Master Thesis

Presence of *esxA* (ESAT-6) and *esxB* (CFP-10) by non-tuberculous mycobacterial species

Universiteit Utrecht

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Agnes Bouw (3578739), 2020
Research period: from 09-04-2018 till 29-06-2018
Farm animal health and veterinary public health, Utrecht University, Utrecht, The Netherlands

Supervised by
Dr. Nomakorinte Gcebe (PhD)
Agricultural Research Council-OVI, Onderstepoort, South Africa

Prof. Anita Michel and dr. E.M.D.L. (Elise) van der Heijden
Department of Veterinary Tropical Diseases, University of Pretoria, Onderstepoort, South Africa

Prof. dr. V.P.M.G. Rutten and dr. E.M.D.L. (Elise) van der Heijden
Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands

TABLE OF CONTENTS

ABSTRACT

Introduction: Tuberculosis (TB) is a widespread chronic infectious granulomatous disease of many animal species as well as humans that is caused by mycobacteria. Several TB diagnostics use the immunogenic proteins Early Secreted Antigenic Target-6 (ESAT-6) and Culture Filtrate Protein 10 (CFP-10) as diagnostic markers. These proteins are thought to be important for virulence and specific for pathogenic mycobacterial species such as members of the *Mycobacterium tuberculosis complex* (MTBC). However, an increasingly number of non-tuberculous mycobacteria (NTM) species, most of wich are considered to be only environmental contaminants or commensals, have been found to contain gene orthologues of *esxA* and *esxB* genes coding for ESAT-6 and CFP-10 respectively. If these NTM species express ESAT-6 and CFP-10, cross-reactive immune responses between MTBC and NTM species may occur. NTM may therefore interfere with diagnostic tests for and vaccination against TB. The research aim was to investigate whether *M. moriokaense, M. malmesburri* sp. nov.*, M. vaccae/M. vanbaalenii, M. madagascariense, M. acapulcensis, M. elephantis, M. chitae, M. confluentis, M. paraffinicum, M. neonarum, M. engbackii, M. parafortuitum, M. septicum/peregrinum, M. smegmatis, M. fortuitum, M. komanii* sp. nov. and *M. nonchromogenicum* possess orthologue *esxA* and/or *esxB* genes.

Materials & methods: Polymerase chain reactions (PCR) were performed using *M. bovis* derived primers for the *esxA* and *esxB* genes. Amplification products sized approximately 300 basepairs were extracted and submitted for sequencing. BLAST (Basic Local Alignment Search Tool) searches were performed on the sequeces to identify if sequences were coding for ESAT-6 or CFP-10. *EsxA* and *esxB* sequences next to 16S rRNA sequences of NTM were used for phylogenetic analysis. The three phylogentic trees were compared by similarity and a consensus tree was created.

Results: EsxB orthologues were detected in *M. engbackii* isolates and in *M. paraffinicum*. The *esxB* orthologues were 86-88% similar at nucleotide level to the genes coding for CFP-10 in other NTM-like *M. szulgai, M. marinum, M. pseudohottsi, M. ulcerans, M. liflandi* and *M. kansasii.* BLAST searches for the 321 bp sized sequence potentially coding for the *esxA* gene of *M. acapulcensis* resulted in 100% similarity with two short sequences of the type VII secretion system ESX-1 WXG100 family target ESAT-6 of MTBC species (*M. tuberculosis*). The two short sequences (31 and 29 base pairs respectively) overlapped with the primer annealing regions. Therefore the presence of an *esxA* gene orthologue in *M. acapulcensis* remains questionable. Phylogenetic profiling based on *esxA* and *esxB* sequences resulted in slow growing mycobacteria (SGM) roughly separating from rapid growing mycobacteria (RGM) as was shown for 16SrRNA sequences. SGM are mostly pathogenic mycobacterial species, while RGM are not known to be pathogenic or are sometimes known to cause opportunistic infections only. Phylogenetic trees based on *esxA* and *esxB* appeared more similar to each other than these trees were compared to the 16SrRNA-based phylogenetic tree.

Discussion: Amplification of *esxA* and *esxB* gene sequences may be dependent on specific combinations of primers and NTM species, as different NTM were found to possess *esxA* and/or *esxB* orthologues than reported previously. Identification of *esxB* gene orthologues in *M. engbackii* and M. *paraffinicum* isolates confirms again that these genes do occur in non-pathogenic RGM species. In case of expression of ESAT-6 and CFP-10 by these mycobacterial species, infections with those NTM can cause false positive results in TB diagnostics when the antigens will be sufficient immunogenic. Therefore, the expression of ESAT-6 and CFP-10 should be further investigated and kept in mind when designing immunodiagnostic assays for TB.

The phylogenetic classification of mycobacterial species based on *esxA* and *esxB* sequences can predict the potential pathogenicity of NTM species, but this classification does not predict NTM pathogenicity completely correct. This could be due to the fact that the proteins of the type VII secretion system are not the only virulence factors in mycobacterial species. *EsxA*- and *esxB*-based phylogenetic trees resembled each other more than the 16SrRNA-based phylogenetic tree, probably because sequences of different strains of the same mycobacterial species were used for the 16SrRNA-based tree.

Keywords: Tuberculosis, nontuberculous mycobacteria, ESAT-6/*esxA*, CFP-10/*esxB*

Abbreviations

ARC-OVI	=	Agricultural Research Council, Onderstepoort Veterinary Institute
BCG	=	Bacillus Calmette and Guérin
BLAST	=	Basic Local Alignment Search Tool
bp	=	base pairs
CFP-10	=	Culture Filtrate Protein-10
CMI	=	Cell-mediated immunity
ELISA	=	Enzyme-Linked Immuno Sorbent Assay
ESAT-6	=	Early Secreted Antigenic Target-6
IFN-γ	=	Interferon gamma (IFN-γ) (assay)
MOTT	=	Mycobacteria other than tuberculosis
MTBC	=	*Mycobacterium tuberculosis complex*
NTM	=	Non-tuberculous mycobacteria
PCR	=	Polymerase chain reactions
PPD-A	=	Purified protein derivate of *M. bovis*
PPD-B	=	Purified protein derivate of *M. avium*
SGM	=	Slow growing mycobacteria
TB	=	Tuberculosis
TST	=	Tuberculin skin test

1. INTRODUCTION

Tuberculosis (TB) is a widespread chronic infectious granulomatous disease of many animal species as well as humans that is caused by mycobacteria[1]. Tuberculosis has a large impact worldwide: approximately 10.0 million people are estimated to suffer from TB and 1.4 million people died of TB in 2018 alone[2]. Bovine TB causes significant economic losses in animals through reduced milk and meat production; infected cattle lose 10-25% of their productive efficiency for example[1,3,4]. The reduction and future eradication of TB is impaired by several challenges such as appropriate diagnostics, measurements for control and prevention of the disease, due to suboptimal test performance, presence of wildlife reservoirs and the lack of an effective vaccine[2,4,5].
In addition, suboptimal therapy compliance and the rise of antibiotic resistance aggravate the difficulty of eradication of TB in humans[2,4].

1.1 Mycobacteria

Mycobacteria are aerobic, rod shaped and acid-fast (Ziehl-Neelsen positive) bacteria. The genus can grossly be subdivided in environmental mycobacteria and pathogenic mycobacteria. Pathogenic mycobacteria are divided into the genetically highly related pathogenic mycobacteria belonging to the *Mycobacterium tuberculosis complex* (MTBC) group and *Mycobacterium leprae*. *M. leprae* is the causative agent of leprosy, while members of the MTBC group can cause tuberculosis in mammals[1,6-8]. Members of the MTBC are: *Mycobacterium tuberculosis*, *Mycobacterium bovis* and its vaccine strain *Mycobacterium bovis* Bacillus Calmette and Guérin (BCG), *Mycobacterium africanum*, *Mycobacterium canetti* (although debatable because of genetic differences[9]), *Mycobacterium microti*, *Mycobacterium caprae*, *Mycobacterium pinnipedii,* the Dassie bacillus, *Mycobacterium suricattae* and the newly described *Mycobacterium mungi* and *Mycobacterium* orygis[6-12]. Bacteria belonging to the MTBC group show host preferences: although they are able to infect a range of mammals, they cannot sustain within the population in all those species[9]. For example, *M. tuberculosis* is the main cause of TB cases in humans (next to *M. africanum*[9]), but it is able to infect many other mammalian species such as elephants and cattle (in cattle it is sometimes called a 'reverse zoonosis'). *M. bovis* ('bovine TB') is able to infect a range of mammalian species. Cattle and buffaloes are known as the main wildlife reservoir species of *M. bovis* infections in South Africa[13,14]. *M. bovis* is an important zoonosis and transmission to humans can occur through consumption of infected raw milk (products) or inhalation[1,3,4,7-9,15,16].

Environmental mycobacteria are traditionally also referred to as atypical, non-tuberculous mycobacteria (NTM) or mycobacteria other than tuberculosis (MOTT)[17].
Non-tuberculous mycobacteria were previously believed to be only environmental contaminants or commensals and therefore not recognized to be of any importance. However, some NTM have shown to be able to cause opportunistic infections[18]. Currently there are more than 150 species of NTM known, a third of which has been involved in human diseases[17,19]. Besides infections of immune compromised hosts (like AIDS patients[20]) and infection with members of the *Mycobacterium avium* subsp. *paratuberculosis* (causing Johne's disease in ruminants), NTM infections can result in skin lesions, TB-like pulmonary lesions, lymphadenitis, disseminated disease and bovine farcy[17,21-23].
The reader is referred to the article of Gcebe et al. (2017)[17] for an enumeration of identified NTM in South Africa. An overview is given of NTM isolated from animals and humans without any pathology, from diseased animals and humans as well as NTM found in the environment. The distribution of NTM is assumed to be regionally different and exchange between the environment and animals has been noticed[13,17,24,25]. Another study by Gcebe et al. (2013)[13] conducted to decipher the prevalence and distribution of NTM in buffaloes, cattle and the environment in South Africa, reported that cattle and buffaloes are more exposed to some NTM species than to others[13].

The need to recognize the importance of NTM is further exemplified by the fact that they have been found to interfere with the diagnosis of TB (for example *M. kansasii*, *M. marinum*, *M. fortuitum*[17,26,27]and *M. simiae*[17]), giving rise to false-positive results[13,27-29]. Moreover, some studies also report a possible detrimental effect of NTM exposure on the efficacy of BCG vaccination[13,17,30-33].

1.2 Diagnosis of TB in animals and humans

Clearly defined symptoms are frequently absent in TB and may only be seen in advanced stages of disease[3,15]. One cannot rely solely on clinical presentation and additional diagnostic methods are thus necessary. Culture-based methods and molecular assays are direct test assays that detect the pathogen or mycobacterial antigens directly. These diagnostic methods are useful for the characterization of mycobacterial species and may give information about drug resistancy[34]. However, the diagnosis of TB relies mostly on indirect tests which measure the immune response to mycobacteria, possibly because culture-based tests require post mortem samples while immuno-assays offer antemortem diagnosis of TB and provide higher throughput options and lower costs.

In bovine TB, an early and strong cell-mediated immune (CMI) response is elicited, while the humoral response is weaker and occurs later after disease progression[12,29,35,36]. The use of CMI-based tests is therefore preferred to serological assays for early detection of TB. The tuberculin skin test (TST, or 'Mantoux test' in humans[34]) and interferon gamma (IFN-γ) assays (Bovigam® in cattle or QuantiFERON®-TB Gold (in Tube) in humans[34]) are examples of CMI-based tests. These tests measure a T_H1-cell orchestrated (type IV) hypersensitivity reaction in individuals that are previously sensitized to or infected with MTBC species. Briefly, in the TST, purified protein derivate (PPD) or tuberculin of *M.* bovis (PPD-B) (as well as that of *M.* avium (PPD-A), in the comparative TST) is injected intradermally. A skin swelling will occur if the animal was previously sensitized, which is measured 72 hours post injection. Heparinized whole blood is incubated (*in vitro*) with PPD in the IFNγ-assay, in response to which previously sensitized T-cells will produce and release IFNγ. The IFNγ is subsequently captured by an ELISA (Enzyme-Linked Immuno Sorbent Assay) system[5,29,37].

As TB progresses in bovines, the cell-mediated immunity gradually wanes while a humoral response is generated[38-40]. Therefore, serological assays become more sensitive in advanced TB. Several serodiagnostic tests - as for example several ELISAs, the TB StatPak and a point of care serology kit - have been developed for the detection of TB-specific antibodies in animals[5,41]. Like CMI-based assays, these tests are based on different combinations of antigens: PPD-B, PPD-A, MPB70, MPB83 and/or ESAT-6:CFP-10 fusion proteins[5,41].

1.3 Purified Protein Derivates

Purified protein derivates (PPD's) are a crude mixture of proteins, lipids and carbohydrates derived from mycobacteria and used for antigenic stimulation of immune cells in CMI-based assays. For example, PPD-B is derived from *M. bovis*, a member of the MBTC. The preparation of PPD's involves heat treatment or inactivation of the mycobacterial culture, followed by filter-sterilization, precipitation, washing and re-suspension of the pellet. It is a crude method and therefore no selection of specific antigens is possible[29,37].

Animals are frequently infected with or exposed to NTM, as such the immune system of the host that encounters PPD-B might react to epitopes which are shared between *M. bovis* and NTM. In diagnostic testing, this cross-reactivity may lead to false positive test results and a lower diagnostic specificity[26,29,37].

Therefore, animals are often simultaneously tested for an immune reaction to PPD-A, which is derived from *M. avium* subsp. *avium* (an NTM species), in order to discern between MTBC and NTM reactivity[29,37]. However, infection with an MTBC-member may be masked in the case of a co-infection with NTM (for example with *M. avium* subsp. paratuberculosis there may be a high response to both PPD-A and PPD-B)[29,42]. Also, inconsistencies in test results have been reported as a result of different PPD preparations of the same mycobacterial species[6,38,43].

Attempts to increase test specificity are made by replacing PPD's with defined antigens[29,44,45]. Immunogenic antigens which were previously thought to be MTBC-specific and absent from NTM are needed for this purpose. Examples of such antigens are ESAT-6[26,39,40,46-50] and CFP-10[26,39,40,46-50], TB9.8, TB10.3, TB10.4, CanA, MPB70,[51-53] MPB83[51,54] and EspC[51,54].

1.4 Cross-reactivity of NTM and MTBC: ESAT-6 and CFP-10

Early Secreted Antigenic Target-6 (ESAT-6 or *esxA*) and Culture Filtrate Protein 10 (CFP-10 or *esxB*) are widely used and investigated as diagnostic markers in the whole blood IFNγ-assays[26,39,40,46-50]. These proteins are part of the so-called ESX-1 secretion system - which in turn is part of the type VII secretion system - that is involved in protein secretion and is important for virulence of *M. tuberculosis* [55,56]. ESAT-6 and CFP-10 were considered to be absent from most (investigated) NTM species as well as *M. bovis* BCG, but present in the other MTBC members[6,57]. However, more recently, several NTM (for example *M. kansasii*[25,26,58,59], *M. marinum*[25,26,58], *M. szulgai*[25,26], *M. smegmatis*[6,25,33,60], *M. riyadhense*[25,57], *M. flavescens*[25,26,57], *M. gastri*[6,26,61], *M. gordonae*[25,26], *M. avium*[60], *M. mageritense*[6], *M. vulneris*[6], *M. farcinogenes*[6], and *M. sp.* JLS[6]) and *M. leprae*[58,60,62], have been found to possess orthologues for the genes encoding for ESAT-6 and/or CFP-10[6]. In addition, also *M. septicum/M. peregrinum*[6] *and M. fortuitum* (the PPD of which is used in a modified Bovigam assay[27]), *M. nonchromogenicum* and the novel species *M. malmesburri* subsp. *nov*. and *M. komanii* subsp. *nov*., all frequently found in cattle, buffaloes and their environment in South Africa,[13] were discovered to possess *M. bovis* orthologues of *esxA* and *esxB* genes[6,33].

T-cell receptors and antibodies recognize mostly only a small region on the surface of an antigen-molecule called an epitope[63]. Epitope sequences of some NTM species have shown high similarity with epitope sequences of *M. bovis* genes[33]. Therefore, the presence of the *esxA* and *esxB* orthologues, as well as others encoding immunogenic proteins, is hypothesized to be the cause of cross-reactivity between NTM and MTBC species[33]. This hypothesis can only explain the cross-reactivity between NTM and MTBC species when the orthologue *esxA* and *esxB* genes are really being expressed in the NTM species, as is described for *M. kansasii* and *M. marinum*[26,64,65].

1.5 Problem statement and research aims

Humans and animals can be exposed to NTM. Several NTM species have been found to contain gene orthologues of *esxA* and *esxB* genes, coding for the immunogenic proteins ESAT-6 and CFP-10 respectively. If these NTM species express ESAT-6 and CFP-10, which is found to be immunogenic (capable of inducing an immune response) cross-reactive immune responses may occur (recognition by T-cell receptors and/or antibodies, due to high sequence identity at epitope level as compared to *M. bovis* proteins). Non-tuberculous mycobacteria may therefore interfere with diagnostic tests for and vaccination against TB. Therefore, the research aim is to investigate whether the following NTM species possess orthologue *esxA* and *esxB* genes: *M. moriokaense, M. malmesburri* sp. nov.*, M. vaccae/M. vanbaalenii, M. madagascariense, M. acapulcensis, M. elephantis, M. chitae, M. confluentis, M. paraffinicum, M. neonarum, M. engbackii, M. parafortuitum, M. septicum/peregrinum, M. smegmatis, M. fortuitum, M. komanii* sp. nov. and *M. nonchromogenicum.*

2. MATERIAL AND METHODS

2.1 Bacterial cultures

The mycobacterial isolates used in this research were derived from ARC-OVI (Agricultural Research Council, Onderstepoort Veterinary Institute, Pretoria). They were collected earlier from different sources in South Africa and mycobacterial species were identified previously[33]. The isolates have been stored in freezers (-20°C at Löwenstein-Jensen slopes supplemented with pyruvate or glycerol and antibiotics) at the ARC-OVI. These are shown in Table 1.

2.2 DNA extraction (from cultures)

Extraction was performed using the heating method. For every isolate approximately half a loop of colony (or liquid from the bottom of the tubes in case no colony was present) was added to 200 µL sterile distilled water (Invitrogen, ultrapure DNase/RNase free) and mixed in microcentrifuge tubes (QSP 508-GRD-Q, flat top cap, natural, 2mL). Tubes were heated in a heating block (Accublock™ Digital Dry Bath, Labnet International, Inc) for 25 minutes at 95°C. The culture lysates have been stored at −20°C or−70°C until the culture lysates were used as DNA template in subsequent PCR protocols.

2.3 Polymerase Chain Reactions (PCR)

Before performing PCR on the extracted DNA samples, a trial PCR was run to evaluate the efficacy of the combination of the mastermix, primers and base pair ladder.
A 'CFP-10-PCR mixture' was prepared by adding 12.5 µL DreamTaq Green PCR MasterMix (Thermo Fisher Scientific), 1 µL *M. bovis* CFP-10 forward primer (50 µM) and 1 µL *M. bovis* CFP-10 reverse primer (50 µM) to 6.5 µL distilled water (Promega). An 'ESAT-6-PCR mixture' was prepared identically with *M. bovis* ESAT-6 forward and reverse primers (50 µM).
The primers consist of the following nucleotide sequences: 5'ATGACAGAGCAGCAGTGGAA 3' (ESAT-6 forward primer), 5'CTATGCGAACATCCCAGTGA 3' (ESAT-6 reverse primer), 5'ATGGCAGAGATGAAGACCGA 3' (CFP-10 forward primer) and 5' TCAGAAGCCCATTTGCGAGG 3' (CFP-10 reverse primer).
Subsequently, 5 µL of the extracted DNA from the NTM-isolates, four positive controls (*M. tuberculosis* TB 8993A, *M. bovis* TB 1054C and TB 9243D, *M. kansasii* ATCC 12487) and water (negative control) were added to 20 µL of CFP-10-PCR mixture and to 20 µL of ESAT-6-PCR mixture. The applicated PCR protocol consisted of 35 cycles of denaturation (94°C, 1 min), annealing (60°C, 1 min), elongation (72°C, 1 min) and final extension (72°C, 5 min) (in Biorad T100™ Thermal Cycler).
One-dimensional gel electrophoresis was performed to visualize amplification by preparing 1.5% agarose gels: 1.5 gram Seakern® LE Agarose powder for gelelectrophoresis (Lonza) has been added to 300 mL Tris acetaat-EDTA buffer 1x from Biograd. The liquid was heated, followed by adding 15 µL of 1% ethidium bromide (Merck) and pouring the mixture into the gel template to polymerize.
The gels were loaded in the gel electrophoresis box with 10 µL amplification product and 10 µL 6x DNA loading dye (Thermo Fisher) in each well. Therafter, 5 µL of a 100 base pair DNA ladder (Promega) was added to one of the wells in each row. Gel electrophoresis was run on 100V for 90 minutes and results have been visualized using ultra violet light in a gel documentation system (Vacutec SynGene G-box).
The gels were screened for bright DNA amplicons with the size of approximately 300 basepairs (bp) as *esxA* and *esxB* were expected to be between 265-288 bp and 255-303 bp respectively[6,66,67]. In preparation of DNA extraction from the gels, PCR was repeated on the isolates that show DNA bands around 300 bp, but this time without the loading dye to gain purified DNA for sequencing. The PCR and gel electrophoresis were furthermore performed according to the description in §2.1.3, except for the use of a different mastermix: Taq2x Mastermix RED 1,5 mM MgCl2.

Table 1: In this table an overview is given of the bacterial cultures used in the screening for *esxA* and *esxB.* Background information about collection of these isolates can be found in Gcebe et al. (2013, 2015 and 2017)[6,13,17].

Mycobacterial species	**Isolate ID (or strain)**	**Isolate origin**	**Number of slopes**
Examined NTM species			
M. moriokaense	Lucingweni swab 4	Swab	1
M. moriokaense	Smithfied H2O	Water	3
M. moriokaense	ACTT 43059	Reference strain	2
M. moriokaense-like mycobacterial species	TB 6607	Buffalo pharyngeal swab	1
M. malmesburri sp. *nov*. strain WCM 7299	C4	Buffalo pharyngeal swab	9
M. malmesburii sp. *nov*.	Western cape swab 5	Bovine nasal swab	1
M. vaccae/M. vanbaalenii	Xhongora swab 5	Swab	2
M. vaccae/M. vanbaalenii	Botharust swab 3	Swab	1
M. vaccae/M. vanbaalenii	Kabisa S3	Soil	3
M. vaccae/M. vanbaalenii	Pan2S3	Soil	2
M. madagascariense	Ladybrand swab 3	Swab	1
M. madagascariense	Virginia swab 3	Swab	2
M. acapulcensis	Kanaland swab 2	Swab	2
M. acapulcensis	Honing S4	Soil	2
M. elephantis	Rustenburg kraal swab 1	Swab	9
M. elephantis	Rustenburg welgevonden swab 3	Swab	6
M. chitae	Honing r116 (yellow)	Swab	1
M. confluentis	Jakkalsfarm S2	Soil	2
M. paraffinicum	Molabedju H2O	Water	1
M. paraffinicum	Langkul S1	Soil	2
M. neonarum	Mahlamvu swab 2	Swab	1
M. neonarum	Xhongora swab 4	Swab	2
M. engbackii	Xhongora S1	Soil	1
M. engbackii	Kabisa S1	Soil	2
M. parafortuitum	Ncambele swab 4	Swab	2
M. parafortuitum	Cacadu swab 3	Swab	1
M. septicum/peregrinum	Ladybrand S2	Soil	3
M. smegmatis	ATCC 14468	Reference strain	2
M. fortuitum	ATCC 6481	Reference strain	2
M. komanii sp. *nov*. strain GPK 1020	Trigaarspoort swab 03027	Bovine nasal swab	1
M. nonchromogenicum strain NCK 8460	KonoS1	Soil	1
Positive controls			
M. tuberculosis	TB 8993A/E	Vervet monkey	1
M. bovis	TB 1054C	Cattle	1
M. bovis	TB 9243D	Buffalo	1
M. kansasii	ATCC 12487	Reference strain	1

2.4 DNA extraction (from gels)

For the purpose of sequencing, selected amplificated products were as accurate as possible excised from the gel using a clean scalpel during transillumination (Spectroline UV Transulliminator, Model T312A). Weight of every excised DNA gel sample was calculated by substracting the weight of the empty tubes and samples were stored overnight in the fridge until further processing.
DNA was extracted from the gel slides following a Quick-Start protocol (QIAquick® Gel Extraction Kit). The protocol is as follows: Add 3 volumes Buffer QG to 1 volume gel (100 mg gel is approximately 100 µL) with maximally 400 mg gel per spin column; incubate in a heating block at 50°C for 10 min while vortexing every 2-3 minutes to dissolve the gel; add and mix 1 volume isopropanol (100%); apply the sample to a QIAquick column that is placed in a 2 mL collection tube; centrifuge for 1 minute at 17900 x g (13000 rpm) in a conventional table-top microcentrifuge; discard flow-through; samples that are bigger than >800 µL should be loaded and spinned again; adding of 500 µL Buffer QG to QIAquick column; centrifuge 1 min (13000 rpm); discard flow-through; add 750 µL Buffer PE to column (add 96-100% ethanol to Buffer PE before use); let it stand for 2-5 min after addition of Buffer PE and centrifuge again (1 min, same rpm) to wash sample; discard flow-through; centrifuge 1 min (13000 rpm) to remove residual wash buffer; put a QIAquick column into a clean 1.5 mL microcentrifuge tube; add 50 µL Buffer EB (10 mM Tris Cl, pH 8.5) to the center of QIAquick membrane; wait for 1-4 min; centrifuge the column for 1 min (13000 rpm); store extracted DNA in a freezer at -20°C (50 µL in each tube).
One-dimensional gel electrophoresis on the extraction product has been performed to check whether DNA was truly extracted (following the same protocol as described previously). Samples (10 µL) were mixed with 10 µL loading dye (Gel loading dye, purple 6x, no SDS, BioLabs) and loaded into the wells of the gel. Gels were run at 90V for 2 hours.

2.5 Sequencing DNA fragments

Purified amplicons of selected DNA samples (Table 1) were sent to Inqaba Biotechnical Industries (Pty) Ltd. (Pretoria, South Africa) for Sanger sequencing using both reverse and forward primers that were used for amplification. This resulted in sequences visualized by chromatograms which were manually cleaned and edited. For the sequences derived by amplification with reverse primers, the reverse complement sequences were created with help of BioEdit Sequence Alignment Editor (version 7.0.5.3). Subsequently, these reverse complement sequences were aligned pairwise with the sequences derived by amplification with forward primers and consensus sequences were established with BioEdit as well. The final sequences were analyzed with the Basic Local Alignment Search Tool (BLAST, from the NCBI, available via https://blast.ncbi.nlm.nih.gov/Blast.cgi) to search for regions of genetic similarity between biological sequences.

2.6 Phylogenetic analyses

Three phyogenetic trees have been constructed: one based on ESAT-6 sequences (265-288 bp), one on CFP-10 sequences (255-303 bp) and one tree based on the partial 16S rRNA gene sequences (577 bp)[6,66,67]. Molecular Evolutionary Genetics Analysis (MEGA) platform (version 7.0.26, https://www.megasoftware.net/)[68] was used to perform phylogenetic analyses of isolates. Sequences of some other isolates that have been identified previously were included as well[55]. Sequences were trimmed at both ends and multiple alignments of sequences were performed by Clustalw[69]. Construction of trees was done by the neighbor-joining method[70]. One thousand bootstrap replicates were run and evolutionary distances were calculated using the Maximum Composite Likelihood method[71]. The sequence of *M. leprae* TN (AL583917.1) has been used as an out-group species for the phyogenetic trees based on ESAT-6 and CFP-10, while the sequence of *Nocardia farcinica* IFM 11285 (AB638765) has been used as te out-group in the partial 16S rRNA based tree. The phylogenetic relationship of the investigated mycobacteria and those of other mycobacterial species whose ESAT-6 and CFP-10 sequences could be retrieved from the NCBI nucleotide database have been explored as was done previously[55]. The 16S rRNA sequences of mycobacterial species were derived from the NCBI nucleotide database

(https://www.ncbi.nlm.nih.gov/nucleotide), while ESAT-6 and CFP-10 sequences from additional mycobacterial species were derived from Gcebe et al. (2018)[55].
Since no software was accessible to compare the phylogenetic trees on similarity within subclusters or clades in trees with different taxa numbers[72-75], DNA sequences of mycobacterial species that did not occur in all three phylogenetic trees were deleted. Only 33 mycobacterial species were left that overlap between the trees and new phylogenetic trees were created with MEGA as described above. These smaller phylogenetic trees were exported from MEGA, converted into Newick tree format and imported to the software Mesquite (version 3.61, https://www.mesquiteproject.org/home.html)[76]. Similarity of the smaller trees was expressed by several (dis)similarity indices: shared partitions and shared clades, patristic distance correlation, average differences of values associated with nodes and fraction of clades that are shared. A concensus tree of the three phylogenetic trees was created by majority rule consensus, semistrict consensus and strict consensus in Mesquite.

2.7 Data analysis

BLAST searches for regions of genetic similarity between the derived potential *esxA* or *esxB* sequences and sequences in the NCBI database were performed. The highest BLAST hits were selected based on the percentage identity and the expect value (e-value). The percentage identity is a number that describes the similarity of the query sequence to the target sequence based on the number of identical characters (bases) in each sequence. Therefore, the higher the percentage identity (or similarity), the more significant the match is.
The e-value is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size. (For example, an e-value of '1' assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance). The e-value could be considered to describe the random background noise. The lower the e-value, or the closer it is to zero, the more "significant" the match is. However, short alignments have relatively high e-values, because the shorther the length of the sequences, the higher the probability of occurig in the database by chance. (Cited from National Center for Biotechnology Information (NCBI), Basic Local Alignment Search Tool (BLAST). Frequently Asked Questions Website. Accessed in May 2018 and available via: https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect [77])

3. RESULTS

3.1 Polymerase Chain Reactions (PCR)

Positive controls appeared clearly positive and negative controls appeared negative, indicating the suitability of the PCR protocol.
The regions of interest were approximately 300 bp (265-288 bp and 255-303 bp for *esxA* and *esxB* respectively)[6,66,67]. Samples from *M. elephantis* (Rustenburg kraal swab 1, no.1-3,6-9, Rustenburg welgevonden swab 3, no. 1,2,4-6)*, M. acapulcensis* (Kanaland swab 2, no. 1 and 2)*, M. neonarum* (Xhongora swab 4 no. 1 and 2) doubtful, *M. vaccae/M.vanbaalenii* (Xhongora swab 5, no. 1 and 2)*, M. madagascariense* (Virginia swab 3, no. 2, Ladybrand swab 3)*, M. engbackii* (Kabisa S1 no.1)*, M. moriokaense* (ACTT 43059 no. 1 and 2)*, M. smegmatis* (ACTT 14468, no. 1 and 2)*, M. malmesburii* sp. *nov.* (Western cape swab 5)*, M. paraffinicum* (Langkul no. 1 and 2)*, M. moriokaense*-like mycobacterial species (TB 6607), *M.tuberculosis* (TB 8893A)*, M. bovis* (TB 9243D and TB 1054C) all amplified DNA fragments of approximately 300 bp in the *esxA*-PCR.

In the *esxB*-PCR, samples from *M. malmesburri* (C4. no. 1-4)*, M. elephantis* (Rustenburg kraal no.5-7 and 9) doubtful, *M. elephantis* (Rustenburg welgevonden no. 1 and 2) doubtful, *M. neonarum* (Xhongora swab 4, no. 1 and 2 and possibly isolate Mahlamvu)*, M. engbackii* (Xhongora S1 and Kabisa S1 no. 2)*, M. vaccae/M. vanbaalenii* (Xhongora swab 5 no. 2, Botharust swab 3)*, M. acapulcensis* (Kanaland swab 2 no. 1 and 2 and isolate Honing S4 no. 2)*, M. moriokaense* (ACTT 43059 no. 1-3)*, M. fortuitum* (ATCC 6481 no. 1 and 2)*, M. nonchromogenicum* (KonoS1)*, M. parafortuitum* (Ncambele swab 4 no. 2)*, M. madagascariense* (Ladybrand swab 3) doubtful, *M. septicum/peregrinum* (Ladybrand S2 no. 3)*, M. paraffinicum* (Langkul no. 1 and 2)*, M. moriokaense*-like mycobacterial species (TB 6607) doubtful, *M.tuberculosis* (TB 8993A), *M. bovis* (TB 1054C and 9243D) amplified DNA fragments of approximately the same size.
Therefore, these isolates could be potentially positive for the *esxA* and *esxB* genes.

Furthermore, in both the *esxA*- and *esxB*-PCRs, many samples showed several additional amplification products. These appeared sometimes vague and sometimes more intense. The PCR products occured around or below 100-200 bp as well as higher along the base pair DNA ladder. Sometimes the amplified fragments were a multiple of 300 bp in size (600 bp or 900 bp), but PCR products of different sizes occurred as well.

3.2 Selected amplification products for sequencing

To investigate which DNA sequences has been amplified, some PCR products were selected for sequencing.
Therefore, PCRs were first repeated without adding the dye necessary for DNA sequencing. Results of these repeated *esxA*- and *esxB*-PCRs were quite similar to the previous performed *esxA*- and *esxB*-PCRs.
Table 2 shows the selected mycobacterial species, isolates and DNA amplicons. Amplification products were chosen to be submitted for sequencing based on their size and intensity: DNA samples derived from the 300 bp sized PCR products that gave a clear signal were selected to send in for sequencing as this DNA might include *esxA* or *esxB* sequences. Also, some DNA amplicons with a size that is a multiple of 300 bp, were submitted for sequencing to check for gene duplications. As it was expected that insufficient DNA would be present for sequencing in samples that were visualized by very faint PCR amplicons, these samples were excluded for sequencing. Sometimes, very intense PCR amplicons at other sizes were included too to figure out which DNA has been amplified in the PCRs. It was aimed to include as many as possible different mycobacterial species in the selection. Finally, 13 different *esxA*-PCR products and 14 *esxB*-PCR amplicons (27 amplicons in total) were selected for sequencing.

Table 2: NTM-isolates and amplification products selected for sequencing. # = amplicons selected for sequencing

Mycobacterial species **(isolate)**	**Produced amplicons of interest in *esxA*-PCR** (estimated size in base pairs)	**Produced amplicons of interest in *esxB*-PCR** (estimated size in base pairs)
M. moriokaense (Smithfield H2O no. 2)		500 bp
M. malmesburri sp. *nov*. strain WCM 7299 (C4 no. 3)	600 bp #	<300 bp >300 bp
M. elephantis (Rustenburg kraal 1 swab 1 no. 7)	300 bp 750 bp #	300 bp
M. elephantis (Rustenburg welgevonden swab 3 no. 3)	600 bp # 800 bp #	
M. elephantis (Rustenburg welgevonden swab 3 no. 2)		250 bp 300 bp # 600 bp
M. acapulcensis (Kanaland swab 2 no. 2)	300 bp # 950 bp	300 bp #
M. neonarum (Xhongora swab 4 no. 1)	300 bp 600 bp 700 bp #	300 bp # 650 bp
M. vaccae/M. vanbaalenni (Xhongora swab 5 no. 2)	300 bp #	300 bp #
M. engbackii (Xhongora S1)		300 bp #
M. madagascariense (Virginia swab 3 no. 2)	300 bp 500 bp #	400 bp # 600 bp #
M. engbackii (Kabisa S1 no. 1)	300 bp	
M. engbackii (Kabisa S1 no. 2)		300 bp #
M. moriokaense (ACTT 43059 no. 1)		300 bp 400 bp #
M. moriokaense (ACTT 43059 no. 2)	300 bp # 500 bp # 900 bp #	
M. vaccae/M. vanbaalenni (Botharust swab 3)	600 bp 700 bp	450 bp #
M. komanii sp. *nov*. strain GPK 1020 (Trigaarspoort swab 03027)		450 bp
M. malmesburii sp. *nov.* (Western cape swab 5)	300 bp # 600 bp #	450 bp #
M. nonchromogenicum (KonoS1)	400 bp 600 bp	500 bp 600 bp
M. parafortuitum (Ncambele swab 4 no. 1)	650 bp	
M. parafortuitum (Ncambele swab 4 no. 2)		300 bp #
M. paraffinicum (Langkul S1 no.1)	450 bp 750 bp	300 bp #
M. paraffinicum (Langkul S1 no.2)	300 bp	300 bp
M. madagascariense (Ladybrand swab 3)	300 bp 500 bp 600 bp 800 bp	350 bp #
M. moriokaense-like mycobacterial species (TB 6607)	300 bp	300 bp

3.3 One-dimensional gel electrophoresis at extracted DNA from gels

Finally, one-dimensional gel electrophoresis has been run at the DNA extracted from gels to confirm whether DNA extraction from the gels had been successful. The results are shown in figures 1 and 2. Visualization with the gel documentation system showed clearly separated PCR products from variable intensity. Some samples resulted in very intense DNA amplicons whereas others resulted in very faint DNA amplicons. Nearly always one DNA amplicon per sample was visible and PCR products appeared approximately at the expected levels (corresponding to selected DNA base pair sizes). Some samples did not show any clear amplification product. However, DNA extraction had been successful for most of the samples.

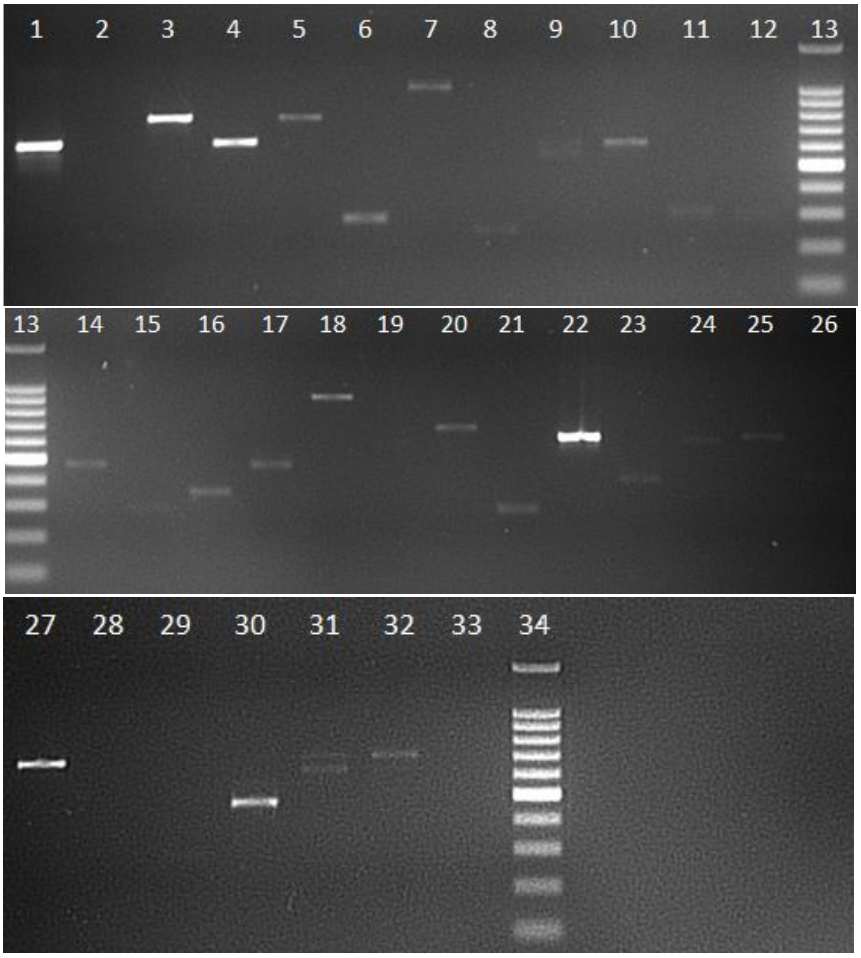

Figure 1: Agarose gel (1.5%) depicting DNA products following PCR amplification of 13 NTM isolates (32 amplicons) with *M. bovis esxA* primers.
1: *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no. 3: 600 bp); 2: *M. elephantis* (Rustenburg kraal 1 no. 7: 300 bp); 3: *M. elephantis* (Rustenburg kraal 1 no. 7: 750 bp); 4: *M. elephantis* (Rustenburg welgevonden no. 3: 600 bp); 5: *M. elephantis* (Rustenburg welgevonden no. 3: 800 bp); 6: *M. acapulcensis* (Kanaland no. 2: 300 bp); 7: *M. acapulcensis* (Kanaland no. 2: 950 bp); 8: *M. neonarum* (Xhongora swab 4 no. 1: 300 bp); 9: *M. neonarum* (Xhongora swab 4 no. 1: 600 bp); 10: *M. neonarum* (Xhongora swab 4 no. 1: 700 bp); 11: *M. vaccae/M. vanbaalenii* (Xhongora swab 5 no. 2: 300 bp); 12: *M. madagascariense* (Virginia swab 3 no. 2: 300 bp); 13: 100 base pair DNA ladder (Promega); 14: *M. madagascariense* (Virginia swab 3 no. 2: 500 bp); 15: *M. engbackii* (Kabisa S1 no. 1: 300 bp); 16: *M. moriokaense* (ACTT no. 2: 300 bp); 17: *M. moriokaense* (ACTT no. 2: 500 bp); 18: *M. moriokaense* (ACTT no. 2: 900 bp); 19: *M. vaccae/M. vanbaalenii* (Botharust: 600 bp); 20: *M. vaccae/M. vanbaalenii* (Botharust: 700 bp); 21: *M. malmesburii* sp. *nov*. (Western Cape swab 5: 300 bp); 22: *M. malmesburii* sp. *nov*. (Western Cape swab 5: 600 bp); 23: *M. nonchromogenicum* strain NCK 8460 (KonoS1: 400 bp); 24: *M. nonchromogenicum* strain NCK 8460 (KonoS1: 600 bp); 25: *M. parafortuitum* (Ncambele no. 1: 650 bp); 26: *M. paraffinicum* (Langkul S1 no. 1: 450 bp); 27: *M. paraffinicum* (Langkul S1 no. 1: 600 bp); 28: *M. paraffinicum* (Langkul S1 no. 2: 300 bp); 29: *M. madagascariense* (Ladybrand swab 3: 300 bp); 30: *M. madagascariense* (Ladybrand swab 3: 500 bp); 31: *M. madagascariense* (Ladybrand swab 3: 600 bp); 32: *M. madagascariense* (Ladybrand swab 3: 800 bp); 33: *M. moriokaense*-like mycobacterial species (TB 6607: 300 bp); 34: 100 base pair DNA ladder (Promega).

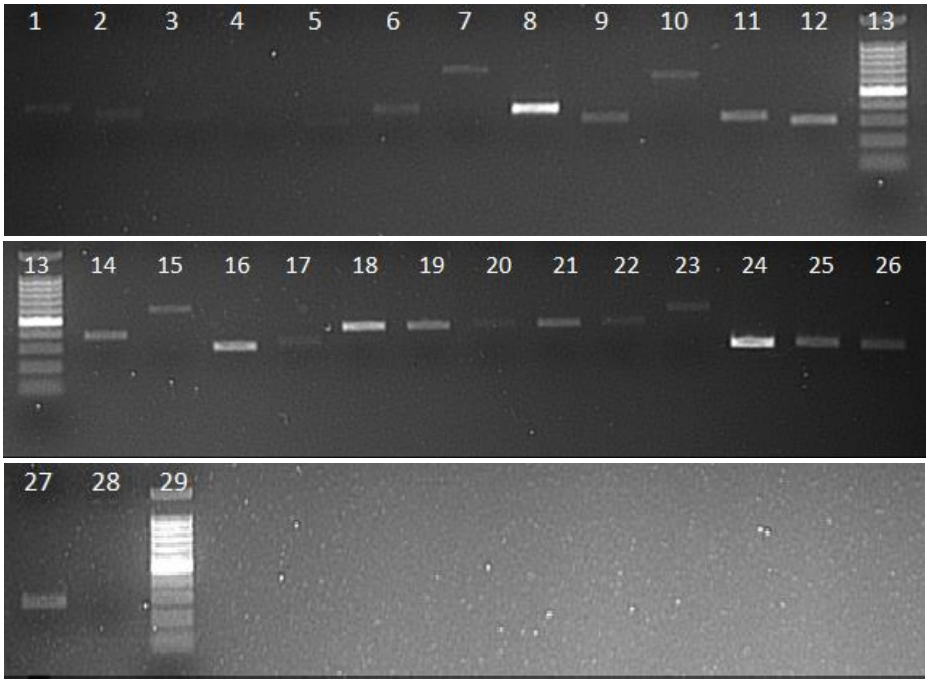

Figure 2: Agarose gel (1.5%) depicting DNA products following PCR amplification of 19 NTM isolates (27 amplicons) with *M. bovis esxB* primers. 1: *M. moriokaense* (Smithfield no. 2: 500 bp); 2: *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no. 3: >300 bp); 3: *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no. 3: <300 bp); 4: *M. elephantis* (Rustenburg kraal 1 no. 7: 300 bp); 5. *M. elephantis* (Rustenburg welgevonden no. 2: 250 bp); 6. *M. elephantis* (Rustenburg welgevonden no. 2: 300 bp); 7. *M. elephantis* (Rustenburg welgevonden no. 2: 600 bp); 8. *M. acapulcensis* (Kanaland no. 2: 300 bp); 9. *M. neonarum* (Xhongora swab 4 no. 1: 300 bp); 10. *M. neonarum* (Xhongora swab 4 no. 1: 650 bp); 11. *M. vaccae/M. vanbaalenii* (Xhongora swab 5 no. 2: 300 bp); 12: *M. engbackii* (Xhongora S1: 300 bp); 13: 100 base pair DNA ladder (Promega); 14: *M. madagascariense* (Virginia swab 3 no. 2: 400 bp); 15: *M. madagascariense* (Virginia swab 3 no. 2: 600 bp); 16: *M. engbackii* (Kabisa S1 no. 2: 300 bp); 17: *M. moriokaense* (ACTT no. 1: 300 bp); 18: *M. moriokaense* (ACTT no. 1: 400 bp); 19: *M. vaccae/M. vanbaalenii* (Botharust 450 bp); 20: *M. komanii* sp. *nov*. strain GPK 1020 (Trigaarspoort swab 03027: 450 bp); 21: *M. malmesburii* sp. *nov*. (Western cape swab 5: 450 bp); 22: *M. nonchromogenicum* strain NCK 8460 (KonoS1: 500 bp); 23: *M. nonchromogenicum* strain NCK 8460 (KonoS1: 600 bp); 24: *M. parafortuitum* (Ncambele swab 4 no. 2: 300 bp); 25: *M. paraffinicum* (Langkul S1 no. 1: 300 bp); 26: *M. paraffinicum* (Langkul S1 no. 2: 300 bp); 27: *M. madagascariense* (Ladybrand swab 3: 350 bp); 28: *M. moriokaense*-like mycobacterial species (TB 6607: 300 bp); 29: 100 base pair DNA ladder (Promega).

3.4 Sequencing and BLAST searches

Inqaba Biotechnical Industries reported that the sequence reaction of a few DNA samples failed (*M. elephantis* isolate Rustenburg welgevonden swab 3 no. 3 800 bp and *M. vaccae/M. vanbaalenii* isolate Xhongora swab 5 no. 2 300 bp with both forward and reverse ESAT-6 primers, *M. moriokaense* ACTT 43059 no. 2 500 bp with only the forward ESAT-6 primer and *M. malmesburii* sp. *nov*. Western cape swab 5 450 bp with the forward CFP-10 primer).
The final cleaned, aligned and created consensus sequences from the other DNA samples were analyzed with the Basic Local Alignment Search Tool (BLAST, from the NCBI, accessed in May 2018 and available via https://blast.ncbi.nlm.nih.gov/Blast.cgi). Complete results of the BLAST searches are shown in Tables A and B in appendices. Tables 3 and 4 show the BLAST results of sequences that appeared more or less similar to *esxA* or *esxB* in other mycobacterial species.

For *M. acapulcensis* (isolate Kanaland swab 2 no. 2, 300 bp) congruent results were found when BLAST searches were performed with the consensus sequence, reverse complement sequence and forward sequence from the *esxA*-PCR. (See appendix Table A). Highest hits were found in several strains of MTBC species (*M. tuberculosis* and *M. bovis* strains): two successive positioned hits of approximately 30 base pairs with 100% similarity to parts of the "type VII secretion system ESX-1 WXG100 family target ESAT-6". The two short sequences (31 and 29 base pairs respectively) were separated by 227 base pairs in the genomes of MTBC species. The consensus sequence of *M. acapulcensis* (Kanaland swab 2 no. 2) is given in appendix B.
The *M. acapulcensis* (Kanaland swab 2 no. 2) concensus sequence and *esxA* genes of *M. tuberculosis* (H37Rv) and *M. smegmatis (MC²-155)* were aligned by MEGA and identity is shown in appendices D and E.

Table 3: BLAST results of sequences derived from the *esxA*-PCR
* In several other *M. tuberculosis* strains and one *M. bovis* strain (2002/0476 complete genome) very similar results were found: two hits of only around 30 base pairs in size in each genome with 100% similarity and an E-value of 1e-04.

Mycobacterium species (isolate ID)	**Highest BLAST hit (accession number)**	**Sequence position in the genome of the highest hit**	**Identity/ similarity (%)**	**E-value**	**Product**
M. acapulcensis (Kanaland swab 2 no.2); 300 bp; consensus	*Mycobacterium tuberculosis* strain CAS chromosome (CP028428.1)*	4336363 to 4336394	100%	1e-04	type VII secretion system ESX-1 WXG100 family target ESAT-6
		4336621 to 4336650	100%	0.002	type VII secretion system ESX-1 WXG100 family target ESAT-6

Table 4: BLAST results of sequences derived from the *esxB*-PCR
a) Next to the three mycobacterial species mentioned in this table, another hundred very similar BLAST hits were found for several strains and isolates from *M. pseudoshottsii, M. liflandii, M. ulcerans, M. riyadhense, M. tuberculosis* and *M. bovis*.
b) *Mycobacterium ulcerans* isolate ITM842 does have two other titles in the NCBI database: *Mycobacterium ulcerans* isolate ITM7922 10 kDa culture filtrate protein (*esxB*) gene, complete cds

Mycobacterium species (isolate ID)	**Highest BLAST hit (accession number)**	**Sequence position in the genome of the highest hit**	**Identity/ similarity (%)**	**E-value**	**Product**
M. engbackii (Xhongora S1); 300 bp; forward	*Mycobacterium szulgai* strain NLA000501479 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014490.1)*[a]	39 to 303	87%	2e-77	10 kDa culture filtrate protein
	Mycobacterium kansasii strain NLA000601759 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014492.1)	43 to 303	87%	1e-73	10 kDa culture filtrate protein
	Mycobacterium marinum ATCC 927 DNA, complete genome (AP018496.1)	213621 to 213881	86%	6e-72	ESAT-6-like protein *esxB*
M. engbackii (Kabisa S1 no. 2); 300 bp; forward	*Mycobacterium szulgai* strain NLA000501479 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014490.1)*[a]	39 to 303	88%	4e-79	10 kDa culture filtrate protein
M. paraffinicum (Langkul S1 no. 2); 300 bp; forward	*Mycobacterium marinum* ATCC 927 DNA, complete genome (AP018496.1)	213619 to 213881	86%	5e-73	ESAT-6-like protein *esxB*
		6403504 to 6403766	86%	5e-73	ESAT-6-like protein *esxB*
	Mycobacterium marinum strain 1123 10 kDa culture filtrate protein (*esxB*) gene, complete cds (MF411158.1)	41 to 303	86%	5e-73	10 kDa culture filtrate protein
	Mycobacterium pseudoshottsii JCM 15466 DNA, complete genome (AP018410.1)	6008615 to 6008877	86%	5e-73	ESAT-6-like protein *esxB*
	Mycobacterium liflandii 128FXT, complete genome (CP003899.1)	6154651 to 6154913	86%	5e-73	Esat-6 like protein *esxB*
	Mycobacterium ulcerans isolate ITM842 10 kDa culture filtrate	41 to 303	86%	5e-73	10 kDa culture filtrate protein

	protein (*esxB*) gene, complete cds (EU257148.1)*[b]				

The sequences that were amplified in the *esxB*-PCR for both isolates of *M. engbackii* (Xhongora S1 and Kabisa S1 no. 2) resulted in the BLAST search to highest hits in several mycobacterial species for the 10kDA culture filtrate protein (CFP-10), also called "ESAT-6 like protein *esxB*". The mycobacterial species with highest similarity were: *M. tuberculosis, M. bovis, M. kansasii, M. szulgai, M. marinum, M. pseudoshottsii, M. liflandii, M. ulcerans* and *M. riyadhense.* The DNA fragments showed 86-88% similarity in sequence and were sized between 260-264 bp.
Amplified sequences in the *esxB*-PCR from *M. paraffinicum* (Langkul S1 no.2) led to comparable results in the BLAST search. Highest hits were found for *M. marinum* isolates, *M. pseudoshottsii, M. liflandii* and *M. ulcerans* for fragments of 262 bp (86% similarity) that were identified as the 10kDA culture filtrate protein (CFP-10) (or "ESAT-6 like protein *esxB*").
Sequences from *M. engbackii* (Xhongora S1 and Kabisa S1 no. 2) and *M. paraffincium* (Langkul S1 no.2) amplified with forward *esxB M. bovis* primers are given in appendix B.

3.5 Phylogenetic analysis

Three phylogenetic trees have been created from investigated and previous identified NTM. The phylogenetic trees are based on *esxA*, *esxB* and 16S rRNA sequences and are shown in figures 3,4 and 5 respectively.

As was found previously, RGM and SGM clustered together in the partial 16S rRNA based phylogenetic tree with the exception of *M. doricum* and *M. tusciae* (which are SGM but clustered with the RGM)[55]. In addition, the SGM *M. terrae* and *M. nonchromogenicum* clustered together with *M. engbackii, M. elephantis, M. holsaticum* and *M. malmesburii* within the RGM.
In the phylogenetic tree based on the *esxA* gene sequence the RGM and SGM clustered together as well, but only *M. vulneris* clustered with the RGM while it is classified as a SGM. In addition, the RGM *M. flouranthenivorans* clustered with the SGM.
In the phylogenetic tree based on the *esxB* gene sequence the RGM and SGM clustered together as well, with the following exceptions: *M. doricum, M. tusciae, M. nonchromogenicum* and *M. vulneris* which are SGM clustered with the RGM. In addition, *M. engbackii* clustered within the SGM while this mycobacterial species is classified as RGM.

Although the SGM and the RGM both roughly clustered together in all phylogenetic trees, there were similarities and differences in organization of the mycobacterial species within clades when the trees were studied in more detail. *Mycobacterium* sp. *JLS, mycobacterium* sp. *KMS* and *Mycobacterium* sp. *MCS* clustered together in all trees, as did *M. smegmatis* and *M. goodii* and did *M. elephantis* and *M. holstaticum*. However, there were differences in the subclustering of mycobacterial species as well. For example, *M. kansassi* and *M. riyadhense* clustered together in the *esxA*-based tree, but not in the *esxB*-based tree.

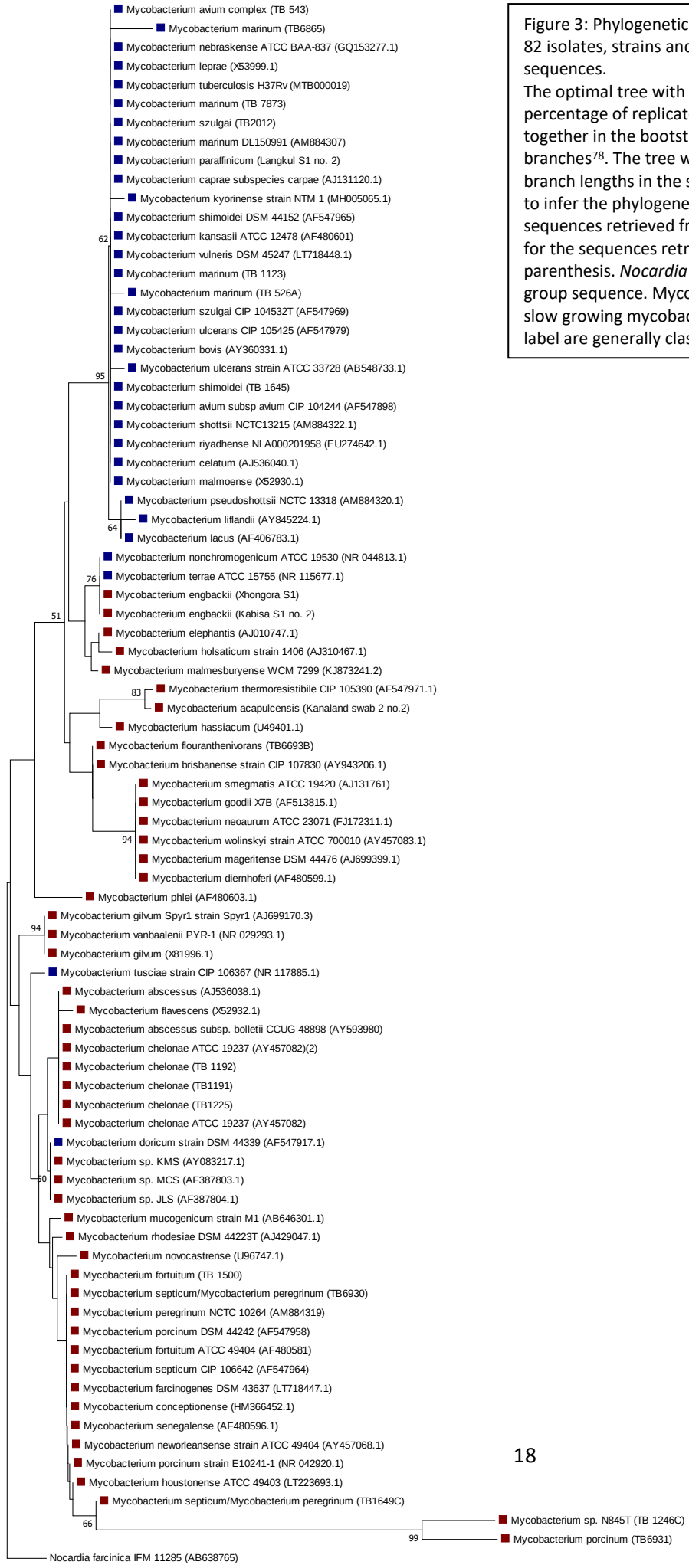

Figure 3: Phylogenetic relationship of 62 mycobacterial species (including in 82 isolates, strains and subspecies) based on the partial 16S rRNA gene sequences.
The optimal tree with the sum of branch length = 0.56654643 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree was drawn to scale (bar 0.02 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Either genebank accession numbers for the sequences retrieved from the database, or isolate identification numbers for the sequences retrieved from the ARC-OVI database are shown in parenthesis. *Nocardia farcinica* IFM 11285 (AB638765) was used as an out-group sequence. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

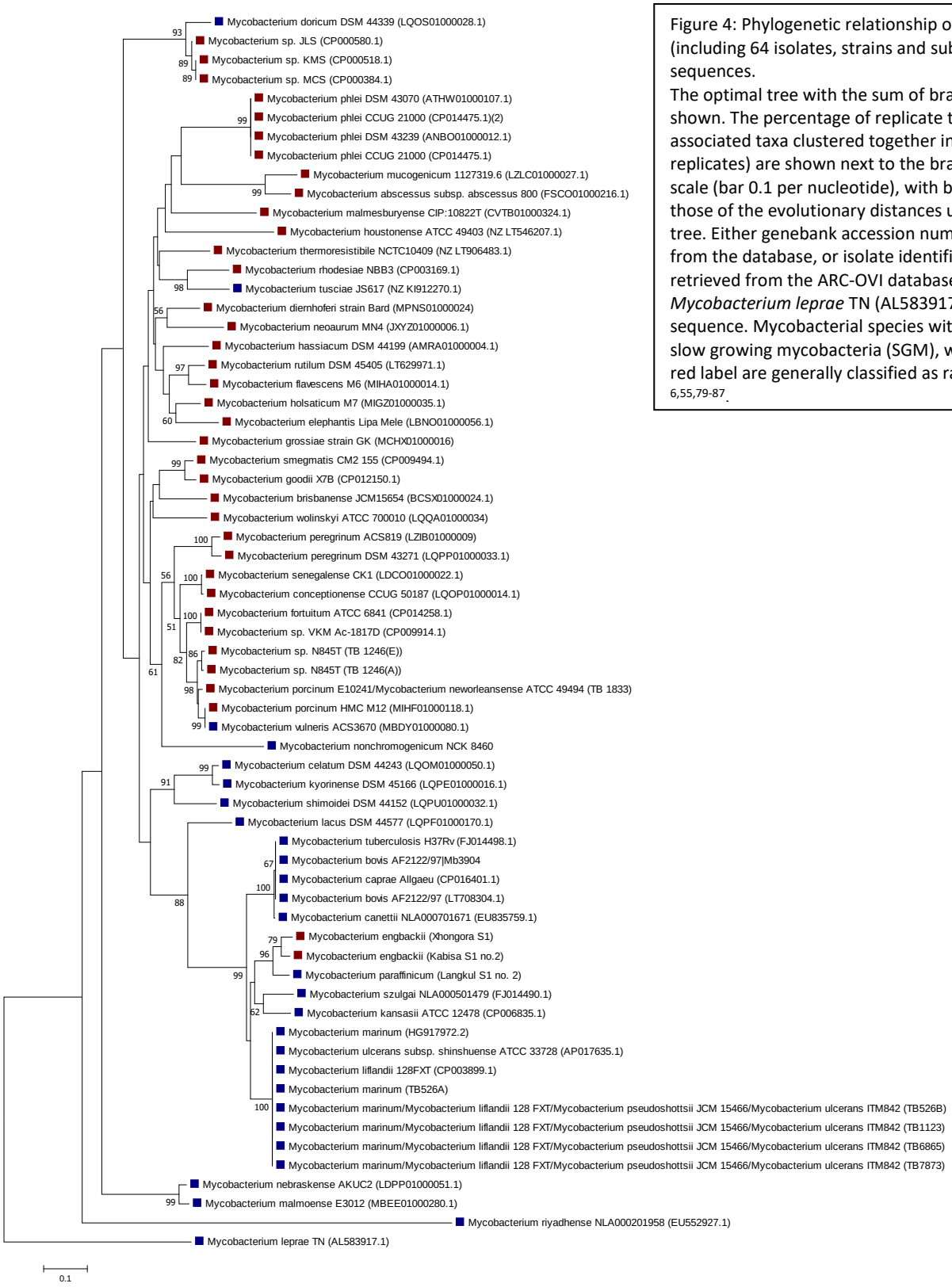

Figure 4: Phylogenetic relationship of 51 mycobacterial species (including 64 isolates, strains and subspecies) based on the *esxB* gene sequences.
The optimal tree with the sum of branch length = 6.20058143 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree was drawn to scale (bar 0.1 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Either genebank accession numbers for the sequences retrieved from the database, or isolate identification numbers for the sequences retrieved from the ARC-OVI database are shown in parenthesis. *Mycobacterium leprae* TN (AL583917.1) was used as an out-group sequence. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM) [6,55,79-87].

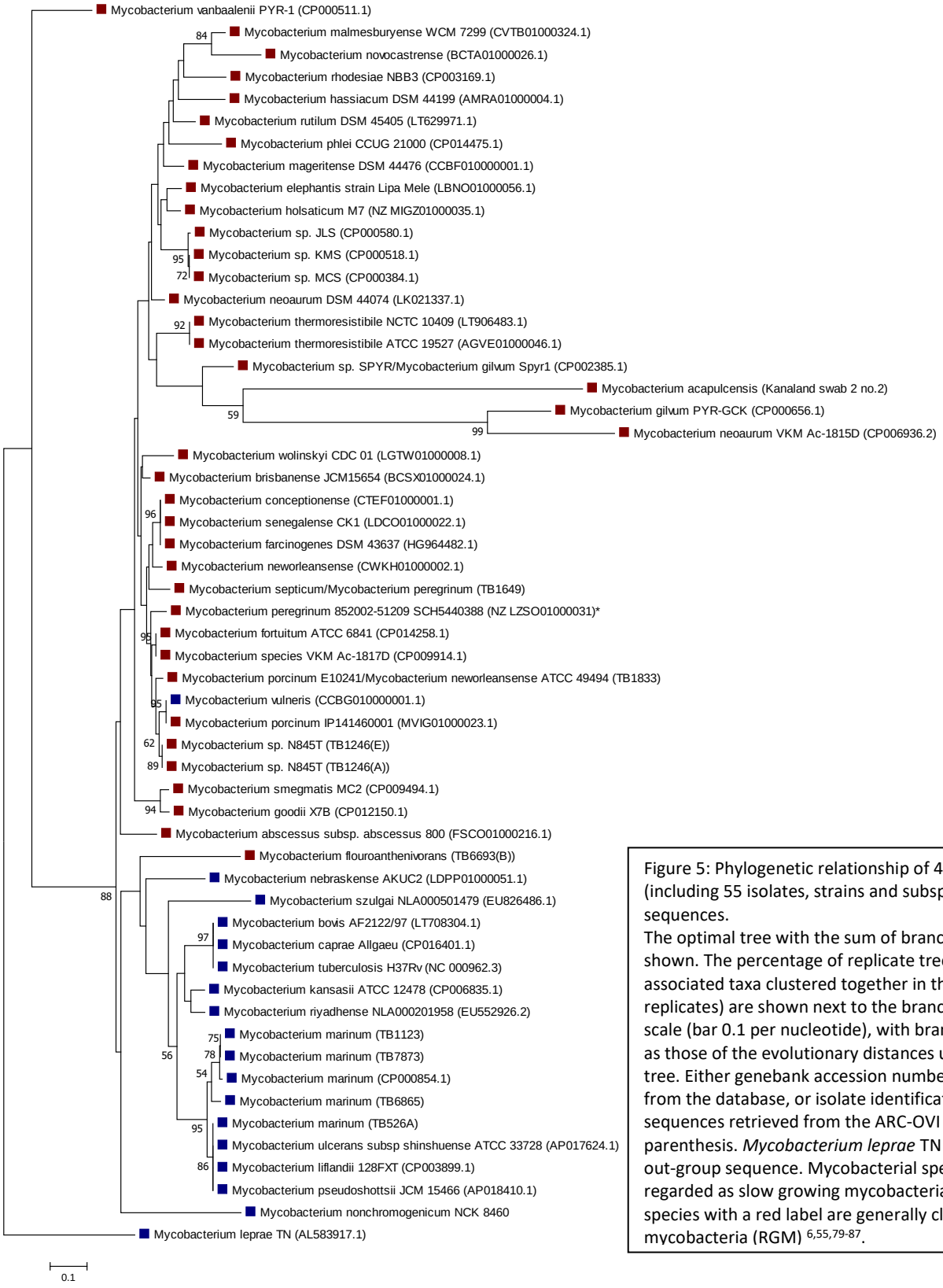

Figure 5: Phylogenetic relationship of 47 mycobacterial species (including 55 isolates, strains and subspecies) based on the *esxA* gene sequences.
The optimal tree with the sum of branch length = 6.59254275 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree was drawn to scale (bar 0.1 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Either genebank accession numbers for the sequences retrieved from the database, or isolate identification numbers for the sequences retrieved from the ARC-OVI database are shown in parenthesis. *Mycobacterium leprae* TN (AL583917.1) was used as an out-group sequence. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM) [6,55,79-87].

To estimate similarity or correlation between the phylogenetic trees based on the 16S rRNA, *esxA* and *esxB* genes, taxa that did not occur in all three trees were excluded. The phylogenetic trees consisting of 33 taxa are shown in figures D, E and F in appendix C.
Similarity of the smaller trees was expressed by several (dis)similarity indices which are shown in Tables 5a-d.

Table 5a: Patristic distance correlation refers to the correlation coefficient of the patristic distances between the two different trees. The patristic distance between two taxa reflects the sum of the branch lengths between them, also: the path-length distance along branches of the tree from one taxon to another. More similar trees have a correlation coefficient closer to 1, which is the maximum for identical trees[88].

Patristic distance correlation	16S rRNA based tree	*esxA* based tree	*esxB* based tree
16S rRNA based tree	1	0.43818924	0.34723786
esxA based tree	0.43818924	1	0.67043899
esxB based tree	0.34723786	0.67043899	1

Table 5b: Shared partitions refers to the number of shared partitions between two trees. A shared partition exists when branches in both trees separate the taxa into the same two groups. These are the same as shared clades on an unrooted tree. Shared clades refers to the number of shared clades between two trees (excluding the clade consisting of all taxa). More shared partitions and clades represent more similar phylogenetic trees[88].

Shared partitions/ clades	16S rRNA based tree	*esxA* based tree	*esxB* based tree
16S rRNA based tree	30	1	1
esxA based tree	1	30	13
esxB based tree	1	13	30

Table 5c: The fraction of clades that are shared refers to the fraction of the total number of clades in two trees that are shared between the trees (excluding the clade consisting of all taxa)[88].

Fraction of clades that are shared	16S rRNA based tree	*esxA* based tree	*esxB* based tree
16S rRNA based tree	1	0.033333333	0.033333333
esxA based tree	0.033333333	1	0.43333333
esxB based tree	0.033333333	0.43333333	1

Table 5d: Average differences of values associated with nodes refers to the average differences in values associated with nodes across those clades shared between two trees (excluding the clade consisting of all taxa)[88].

Average differences of values associated with nodes	16S rRNA based tree	*esxA* based tree	*esxB* based tree
16S rRNA based tree	0	0.274	0.257
esxA based tree	-0.2	0	0.05230769
esxB based tree	-0.257	-0.05230769	0

A concensus tree was created of the phylogenetic trees based on 16S rRNA-, *esxA*- and *esxB*- gene sequences including 33 mycobacterial species by majority rule consensus, semistrict consensus and strict consensus in Mesquite. The structure of the phylogenetic tree was the same if constructed using strict, semistrict or majority-rule consensus (figure 6).

Figure 6: Consensus tree based on the gene sequences of 16S rRNA, *esxA* and *esxB* created by Mesquite using strict, semistrict and majority-rule consensus. Mesquite creates strict concensus trees by presenting only those clades that are present in all of the trees in the tree source. Semistrict consensus trees present only those clades that occured in at least one of the trees and are not contradicted by any of the other trees. Majority-rule consensus trees present only those clades in the specified fraction (>0.50) of the trees. Tree weights are considered, so each tree contributes the value of the tree weight rather than 1, to the tabulation of the frequency of clades contained therein. In this case, each bootstrap replicate – rather than each tree – counts equally toward the clade frequencies. Because all the trees contained the same taxa, the results of the strict, semistrict and majority-rule concensus trees were the same[76]. Mycobacterial species in blue are regarded as slow growing mycobacteria (SGM), while mycobacterial species in red are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

4. DISCUSSION

To understand the role of NTM in immune responses in the context of TB diagnostics more fully, several NTM species have been screened to posess orthologues of *esxA* and *esxB* genes. These genes are coding for the immunogenic proteins ESAT-6 and CFP-10 respectively. The NTM species were screened for the presence of *esxA* and *esxB* orthologues using primers designed from *M. bovis.*
PCR results showed additional amplification products. DNA amplicons that appeared around or below 100-200 bp could be primer dimers. Primer dimers are primers that hybridize with each other and they thus produce low molecular weight amplicons at PCRs (short artifacts; an artifact is unnoticed amplification of a nonspecific product)[89]. Primer dimers are mainly due to homology between the primer sequences, but can also be influenced by primer concentrations, template concentrations and temperatures during the PCR[89,90]. However, also longer artifacts were visible at the PCR gels. Long artifacts consist of off-target amplification products which contain additional sequences that do not or only partially overlap with the targeted sequence. This nonspecific amplification could be influenced by primer design, concentrations of primers, low or too high template (or non-template DNA) concentrations, low annealing temperature and time between the steps of the PCR process[89]. Although artifacts can result in false positive results in case artifacts occur at the exact spot where a positive result would be expected to appear, possible positive results were checked by sequencing and BLAST searches in this NTM screening. BLAST searches revealed that indeed DNA coding for many other different proteins was amplified as well. Some of these proteins are present in the ESX-loci as well, so possibly more of the mycobacterial ESX-locus than only the *esxA*/*esxB* genes appeared to be amplified.

It was hypothesized that amplification products sized 600 bp (or even 900 bp) could be fused gene duplicates (or triplicates) of the potentially present *esxA* or *esxB* orthologues. This was thought to be possible because separate gene duplicates occur in some mycobacterial species as was found in any case for *M. marinum*; see the BLAST results in Table 4. In the complete genome of *M. marinum* (ATCC 927 DNA) wherein the 262 bp sized 'ESAT-6-like protein *esxB*' occurs at two positions in the genome (213619-213881 and 6403504-6403766). Also, the presence of 23 homologues of the *esxAB* genes in 11 different loci is described after whole genome sequencing of the genome of *M. tuberculosis* (H37Rv)[56]. The *esx* gene couples are organized functionally as five duplications that are proposed to form the *esxA*B secretion apparatus called ESX- or type VII secretion system. The five ESX systems of *M. tuberculosis* were likely derived from gene duplication events (ESX-4, ESX-1, ESX-3, ESX-2 and finally ESX-5) and each ESX-system appears to have its own role in mycobacterial virulence, but the genetic composition and organization of the five ESX clusters is similar[56].
However, the hypothesis of fused gene duplicates (or triplicates) could not be confirmed by sequencing and BLAST searching in the following and other amplicons: *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no.3, 600 bp), *M. elephantis* (Rustenburg welgevonden swab 3, no. 2, 600 bp), *M. neonarum* (Xhongora swab 4, no. 1, 700 bp). *M. madagascariense* (Virginia swab 3, no. 2, 500 and 600 bp), *M. moriokaense* (ACTT 43059 no. 2, 500 and 900 bp) and *M. malmesburii* sp. *nov*. (Western cape swab 5 600 bp).

With the help of *M. bovis* derived *esxB*-primers, *esxB* orthologues were detected in both *M. engbackii* isolates (Xhongora S1 and Kabisa S1 no. 2) and in *M. paraffinicum* (isolate Langkul S1 no. 2). The *esxB* orthologues were 86-88% similar at nucleotide level to the genes coding for CFP-10 in other NTM-like *M. szulgai, M. marinum, M. pseudohottsi, M. ulcerans, M. liflandi* and *M. kansasii.* In previous research, these mycobacterial species (and exact the same isolates) were found to be negative for *esxB* (and *esxA*) orthologues when evaluated using *M. smegmatis* primers encoding the *esxB* gene[6].

BLAST searches for the 321 bp sized concensus sequence potentially coding for the *esxA* gene of *M. acapulcensis* (isolate Kanaland swab 2 no. 2) resulted in 100% similarity with two short sequences of the type VII secretion system ESX-1 WXG100 family target ESAT-6 of MTBC species (*M. tuberculosis*).

The two short sequences (31 and 29 base pairs respectively) are separated by 227 base pairs in the genomes of MTBC species (Table 3).
The *esxA* gene of *M. tuberculosis* (H37Rv) and *M. acapulcensis* (Kanaland swab 2 no. 2) concensus sequence were aligned by MEGA and identity is shown in appendix D. The 31 and 29 bp sized regions of the sequence of *M. acapulcensis* that are 100% similar with *M. tuberculosis* strain CAS (Table 3) overlap with the regions were the *M. bovis* derived *esxA* primers annealed (appendix D). The forward primer sequence (comprising 20 base pairs) was 100% identical with the first 20 bp of the (potential) *esxA* gene sequences of both *M. acapulcensis* and *M. tuberculosis.* However, next to the primer sequence an additional 11 base pairs were identical between the particular sequences of both mycobacterial species. The complement reverse primer sequence (comprising 20 base pairs) was 100% identical with the last 20 bp of the (potential) *esxA* gene sequences of *M. acapulcensis* and *M. tuberculosis* as well. An additional 9 base pairs were identical between the sequences of both mycobacterial species.
The DNA sequence of *esxA* of *M. tuberculosis* is 287 base pairs, while the potential *esxA* sequence of *M. acapulcensis* is 321 base pairs. When gaps (defined as missing nucleotides in the aligned sequence) are not counted and the measurement is related to the shortest sequence[91] (which is the *M. tuberculosis* H37Rv *esxA* sequence of 287 bp in this case), sequence identity is 48%. Sequence identity is 62% when the identical primer annealing regions are taken into account as well. However, with the exception of the primer annealing regions, no more than six consecutive bases were identical between the two sequences anywhere. Therefore, it is questionable whether *M. acapulcensis* possesses an *esxA* orthologue or not.
The 37 additional nucleotides in the *M. acapulcensis* consensus sequence, as compared to the *esxA* sequence of *M. tuberculosis*, could be the result of cleaning and editing the sequence manually from the chromatogram. The chromatogram contained many entangled hyperboles and the middle part of the sequence of *M. acapulcensis* (isolate Kanaland swab 2 no. 2) contained many ambigue bases that could not be identified with certainty (it contained nucleotide codes like R,Y,S,W,K,M or N instead of A,T,C and G)[92].
Alignment with *M. bovis* AF 2122/97 (Mb3905) *esxA* sequence (287 nucleotides)[93] resulted in similar sequence identity percentages (alignment not shown). Alignment with *M. smegmatis* MC^2-155 *esxA* sequence (287 nucleotides)[94] resulted in maximum sequence identity of 58% (appendix E).
So hypothetically, DNA samples of bad quality could have resulted in adding additional and ambiguous nucleotides to the sequence and by that way into an extended and altered nucleotide sequence of the potential original *esxA* gene orthologue in *M. acapulcensis* (isolate Kanaland swab 2 no. 2).
An alternative explanation of the BLAST search results (Table 3) is that the primers annealed to random regions of the genome of *M. acapulcensis* which are not coding for the ESAT-6 protein, covering a 321 bp region of *M. acapulcensis'* DNA.
Therefore, it will remain questionable whether *M. acapulcensis* possesses an *esxA* orthologue or not. Screening for an *esxA* gene orthologue with the help of *M. bovis* derived *esxA* primers could be repeated for *M. acapulcensis,* if possible in other isolates and cultures of *M. acapulcesis* as well.
In previous research, *M. acapulcensis* (this exact same isolate) was found to be negative for *esxA* (and *esxB*) orthologues when evaluated using *M. smegmatis esxA* primers[6].

Amplification of *esxA* and *esxB* gene sequences may thus be dependent on specific combinations of primers and NTM species, because the primer sequence derived from a particular mycobacterial species may not be specific for all NTM species. In appendix F the comparison of *esxA* and *esxB* primer sequences from *M. smegmatis* and *M. bovis* derived forward and reverse primers are shown. Identity percentages of *M. smegmatis* and *M. bovis* primer sequences differ from 35% to 70%, so differences in primer sequences could lead to failure of primer annealing in some primer and NTM species combinations. Therefore, negative results in the screening for *esxA* and *esxB* orthologues in these NTM species with the use of *esxA* or *esxB M. bovis* primers will not rule out the existence of these gene orthologues definitely. For future research, NTM screening on *esxA* and *esxB* gene

orthologues with primers derived from both slow growing (*M. bovis*) and rapid growing (*M. smegmatis*) derived primers could be performed. Alternatively, whole genome sequencing can be used to avoid primer bias, although this is a more expensive approach.
To disminish aspecific amplification and to improve the specificity of the *esxA*- and *esxB*-PCR's in the future, the PCR protocol could be adapted with regard to (non-)template DNA or primer concentrations, annealing temperature or otherwise primer design.

Other explanations for potential false negative results could be failure of DNA extraction from solid cultures required for PCR or suboptimal excision of ampificated products from the PCR gels to send in for sequencing This applied in any case to *M. elephantis* (Rustenburg kraal swab 1 no. 7 300 bp), *M. vaccae/M. vanbaalenii* (Botharust swab 3 600 bp), *M. paraffinicum* (Langkul S1 no. 1 and no. 2 300 bp) and MOTT (TB 6607 300 bp) in the *esxA*-PCR and to *M. elephantis* (Rustenburg kraal swab 1 no. 7 300 bp), *M. moriokaense* (ACTT 43059 no. 1 300 bp) and MOTT (TB 6607 300 bp) in the *esxB*-PCR. Amplification products at the original PCRs appeared quite faint or just above or below the expected size (300 or 600 bp).

Sequencing failure can lead to failure to detect potentially present *esxA* or *esxB* gene orthologues as well. Sanger sequencing of potential *esxB* DNA fragments was performed with the forward primer only, to perform the research cost-effectively. However, the reverse primer was expected not to be necessary as all potential *esxB* DNA fragments were sized 400 bp or shorter. Sequencing reaction failed in *M. elephantis* (Rustenburg welgevonden swab 3 no. 3 800 bp with both forward and reverse primers in the *esxA*-PCR), *M. vaccae/M. vanbaalenii* (Xhongora swab 5 no. 2 300 bp with both forward and reverse primers in the *esxA*-PCR), *M. moriokaense* (ACTT 43059 no. 2 500 bp with only the forward primer in the *esxA*-PCR) and *M. malmesburii* sp. *nov*. (Western cape swab 5 450 bp with the forward primer in the *esxB*-PCR). Sequencing failed probably because of insufficient amount of DNA template (could be the case for the samples that resulted in faint signals at the PCRs) or inhibitory contaminants in the samples[95].

Identification of the *esxA* gene in a *M. acapulcensis* isolate and *esxB* gene in *M. engbackii* and M. *paraffinicum* isolates confirms again that these genes do occur in non-pathogenic RGM species. Occurrence of *esxA* and *esxB* genes in non-MTBC mycobacteria was demonstrated previously in at least *M. kansasii*[25,26,58,59], *M. marinum*[25,26,58], *M. szulgai*[25,26], *M. smegmatis*[6,25,33,60], *M. riyadhense*[25,57], *M. flavescens*[25,26,57], *M. gastri*[6,26,61], *M. gordonae*[26], *M. avium*[60], *M. mageritense*[6], *M. vulneris*[6], *M. farcinogenes*[6], *M. leprae*[58,60,62], *M. septicum/M. peregrinum*[6], *M. fortuitum*[6,33], *M. nonchromogenicum*[6,33], *M. malmesburri* subsp. *nov.*[6,33] and *M. komanii* subsp. *nov.*[6,33]. *Mycobacterium* sp. JLS was mentioned to possess an *esxA* or *esxB* orthologue as well[6], although the original research paper could not be retrieved.
Protein products of *esxA* and *esxB*, ESAT-6 and CFP-10 respectively, are broadly applied as MTBC-specific markers in several TB diagnostics. The detection of *esxA* and/or *esxB* presence in non-pathogenic NTM species could cause theoreticaly false positive results in TB diagnostics in case ESAT-6 and/or CFP-10 are produced by these NTM species. For example, a study that used the Bovigam® assay in cattle showed decreased specificity of the assay in African buffalo.[28] Therefore ESAT-6 and CFP-10 expression in *M. engbackii, M. paraffinicum* and *M. acapulcensis* and their immunogenicity[26] need to be further investigated.
However, *M. enbackii* and *M. paraffinicum* have been isolated from soil samples (Table 1). Their presence in the environment does not necessarily signify their importance in animal infections and possible interference with TB diagnostics. *M. acapulcensis* originated from animal source (isolated from a swab). Animal exposure to *M. acapulcensis*, makes the questionable prescence of the *esxA* gene in this NTM species more relevant with regard to potential cross reactivity in TB diagnostics in case ESAT-6 will be expressed in *M. acapulcensis.*

In previous studies[55] sequence differences in *esxA* and *esxB* genes of different mycobacterial species have been assessed for the usefulness of phylogenetic classification and prediction of potential pathogenicity of mycobacterial species. Phylogenetic profiling based on *esxA* and *esxB* sequences resulted in SGM roughly separating from RGM. SGM are mostly pathogenic mycobacterial species, while RGM are not known to be pathogenic or are sometimes known to cause opportunistic infections only[10].
As can be seen in figure 3 and as found before[55],RGM and SGM clustered together in the phylogenetic tree based on the partial 16S rRNA sequences with the exception of *M. doricum* and *M. tusciae* (which are SGM but clustered with the RGM). The SGM *M. terrae* and *M. nonchromogenicum* clustered together with *M. engbackii, M. elephantis, M. holsaticum* and *M. malmesburii* within the RGM as well. In the study of Gcebe et al. (2018)[55] *M. nonchromogenicum* and *M. terrae* appeared at the border between SGM and RGM as well, whereas *M. engbackii, M. elephantis, M. holstaticum* and *M. malmesburii* were not included in that phylogenetic tree. Thus, the organization of the 16S rRNA phylogenetic tree is congruent with the previous study.

In the phylogenetic tree based on the *esxA* gene sequence depicted in figure 5 the RGM and SGM clustered together as well, but only *M. vulneris* clustered with the RGM while it is classified as a SGM. In addition, the RGM *M. flouranthenivorans* clustered with the SGM. In the study of Gcebe et al. (2018)[55], in addition to *M. vulneris*, the SGM *M. nonchromogenicum* clustered together with the RGM as well in the *esxA*-based tree. In figure 5 *M. nonchromogenicum* clustered to the SGM, but its branch could evenly have been swapped so that it seemed to cluster with the RGM, as was found in the study of Gcebe et al. (2018)[55]. In this last mentioned study the RGM *M. abscessus* subsp. *abscessus* and *M. phlei* were placed at the border between SGM and RGM. In figure 5 *M. abscessus* subsp. *abscessus* was placed at the border as well and *M. phlei* clustered within the RGM. *M. abscessus* subsp. *abscessus* and *M. phlei* clustered within the RGM in the 16SrRNA-based tree (figure 3), while they were not included in the 16S rRNA-based tree in Gcebe et al. (2018)[55]. The RGM *M. fluoranthenivorans* was clustered with the SGM at the border in the *esxA*-based tree in figure 3, while it was not included in the *esxA*- or 16S rRNA based tree in Gcebe et al. (2018)[55]. Thus, the organization of the *esxA*-based phylogenetic tree is congruent with the previous study.

In the phylogenetic tree based on the *esxB* gene sequence shown in figure 4 the RGM and SGM clustered together as well, with the following exceptions: the SGM *M. doricum, M. tusciae, M. nonchromogenicum* and *M. vulneris* clustered with the RGM. In addition, *M. engbackii* clustered within the SGM while this mycobacterial species is classified as RGM. In the study of Gcebe et al. (2018)[55] the same four SGM mycobacterial species clustered with the RGM. *M. engbackii* was not included in the *esxB*-based tree from Gcebe et al. (2018)[55]. Thus, the organization of the *esxB*-based phylogenetic tree is congruent with the previous study.

So, the organization of phylogenetic trees based on 16S rRNA, *esxA* and *esxB* sequences, found in this project is similar to the organization of those trees in Gcebe et al. (2018)[55]. Besides, phylogenetic trees based on 16S rRNA, *esxA* and *esxB* sequences resulted in quite similar classification of SGM and RGM, although they differed somewhat in the individual exceptions to this classification.
Sequence differences of *esxA* and *esxB* between mycobacterial species can be useful in their phylogenetic organization with regard to their growth rate as do 16S rRNA sequence differences. In addition, the phylogenetic classification of mycobacterial species based on *esxA* and *esxB* sequences can predict the potential pathogenicity of NTM species correctly in many cases. However, they do not predict NTM pathogenicity completely correct[55]. This could be due to the fact that the proteins of the type VII secretion system are not the only virulence factors in mycobacterial species[55,56].

Although the SGM and the RGM both clustered together in all phylogenetic trees, there are similarities and differences in organization of the mycobacterial species within subclusters when the trees are studied in more detail. For example, *Mycobacterium* sp*. JLS, Mycobacterium* sp*. KMS* and *Mycobacterium* sp*. MCS* cluster together in all trees, as do *M. smegmatis* and *M. goodii* and do *M. elephantis* and *M. holstaticum*. However, there are also differences in the subclustering of

mycobacterial species. *M. kansasii* and *M. riyadhense* clustered together in the *esxA*-based tree, but not in the *esxB*-based tree for instance. Therefore, subclustering within clades was analysed with the help of Mesquite software[76].
With the exception of the excluded taxa, the smaller phylogenetic trees (33 mycobacterial species) based on 16S rRNA, *esxA* and *esxB* gene sequences appear similar to their original bigger sized phylogenetic trees (that included 83, 56 and 65 mycobacterial species) (figures D, E and F in appendix C). This only supports the correctness of the performed analysis.
To express similarity of phylogenetic trees based on 16S rRNA, *esxA* and *esxB* gene sequences, several (dis)similarity indices were calculated. As can be seen in tables 5a-d phylogenetic trees based on *esxA* and *esxB* are more similar to each other than when these trees are compared to the 16S rRNA based phylogenetic tree.
Table 5a shows that the *esxA*- and *esxB*-based phylogenetic trees are highest correlated with regard to patristic distances (r = 0.67, which is a moderate-high to strong correlation). Also, the absolute number of shared partitions and clades between the trees are higher between *esxA*- and *esxB*-based trees than for one of these trees and the 16S rRNA based tree (Table 5b). The absolute numbers can be expressed as fractions of shared clades (Table 5c). Even *esxA*- and *esxB*-based trees share less than 50% of their clades with each other (0.43 shared clades). Table 5d includes the average differences of values associated with nodes (across those clades shared between two trees) and differences are again lowest between *esxA*- and *exB*-based trees (0,05).
The *esxA*- and *esxB*-based trees result in the highest similarity. The higher similarity between *esxA*- and *esxB*-based trees might be the result of the included strains of mycobacterial species included in the trees. Included strains in each phylogenetic tree with 33 taxa can be found in the description of figures D, E and F in appendix C. Because of availability of gene sequences, identical strains of a particular mycobacterial species are used more often in *esxA*- and *esxB*- based trees than in the 16S rRNA-based tree.

Figure 6 shows the concensus tree created from the phylogenetic trees based on 16S rRNA-, *esxA*- and *esxB*- gene sequences including 33 mycobacterial species by majority rule consensus, semistrict consensus and strict consensus. Only three mycobacterial species clustered together (*Mycobacterium* sp. *KMS*, *Mycobacterium* sp. *MCS* and *Mycobacterium* sp. *JLS*), which points at the lack of similar clades in all tree phylogenetic trees. This fits with the results of the comparison of each set of two trees above; as the 16S rRNA-based tree resemble less to any of the two other trees, the consensus tree based on the three trees will not show shared clades. Accordingly, in figure G in the appendix is visualized that more small clades occur in the consensus tree based on only *esxA* and *esxB* gene sequences. (The formed clades are: *M. conceptionense* and *M. senegalense*; *M. smegmatis* and *M. goodii*; *Mycobacterium* sp. *JLS*, *Mycobacterium* sp. *KMS* and *Mycobacterium* sp. *MCS*; *M. hassiacum, M. elephantis* and *M. holstaticum*; *M. marinum, M. ulcerans* and *M. liflandii*; *M. fortuitum, Mycobacterium* sp. *N845T, M. vulneris* and *M. porcinum*; *M. kansasii, M. tuberculosis, M. bovis* and *M. caprae*).

To be able to better compare phylogenetic trees based on the 16S rRNA, *esxA* and *esxB* gene sequences, gene sequences of similar mycobacterial strains should be unraveled and included in the tree analysis in the future.

5. CONCLUSION

To understand the role of NTM in immune responses in the context of TB diagnostics more fully, several NTM species have been screened to posess orthologues of *esxA* and *esxB* genes. These genes are coding for the immunogenic proteins ESAT-6 and CFP-10 respectively. With the help of *M. bovis esxA* and *esxB* primers, the occurrence of *esxB* orthologues were detected in *M. engbackii* and in *M. paraffinicum*. These mycobacterial species can thus be added to the list of non-pathogenic RGM species that possess *esxA* and/or *esxB* orthologues which were tought to be specific for pathogenic MTBC bacteria. In case of expression of the protein products of *esxA* and *esxB*, ESAT-6 and CFP-10 respectively, by these mycobacterial species, infections with those NTM can cause false positive results in TB diagnostics when the antigens will be sufficient immunogenic. ESAT-6 and CFP-10 are broadly applied as MTBC-specific markers in several TB diagnostics namely. Therefore, the expression of ESAT-6 and CFP-10 should be further investigated and kept in mind when designing immunodiagnostic assays for TB.

As *M. engbackii* and *M. paraffinicum* (and even the exact same isolates) were found to be negative for *esxA* or *esxB* orthologues when evaluated using *esxA* and *esxB M. smegmatis* primers in previous research, amplification of *esxA* and *esxB* gene sequences may thus be dependent on specific combinations of primers and NTM species. Therefore, negative results in the screening for *esxA* and *esxB* orthologues in these NTM species with the use of *M. bovis* primers will not rule out the existence of these gene orthologues definitely. Further screenings of NTM for *esxA* and *esxB* orthologues are recommended to be performed by primers derived from different mycobacterial species.

In previous studies sequence differences in *esxA* and *esxB* genes of different mycobacterial species have been assessed for the usefulness of phylogenetic classification and prediction of potential pathogenicity of mycobacterial species. Phylogenetic profiling based on *esxA* and *esxB* sequences resulted roughly in mostly pathogenic SGM separating from RGM which are not known to be pathogenic or are sometimes known to cause opportunistic infections only. Sequence differences of *esxA* and *esxB* between mycobacterial species can therefore be useful in their phylogenetic organization with regard to their growth rate as do 16S rRNA sequence differences. In addition, the phylogenetic classification of mycobacterial species based on *esxA* and *esxB* sequences can predict the potential pathogenicity of NTM species correctly in many cases. However, they do not predict NTM pathogenicity completely correct which could be due to the fact that the proteins of the type VII secretion system are not the only virulence factors in mycobacterial species.

Acknowledgements

The project was realized through a collaboration of three institutions: the Agricultural Research Council-OVI (Onderstepoort, South Africa), Department of Veterinary Tropical Diseases, University of Pretoria (Onderstepoort, South Africa) and Department of Infectious Diseases and Immunology, Utrecht University (Utrecht, The Netherlands).

The help of the following persons is gratefully acknowledged: prof. Victor Rutten arranged the possibility to participate in this research project, whereas Nomakorinte Gcebe, prof. Anita Michel and Elise van der Heijden committed to guiding me in conducting the research, discussing the interim results and think along in solving some problems.

Victor and Kuda were always willing to help with practical matters in the lab of the ARC-OVI.

References

1) Quinn, P.J., Markey, B.K., Leonard, F.C., FitzPatrick, E.S., Fanning, S., Hartigan, P.J. Chapter 23: Mycobacterium species. In: *Veterinary microbiology and microbial disease*. 2th ed. Chichester: John Wiley & Sons Ltd; 2011: 250-258.
2) World Health Organization. Global tuberculosis report 2019. *Geneva (Switzerland).* 2019.
3) Radostits, O.M., Gay, C.C., Hinchcliff, K.W., Constable, P.D. Chapter 19: Diseases associated with bacteria (IV): Diseases associated with *Mycobacterium* spp. In: *Veterinary Medicine: A textbook of the diseases of cattle, horses, sheep, pigs and goats*. 10th Edition. Philadelphia: Saunders Elsevier Ltd.; 2006: 1007-1044.
4) World Health Organization, Food and Agriculture Organization of the United Nations (FAO) and World Organization for Animal Health (OIE). Roadmap for zoonotic tuberculosis. 2017. Available via: http://www.who.int/tb/publications/2017/zoonotic_TB/en/
5) [OIE] - World Organization for Animal Health. (2012). Manual of diagnostic tests and vaccines for terrestrial animals: Mammals, birds and bees. *Biological Standards Commission, World Organization for Animal Health, Paris, France.* 2012. 7th ed.
6) Gcebe, N. The occurrence and molecular characterization of non-tuberculous mycobacteria in cattle, African buffalo (*Syncercus Caffer*) and their environments in South Africa and genomic characterization and proteomic comparison with *Mycobacterium bovis. [Thesis PhD]*. *Pretoria*, *SA*; *University of Pretoria*. 2015. doi: https://repository.up.ac.za/handle/2263/58682
7) Shamputa, I.C., Cho SangNae Lebron, J., Via, L.E. Chapter 1: Introduction and epidemiology of *Mycobacterium tuberculosis* complex in humans. In: *Tuberculosis, Leprosy and Mycobacterial Diseases of Man and Animals: The Many Hosts of Mycobacteria*; CABI. 2015: 1-16.
8) Rodriguez-Campos, S., Smith, N. H., Boniotti, M. B., & Aranaz, A. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: Implications for diagnostics and legislation of bovine tuberculosis. *Research in veterinary science*. 2014; *97*:S5-S19.
9) Gordon, S.V., Behr, M.A. Chapter 2: Comparative Mycobacteriology of the *Mycobacterium tuberculosis* Complex. In: *Tuberculosis, Leprosy and Mycobacterial Diseases of Man and Animals: The Many Hosts of Mycobacteria, Wallingford: CABI.* 2015: 17-29.
10) Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M., & van Pittius, N. C. G. Novel Mycobacterium tuberculosis complex pathogen, M. mungi. *Emerging infectious diseases*. 2010; *16*(8): 1296-1299. doi: 10.3201/eid1608.100314 [doi].
11) van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeone, R., Brosch, R., & van Soolingen, D. Characterization of *Mycobacterium orygis* as M. tuberculosis complex subspecies. *Emerging infectious diseases*. 2012;*18*(4), 653-655. doi: 10.3201/eid1804.110888 [doi]
12) Parsons, S. D., Drewe, J. A., van Pittius, N. C. G., Warren, R. M., & Van Helden, P. D. Novel cause of tuberculosis in meerkats, South Africa. *Emerging infectious diseases*. 2013; *19*(12): 2004-2007. doi: 10.3201/eid1912.130268 [doi]

13) Gcebe, N., Rutten, V., Gey van Pittius, N. C., & Michel, A. Prevalence and Distribution of Non-Tuberculous Mycobacteria (NTM) in Cattle, African Buffaloes (*Syncerus caffer*) and their Environments in South Africa. *Transboundary and emerging diseases*. 2013;*60*:74-84.
14) Cousins, D. V. Mycobacterium bovis infection and control in domestic livestock. *Revue Scientifique et Technique-Office International des Epizooties*. 2001; *20*(1), 71-85.
15) Kolb, N.R. Chapter 54: Tuberculosis. In: South-Paul, J.E., Matheny, S.C, Lewis, E.L. *Current diagnosis & treatment in family medicine.* 4th ed. Lange textbooks: McGraw-Hill Education; 2015: 570-581.
16) Wedlock, D. N., Skinner, M. A., de Lisle, G. W., & Buddle, B. M. Control of *Mycobacterium bovis* infections and the risk to human populations. *Microbes and infection*. 2002; *4*(4), 471-480.
17) Gcebe, N., & Hlokwe, T. M. Non-tuberculous mycobacteria in South African wildlife: neglected pathogens and potential impediments for bovine tuberculosis diagnosis. *Frontiers in Cellular and Infection Microbiology*. 2017; *7*, 15.
18) Mirsaeidi, M., Farshidpour, M., Allen, M.B., Ebrahimi, G., Falkinham, J.O. Highlights on advances in non-tuberculosis-like mycobacterial disease in North America. *BioMed Research International.* 2014.
19) Botha, L., Gey van Pittius, N. C., & Van Helden, P. D. Mycobacteria and disease in southern Africa. *Transboundary and Emerging diseases.* 2013; *60*, 147-156.
20) Gopinath, K., Singh, S. Non-tuberculosis-like mycobacteria in TB-endemic countries: are we neglecting the danger? *PloS Neglected Tropical Diseases*. 2010;4:e615.10.1371.
21) Mwikuma, G., Kwenda, G., Hang'Ombe, B. M., Simulundu, E., Kaile, T., Nzala, S., Suzuki, Y. Molecular identification of non-tuberculous mycobacteria isolated from clinical specimens in Zambia. *Annals of clinical microbiology and antimicrobials*. 2015;*14*(1): 1.
22) Primm, T. P., Lucero, C. A., & Falkinham, J. O. Health impacts of environmental mycobacteria. *Clinical microbiology reviews*. 2004;*17*(1), 98-106. doi: 10.1128/cmr.17.1.98-106.2004 [doi].
23) Bercovier, H., & Vincent, V. Mycobacterial infections in domestic and wild animals due to Mycobacterium marinum, M. fortuitum, *M. chelonae, M. porcicum, M. farcinogenes, M. smegmatis, M. scrofulaceum, M. xenopi, M. kansasii, M. simiae* and *M. genavense*. Revue scientifique et technique (International Office of Epizootics). 2001; *20*(1), 265-290.
24) Martín-Casabona, N., Bahrmand, A. R., Bennedsen, J., Thomsen, V.O., Curcio, M., Fauville-Dufaux, M., Feldman, K., Havelkova, M., Katila, M.L., Koksalan, K., Pereira M.F., Rodrigues, F., Pfyffer, G.E., Portaels, F., Urgell, J.R., Rusch-Gerdes, S., Tortoli, E., Vincent, V., Watt, B. (2004). Non-tuberculous mycobacteria: patterns of isolation. A multi-country retrospective survey. *The International Journal of Tuberculosis and Lung Disease*. 2004;*8*(10), 1186-1193.
25) van Ingen, J., Boeree, M. J., Kösters, K., Wieland, A., Tortoli, E., Dekhuijzen, P. R., & van Soolingen, D. Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. *International journal of systematic and evolutionary microbiology*. 2009; *59*(9), 2277-2282.
26) Vordermeier, H. M., Brown, J., Cockle, P. J., Franken, W. P., Drijfhout, J.W., Arend, S. M., Ottenhoff, T. H., Jahans, K., Hewinson, R. G. Assessment of cross-reactivity between Mycobacterium bovis and M. kansasii ESAT-6 and CFP-10 at the T-cell epitope level. *Clin. Vaccine Immunol.* 2007;*14*(9), 1203-1209. doi: CVI.00116-07 [pii].
27) Michel, A. L., Cooper, D., Jooste, J., De Klerk, L. M., & Jolles, A. Approaches towards optimising the gamma interferon assay for diagnosing Mycobacterium bovis infection in African buffalo (*Syncerus caffer*). *Preventive veterinary medicine.* 2011; *98*(2-3), 142-151.
28) Michel, A. L. *Mycobacterium fortuitum* infection interference with *Mycobacterium bovis* diagnostics: natural infection cases and a pilot experimental infection. *Journal of veterinary diagnostic investigation*. 2008; *20*(4), 501-503.
29) Schiller, I., Oesch, B., Vordermeier, H. M., Palmer, M. V., Harris, B. N., Orloski, K. A., Waters, W. R. Bovine tuberculosis: a review of current and emerging diagnostic techniques in view of

their relevance for disease control and eradication. *Transboundary and emerging diseases*. 2010; *57*(4), 205-220.

30) Poyntz, H. C., Stylianou, E., Griffiths, K. L., Marsay, L., Checkley, A. M., & McShane, H. Non-tuberculous mycobacteria have diverse effects on BCG efficacy against *Mycobacterium tuberculosis*. *Tuberculosis*. 2014; *94*(3), 226-237.
31) Brandt, L., Cunha, J. F., Olsen, A. W., Chilima, B., Hirsch, P., Appelberg, R., & Andersen, P. Failure of the *Mycobacterium bovis* BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis. *Infection and immunity*. 2002; *70*(2), 672-678. doi: 10.1128/iai.70.2.672-678.2002 [doi].
32) Weir, R. E., Black, G. F., Nazareth, B., Floyd, S., Stenson, S., Stanley, C., Crampin, A. C. The influence of previous exposure to environmental mycobacteria on the interferon-gamma response to bacille Calmette–Guérin vaccination in southern England and northern Malawi. *Clinical & Experimental Immunology*. 2006; *146*(3), 390-399.
33) Gcebe, N., Michel, A., Gey van Pittius, N. C., & Rutten, V. Comparative genomics and proteomic analysis of four non-tuberculous Mycobacterium species and *Mycobacterium tuberculosis* complex: occurrence of shared immunogenic proteins. *Frontiers in microbiology*. 2016; *7*:795.
34) Sakamuri, R.M., Moodley, P., Yusim, K., Feng, S.H., Sturn, A.W., Korber, B.T.M., Mukundan, H. Chapter 4: Current methods for diagnosis of human tuberculosis and considerations for global surveillance. In: Mukundan, H., Chambers, M., Waters, R., & Larsen, M. (Eds.). Introduction and epidemiology of *Mycobacterium tuberculosis* complex in humans. *CAB International.* 2015; 72-102.
35) Pollock, J. M., McNair, J., Welsh, M. D., Girvin, R. M., Kennedy, H. E., Mackie, D. P., & Neill, S. D. Immune responses in bovine tuberculosis. *Tuberculosis*. 2001; *81*(1-2), 103-107.
36) Welsh, M. D., Cunningham, R. T., Corbett, D. M., Girvin, R. M., McNair, J., Skuce, R. A., Pollock, J. M. Influence of pathological progression on the balance between cellular and humoral immune responses in bovine tuberculosis. *Immunology*. 2005; *114*(1), 101-111.
37) De la Rua-Domenech, R., Goodchild, A. T., Vordermeier, H. M., Hewinson, R. G., Christiansen, K. H., & Clifton-Hadley, R. S. Ante mortem diagnosis of tuberculosis in cattle: a review of the tuberculin tests, γ-interferon assay and other ancillary diagnostic techniques. *Research in veterinary science*. 2006; *81*(2), 190-210.
38) Lifson, A. R., Watters, J. K., Thompson, S., Crane, C. M., & Wise, F. Discrepancies in tuberculin skin test results with two commercial products in a population of intravenous drug users. *Journal of Infectious Diseases*. 1993; *168*(4), 1048-1051.
39) Buddle, B. M., McCarthy, A. R., Ryan, T. J., Pollock, J. M., Vordermeier, H. M., Hewinson, R. G., De Lisle, G. W. Use of mycobacterial peptides and recombinant proteins for the diagnosis of bovine tuberculosis in skin test-positive cattle. *Veterinary Record*. 2003; *153*(20), 615-620. doi: 10.1136/vr.153.20.615 [doi].
40) Cockle, P. J., Gordon, S. V., Hewinson, R. G., & Vordermeier, H. M. Field evaluation of a novel differential diagnostic reagent for detection of Mycobacterium bovis in cattle. *Clin. Vaccine Immunol.* 2006; *13*(10), 1119-1124. doi: CVI.00209-06 [pii].
41) Veerasami, M., Venkataraman, K., Karuppannan, C., Shanmugam, A. A., Prudhvi, M. C., Holder, T., Subramanian, B. M. Point of Care Tuberculosis Sero-Diagnosis Kit for Wild Animals: Combination of Proteins for Improving the Diagnostic Sensitivity and Specificity. *Indian journal of microbiology*. 2018; *58*(1), 81-92.
42) Aranaz, A., De Juan, L., Bezos, J., Álvarez, J., Romero, B., Lozano, F., Paramio, J. L., Lopez-Sanchez, J., Mateos, A., Domínguez, L. Assessment of diagnostic tools for eradication of bovine tuberculosis in cattle co-infected with *Mycobacterium bovis* and *M. avium* subsp. paratuberculosis. *Veterinary research*. 2006; *37*(4), 593-606.

43) Rupp, M. E., Schultz Jr, A. W., & Davis, J. C. Discordance between tuberculin skin test results with two commercial purified protein derivative preparations. *Journal of Infectious Diseases*. 1994;*169*(5), 1174-1175.
44) Pollock, J. M., Girvin, R. M., Lightbody, K. A., Neill, S. D., Clements, R. A., Buddle, B. M., & Andersen, P. Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Veterinary Record*. 2000;*146*(23), 659-665. doi: 10.1136/vr.146.23.659 [doi].
45) Buddle, B. M., Ryan, T. J., Pollock, J. M., Andersen, P., & de Lisle, G. W. Use of ESAT-6 in the interferon-γ test for diagnosis of bovine tuberculosis following skin testing. *Veterinary microbiology*. 2001; *80*(1), 37-46.
46) Vordermeier, H. M., Whelan, A., Cockle, P. J., Farrant, L., Palmer, N., & Hewinson, R. G. Use of synthetic peptides derived from the antigens ESAT-6 and CFP-10 for differential diagnosis of bovine tuberculosis in cattle. *Clin. Diagn. Lab. Immunol.* 2001; *8*(3), 571-578. doi: 10.1128/CDLI.8.3.571-578.2001 [doi].
47) Aagaard, C., Govaerts, M., Okkels, L. M., Andersen, P., & Pollock, J. M. Genomic approach to identification of Mycobacterium bovis diagnostic antigens in cattle. *Journal of clinical microbiology*. 2003; *41*(8), 3719-3728. doi: 10.1128/jcm.41.8.3719-3728.2003 [doi].
48) Waters, W. R., Nonnecke, B. J., Palmer, M. V., Robbe-Austermann, S., Bannantine, J. P., Stabel, J. R., Minion, F. C. Use of recombinant ESAT-6: CFP-10 fusion protein for differentiation of infections of cattle by *Mycobacterium bovis* and by *M. avium* subsp. avium and *M. avium* subsp. paratuberculosis. *Clin. Diagn. Lab. Immunol.* 2004; *11*(4), 729-735. doi: 10.1128/CDLI.11.4.729-735.2004 [doi].
49) Vordermeier, H. M., Pontarollo, R., Karvonen, B., Cockle, P., Hecker, R., Singh, M., Babjuk, L.A., Hewinson, R.G., van Drunen Littel-van Den, S. Synthetic peptide vaccination in cattle: induction of strong cellular immune responses against peptides derived from the *Mycobacterium bovis* antigen Rv3019c. *Vaccine*. 2005; *23*(35), 4375-4384.
50) Aagaard, C., Govaerts, M., Meikle, V., Vallecillo, A. J., Gutierrez-Pabello, J. A., Suarez-Güemes, F., McNair, J., Cataldi, A., Espitia, C., Andersen, P., Pollock, J. M. Optimizing antigen cocktails for detection of *Mycobacterium bovis* in herds with different prevalences of bovine tuberculosis: ESAT6-CFP10 mixture shows optimal sensitivity and specificity. *Journal of clinical microbiology*. 2006; *44*(12), 4326-4335. doi: JCM.01184-06 [pii].
51) Wiker, H. G. MPB70 and MPB83–major antigens of *Mycobacterium bovis*. *Scandinavian journal of immunology*. 2009; *69*(6), 492-499.
52) Vordermeier, H. M., Cockle, P. C., Whelan, A., Rhodes, S., Palmer, N., Bakker, D., & Hewinson, R. G. Development of diagnostic reagents to differentiate between *Mycobacterium bovis* BCG vaccination and *M. bovis* infection in cattle. *Clin. Diagn. Lab. Immunol.* 1999; *6*(5), 675-682.
53) Pollock, J. M., Douglas, A. J., Mackie, D. P., & Neill, S. D. Identification of bovine T-cell epitopes for three *Mycobacterium bovis* antigens: MPB70, 19,000 MW and MPB57. *Immunology*. 1994; *82*(1), 9-15.
54) Millington, K. A., Fortune, S. M., Low, J., Garces, A., Hingley-Wilson, S. M., Wickremasinghe, M., Lalvani, A. Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Sciences*. 2011; *108*(14), 5730-5735. doi: 10.1073/pnas.1015153108 [doi].
55) Gcebe, N., Michel, A. L., & Hlokwe, T. M. Non-tuberculous Mycobacterium species causing mycobacteriosis in farmed aquatic animals of South Africa. *BMC microbiology*. 2018; *18*(1):32.
56) Pieters, J., McKinney, J.D. *Pathogenesis of mycobacterium tuberculosis and its interaction with the host organism.* Vol.374. Heidelberg: Springer Science & Business Media; 2013; 120-125.
57) Harboe, M., Oettinger, T., Wiker, H. G., Rosenkrands, I. D. A., & Andersen, P. Evidence for occurrence of the ESAT-6 protein in *Mycobacterium tuberculosis* and virulent *Mycobacterium*

bovis and for its absence in *Mycobacterium bovis* BCG. *Infection and immunity.* 1996; *64*(1), 16-22.

58) Geluk, A., Van Meijgaarden, K. E., Franken, K. L. M. C., Wieles, B., Arend, S. M., Faber, W. R., Naafs, B., Ottenhoff, T. H. M. Immunological Crossreactivity of the *Mycobacterium leprae* CFP-10 with its Homologue in *Mycobacterium tuberculosis*. *Scandinavian journal of immunology*. 2004; *59*(1), 66-70.
59) Hughes, M.S., Ball, N.W., McCarroll, J., Erskine, M., Taylor, M.J., Pollock, J.M., Skuce, R.A., Neill, S.D. Molecular analysis of mycobacteria other than the *M. tuberculosis* complex isolated from Northern Ireland cattle. *Veterinary Microbiology*. 2005; 108 (1-2):101-112.
60) van Pittius, N. C. G., Gamieldien, J., Hide, W., Brown, G. D., Siezen, R. J., & Beyers, A. D. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+ C Gram-positive bacteria. *Genome biology*. 2001; *2*(10), research0044-1.
61) Colangeli, R., Spencer, J.S., Bifani, P., Williams, A., Lyashchenko, K., Keen, M.A., Hill, P.J., Belisle, J., Gennaro, M.L. MTSA-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis specific delayed type hypersensitivity in guinea pigs. *Infection and Immunity.* 2000; 68: 990:993. doi: 10.1128/iai.68.2.990-993.2000 [doi].
62) Geluk, A., Van Meijgaarden, K.E., Franken, K.L.M.C., Subronto, Y.W., Wieles, B., Arend, S.M., Sampiaio, E.P., De Boer, T., Faber, W.R., Naafs, B., Ottenhogff, T.H.M. Identification and characterisation ESAT-6 homologue of *Mycobacterium leprae* and T-cell cross reactivity with *Mycobacterium tuberculosis. Infection and immunity.* 2002; 70:2544-2548. doi: 10.1128/iai.70.5.2544-2548.2002 [doi].
63) Murphy, K., Weaver, C. *Janeway's Immunobiology.* 8th Edition; Garland Science. Oxford: Taylor & Francis Inc. 2016.
64) Arend, S. M., de Haas, P., Leyten, E., Rosenkrands, I., Rigouts, L., Andersen, P., van Soolingen, D. ESAT-6 and CFP-10 in clinical versus environmental isolates of *Mycobacterium kansasii. The Journal of infectious diseases*. 2005;*191*(8), 1301-1310.
65) Arend, S. M., van Meijgaarden, K. E., de Boer, K., de Palou, E. C., van Soolingen, D., Ottenhoff, T. H., & van Dissel, J. T. Tuberculin skin testing and in vitro T cell responses to ESAT-6 and culture filtrate protein 10 after infection with *Mycobacterium marinum* or *M. kansasii*. *The Journal of infectious diseases*. 2002; *186*(12), 1797-1807.
66) Author unknown: National Center for Biotechnology Information (NCBI). Nucleotide database. Searches: "Mycobacterium tuberculosis esat 6" and "Mycobacterium tuberculosis cfp 10". https://www.ncbi.nlm.nih.gov/gene/886209 and https://www.ncbi.nlm.nih.gov/gene/886194. Accessed May 2018.
67) Author unknown: Pasteur Institute. TubercuList database. Searches: "Rv3874" and "Rv3875". http://svitsrv8.epfl.ch/tuberculist/quicksearch.php?gene+name=Rv3875&submit=Search and http://svitsrv8.epfl.ch/tuberculist/quicksearch.php?gene+name=Rv3874&submit=Search
68) Kumar, S., Stecher, G., & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*. 2016; *33*(7): 1870-1874.
69) Thompson, J. D., Higgins, D. G., & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*. 1994; *22*(22): 4673-4680.
70) Saitou, N., & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. 1987; *4*(4): 406-425. doi: 10.1093/oxfordjournals.molbev.a040454 [doi].
71) Tamura, K., Nei, M., & Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences*. 2004; *101*(30): 11030-11035. doi: 10.1073/pnas.0404206101 [doi].
72) van Koonin, E., Puigbo, P., Wolf, Y.I. Comparison of Phylogenetic Trees and Search for a Central Trend in the ''Forest of Life''. *Journal of computational biology.* 2011; 18(7): 917-924.
73) Prasanna, A.N., Mehra, S. Comprehensive Phylogenetic Analysis of Mycobacteria. *12th IFAC Proceedings Volumes.* 2013;46(31):101-106.

74) Bremm, S., Von Landesberger, T., Hess, M. et al. Interactive Visual Comparison of Multiple Trees. *IEEE Conference on Visual Analytics Science and Technology (VAST)*; 2011: 31-40.
75) Robinson, O., Dylus, D., Dessimoz, C. Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Molecuar Biology and Evolution*. 2016; 33(8):2163-2166.
76) Maddison, W.P., Maddison, D.R. Mesquite: a modular system for evolutionary analysis. Version 3.16. 1997-2019. Available via: https://www.mesquiteproject.org/home.html http://www.mesquiteproject.org/Trees.html#consensus
77) Author unknown: National Center for Biotechnology Information (NCBI). Basic Local Alignment Search Tool (BLAST). Frequently Asked Questions Website. https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect Accessed May 2018
78) Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985; *39*(4), 783-791.
79) Chan, A. W., Kabbani, S., Staton, G., & Kraft, C. S. *Mycobacterium paraffinicum* causing symptomatic pulmonary infection. *Journal of clinical microbiology*. 2014; *52*(4), 1281-1283. doi: 10.1128/JCM.03107-13 [doi].
80) Toney, N., Adekambi, T., Toney, S., Yakrus, M., & Butler, W. R. Revival and emended description of '*Mycobacterium paraffinicum*'Davis, Chase and Raymond 1956 as *Mycobacterium paraffinicum* sp. nov., nom. rev. *International journal of systematic and evolutionary microbiology.* 2010; *60*(10): 2307-2313.
81) Dostal, S., Richter, E., & Harmsen, D. *Concise guide to mycobacteria and their molecular differentiation*. BoD-Books on Demand. Ridom Press. 2003: 206.
82) Kazda, J., Cooney, R., Monaghan, M., Quinn, P. J., Stackebrandt, E., Dorsch, M., Tarnok, Z. S. *Mycobacterium hiberniae* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. 1993; *43*(2), 352-357.
83) Hormisch, D., Brost, I., Kohring, G. W., Giffhorn, F., Kroppenstedt, R. M., Stackebradt, E., Holzapfel, W. H. *Mycobacterium fluoranthenivorans* sp. nov., a fluoranthene and aflatoxin B1 degrading bacterium from contaminated soil of a former coal gas plant. *Systematic and applied microbiology*. 2004; *27*(6), 653-660.
84) Asmar, S., Rascovan, N., Robert, C., & Drancourt, M. Draft genome sequence of *Mycobacterium acapulcensis* strain CSURP1424. *Genome Announc.* 2016; *4*(4): e00836-16. doi: 10.1128/genomeA.00836-16 [doi].
85) Stanford, J. L., & Gunthorpe, W. J. A study of some fast-growing scotochromogenic mycobacteria including species descriptions of *Mycobacterium gilvum* (new species) and *Mycobacterium duvalii* (new species). *British journal of experimental pathology*. 1971; *52*(6), 627-637.
86) Wee, W. Y., Dutta, A., Choo, S. W. Comparative genome analyses of mycobacteria give better insights into their evolution. *PloS one*. 2017; *12*(3) :e0172831. doi: 10.1371/journal.pone.0172831 [doi].
87) Paniz-Mondofi, A.E., Greninger, A.L., Ladutko, L., Brown-Elliott, B.A., Vasireddy, R., Jukabiec, W., Vasireddy, S., Wallace, R.J., Simmon, K.E., Dunn, B.E., Jackoway, G., Vora, S.B., Quinn, K.K., Qin, X., Campbell, S. *Mycobacterium grossiae* sp. Nov., a rapidly growing, scotochromogenic species isolated from human clinical respiratory and blood culture specimens. *International Journal of Systematic and Evolutionary Microbiology.* 2017; 67(11):4345-4351.
88) Maddison, W.P., Maddison, D.R. Mesquite online manual. 2002-2011. http://mesquiteproject.org/mesquiteArchives/mesquite2.75/Mesquite_Folder/docs/mesquite/moduleIndex.html
89) Ruiz-Villalba, A., van Pelt-Verkuil, E., Gunst, Q.D., Ruijter, J.M., van den Hoff, M.J. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). *Biomolecular detection and quantification*. 2017; *14*:7–18.

90) Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C., Little, S. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Research*. 1997; 25(16);3235–3241.
91) Jünemann, S. Homology, similarity and identity. *Research gate.* 2013. https://www.researchgate.net/post/Homology_similarity_and_identity-can_anyone_help_with_these_terms/5278c854cf57d7ba368b45ed/citation/download Accessed May 2020.
92) Author unknown: International Union of Pure and Applied Chemistry (IUPAC). Bioinformatics IUPAC Nucleotide codes. https://www.bioinformatics.org/sms/iupac.html Accessed May 2018
93) Author unknown: Institut Pasteur. Bovilist. Search: "esxA". http://genolist.pasteur.fr/BoviList/genome.cgi Accessed May 2020.
94) Author unknown: Mycobrowser. Gene MSMEG_0066 in *Mycobacterium smegmatis* MC^2-155. https://mycobrowser.epfl.ch/genes/MSMEG_0066 Accessed May 2020.
95) Author unknown: The genepool. Sanger Sequencing Troubleshooting Guide. 2009. http://genepool.bio.ed.ac.uk/sanger/Sanger_troubleshooting_guide_v1.pdf Accessed June 2018

Appendices

Appendix A: Complete BLAST results of all obtained sequences

Table A: Complete BLAST results of all sequences obtained from *esxA*-PCR
a) *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no. 3; 600 bp); forward sequence product can be many proteins: NUDIX hydrolase, dolichyl-phosphate-mannose--protein O-mannosyl transferase, glycerophosphoryl diester phosphodiesterase, transcription termination factor Rho, non-ribosomal peptide synthetase, extradiol dioxygenase, keto-deoxy-phosphogluconate aldolase , FAD-binding dehydrogenase, nitronate monooxygenase, L,D-peptidoglycan transpeptidase YkuD, ErfK/YbiS/YcfS/YnhG family, alpha-L-rhamnosidase, amidohydrolase, N-acetylglucosamine-6-phosphate deacetylase, Mce-associated membrane protein, FdhE-like protein, epoxide hydrolase, aconitate hydratase, DNA repair ATPase, cell division protein, sensor histidine kinase, DNA helicase, TadE-like protein, long-chain acyl-CoA synthetase, Adenylate kinase-related kinase, ABC transport protein, ATP-binding component, putative cytochrome P450 hydroxylase
b) *M. malmesburri* sp. *nov*. strain WCM 7299 (C4 no. 3; 600 bp); forward product can be many proteins: family 1 extracellular solute-binding protein, (3S)-malyl-CoA thioesterase, HTH-type transcriptional regulator GltC, citryl-CoA lyase, LysR family transcriptional regulator, methyl-accepting chemotaxis protein, multidrug transporter, CoA ester lyase, Na(+)-translocating NADH-quinone reductase subunit C, 9326 bp at 5' side: cytochrome P450, 31947 bp at 3' side: LuxR family transcriptional regulator, RND transporter, Secreted protein, MurR/RpiR family transcriptional regulator, S1 family peptidase, acetoin utilization protein AcuC, ribonuclease D, glucosamine-6-phosphate deaminase, DNA helicase IV, RpiR family transcriptional regulator , ferrichrome-iron receptor, threonyl-tRNA synthetase , "fumarase, class II", DNA-directed RNA polymerase subunit beta', thioredoxin domain-containing protein , Antibiotic biosynthesis monooxygenase, glycogen debranching protein, cysteine synthase A, DUF2277 domain-containing protein, KR domain-containing protein, chromate transporter, apolipoprotein N-acyltransferase
c) Only a few of the highest BLAST hits are shown in this table.

Mycobacterium species (isolate ID)	**Highest BLAST hit (accession number)**	**Sequence position in the genome of the highest hit**	**Identity/similarity (%)**	**E-value**	**Product**
M. malmesburri sp. *nov*. strain WCM 7299 (C4 no. 3); 600 bp; consensus	PREDICTED: Hipposideros armiger Kruppel like factor 13 (KLF13), mRNA (XM_019625163.1)	2626 to 2651	100%	0.30	(…)
	Pseudomonas parafulva strain CRS01-1, complete genome(CP009747.1)	2916681 to 2916707	96%	3.7	DNA polymerase IV
M. malmesburri sp. *nov*. strain WCM 7299 (C4 no. 3); 600 bp; reverse complement	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	3292085 to 3292166	87%	5e-16	2-hydroxycyclohexanecarboxyl-CoA dehydrogenase
		3323671 to 3323743	76%	0.018	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family
		4064615 to 4064655	85%	0.064	3alpha(or 20beta)-hydroxysteroid dehydrogenase
M. malmesburri sp. *nov*. strain WCM 7299 (C4 no. 3); 600 bp; forward	*Mycobacterium* sp. *MS1601,* complete genome (CP019420.1)	5654878 to 5654976	78%	2e-10	hypothetical protein; protein id: AQA06765.1 -> can be many proteins[a]

M. elephantis (Rustenburg kraal swab 1 no.7); 750 bp; consensus	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	1680305 to 1680962	83%	9e-174	ABC-2 type transport system permease protein
	Mycobacterium vaccae 95051, complete genome (CP011491.1)*c	5760804 to 5761400	83%	6e-156	sugar ABC transporter permease
M. elephantis (Rustenburg kraal swab 1 no.7); 750 bp; reverse complement	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	1680305 to 1680962	85%	0.0*	ABC-2 type transport system permease protein
M. elephantis (Rustenburg kraal swab 1 no.7); 750 bp; forward	PREDICTED: Chrysemys picta bellii transmembrane protein 178B-like (LOC101933992), mRNA (XM_005302359.3)	130 to 150	100%	5.2	transmembrane protein 178B-like
M. elephantis (Rustenburg welgevonden swab 3 no.3); 600 bp; consensus	*Mycobacterium phlei* strain CCUG 21000, complete genome (CP014475.1)*c	4079320 to 4079651	64%	1e-32	Copper-exporting P-type ATPase A
M. elephantis (Rustenburg welgevonden swab 3 no.3); 600 bp; reverse complement	*Mycobacterium fortuitum* strain CT6, complete genome (CP011269.1)	4549680 to 4549942	78%	5e-36	ATPase
M. elephantis (Rustenburg welgevonden swab 3 no.3); 600 bp; forward	*Mycobacterium phlei strain* CCUG 21000, complete genome (CP014475.1)	4079321 to 4079643	69%	3e-33	Copper-exporting P-type ATPase A
M. acapulcensis (Kanaland swab 2 no.2); 300 bp; consensus	*Mycobacterium tuberculosis strain* CAS chromosome (CP028428.1)*c	4336363 to 4336394	100%	1e-04	**type VII secretion system ESX-1 WXG100 family target ESAT-6**
		4336621 to 4336650	100%	0.002	**type VII secretion system ESX-1 WXG100 family target ESAT-6**
M. acapulcensis (Kanaland swab 2 no.2); 300 bp; reverse complement	*Mycobacterium tuberculosis* strain CAS chromosome (CP028428.1)*c	4336363 to 4336394	100%	7e-05	**type VII secretion system ESX-1 WXG100 family target ESAT-6**
M. acapulcensis (Kanaland swab 2 no.2); 300 bp; forward	*Mycobacterium tuberculosis* strain CAS chromosome (CP028428.1)*c	4336621 to 4336650	100%	0.001	**type VII secretion system ESX-1 WXG100 family target ESAT-6**

M. neonarum (Xhongora swab 4 no. 1); 700 bp; consensus	*Mycobacterium* sp. *NRRL B-3805*, complete genome (CP011022.1)	819525 to 820057	85%	5e-166	peptidase
	Mycobacterium neoaurum VKM Ac-1815D, complete genome (CP006936.2)	819526 to 820058	85%	5e-166	peptidase
M. neonarum (Xhongora swab 4 no. 1); 700 bp; reverse complement	*Mycobacterium* sp. *NRRL B-3805*, complete genome (CP011022.1)	819525 to 820018	90%	0.0*	peptidase
	Mycobacterium neoaurum VKM Ac-1815D, complete genome (CP006936.2)	819526 to 820019	90%	0.0*	peptidase
M. neonarum (Xhongora swab 4 no. 1); 700 bp; forward	*Mycobacterium* sp. *NRRL B-3805*, complete genome (CP011022.1)	819606 to 820057	88%	2e-144	peptidase
	Mycobacterium neoaurum VKM Ac-1815D, complete genome (CP006936.2)	819607 to 820058	88%	2e-144	peptidase
M. madagascariense (Virginia swab 3 no.2); 500 bp; consensus	*Mycobacterium* sp. *EPa45*, complete genome (CP011773.1)	3030317 to 3030770	75%	4e-74	amidohydrolase
		3019319 to 3019456	72%	2e-08	amidohydrolase
		3024741 to 3024896	69%	1e-05	amidohydrolase
		4001612 to 4001692	75%	5e-04	luciferase
		6035420 to 6035573	68%	0.006	amidohydrolase
		4003656 to 4003701	83%	0.25	luciferase
		4482686 to 4482818	69%	0.86	amidohydrolase
M. madagascariense (Virginia swab 3 no.2); 500 bp; reverse complement	*Mycobacterium* sp. *EPa45*, complete genome (CP011773.1)	3030317 to 3030736	74%	5e-60	amidohydrolase
		3019319 to 3019422	75%	2e-08	amidohydrolase
		3024822 to 3024896	77%	1e-04	amidohydrolase
		4001612 to 4001692	75%	4e-04	luciferase
		4003656 to 4003701	83%	0.22	luciferase
		4482686 to 4482818	69%	0.78	amidohydrolase

M. madagascariense (Virginia swab 3 no.2); 500 bp; forward	*Mycobacterium* sp. *EPa45*, complete genome (CP011773.1)	3030346 to 3030770	75%	1e-67	luciferase & amidohydrolase
		3019319 to 3019456	72%	2e-08	amidohydrolase
		3024741 to 3024896	69%	1e-05	amidohydrolase
		4001612 to 4001692	75%	4e-04	luciferase
		6035420 to 6035573	68%	0.005	amidohydrolase
		4003656 to 4003701	83%	0.22	luciferase
		4482686 to 4482818	69%	0.78	amidohydrolase
M. moriokaense (ACTT 43059 no. 2); 300 bp; consensus	Message ID#24 Error: Failed to read the Blast query: Protein FASTA provided for nucleotide sequence				
M. moriokaense (ACTT 43059 no. 2); 300 bp; reverse complement	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	4155547 to 4155751	67%	1e-04	CAAX protease self-immunity
M. moriokaense (ACTT 43059 no. 2); 300 bp; forward	Lupinus angustifolius cultivar Tanjil chromosome LG-04 (CP023116.1)	17719223 to 17719265	86%	0.21	(…)
M. moriokaense (ACTT 43059 no. 2); 500 bp; reverse complement	*Mycobacterium* sp. *PYR15* chromosome, complete genome (CP023435.1)	235118 to 235144	100%	0.068	(…)
M. malmesburii sp. *nov*. (Western cape swab 5); 300 bp; consensus	No significant similarity found (even when ambigue bases were deleted)				
M. malmesburii sp. *nov*. (Western cape swab 5); 300 bp; reverse complement	*Mycobacterium* sp. *djl-10*, complete genome (CP016640.1)	5946702 to 5946743	93%	6e-06	3-alpha-hydroxysteroid dehydrogenase
	Mycobacterium gilvum Spyr1, complete genome (CP002385.1)	2681217 to 2681258	93%	6e-06	dehydrogenase of unknown specificity, short-chain alcohol dehydrogenase like protein
	Mycobacterium gilvum PYR-GCK, complete genome (CP000656.1)	3469424 to 3469465	93%	6e-06	short-chain dehydrogenase/reductase SDR

M. malmesburii sp. *nov*. (Western cape swab 5); 300 bp; forward	*Bradyrhizobium oligotrophicum* S58 DNA, complete genome (AP012603.1)	6881698 to 6881741	84%	0.39	tRNA-specific 2-thiouridylase MnmA
M. malmesburii sp. *nov*. (Western cape swab 5); 600 bp; consensus	Message ID#24 Error: Failed to read the Blast query: Protein FASTA provided for nucleotide sequence				
M. malmesburii sp. *nov*. (Western cape swab 5); 600 bp; reverse complement	*Mycobacterium gilvum Spyr1*, complete genome (CP002385.1)	1902989 to 1903362	86%	7e-109	predicted TIM-barrel fold metal-dependent hydrolase
	Mycobacterium gilvum PYR-GCK, complete genome (CP000656.1)	2544440 to 2544813	86%	7e-109	amidohydrolase 2
		4595705 to 4596078	86%	3e-107	amidohydrolase 2
	Mycobacterium vanbaalenii PYR-1, complete genome (CP000511.1)	4518238 to 4518620	85%	7e-109	amidohydrolase 2
		2161960 to 2162342	85%	3e-107	amidohydrolase 2
M. malmesburii sp. *nov*. (Western cape swab 5); 600 bp; forward	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	1131935 to 1131998	81%	2e-04	hypothetical protein; protein id: SEH53507.1 → can be many proteins*[b]

Table B: Complete BLAST results of all sequences obtained from *esxB*-PCR
a) Sequence of *M. moriokaense* (ACTT 43059 no. 1; 400 bp) can be many hypothetical proteins: "CMGC kinase, CDK family, putative", electron transport complex subunit RsxE, Putative acyl-peptide hydrolase, Spermidine synthase, Carbonic anhydrase, iron ABC transporter, MFS transporter , electron transport complex subunit E, S9 family peptidase, PAC2 family protein, ABC transporter ATP-binding protein, putative amino acid permease YhdG, D-beta-hydroxybutyrate dehydrogenase and/or membrane transport protein.
b) Only a few of the highest BLAST hits are shown in this table.

Mycobacterium species (isolate ID)	**Highest BLAST hit (accession number)**	**Sequence position in the genome of the highest hit**	**Identity/similarity (%)**	**E-value**	**Product**
M. elephantis (Rustenburg welgevonden swab 3 no.2); 300 bp; forward	*Anaplasma marginale* str. Dawn genome (CP006847.1)	950922 to 950944	100%	6.8	hypothetical protein; protein id: AGZ80133.1 = Possibly _ankyrin repeat domain-containing protein 13C
	Anaplasma marginale str. Gypsy Plains genome (CP006846.1)	950877 to 950899	100%	6.8	hypothetical protein; disrupted = Possibly _ankyrin repeat domain-containing protein 13C
	Anaplasma marginale str. Florida, complete genome (CP001079.1)	954436 to 954458	100%	6.8	Conserved hypothetical protein; Protein id: ACM49626.1 = Possibly _ankyrin repeat domain-containing protein 13C
M. acapulcensis (Kanaland swab 2 no. 2); 300 bp; forward	*Mycobacterium* sp. *JLS*, complete genome (CP000580.1)	2798732 to 2799035	67%	7e-19	pyridoxamine 5'-phosphate oxidase-related; FMN-binding protein;
	Mycobacterium sp. *KMS*, complete genome (CP000518.1)	2824656 to 2824959	67%	7e-19	pyridoxamine 5'-phosphate oxidase-related; FMN-binding protein;
	Mycobacterium sp. *MCS*, complete genome (CP000384.1)	2806768 to 2807071	67%	7e-19	pyridoxamine 5'-phosphate oxidase-related; FMN-binding protein;
M. neonarum (Xhongora swab 4 no. 1); 300 bp; forward	*Mycobacterium* sp. *NRRL B-3805*, complete genome (CP011022.1)	3574898 to 3575176	76%	5e-45	channel-forming protein
	Mycobacterium neoaurum VKM Ac-1815D, complete genome (CP006936.2)	3574899 to 3575177	76%	5e-45	channel-forming protein

M. vaccae/M. vanbaalenni (Xhongora swab 5 no. 2); 300 bp; forward	*Mycobacterium* sp. *NRRL B-3805*, complete genome (CP011022.1)	3574898 to 3575184	84%	3e-70	channel-forming protein
	Mycobacterium neoaurum VKM Ac-1815D, complete genome (CP006936.2)	3574899 to 3575185	84%	3e-70	channel-forming protein
M. engbackii (Xhongora S1); 300 bp; forward	*Mycobacterium szulgai* strain NLA000501479 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014490.1)*[b]	39 to 303	87%	2e-77	**10 kDa culture filtrate protein**
	Mycobacterium kansasii strain NLA000601759 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014492.1)*[b]	43 to 303	87%	1e-73	**10 kDa culture filtrate protein**
	Mycobacterium marinum ATCC 927 DNA, complete genome (AP018496.1)*[b]	213621 to 213881	86%	6e-72	**ESAT-6-like protein *esxB***
M. madagascariense (Virginia swab 3 no. 2); 400 bp; forward	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	206498 to 206674	84%	8e-37	Pyridoxamine 5'-phosphate oxidase
M. madagascariense (Virginia swab 3 no. 2); 600 bp; forward	*Mycobacterium rutilum* strain DSM 45405 genome assembly, chromosome: I (LT629971.1)	2855594 to 2855845	70%	6e-21	phospholipid/cholesterol/gamma-HCH transport system permease protein
M. engbackii (Kabisa S1 no. 2); 300 bp; forward	*Mycobacterium szulgai* strain NLA000501479 10 kDa culture filtrate protein (cfp-10) gene, complete cds (FJ014490.1)*[b]	39 to 303	88%	4e-79	**10 kDa culture filtrate protein**
M. moriokaense (ACTT 43059 no. 1); 400 bp; forward	*Penicilliopsis zonata* CBS 506.65 hypothetical protein (ASPZODRAFT_130697), mRNA (XM_022722566.1)	402 to 428	96%	2.4	hypothetical protein; protein id: XP_022583117.1→ can be many proteins[a]
M. vaccae/M. vanbaalenni (Botharust swab 3); 450 bp; forward	*Mycobacterium vaccae* 95051, complete genome (CP011491.1)	353493 to 353789	79%	3e-63	Membrane protein/hypothetical protein; protein id: ANI37607.1
M. parafortuitum (Ncambele swab 4 no. 2); 300 bp; forward	Bos mutus isolate yakQH1 chromosome 16 (CP027084.1)	70005168 to 70005200	88%	5.7	(…)
M. paraffinicum (Langkul S1 no. 2); 300 bp; forward	*Mycobacterium marinum* ATCC 927	213619 to 213881	86%	5e-73	**ESAT-6-like protein *esxB***

	DNA, complete genome (AP018496.1)*b				
		6403504 to 6403766	86%	5e-73	**ESAT-6-like protein *esxB***
	Mycobacterium marinum strain 1123 10 kDa culture filtrate protein (*esxB*) gene, complete cds (MF411158.1)*b	41 to 303	86%	5e-73	**10 kDa culture filtrate protein**
	Mycobacterium marinum strain 7873 10 kDa culture filtrate protein (*esxB*) gene, complete cds (MF411156.1)*b	41 to 303	86%	5e-73	**10 kDa culture filtrate protein**
	Mycobacterium marinum strain 6865 10 kDa culture filtrate protein (*esxB*) gene, complete cds (MF411155.1)*b	41 to 303	86%	5e-73	**10 kDa culture filtrate protein**
	Mycobacterium marinum strain 526A 10 kDa culture filtrate protein (*esxB*) gene, complete cds (MF411154.1)*b	41 to 303	86%	5e-73	**10 kDa culture filtrate protein**
	Mycobacterium pseudoshottsii JCM 15466 DNA, complete genome (AP018410.1)*b	6008615 to 6008877	86%	5e-73	**ESAT-6-like protein *esxB***
	Mycobacterium marinum E11 main chromosome genome (HG917972.2)*b	213753 to 214015	86%	5e-73	**10 kDa culture filtrate antigen *esxB*_1**
		6287814 to 6288076	86%	5e-73	**ESAT-6 like protein *esxB***
	Mycobacterium liflandii 128FXT, complete genome (CP003899.1)*b	6154651 to 6154913	86%	5e-73	**ESAT-6 like protein e*sxB***
	Mycobacterium marinum, complete genome (CP000854.1)*b	6591198 to 6591460	86%	5e-73	**ESAT-6 like protein e*sxB***
		222129 to 222391	86%	2e-71	**10 kDa culture filtrate antigen e*sxB*_1**
	Mycobacterium ulcerans isolate ITM842 10 kDa culture filtrate protein (*esxB*) gene, complete cds (EU257148.1)*b	41 to 303	86%	5e-73	**10 kDa culture filtrate protein**

M. madagascariense (Ladybrand swab 3); 350 bp; forward	*Agrobacterium tumefaciens* str. C58 circular chromosome, complete sequence (AE007869.2)	1371922 to 1372029	86%	1e-21	undecaprenyl pyrophosphate synthase
		1371889 to 1372017	77%	6e-12	undecaprenyl pyrophosphate synthase

Appendix B: Sequences of esxA- or esxB-positive DNA fragments (results of sequencing)

- *M. acapulcensis* (Kanaland swab 2 no.2); 300 bp; consensus (ESAT-6):

5'ATGACAGAGCAGCAGTGGAATTTCGCGGGTATGACAGAGCGCAGTGGACCGGMCTGGCCCGGGCCCGCcC
CGTGGTcggcCaGcATAGATGaGGSGAASCaTATCYTTACgcTTCgTCGAGCGCGCGATAAGTCCACYaacGGCcTT
CACACGTCAGcGCTATcgCCTCGGGGGCACGCGCTaCtTGCGCGGCGAACAgCtGTGTGATGGACACGGSCGCG
GCGGCTTGGgCCagCACcGCACGGTTGCCCCaCCGGTCCAGTTGTGCGCGTTCACTGGGAGGTTCGCATAGAAG
GCAACGTCACTGGGATGTTCGCATAG 3'= 321 bp

- *M. engbackii* (Xhongora S1); 300 bp; forward (CFP-10):

5'xCxagAxGCAGGTACTTCGAGCGGATCTCGAGCGATCTGAAGACCCAGATCGACCAGATCGAGTCCACCGCA
GGTTCACTGCAGGGTCAGTGGCGTGGCGTCGCCGGTCAGGCCGCCCAAGCCGCAGTCGTGCGCTTCCAGGAA
GCCGCCAACAAGCAAAAGGCCGAGCTCGACGAGATTTCGACCAACATCCGTCAGGCCGGCGTCCAGTACCAG
CGGGCCGACGAGGAGCAGCAGCAGGCGCTGTCCTCGCAAATGGGCTTCTGAA 3' = 268 bp

- *M. engbackii* (Kabisa S1 no. 2); 300 bp; forward (CFP-10):

5'xCAgAGCAgGTACTTCGAGCGGATCTCGGGCGACCTGAAGACCCAGATCGACCAGATCGAGTCCACTGCAGC
CTCGCTGCAGGGTCAGTGGCGTGGTGTCGCCGGTCAGGCCGCCCAAGCCGCAGTCGTGCGCTTCCAGGAAGC
CGCCAACAAGCAGAAGGCCGAGCTCGACGAGATTTCAACCAACATCCGTCAGGCCGGCGTCCAGTACCAGCG
GGCCGACGAGGAGCAGCAGCAGGCGCTGTCCTCGCAAATGGGCTTCTGAA 3'= 266 bp

- *M. paraffinicum* (Langkul S1 no. 2); 300 bp; forward (CFP-10):

5'xcaxAgGCAGGTACTTCGAGCGGATCTCGAGCGATCTGAAGACCCAGATCGACCAGATCGAGTCCACCGCCG
GTTCGCTCCAGGGGCAGTGGCGCGGCGTCGCCGGCCAGGCCGCGCAGGCCGCCGTCGTGCGCTTCCAGG
AAGCCGCCAACAAGCAGAAGGCTGAGCTCGACGAGATTTCGACCAACATCCGTCAGGCCGGCGTTCAGTACC
AGCGTGCCGACGAGGAGCAGCAGCAGGCGCTGTCCTCGCAAATGGGCTTCTGAA 3' = 267

Appendix C: Phylogenetic trees based on 16SrRNA, esxA, esxB and the consensus tree (33 taxa)

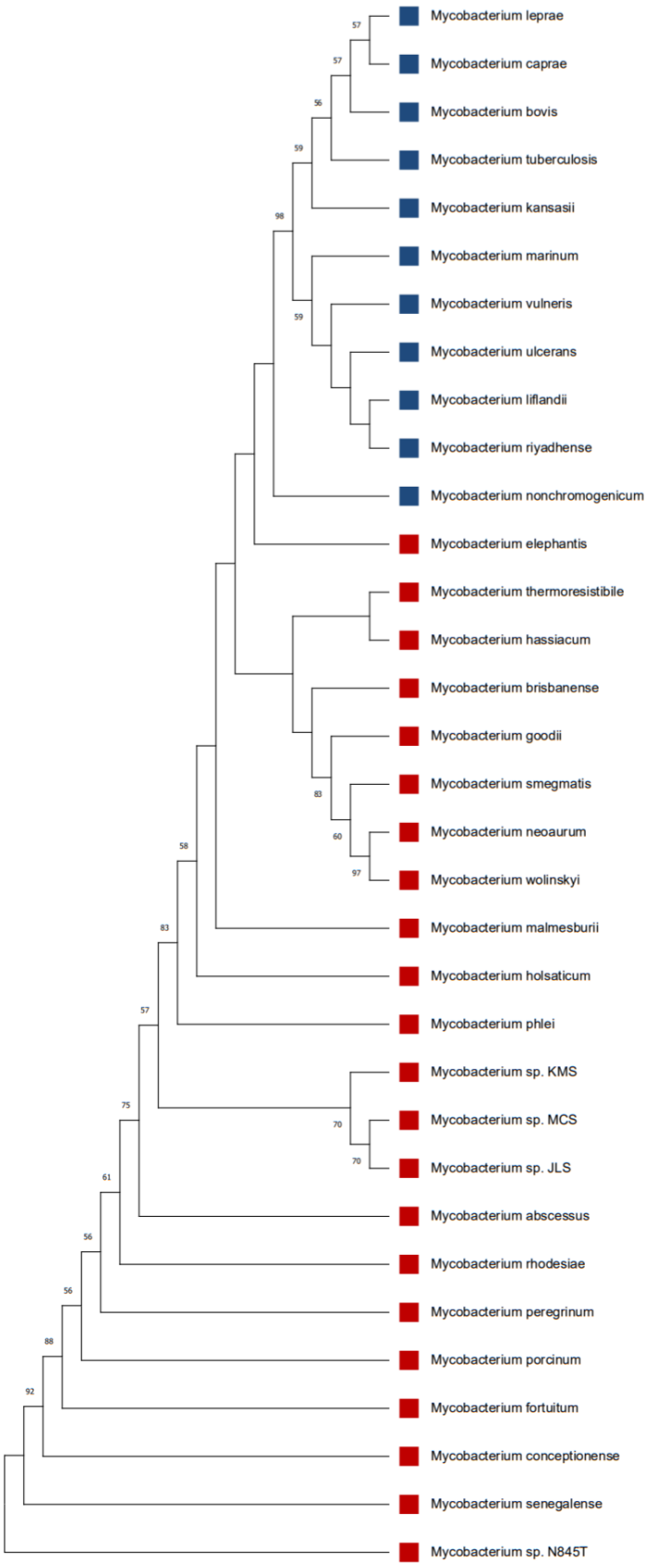

Figure D: Phylogenetic relationship of 33 mycobacterial species based on the 16S rRNA gene sequences shown by MEGA software. The optimal tree with the sum of branch length = 6.59254275 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree is drawn to scale (bar 0.1 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Gene sequences of the following specific strains of mycobacterial species used in the 16S rRNA-based tree, are: *M.* sp. *N845T* (TB 1246C), *M. marinum* DL150991 (AM884307), *M. porcinum* DSM 44242 (AF547958), *M. fortuitum* ATCC 49404 (AF480581), *M. peregrinum* NCTC 10264 (AM884319), *M. kansasii* ATCC 12478 (AF480601), *M. smegmatis* ATCC 19420 (AJ131761), *M. tuberculosis* H37Rv (MTB000019), *M. vulneris* DSM 45247 (LT718448.1), *M. nonchromogenicum* ATCC 19530 (NR_044813.1), *M. thermoresistibile* CIP 105390 (AF547971.1), *M. bovis* (AY360331.1), *M. leprae* (X53999.1), *M. liflandii* (AY845224.1), *M. riyadhense* NLA000201958 (EU274642.1), *M. goodii* X7B (AF513815.1), *M.* sp. *KMS* (AY083217.1), *M.* sp. *MCS* (AF387803.1), *M.* sp. *JLS* (AF387804.1), *M. rhodesiae* DSM 44223T (AJ429047.1), *M. phlei* (AF480603.1), *M. caprae* sp. *Caprae* (AJ131120.1), *M. neoaurum* ATCC 23071 (FJ172311.1), *M. elephantis* (AJ010747.1), *M. holsaticum* strain 1406 (AJ310467.1), *M. malmesburii* WCM 7299 (KJ873241.2), *M. abscessus* (AJ536038.1), *M. conceptionense* (HM366452.1), *M. senegalense* (AF480596.1), *M. wolinskyi* strain ATCC 700010 (AY457083.1), *M. brisbanense* strain CIP 107830 (AY943206.1), *M. hassiacum* (U49401.1) and *M. ulcerans* strain ATCC 33728 (AB548733.1). Either genebank accession numbers for the sequences retrieved from the database, or isolate identification number for the sequences retrieved from the ARC-OVI database are shown in parenthesis. The trees are unrooted. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

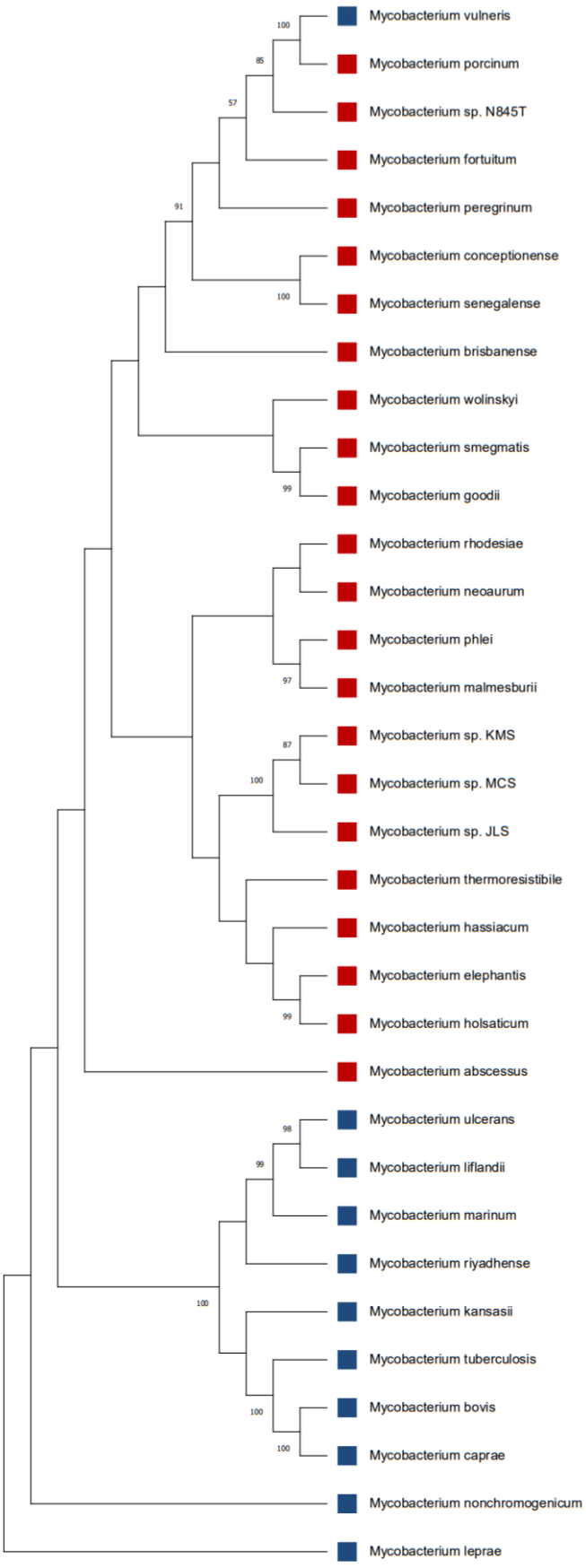

Figure E: Phylogenetic relationship of 33 mycobacterial species based on the ESAT-6 gene sequences shown by MEGA software. The optimal tree with the sum of branch length = 6.59254275 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree is drawn to scale (bar 0.1 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Gene sequences of the following specific strains of mycobacterial species used in the 16S rRNA-based tree, are: *M.* sp. *N845T* (TB 1246A), *M. marinum* (CP000854.1), *M. porcinum* IP141460001 (MVIG01000023.1), *M. fortuitum* ATCC 6841 (CP014258.1), *M. peregrinum* 852002-51209 SCH5440388 (NZ_LZSO01000031), *M. kansasii* ATCC 12478 (CP006835.1), *M. smegmatis* MC^2 (CP009494.1), *M. tuberculosis* H37Rv (NC_000962.3), *M. vulneris* (CCBG010000001.1), *M. nonchromogenicum* strain NCK 8460 (KonoS1), *M. thermoresistibile* ATCC 19527 (AGVE01000046.1), *M. bovis* AF2122/97 (LT708304.1), *M. leprae* TN (AL583917.1), *M. liflandii* 128FXT (CP003899.1), *M. riyadhense* NLA000201958 (EU552926.1), *M. goodii* X7B (CP012150.1), *M.* sp. *KMS* (CP000518.1), *M.* sp. *MCS* (CP000384.1), *M.* sp. *JLS* (CP000580.1), *M. rhodesiae* NBB3 (CP003169.1), *M. phlei* CCUG 21000 (CP014475.1), *M. caprae* Allgaeu (CP016401.1), *M. neoaurum* DSM 44074 (LK021337.1), *M. elephantis* strain Lipa Mele (LBNO01000056.1), *M. holsaticum* M7 (NZ_MIGZ01000035.1), *M. malmesburii* WCM 7299 (CVTB01000324.1), *M. abscessus* sp. abscessus 800 (FSCO01000216.1), *M. conceptionense* (CTEF01000001.1), *M. senegalense* CK1 (LDCO01000022.1), *M. wolinskyi* CDC_01 (LGTW01000008.1), *M. brisbanense* JCM15654 (BCSX01000024.1), *M. hassiacum* DSM 44199 (AMRA01000004.1) and *M. ulcerans* sp. *shinshuense* ATCC 33728 (AP017624.1). Either genebank accession numbers for the sequences retrieved from the database, or isolate identification number for the sequences retrieved from the ARC-OVI database are shown in parenthesis. The trees are unrooted. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

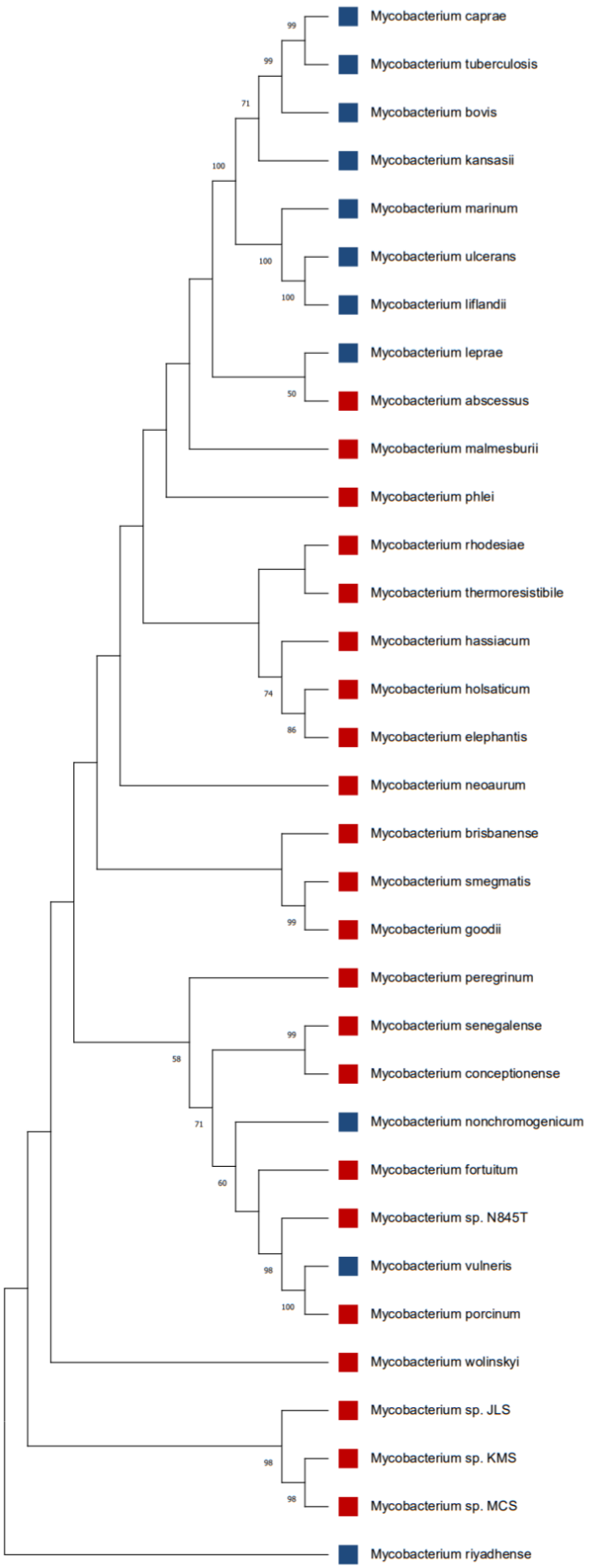

Figure F: Phylogenetic relationship of 33 mycobacterial species based on the CFP-10 gene sequences shown by MEGA software. The optimal tree with the sum of branch length = 6.59254275 is shown. The percentage of replicate trees (>50%) in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches[78]. The tree is drawn to scale (bar 0.1 per nucleotide), with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Gene sequences of the following specific strains of mycobacterial species used in the 16S rRNA-based tree, are: *M.* sp. *N845T* (TB 1246A), *M. marinum* (HG917972.2), *M. porcinum* HMC_M12 (MIHF01000118.1), *M. fortuitum* ATCC 6841 (CP014258.1), *M. peregrinum* ACS819 (LZIB01000009), *M. kansasii* ATCC 12478 (CP006835.1), *M. smegmatis* MC^2 (CP009494.1), *M. tuberculosis* H37Rv (FJ014498.1), *M. vulneris* ACS3670 (MBDY01000080.1), *M. nonchromogenicum* strain NCK 8460 (KonoS1), *M. thermoresistibile* NCTC10409 (NZ_LT906483.1), *M. bovis* AF2122/97 (LT708304.1), *M. leprae* TN (AL583917.1), *M. liflandii* 128FXT (CP003899.1), *M. riyadhense* NLA000201958 (EU552927.1), *M. goodii* X7B (CP012150.1), *M.* sp. *KMS* (CP000518.1), *M.* sp. *MCS* (CP000384.1), *M.* sp. *JLS* (CP000580.1), *M. rhodesiae* NBB3 (CP003169.1), *M. phlei* CCUG 21000 (CP014475.1), *M. caprae* Allgaeu (CP016401.1), *M. neoaurum* MN4 (JXYZ01000006.1), *M. elephantis* strain Lipa Mele (LBNO01000056.1), *M. holsaticum* M7 (NZ_MIGZ01000035.1), *M. malmesburii* CIP:10822T (CVTB01000324.1), *M. abscessus* sp. abscessus 800 (FSCO01000216.1), *M. conceptionense* CCUG 50187 (LQOP01000014.1), *M. senegalense* CK1 (LDCO01000022.1), *M. wolinskyi* ATCC 700010 (LQQA01000034), *M. brisbanense* JCM15654 (BCSX01000024.1), *M. hassiacum* DSM 44199 (AMRA01000004.1) and *M. ulcerans* sp. *shinshuense* ATCC 33728 (AP017635.1). Either genebank accession numbers for the sequences retrieved from the database, or isolate identification number for the sequences retrieved from the ARC-OVI database are shown in parenthesis. The trees are unrooted. Mycobacterial species with a blue label are regarded as slow growing mycobacteria (SGM), while mycobacterial species with a red label are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

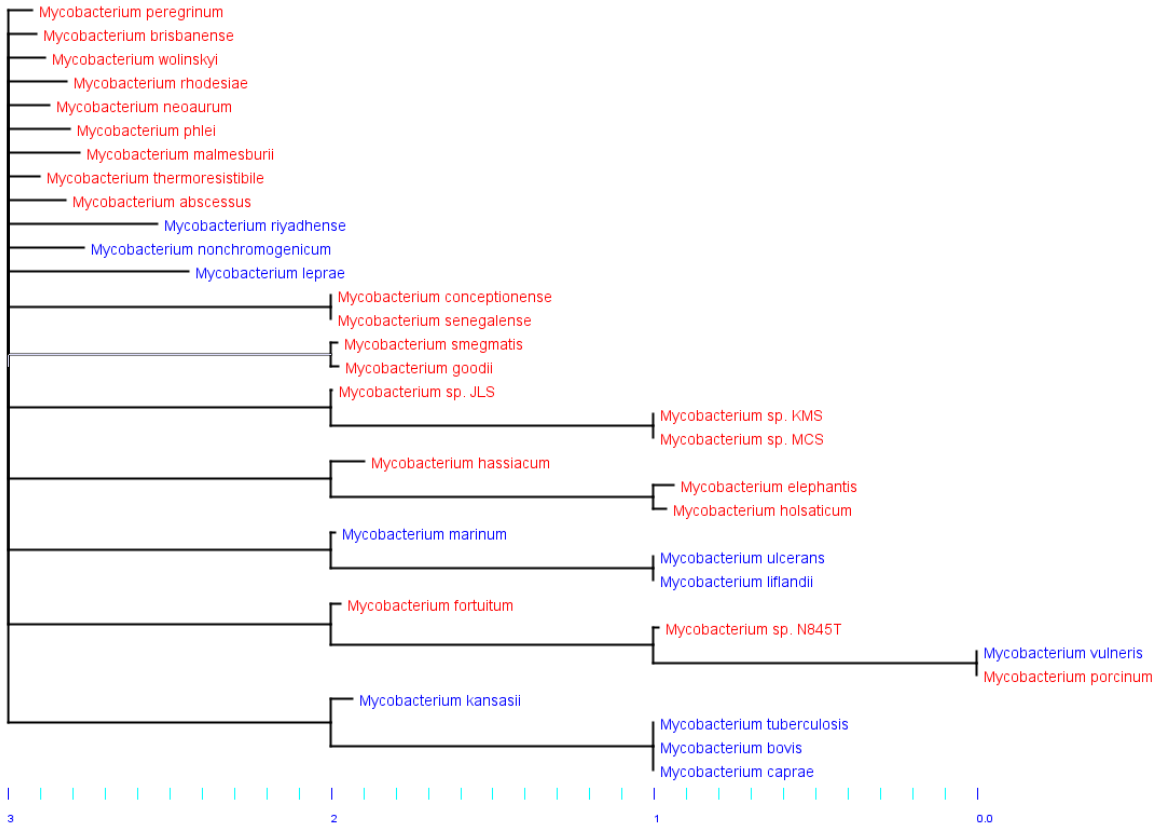

Figure G: Consensus tree based on the gene sequences of only *esxA* and *esxB* created by Mesquite using strict, semistrict and majority-rule consensus. Mesquite creates strict concensus trees by presenting only those clades that are present in all of the trees in the tree source. Semistrict consensus trees present only those clades that are present in at least one of the trees and are not contradicted by any of the other trees. Majority-rule consensus trees present only those clades in the specified fraction (>0.50) of the trees. Tree weights are considered, so each tree contributes the value of the tree weight rather than 1, to the tabulation of the frequency of clades contained therein. In this case, each bootstrap replicate – rather than each tree – counts equally toward the clade frequencies. Because all the trees contained the same taxa, the results of the strict, semistrict and majority-rule concensus trees are the same[76]. Mycobacterial species in blue are regarded as slow growing mycobacteria (SGM), while mycobacterial species in red are generally classified as rapid growing mycobacteria (RGM)[6,55,79-87].

Appendix D: Sequences of M. acapulcensis and M. tuberculosis (potential) esxA gene homologues
Sequences were aligned by MEGA pairwise alignment.

#*M. tuberculosis* H37Rv *esxA* gene sequence[66]
#*M. acapulcensis* consensus sequence
#*M. bovis* forward *esxA* primer
#*M. bovis complement* reverse *esxA* primer

* = identical nucleotide at a certain position (n = 177)
o = ambiguous nucleotide at a certain position, but nucleotide in the aligned sequence could be an option (n = 5)
- = no nucleotide present when compared to the aligned sequence (n = 37 in *M. tuberculosis* sequence, n = 5 in *M. acapulcensis* sequence)

```
5'ATGACAGAGCAGCAGTGGAA3'
5'ATGACAGAGCAGCAGTGGAATTTCGCGGGTATCG-AGGCCGCGGCAAGCG
5'ATGACAGAGCAGCAGTGGAATTTCGCGGGTATGACAGAGCGCAGTGGACC
 *** * ***** * *** *** *** ****** **** * ****    **   *** *  *     *

CAATCCAGGGAAATGTCACGTCCATTCATT--CCCTCCTTGACGAGGGGAAG
GGMCT---GGCCCGGGCCCGCCCCGTGGTCGGCCAGCATAGATGAGGSGAAS
  o      **     *  *  **  **   *  *    **   * * **  ****o***o

CAGTCCCTGAC--------CAAGCTCGC-----------AGCGGCCTGGG-
CATATCYTTACGCTTCGTCGAGCGCGCGATAAGTCCACYAACGGCCTTCAC
**    *o*  **          * ***  ***              *  *******

------GCGGTAGCGGTTCGGAGGCG-----TACCAGGGTGTCCAGCAAAAA
ACGTCAGCGCTATCGCCTCGGGGGCACGCGCTACTTGCGCGGCGAACAGCTG
        *** **  **  ****  ***       ***  * * * * * **

TGGGACGC-CACGGCTAC--CGAGCTGAACAACGCGCTGCAGAACCTGGCGC
TGTGATGGACACGGSCGCGGCGGCTTGGGCCA-GCACCGCACGGT-TGCCCC
**  **  *   *****o  *  **    **   * * * * * ***      ** * *

GGACGATCAGCGAAGCCGGTCAGGCAATGGCTTCG-ACCGAAGGCAACGTC
ACCGGTCCAGTTGTGCGCGTTCACTGGGAGGTTCGCATAGAAGGCAACGTC
      *  ***      **  **           * **** *  * ********** ***
                                                    5'TC

ACTGGGATGTTCGCATAG3'
ACTGGGATGTTCGCATAG3'
******** ******** ** ***
ACTGGGATGTTCGCATAG3'
```

Appendix E: Sequences of M. acapulcensis and M. smegmatis (potential) esxA gene homologues
Sequences were aligned by MEGA pairwise alignment.

#*M. smegmatis* MC^2-155[94]
#*M. acapulcensis* consensus sequence

* = identical nucleotide at a certain position (n = 161)
o = ambiguous nucleotide at a certain position, but nucleotide in the aligned sequence could be an option (n = 5)
- = no nucleotide present when compared to the aligned sequence (n = 38 in *M. smegmatis* sequence, n = 4 in *M. acapulcensis* sequence)

```
5'ATGACAGAACAGGTATGGAATTTCGCCGGTATCGAGGGCGGCGCGTCGGA
5'ATGACAGAGCAGCAGTGGAATTTCGCGGGTATGACAGA--GCGCAGTGGA
  ********  ***     ********** ***  **  ***      *     ****     ***

GATCCACGGCGCCGTGTCCACCACG--GCCGGTCTGC-TCGACGAGGGCAA
CCGGMCTGGC-CCGGGCCCCGCCCCGTGGTCGGCCAGCATAGATGAGGSGAA
      o   ** *  *** *  ** **  **   *  *** *  ** *  ** ***** o **

G-----------GCCTCGCTGA-----------CCACTCTCG-CCTCGGC
SCATATCYTTACGCTTCGTCGAGCGCGCGATAAGTCCACYAACGGCCTTCAC
o              **  ***   **                  *** *o  **  ***    *

GTGG-GGCGGCACCGGTTCGGAGGC-----CTACCAGGCCGTC---CAGGC
ACGTCAGCGCTATCGCCTCGGGGGCACGCGCTACTTGCGCGGCGAACAGCT
   *     ***   *  **   **** *  ***       ****   *   **  *     ***

CCGTTGGGACTCCACCTCCAACGAGCTGAACCTGGCACTGCAGAACCTCGCC
GTGTGATGGACACGGSCGCGGCGGCTTGGGCC-AGCACCGCACGGTTGC-CC
   **    *      *  o   *   **    **   **   **** ***       * **

CAGACCATCAGCGAGGCGGGC-CAGACCATGGCGCAGACCGAGGCCGGCGT
CACCGGTCCAGTTGTGCGCGTTCACTGGGAGGTTCGCATAGAAGGCAACGT
**        ***      ***  *   **        **   *      **  *  *   ***

CACGGGAATGTTTGCCTGA3'
CACTGGGATGTTCGCATAG3'
***  **  ***** **  *
```

Appendix F: Alignment of *esxA* and *esxB* primers derived from *M. smegmatis* and *M. bovis*
Primer sequences were aligned with the help of MEGA software[68].

Primer	Primer sequence	Identity percentage
M. smegmatis esxA forward primer	5' A A T T T C G C C G G T A T C G A G G G - 3'	7/20 = 35%
M. bovis esxA forward primer	5' - A T G A C A G A G C A G C A G T G G A A 3'	
	* * * * * * *	
M. smegmatis esxA reverse primer	5' - C A G G C A A A C A T T C C C G T G A C 3'	14/20 = 70%
M. bovis esxA reverse primer	5' C T A T G C G A A C A T C C C A G T G A - 3'	
	* * * * * * * * * * * * *	
M. smegmatis esxB forward primer	5' - - - G C G A A T T T C G A G C G C A T C T C 3'	7/20 = 35%
M. bovis esxB forward primer	5' A T G G C A G A G A T G A A G A C C G A - - - 3'	
	* * * * * * *	
M. smegmatis esxB reverse primer	5' - - - G A T G T T C A T C G A C G A C G A A A G 3'	10/20 = 50%
M. bovis esxB reverse primer	5' T C A G A A G C C C A T T T G C G A G G - - - - 3'	
	* * * * * * * * * * *	