region identified in Mauve alignment. Subsequently, we searched the 166 *Mycobacterium* genome sequences database to confirm the lack of indicated sequence in other species.

Primer specificity testing

The alignment of the *mmpl* genes across mycobacteria was used to design primers amplifying 391 bp *Mycobacterium-specific* sequence (Table 1) within the 2865 bp *mmpl* gene (Table 1, Supplementary sequences, Supplementary figures). The primers for the *espE_2* fragment were designed based on the alignment of sequences identified in the *M. marinum* database (Table 1) using the Geneious build-in feature. For *mmpl* gene fragment, primer sequence selection was forced by genetic variability within this gene (Supplementary figures).

The specificity of primers was determined using the Primer-BLAST tool (NCBI) (Supplementary primer specificity testing results). We selected a non-redundant database and 4000 bp as the maximum target amplicon size. Primer specificity stringency was set to at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. We ignored targets that have six or more mismatches to the primer.

Next, the specificity of primers was assessed on DNA gels with amplification products for mycobacteria and *M. marinum*. Exemplary specific and non-specific products were sequenced by Sagner sequencing, using same primers. The non-specific products yelled no sequencing results, while the sequences of specific products were confirmed. The intensity of bands on gels was assessed with ImageJ software (National Institutes of Health and the Laboratory for Optical and Computational Instrumentation, Wisconsin). The gels contained 0.4 µg/ml ethidium bromide.

Specificity was calculated as a percentage (Molina-Ruiz et al. 2024). We counted the pixels of proper PCR products and calculated the mean value. Next, we counted the pixels of non-specific PCR products and calculated the mean value. The sum of both values was 100%. Then the quantity of each fraction was calculated.

In vitro validation using DNA isolated from bacterial culture

We used clinical bacterial strains sourced from a collection of the Institute of Tuberculosis and Lung Diseases in Warsaw (clinical strains genotyping by GenoType Mycobacterium common mycobacteria/additional species (CM/AS) assay, Hain Lifescience, Nehren Germany), the University of Lodz (typing by biochemical assays), and the Institute of Medical Biology (purchased from American Type Culture Collection), Polish Academy of Sciences in Lodz (Supplementary_strains). To isolate DNA, we used the thermal method (van Embden et al. 1993, Doig et al. 2002). Firstly, solid bacterial cultures were suspended in 300 µl sterile water. The samples were incubated at 96ºC for 30 minutes and then centrifuged (14,100 g; 4ºC; 15 minutes). In the final step, 200 µl of supernatants were carried to new Eppendorf tubes. The quantity and purity of the obtained chromosomal DNA were verified by spectrophotometric measurement on NanoDrop ND 1000 V 3.5.2 (DeNovix).

First, we ran the polymerase chain reaction (PCR) with *Taq* polymerase (Ready mix, Thermo Scientific) and 50 ng of chromosomal DNA, in the Veriti thermal cycler (Applied Biosystems). The reaction consisted of initial denaturation (95ºC, 2 minutes), amplification (repeated 35 times), and final extension (72ºC, 7 minutes). The amplification reaction included denaturation (95ºC, 30 seconds), primer annealing (63 ºC for *Mycobacterium* and 61.5ºC for *M. marinum* specific sequences, respectively 30 seconds; multiplexing was not feasible), and extension (72ºC, 30 seconds). The final process was cooling (4 ºC, ∞). After PCR, we performed electrophoresis on 2% agarose gel with ethidium bromide and TAE buffer in the presence of 100 bp DNA Ladder (Fermentas). The products were separated by applying 120 V and visualized with Image Detection and Analysis System, Imager2 (VWR).

Next, we checked the possibility of using the designed primers to detect a specific sequence using qPCR. Three bacterial species were selected for analysis: *M. marinum, M. tuberculosis*, and *Escherichia coli*, and we used genomic DNA concentrations in the range of 0.003-50 ng. qPCR was performed using SYBR Green Master Mix (Applied Biosystems) and Quant Studio 5 system (Applied Biosystems). Each reaction (final volume 10 µl) contained 5 µl of SYBR Green Master Mix, 300 nM of each primer (Table 1), the appropriate amount of genomic DNA (geometric series of dilutions of genomic DNA in the range of 0.003-50 ng for *M. marinum* and in the range of 0.39-50 ng for *M. tuberculosis* and *E. coli*) and DNase, RNase free water. To detect the sequence specific for the

M. marinum species and the *Mycobacterium* genus, the following protocol was used: first heated 95°C for 2 min and then subjected to 40 cycles of 95°C for 15 s (denaturation), 60°C for 30 s (annealing) and 72°C for 15 s (extension). Data were acquired during the extension step. A melting curve analysis was performed at the end to verify the specificity and identity of the PCR products (Supplementary figures). Each experiment was performed in triplicate, and the results are presented as means of cycle threshold (Ct) and standard errors. We considered a Ct value of 30 or less as a positive result.

In vitro validation using bacteria-spiked samples

First, the method's sensitivity was tested by checking the number of bacteria we could detect in spiked artificial sputum samples. Artificial sputum was prepared as described previously, by dissolving 2 g of methylcellulose (Sigma) in 200 ml of sterile water and adding one-fifth of emulsified egg (Friedrich et al. 2019). Then, appropriate amounts of bacteria were added from frozen bacterial stocks to 1 ml of artificial sputum. The number of bacteria in the inoculum was checked using the colony-forming unit assay (CFU). 10^1-10^7 *Mycobacterium tuberculosis*, 10^1-10^6 *M. marinum*, and 10^1-10^6 *Escherichia coli* were added in 5 repetitions for each number of bacteria. Following the spiking of artificial sputum, the samples underwent a standard procedure they would follow in the diagnostic laboratory where mycobacteria are detected. 1 ml of each of the samples was decontaminated with 1 ml of 2% NALC-NaOH (BBL Mycoprep, Becton Dickinson) for 20 min at room temperature with shaking at 200 rpm and neutralized with phosphate-buffered saline pH 6.8 added to a final volume of 50 ml. Next, the samples were centrifuged for 20 min at 4°C and 3000 × g, the supernatant was discarded. The remaining pellet was resuspended in 0.5 ml phosphate buffer saline and transferred to 1.5 ml tubes. Then, we isolated DNA using the GenoLyse kit (Bruker, Hain Lifescience), a kit already in use in the diagnostics of mycobacteria, following the protocol developed by the manufacturer. Following DNA isolation, the supernatant was transferred to new tubes and used to perform qPCR using the same parameters as during validation using bacterial culture.

Next, we wanted to check the performance of the primers in the presence of the eukaryotic DNA. The samples were prepared by adding 10^5 and 10^7 *M. marinum* cells (as assessed by CFU plating) to buccal swabs and saliva samples. The volume of bacterial suspension added to the swab and saliva samples did not exceed 1/10 of the volume (Klaschik et al. 2004). Buccal swabs were collected by placing the swab in the mouth and rubbing with a twisting motion for 20 seconds on the inside of the cheek. Saliva samples were diluted in a 1:4 ratio with PBS, and centrifuged for 5 minutes at 14,000 g. Following spiking, DNA was isolated using the QIAamp UCP Pathogen Mini Kit (Qiagen), designed for DNA purification from small volumes of whole blood, swabs, cultures, and body fluids. 40 ng of total isolated DNA was used for the qPCR, performed as before. Apart for checking for the presence of the *espE_2* gene fragment, specific for the *M. marinum* species, and the *mmpl* gene fragment, specific for the *Mycobacterium* genus, we also checked the presence of the *YWHAZ* and *B2M* genes (Table 1), which are human reference genes (Bruce et al. 2012, Jabłońska et al. 2020, Hu et al. 2023).

Sensitivity and specificity calculations

We used MedCalc software (MedCalc Sofware Ltd) to calculate sensitivity, specificity and confidence intervals. True positives were defined as the number of instances where the test correctly identified the bacterium; true negatives were defined as the number of instances where the test correctly excluded the identification of the bacterium, false positives were defined as the number of instances where the test incorrectly identified the bacterium, false negatives were defined as the number of instances where the test incorrectly excluded the identification of the bacterium. Confidence intervals for sensitivity and specificity are "exact" Clopper-Pearson confidence intervals.

Supplementary material

The supplementary material referred in this manuscript can be found at ResearchGate repository. We provided supplementary data that includes the databases (DOI: 10.13140/RG.2.2.29142.56647), figures (DOI: 10.13140/RG.2.2.22805.23527), sequences (DOI: 10.13140/RG.2.2.11526.48964), and list of strains (DOI: 10.13140/RG.2.2.24948.26245), and *in silico* primer specificity testing results (10.13140/RG.2.2.30537.25443).

Results

Identification of *Mycobacterium*-specific sequence

For identification of *Mycobacterium*-specific sequence, we performed the Mauve alignment of five bacterial species, including three mycobacteria (*M. tuberculosis*, *M. smegmatis*, *M. abscessus*) and two closely related species (*Corynebacterium glutamicum*, *Rhodococcus rhodochrous*). The Mauve alignment

divided the genomes into 451 Locally Collinear Blocks (LCBs). We manually curated these blocks and identified the fragment of *mmpl* gene (Supplementary_ sequences) as potentially specific to the *Mycobacterium* genus. The 2865 bp *mmpl* gene encodes mycobacterial membrane protein large (Mmpl) protein, which belongs to topologically complex membrane proteins responsible for transmembrane transport (Melly and Purdy 2019).

Next, we investigated the presence the *mmpl* gene, *in silico,* with BLAST-search, in a custom virtual database of genomic sequences of 166 *Mycobacterium* bacteria (Supplemental figures). We identified the *mmpl* gene sequence in 100% (166 out of 166) of mycobacteria in the *Mycobacterium* database. The sequences varied from 1383 to 2949 bp, and the pairwise identity was 76.6%. Next, we searched for the *mmpl* gene sequence in the *Actinobacteria* phylum database. We identified fragments of the sequence ranging from 72 bp to 547 bp localizing at the 3' end of the reference sequence.

Identification of *M. marinum*-specific sequence

For the *M. marinum* species-specific sequence, the Mauve alignment of ten bacteria strains belonging to the *M. marinum* complex, including two *M. marinum* species strains, divided the genomes into 285 Locally Collinear Blocks (LCBs). We manually curated the LCBs, and selected the 1377 bp *espE_2* gene sequence as potentially specific to *M. marinum* species (Supplementary sequences).

We found the presence of the whole gene sequence in 13 genome sequences of *M. marinum* strains (Supplementary figures). The total number of mutations between the strains in the gene sequence was 34 SNPs, and the homology was from 99.1 to 100%. Next, we used BLAST to search for *espE_2* sequences (1,377 bp) in a sequence database of 166 bacteria of the *Mycobacterium* genus. Apart from the sequence of *M. marinum*, whose genome sequence was also present in the database, we found the presence of partially homologous sequences in ten species. The sequences were 45 to 1,112 bp in length.

Primer specificity testing

We used the alignments to design primers specific to 391 bp fragment of the *mmpl* gene and 202 bp of the *espE_2* gene. First, the specificity of primers was assessed using the Primer- BLAST tool (NCBI). The analysis detected no unintended targets.

Next, primer specificity was assesed in *in vitro* experiments. Exemplary PCR products were confirmed by sequencing. To estimate the specificity of the *mmpl* gene fragment primers in mycobacterial samples, we measured the expected and non-specific products. The specificity for *mmpl* gene in *Mycobacterium* samples was 94.4% for specific and 5.6% for non-specific product, while for *Mycobacterium* samples vs samples other than *Mycobacterium*, it was 97.1% and 2.9%, respectively.

The specificity of the *espE_2* gene fragment primers in *M. marinum* samples was 100%, and we did not observe any non-specific PCR products.

In silico validation

We used the BLAST tool from Geneious Prime (blastn, max e-value 1e-10) to search the bacteria database (n=6263) with the fragment of the *mmpl* gene, identifying *Mycobacteria*. We found the sequence in 359 out of 359 (100%) mycobacterial genome sequences included in the database. There were no false positive identifications. *In silico* sensitivity and specificity for our method of *Mycobacterium* identification was 100% (95% CI 98.9%-100%) and 100% (99.94%-100%), respectively.

Next, we validated the specificity of *espE_2* gene fragment for *M. marinum* species. We performed BLAST-search using custom virtual bacteria database (n=6263). We found three out of three *M. marinum* genomes included in the database, setting both the sensitivity and specificity of the method at 100%.

Sensitivity and specificity of methods using bacterial culture DNA

The *mmpl* fragment, specific for the *Mycobacterium* genus, was amplified by PCR in the collection of 52 bacterial isolates (Fig. 2). 44 strains belonged to the *Mycobacterium* genus, while the rest represented closely and distantly related bacterial species (Supplementary_strains). Three exemplary PCR products were sequenced by Sanger sequencing to ensure the amplification of the proper sequence. The sequence was detected in 42 strains of mycobacteria. We did not identify the *mmpl* fragment (or false positive amplification) in 8 non-mycobacterial species and two strains of the *M. chelonae.* Of note, *M. chelonae* isolates were positive for the *mmpl* fragment in our *in silico* studies, yet these were different strains (Supplementary figures). In this experiment, the sensitivity and specificity of the *mmpl* fragment detection for *Mycobacterium* genus identification were at 95% (85% - 99% CI) and 100% (63%-100% CI), respectively (Table 2).

Next, PCR the *espE_2* fragment was amplified by PCR in 44 mycobacterial isolates- including six *M. marinum* strains and 38 other mycobacteria (Fig. 3). Three exemplary PCR products were sequenced

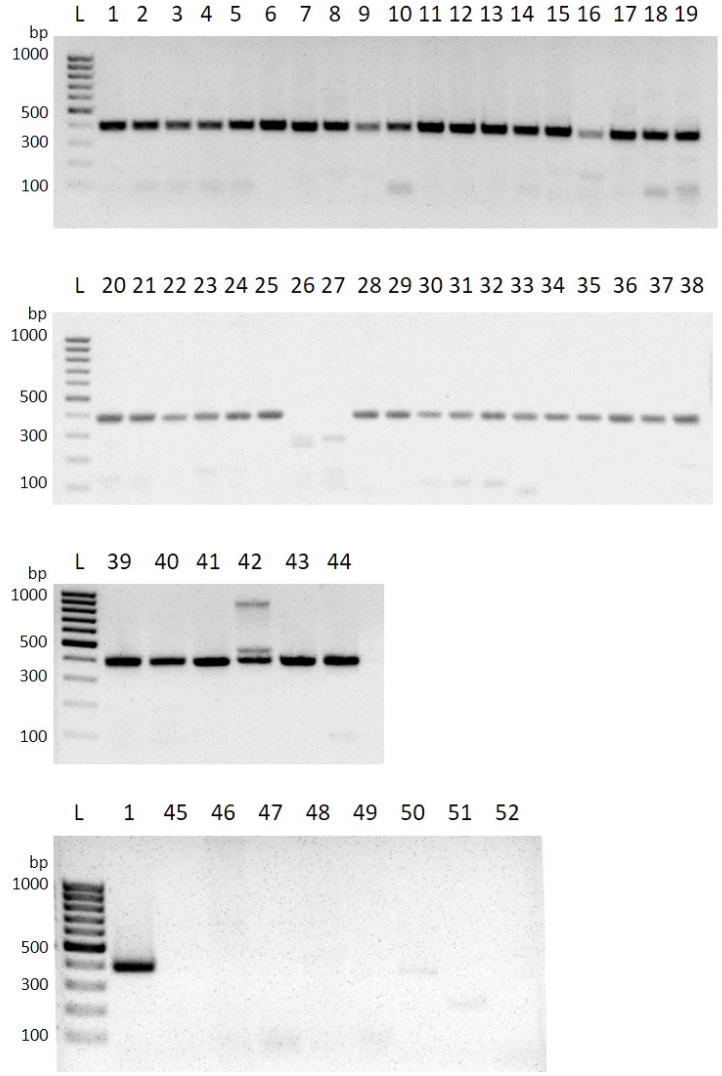

Fig. 2. The *mmpl* fragment (*Mycobacterium* specific) sequence amplification products after electrophoresis in 1% agarose gel. L – DNA ladder, 100 bp; numbers 1-6 *M. marinum* species, 7 – *M. avium*, 8-9 – *M. kansasii*, 10 – *M. xenopi*, 11-15 – *M. gordonae*, 16-17 – *M. intracellulare*, 18-21 – *M. fortuitum*, 22 – *M. chimera*, 23-24, 43 – *M. abscessus*, 25 – *M. malmoense*, 26-27 – *M. chelonae*, 28-29 – *M. lentiflavum*, 30-32 – *M. szulgai*, 33-34 – *M. mucogenicum*, 35-36 – *M. mageritense*, 37 – *M. celatum*, 38-39 – *M. simiae*, 40 – *M. neoaurum*, 41 – *M. sarafulucum*, 42 – *M. tuberculosis*, 44 – *M. smegmatis*, 45 – *Streptomyces lavendulae*, 46 – *Streptomyces aeurofaciens*, 47 – *Streptomyces rimosus*, 48 – *Rhodococcus erythropolis*, 49 – *Arthrobacter spp*., 50 – *Staphylococcus aureus*, 51 – *Salmonella enteritidis*, 52 – *Escherichia coli*.

Table 2. The sensitivity and specificity values of *Mycobacterium* and *M. marinum* specific sequence detection. True positives (TP), true negatives (TN), false positives (FP), false negatives (FN).

	No. strains	TP	TN	FP	FN	Sensitivity	95% CI	Specificity	95% CI
In silico validation									
Bacteria database									
mmpl PCR product (*Mycobacterium*)	6263	359	5904	0	0	100.00%	98.98% to 100.00%	100.00%	99.94% to 100.00%
espE_2 PCR product (*M. marinum*)	6263	3	6260	0	0	100.00%	29.24% to 100.00%	100.00%	99.94% to 100.00%
In vitro validation									
mmpl PCR product (*Mycobacterium*)	52	42	8	0	2	95.45%	84.53% to 99.44%	100.00%	89.72% to 100.00%
espE_2 PCR product (*M. marinum*)	44	6	38	0	0	100.00%	54.07% to 100.00%	100.00%	90.75% to 100.00%

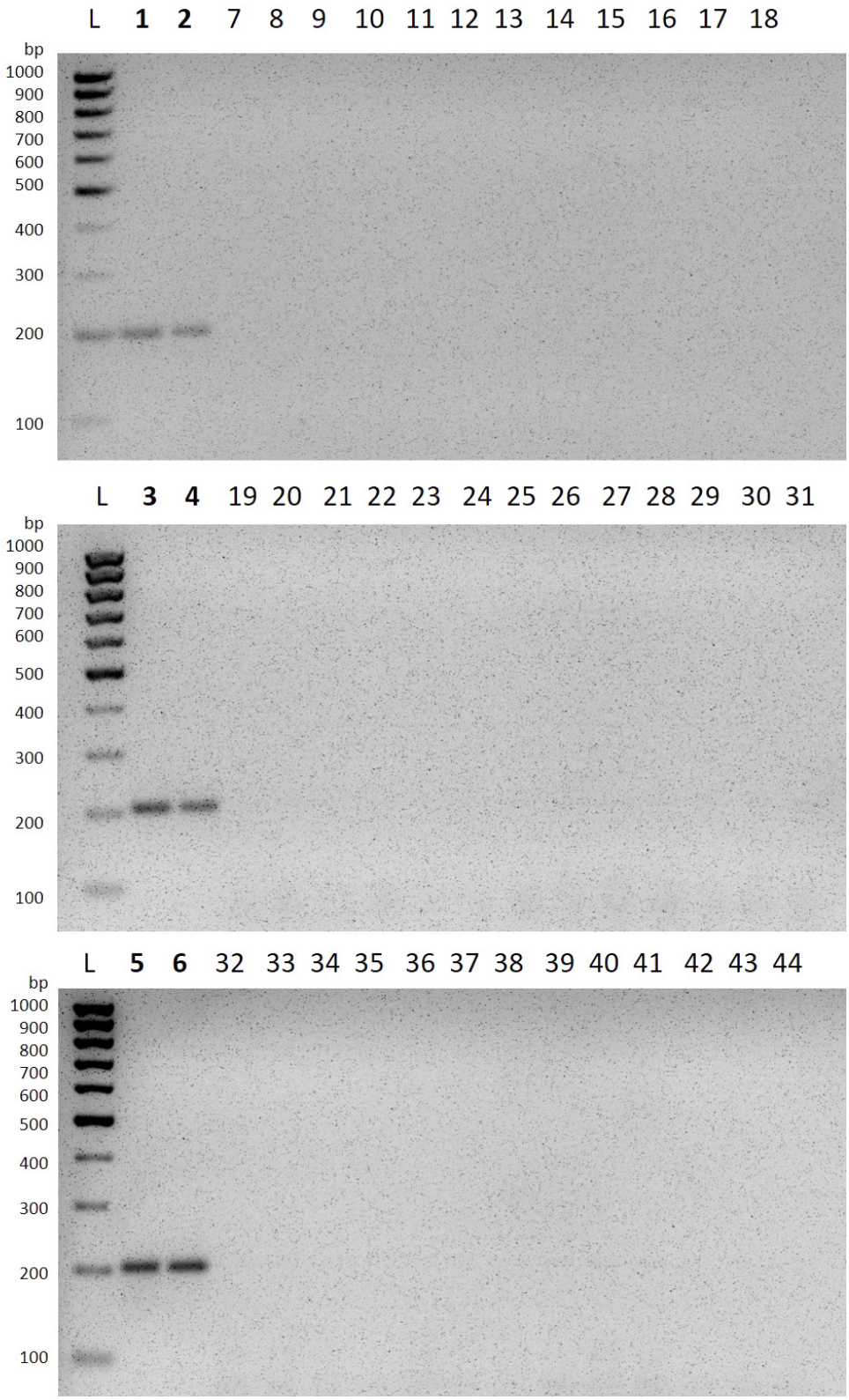

Fig. 3. The *M. marinum* specific sequence amplification products after electrophoresis in 2% agarose gel. L – DNA ladder, 100 bp; numbers 1-6 *M. marinum* species, 7 – *M. avium*, 8-9 – *M. kansasii*, 10 – *M. xenopi*, 11-15 – *M. gordonae*, 16-17 – *M. intracellulare*, 18-21 – *M. fortuitum*, 22 – *M. chimera*, 23-24, 43 – *M. abscessus*, 25 – *M. malmoense*, 26-27 – M. *chelonae*, 28-29 – *M. lentiflavum*, 30-32 – *M. szulgai*, 33-34 – *M. mucogenicum*, 35-36 – *M. mageritense*, 37 – *M. celatum*, 38-39 – *M. simiae*, 40 – *M. neoaurum*, 41 – *M. sarafulucum*, 42 – *M. tuberculosis*, 44 – *M. smegmatis*.

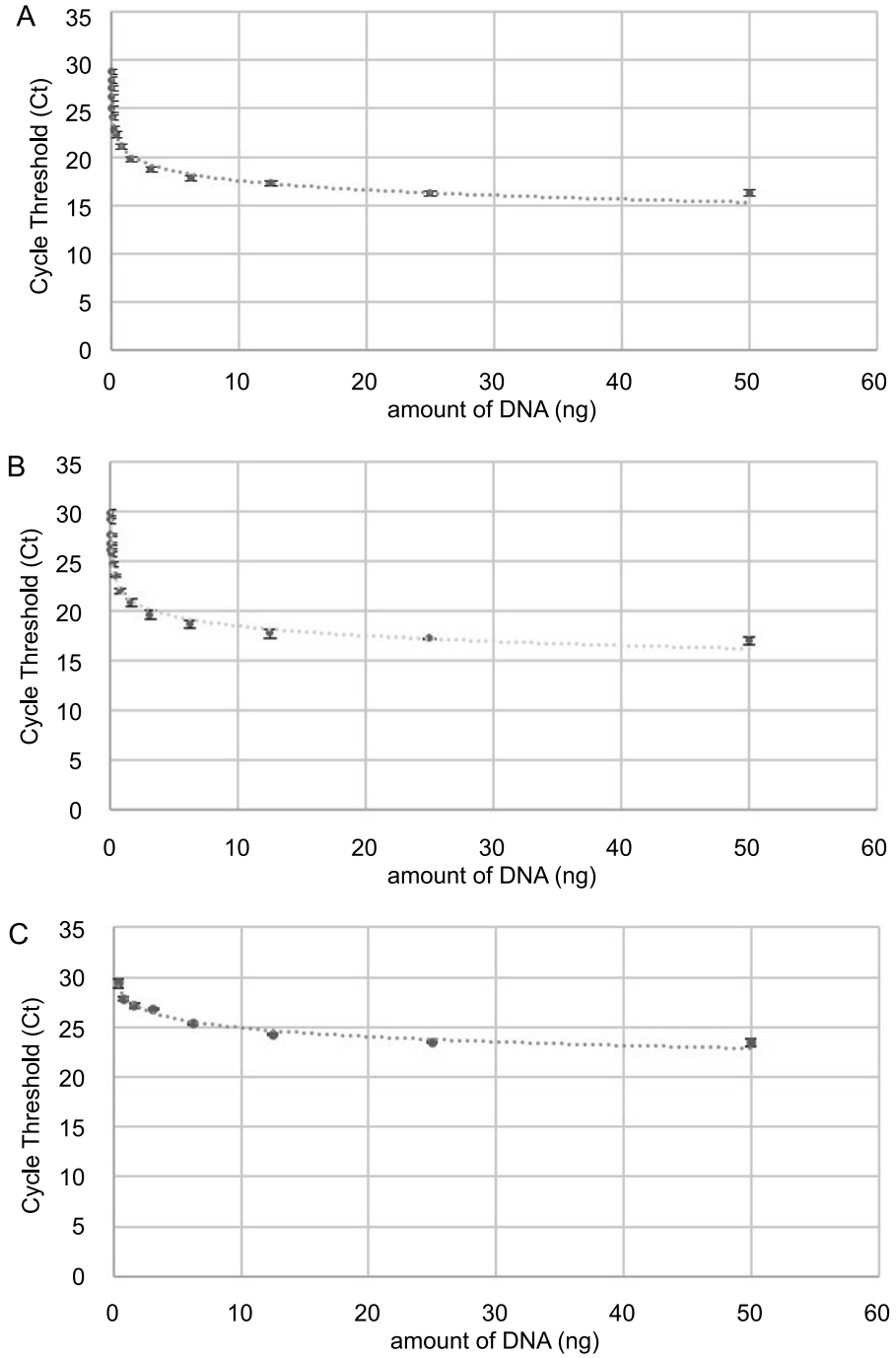

Fig. 4. Detection of specific sequences A: *espE_2* gen fragment in *M. marinum*, B: *mmpl* gen fragment in *M. marinum*, C: *mmpl* gen fragment in *M. tuberculosis* by qPCR. The graphs show the dependence of the cycle threshold value on the amount of DNA present in qPCR reaction.

by Sanger sequencing to ensure the amplification of the proper sequence. The method showed 100% (54%-100% CI) sensitivity and 100% (90%-100 CI) specificity (Table 2).

qPCR identification of Mycobacteria and *M. marinum*

We checked the possibility of using our methods of identification using the qPCR. We detected the *mmpl* fragment gene in the presence of 0.003-50 ng of *M. marinum* DNA, and 0.039-50 ng of *M. tuberculosis* DNA (Fig. 4). The *espE_2* fragment was detected in samples containing 3 pg and up of *M. marinum* DNA (Fig. 4).

Next, we checked the possibility of using our identification methods on artificial sputum samples spiked with bacteria. We detected the presence of *mmpl* gene fragment in samples containing 10^6 CFU/ml *M. tuberculosis* and 10^4 CFU/ml *M. marinum*. We detected an

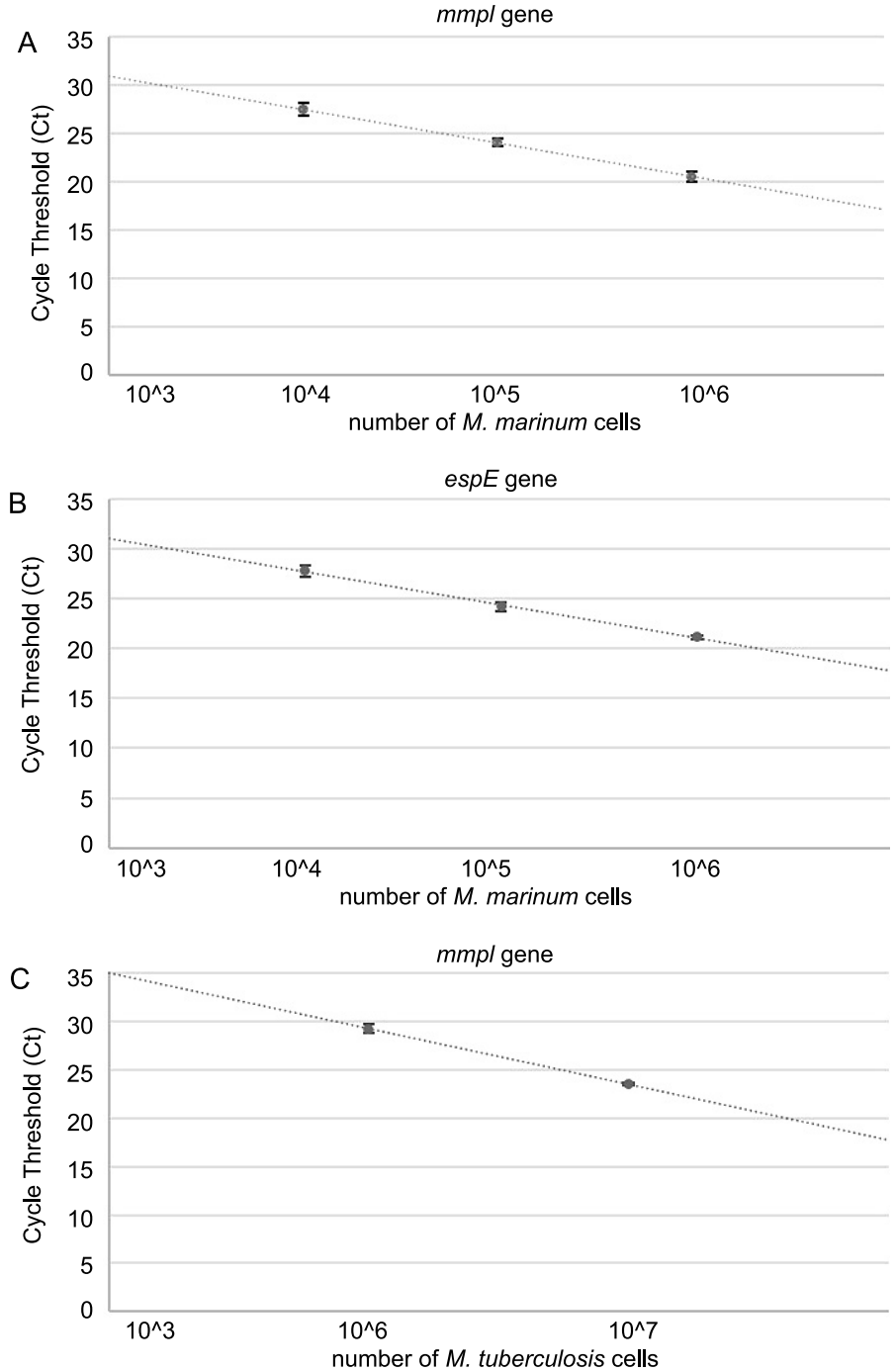

Fig. 5. Detection of: A and C: *mmpl* gene, B: *espE_2* gene in spiked artificial sputum samples by qPCR. The graphs show the dependence of the cycle threshold value in qPCR reaction on the number of *M. marinum* or *M. tuberculosis* cells.

espE_2 gene fragment in samples spiked with 10^4 CFU/ml *M. marinum* (Fig. 5). We detected no amplification in the samples into which we added 10^6 CFU/ml *E. coli*.

We checked the possibility of using qPCR to identify mycobacteria and *M. marinum* in samples containing eukaryotic material. First, we tested sputum samples, that routinely undergo decontamination before DNA isolation in mycobacterial diagnostic laboratories. For the identification of mycobacteria, we used TB-positive sputum samples. We could confirm the presence of the *Mycobacterium* genus only in one out of five samples (data not shown). Further, we also added 10^3, 10^4, and 10^5 *M. marinum* bacteria to TB-negative sputum samples before the decontamination step, in three repetitions, and we were able to confirm the presence of *M. marinum* in one sample with 10^5 bacteria (data not shown).

Finally, we used buccal swabs and saliva samples spiked with *M. marinum* that underwent DNA isolation without the decontamination step. We detected the frag-

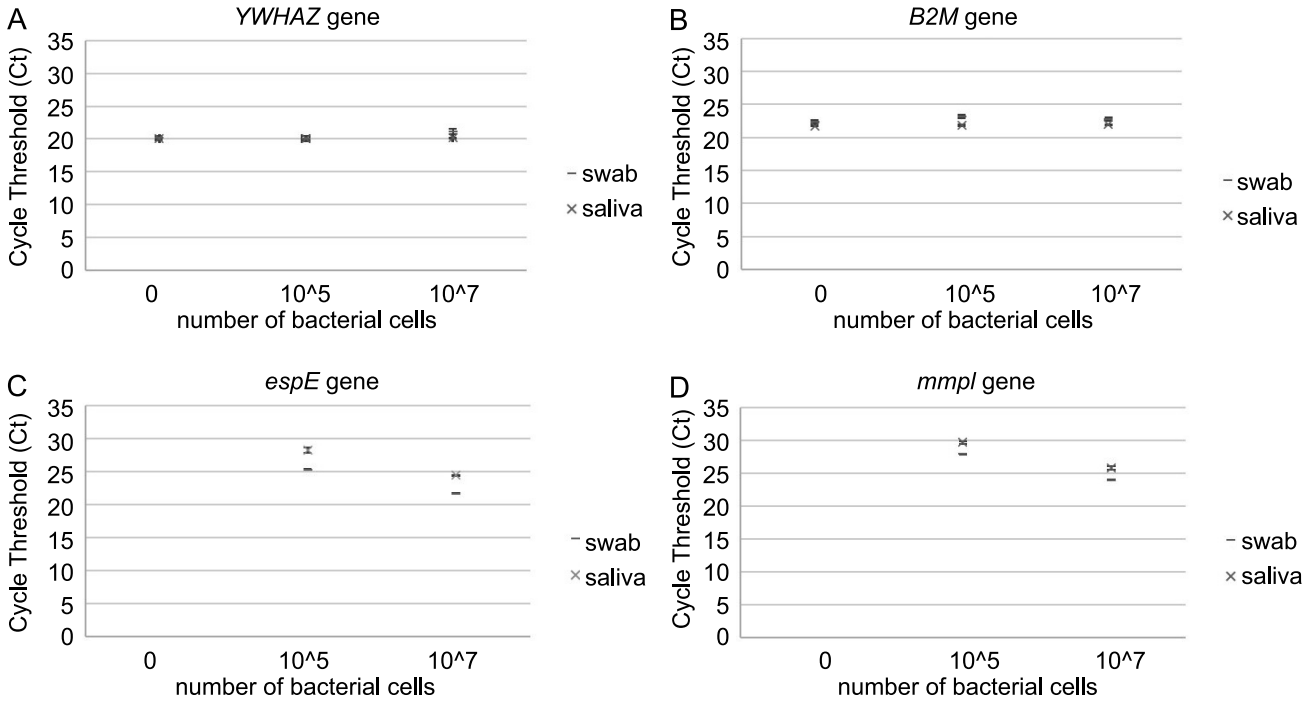

Fig. 6. Detection of: A: *YWHAZ* gene, B: *B2M* gene, C: *espE_2* gene, D: *mmpl* gene in mock clinical samples by qPCR. The graphs show the dependence of the cycle threshold value in qPCR reaction on the number of *M. marinum* cells introduced into swab or saliva sample.

ments of the *espE_2* and *mmpl* genes in samples into which we added at least 10^5 CFU/ml bacteria (Fig. 6). We did not observe the presence of genes in samples without introduced bacteria, so we concluded that there was a lack of cross-reactivity with human genomic DNA.

Discussion

We found a new way of identifying the *Mycobacterium* genus and *M. marinum* species based on identifying a specific sequence. Our identification methods reached excellent diagnostic performance, with 100% sensitivity and specificity in *in silico* validation, and near-perfect sensitivity and specificity in *in vitro* experiments using bacterial cultures. Although qPCR sensitivity was reduced in complex matrices like sputum, likely due to DNA degradation and eukaryotic DNA interference, our method showed strong performance in buccal swabs and saliva.

The biggest advantage of the presented test is that it is based on identifying a specific DNA fragment present only in the selected group of bacteria. Therefore, the method is simple, fast, and cheap for species identification and can be an alternative to other methods that are more expensive, time-consuming, and require specialized equipment. The detection limit, after decontamination of samples, is comparable to commercial GenoType Mycobacterium CM VER 2.0 kit, commonly used for Mycobacterium identification (1.65x 105 CFU/ml). Grant et al. established a minimum detection limit of direct IS900 PCR at 105 CFU of M. avium subsp. paratuberculosis per 50 ml of milk (Grant et al. 2000). The advantage of our method, in comparison for previously reported method of *Mycobacterium* species identification using the internal transcribed spacer 1 (ITS-1) (Mohamed et al. 2005), is that it can be adapted to various molecular biology methods like PCR, real-time PCR, hybridization, and sequencing, while the latter strictly requires utilization of sequencing. In addition, unlike the currently developed tests based on identifying a DNA fragment, our method seems to distinguish between the two species *M. marinum* and *M. ulcerans*. However, this has only been proven by *in silico* analysis and needs to be checked *in vitro* using the genomic DNA of these species.

The major drawback of our study is the low number of analyzed strains in laboratory validation. For *in silico* validation, we used all the sequences of *M. marinum* available in the repositories. For laboratory validation, we used six strains belonging to the *Mycobacterium marinum* and 38 strains belonging to other *Mycobacteria* that were available for this study. The second problem is the lack of a fragment of *mmpl* gene sequence in the genome of two strains of *M. chelonae,* as it indicates certain limitations of bioinformatics analyses and the need to validate the results *in vitro*. The absence of an amplification product may be due to the occurrence of point mutations, deletions or insertions within the

primer sequences. Alternatively, there is a possibility that the entire *mmpl* gene is absent in the genomes of *M. chelonae* isolates in our collection. The *mmpl* gene was shown to be non-essential for growth in *M. tuberculosis*, and hence it is likely dispensable in other species of mycobacteria (Sassetti et al. 2003, Griffin et al. 2011). This is further supported by the observation that some bacteria belonging to the *Mycobacterium* genus, such as *M. avium subsp. paratuberculosis* K10 or *M. massiliense* type II genotype have a deletion in *mmpl* genes (Nessar et al. 2011, Kim et al. 2013). Despite the lack of detection of the *mmpl* fragment in our isolates of *M. chelonae*, it was detected in 42 other species belonging to the *Mycobacterium* genus and in *M. chelonae* genome sequences deposited in repository, supporting the use of our method. The possibility of using our *Mycobacterium*-specific sequence must be tested using a broader collection of clinical bacterial strains belonging to this genus. Method sensitivity could possibly be increased by the use of degenerated primers. Further, in cases where the *mmpl* gene detection fails, the diagnostic algorithm could be extended to alternative sequencing-based techniques.

To sum up, our method of identification of *Mycobacterium* genus and *M. marinum* species, based on the identification of specific sequence, can improve the detection of mycobacteria in laboratories where specialized equipment is not available or can serve as a cross-detection method.

Acknowledgements

This work is part of the research project financed by the National Science Center of Poland, grant number 2019/34/E/NZ6/00221.

References

Akram SM, Aboobacker S (**2023**) Mycobacterium marinum Infection. In: StatPearls. StatPearls Publishing, Treasure Island (FL).

Aubry A, Mougari F, Reibel F, Cambau E (**2017**) Mycobacterium marinum. Microbiol Spectr 5:10.

Bagcchi S (**2023**) WHO's Global Tuberculosis Report 2022. The Lancet Microbe 4: e20.

Biet F, Boschiroli ML (**2014**) Non-tuberculous mycobacterial infections of veterinary relevance. Res Vet Sci 97: 69-77.

Bruce KD, Sihota KK, Byrne CD, Cagampang FR (**2012**) The housekeeping gene YWHAZ remains stable in a model of developmentally primed non-alcoholic fatty liver disease. Liver Int 32: 1315-1321.

Chihota VN, Grant AD, Fielding K, Ndibongo B, van Zyl A, Muirhead D, Churchyard GJ (**2010**) Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. Int J Tuberc Lung Dis 14: 1024-1031.

Chirakos AE, Nicholson KR, Huffman A, Champion PA (**2020**) Conserved ESX-1 substrates EspE and EspF are virulence factors that regulate gene expression. Infect Immun 88: e00289-20.

Coimbra NR, Goes-Neto A, Azevedo V, Ouangraoua A (**2020**) Reconstructing the phylogeny of Corynebacteriales while accounting for horizontal gene transfer. Gen Biol Evol 12: 381-395.

Doig C, Seagar AL, Watt B, Forbes KJ (**2002**) The efficacy of the heat killing of Mycobacterium tuberculosis. J Clin Pathol 55: 778-779.

Francis-Floyd R (**2011**) Mycobacterial infections of fish. SRAC 4706

Friedrich SO, Kolwijck E, Karinja MN, van der Merwe L, Diacon AH (**2019**) Quantification of viable bacterial load in artificial sputum spiked with Mycobacterium tuberculosis. Tuberculosis 115: 140-145.

Grant IR, Pope CM, O'Riordan LM, Ball HJ, Rowe MT (**2000**) Improved detection of Mycobacterium avium subsp. paratuberculosis in milk by immunomagnetic PCR. Vet Microbiol 77: 369-378.

Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti Ch (**2011**) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7: e1002251.

Hashish E, Merwad A, Elgaml S, Amer A, Kamal H, Elsadek A, Marei A, Sitohy M (**2018**) Mycobacterium marinum infection in fish and man: epidemiology, pathophysiology and management; a review. Vet Q 38: 35-46.

Hu Y, Jiang Q, Zhai X, Liu L, Hong Y (**2023**) Screening and validation of the optimal panel of reference genes in colonic epithelium and relative cancer cell lines. Sci Rep 13: 17777.

Jabłońska A, Świerzko AS, Studzińska M, Suski P, Kalinka J, Leśnikowski ZJ, Cedzyński M, Paradowska E (**2020**) Insight into the expression of RIG-I-like receptors in human third trimester placentas following ex vivo cytomegalovirus or vesicular stomatitis virus infection. Mol Immunol 126: 143-152.

Jauhal AA, Newcomb RD (**2021**) Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. Mol Ecol Resour 21: 1416-1421.

Kanabalan RD, Lee LJ, Lee TY, Chong PP, Hassan L, Ismail R, Chin VK (**2021**) Human tuberculosis and Mycobacterium tuberculosis complex: a review on genetic diversity, pathogenesis and omics approaches in host biomarkers discovery. Microbiol Res 246: 126674.

Kim BJ, Kim BR, Lee SY, Kook YH, Kim BJ (**2013**) Rough colony morphology of Mycobacterium massiliense Type II genotype is due to the deletion of glycopeptidolipid locus within its genome. BMC Genomics 14: 890.

Klaschik S, Lehmann LE, Raadts A, Book M, Gebel J, Hoeft A, Stuber F (**2004**) Detection and differentiation of in vitro-spiked bacteria by real-time PCR and melting-curve analysis. J Clin Microbiol 42: 512-517.

Lai LY, Lin TL, Chen YY, Hsieh PF, Wang JT (**2018**) Role of the Mycobacterium marinum ESX-1 secretion system in sliding motility and biofilm formation. Front Microbiol 9: 1160.

Meehan CJ, Barco RA, Loh Y HE, Loh YH, Cogneau S, Rigouts L (**2021**) Reconstituting the genus Mycobacterium. Int J Syst Evol Microbiol 71: 004922.

Melly G, Purdy GE (**2019**) MmpL proteins in physiology and pathogenesis of M. tuberculosis. Microorganisms 7: 70.

Mohamed AM, Kuyper DJ, Iwen PC, Hesham HA, Dhundy RB, Hinrichs SH (**2005**) Computational approach involving use of the internal transcribed spacer 1 region for identification of Mycobacterium species. J Clin Microbiol 43: 3811-3817.

Molina-Ruiz CS, Zamora-Briseño JA, Simón O, Lasa R, Williams T (**2024**) A qPCR Assay for the quantification of selected genotypic variants of Spodoptera frugiperda multiple Nucleopolyhedrovirus (Baculoviridae). Viruses 16: 881.

Nessar R, Reyrat JM, Davidson LB, Byrd TF (**2011**) Deletion of the mmpL4b gene in the Mycobacterium abscessus glycopeptidolipid biosynthetic pathway results in loss of surface colonization capability, but enhanced ability to replicate in human macrophages and stimulate their innate immune response. Microbiology (Reading) 157: 1187-1195.

Peterson TS, Kent ML, Ferguson JA, Wartal VG, Whipps CM (**2013**) Comparison of fixatives, fixation time, and severity of infection on PCR amplification and detection of Mycobacterium marinum and Mycobacterium chelonae DNA in paraffin-embedded zebrafish (Danio rerio). Dis Aquat Organ 104: 113-120.

Sassetti CM, Boyd DH, Rubin EJ (**2003**) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48: 77-84

Sia TY, Taimur S, Blau DM, Lambe J, Ackelsberg J, Yacisin K, Bhatnagar J, Ritter J, Shieh WJ, Muehlenbachs A, Shulman K, Fong D, Kung E, Zaki SR (**2016**) Clinical and pathological evaluation of Mycobacterium marinum group skin infections associated with fish markets in New York City. Clin Infect Dis 62: 590-595.

Singh AK, Maurya AK, Umrao J, Kant S, Kushwaha RAS, Nag VL, Dhole TN (**2013**) Role of GenoType® Mycobacterium Common Mycobacteria/Additional Species assay for rapid differentiation between Mycobacterium tuberculosis complex and different species of non-tuberculous mycobacteria. J Lab Physicians 5: 83-089.

Strobel K, Sickenberger C, Schoen C, Kneitz H, Kolb-Maurer A, Goebeler M (**2022**) Diagnosis and therapy of Mycobacterium marinum: a single-center 21-year retrospective analysis. J Dtsch Dermatol Ges 20: 1211-1218.

van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM (**1993**) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31: 406-409.