Insights into ancestry and adaptive evolution of the *Mycobacterium tuberculosis* complex from analysis of the emerging pathogen *Mycobacterium riyadhense*

Qingtian Guan[1], Musa Garbati[2,#a], Sara Mfarrej[1], Talal AlMutairi[2], Alicia Smyth[3¶], Albel Singh[4¶], Shamsudeen Fagbo[5¶], John A. Browne[3], Muhammad Amin urRahman[2], Alya Alruwaili[2], Anwar Hoosen[2,#b], Conor J Meehan[6], Chie Nakajima[7,8], Yasuhiko Suzuki[7,8], Apoorva Bhatt[4], Stephen V. Gordon[3,7], Faisal AlAsmari[2], A. Pain[1,7*]

[1]Pathogen Genomics Laboratory, BESE Division, King Abdullah University of Science and Technology (KAUST), Thuwal-Jeddah, Kingdom of Saudi Arabia

[2]King Fahad Medical City（KFMC), Riyadh, Saudi Arabia

[3]UCD School of Veterinary Medicine, University College Dublin, Dublin, D04 W6F6, Ireland

[4]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK

[5]One Health Unit, Executive Directorate for Surveillance and Response, Saudi Center for Disease Prevention and Control, Riyadh, Saudi Arabia

[6]School of Chemistry and Biosciences, University of Bradford, Bradford, UK

[7]Global Institution for Collaborative Research and Education, Hokkaido University, Kita 20 Nishi 10, Kita-ku, Sapporo, Japan

[8]Research Center for Zoonosis Control, Hokkaido University, Kita 20 Nishi 10, Kita-ku, Sapporo, Japan

[#a]Current address: Infectious Diseases Unit, Department of Medicine, University of Maiduguri, Nigeria

[¶]These authors contributed equally to the work

[#b]Current address: Pathcare/Vermaak & Partners Pathologists, Pretoria, South Africa.

*Correspondence author

Email: arnab.pain@kaust.edu.sa (AP)

Abstract

Current evolutionary scenarios posit the emergence of *Mycobacterium tuberculosis*, the deadliest bacterial pathogen for humans globally, from an environmental saprophyte through a cumulative process of genome adaptation. *Mycobacterium riyadhense* is a novel non-tuberculous mycobacterium (NTM) that is being increasingly isolated from human clinical cases with tuberculosis (TB)-like symptoms in various parts of the world. We provide evidence here that *M. riyadhense* is likely a 'missing link' in our understanding of the evolution of *M. tuberculosis*. To elucidate the genomic hallmarks that define the evolutionary relationship between *M. riyadhense* and other mycobacterial species, including members of the *Mycobacterium tuberculosis* complex (MTBC), eight clinical isolates of *M. riyadhense* were sequenced and analyzed. We show, among other features, that *M. riyadhense* shares a large number of conserved orthologues with the MTBC; contains linear and circular plasmids carrying type IV and type VII secretion systems; and shows expansion of toxin/anti-toxin pairs. We conclude that *M. riyadhense* is an emerging mycobacterial pathogen that shares a common ancestor with members of the MTBC and that can serve as an experimental model to study the evolution and pathogenesis of tubercle bacilli.

Author summary

Mycobacterium tuberculosis is one of the most prolific infectious killers in humans and is a member of the *Mycobacterium tuberculosis* complex (MTBC) - a group of genetically related pathogens that cause tuberculosis (TB) in mammalian species. It is postulated that MTBC has evolved from a free-living environmental ancestor into an

obligate pathogen. In this evolutionary context, a comprehensive understanding of the genomic hallmarks of the free-living environmental ancestors of the MTBC is of particular scientific interest for better understanding of the evolution of the MTBC. *Mycobacterium riyadhense* is a novel environmental mycobacterium, first isolated in 2009 in a hospital in Riyadh, that is increasingly being isolated from clinical cases with typical tuberculosis (TB)-like symptoms in humans. In this study, we report the characterization of eight clinical isolates of *M. riyadhense*, compare their genomes to members of the MTBC, and provide a comprehensive insight into the adaptive changes associated with the evolution of the MTBC from environmental mycobacteria. We show that *M. riyadhense* is one of the closest known environmental mycobacteria related to the MTBC, and we provide several lines of molecular evidence that *M. riyadhense* is likely the 'missing link' in the evolution of *M. tuberculosis*. It shares a common ancestor with members of the MTBC that have evolved through a process of genome reduction, expansion of toxin/antitoxin (T/A) gene systems, and ultimately host adaptation.

Introduction

The *Mycobacterium tuberculosis* complex (MTBC) is a group of genetically related pathogens that cause tuberculosis (TB) in mammalian species. The hallmark member, *Mycobacterium tuberculosis*, is the single most deadly pathogen, causing over 1.6 million deaths globally in 2017. Current evolutionary scenarios posit the evolution of the MTBC from an environmental saprophyte through a cumulative process of genome adaptation. Such scenarios envisage intermediate mycobacterial species with increasing pathogenic potential for humans, the vestiges of which should be present in extant mycobacterial

92 species. Comparative genomic analyses between the MTBC members and opportunistic
93 mycobacterial pathogens may therefore reveal the key evolutionary steps involved in the
94 emergence of the MTBC, as well as illuminating virulence mechanisms across
95 mycobacterial pathogens as a whole.

96 Non-tuberculous mycobacteria (NTMs), including *Mycobacterium riyadhense* (MR),
97 are ubiquitous, naturally occurring environmental bacteria commonly found in water and
98 soil[1,2]. A wide range of animal and environmental sources (aquaria, swimming pools)
99 act as reservoirs for NTMs, and several human disease outbreaks caused by exposure to
100 environmental NTMs have been described[3–6]. With the ability to cause infections in both
101 immunocompromised[7] and immunocompetent[8] individuals, *M. riyadhense* has
102 positioned itself as a clinically important pulmonary pathogen since its discovery in 2009
103 [2]. The clinical and radiologic characteristics of pulmonary infection caused by
104 *M. riyadhense* are indistinguishable from those caused by *M. tuberculosis*, the most
105 important human pathogen of the MTBC [2,8].

106 Similar to *M. tuberculosis*, *M. riyadhense* grows at 37°C and requires 2~3 weeks[2]
107 to form visible colonies on agar media. However, unlike *M. tuberculosis*, which is a
108 worldwide pathogen transmitted directly from human-to-human with no known
109 environmental reservoirs[9], *M. riyadhense* infections are rare and are transmitted to
110 patients via contact with contaminated water[9] and soil[10], with no evidence of human-
111 to-human transmission yet reported. Infections with *M. riyadhense* have been reported in
112 Asia and Europe in countries including Bahrain, South Korea, France, Italy, and Germany
113 [8,11,12], although most of the recent cases originated in patients from Saudi Arabia[13].

Indeed, the very first case of acute *M. riyadhense* infection was initially misdiagnosed as a case of *M. tuberculosis* infection in a Saudi hospital using commercially available diagnostic tests[14].

It is postulated that *M. tuberculosis* evolved from a free-living environmental ancestor into an obligate pathogen[15,16]. Our current knowledge indicates that *Mycobacterium canettii* is the most closely related obligate pathogenic species to the MTBC[17,18]. Infections with *M. canettii* are rare and found solely in people from the Horn of Africa, with no environmental reservoir defined to date[19].

Previous phylogenetic studies have suggested that *Mycobacterium kansasii*[20], *Mycobacterium marinum*[21], *Mycobacterium lacus*[22], *Mycobacterium decipiens*[23], *Mycobacterium shinjukuense*[24] and *M. riyadhense*[25] are closely related to the free-living ancestor of the MTBC based on a single marker gene (e.g., *hsp65*, 16S). In the evolutionary context, a comprehensive understanding of the genomic hallmarks that drive the close phylogenomic relationship of *M. riyadhense* with the MTBC and NTMs is of particular scientific interest.

In this study, we report the characterization of eight clinical isolates of *M. riyadhense*, compare their genomes to members of the MTBC, and provide comprehensive insight into the adaptive changes associated with the evolution of the MTBC from environmental bacteria. We examined the lipid profiles of both rough and smooth variants of *M. riyadhense*, compared them to those of other related mycobacteria, and analyzed the transcriptional response of immunity-related host genes in a murine macrophage infection model with *M. riyadhense*, *M. kansasii*, and *M. bovis* BCG.

Using genome information, we developed a simple PCR-based diagnostic test for the rapid and accurate identification of *M. riyadhense* to minimize the risk of misdiagnosis in a clinical setting.

Our analyses provide a comprehensive description of the hallmarks of *M. riyadhense* that make it one of the closest known environmental relatives of the MTBC, and that can serve to illuminate studies into the evolution and pathogenesis of the MTBC.

Results

Clinical manifestation and culture characteristics of *M. riyadhense* isolates

Between April 2011 and March 2017, eight clinical cases of infection with *M. riyadhense* were recorded in male patients aged from 8-82 years (Fig 1). In addition to HIV/AIDS, most of the patients had multiple comorbidities, such as pulmonary and/or systemic hypertension, malignancies, and diabetes mellitus (DM), with the lung being the major site of disease (Fig 1).

Bronchoalveolar lavage, endotracheal, sputum, and lymph node biopsy specimens grew *M. riyadhense* colonies when cultured in solid and liquid media using LJ agar and *Mycobacteria* growth indicator tube (MGIT) broth, respectively, with varying times to positivity. Although no consistent findings existed for chest imaging, three out of the eight patients presented with upper lobe consolidation, cavitation, ground-glass opacities, 'tree-in-bud' appearance, hilar lymphadenopathy, and pleural effusion (S1 Fig).

The isolates were subjected to susceptibility testing for antibiotics commonly used to treat both typical and atypical infections. Isolates showed 100% susceptibility to rifampin (RIF), rifabutin (RFB), ethambutol (EMB), clarithromycin (CLR), linezolid (LZD),

159 amikacin (AMK), moxifloxacin (MOXI), and trimethoprim-sulfamethoxazole (TMP-
160 SMX). Three out of the eight isolates were resistant to ciprofloxacin (CIP).
161
162 **Assembly and annotation of the *M. riyadhense* MR226 genome**
163 The comparison of different assemblies of all sequenced *M. riyadhense* strains is listed
164 in Table S1. We obtained chromosomes of all eight isolates in single contiguous sequences
165 for genome comparison at a high resolution. The *M. riyadhense* MR226 strain contains a
166 6,243,587 bp chromosome, a linear plasmid (pMRLP01) of 550,247 bp and a circular
167 plasmid (pMR01) of 94,344 bp. The circular nature of the chromosome and the pMR01
168 plasmid were demonstrated through Gepard[26].
169 As expected of a free-living opportunistic pathogen, the chromosome of *M.*
170 *riyadhense* is significantly larger than the chromosomes of the MTBC (Table 1). The
171 number of genes unique in a species showed that the members of the MTBC have a
172 considerably lower percentage of unique genes than *M. riyadhense* and other closely
173 related NTMs (Table 1)

174

Table 1. Comparison of general features of genome in Mycobacteria

Features	*M. tuberculosis*	*M. africanum*	*M. bovis*	*M. canettii*	*M. riyadhense*	*M. kansasii*	*M. marinum*
Genome Size (M)	4.41	4.39	4.26	4.48	Chromosome: 6.25 Plasmid pMRLP01: 0.55 Plasmid pMR01: 0.09	Chromosome: 6.43 Plasmid pMK12478: 0.14	Chromosome: 6.34 Plasmid pRAW: 0.11
GC content (%)	65.6	65.6	65.6	65.6	Chromosome: 65.35 Plasmid pMRLP01: 67.36 Plasmid pMR01: 65.69	Chromosome: 66.2 Plasmid pMK12478: 65.8	Chromosome: 65.8 Plasmid pRAW: 64.1
CDS	3,906	4,045	4,042	4,064	Chromosome: 5,490 Plasmid pMRLP01: 443 Plasmid pMR01: 88	Chromosome: 5,631 Plasmid pMK12478: 152	Chromosome: 5,149 Plasmid pRAW: 94
Gene Density(bp)	1101	1085	1032	1044	Chromosome: 1126 Plasmid pMRLP01: 1239 Plasmid pMR01: 1072	Chromosome: 1142 Plasmid pMK12478: 954	Chromosome: 1176 Plasmid pRAW: 1166

Continued Table 1. Comparison of general features of genome in Mycobacteria

Conserved protein with assigned function	2855 (73%)	3078 (76%)	3072 (76%)	3125 (77%)	Chromosome: 4079 (74%) Plasmid pMRLP01: 217 (49%) Plasmid pMR01: 22 (25%)	Chromosome: 4468 (79%) Plasmid pMK12478:74 (49%)	Chromosome: 4091 (79%) Plasmid pRAW: 38 (40%)
Unique CDS	62 (1.59%)	25 (0.62%)	40 (0.99%)	83 (2.04%)	Chromosome:1059(19.29%) Plasmid pMRLP01:293 (66.14%) Plasmid pMR01:32 (36.36%)	Chromosome: 1,080 (19.18%) Plasmid pMK12478: 61 (40.13%)	Chromosome: 993 (19.29%) Plasmid pMK12478: 42 (44.68%)
tRNA	45	45	47	45	Chromosome: 51 Plasmid pMRLP01:1	Chromosome: 46 Plasmid: NA	Chromosome: 47 Plasmid: NA
rRNA	3	3	3	3	Chromosome: 3 Plasmid: NA	Chromosome: 3 Plasmid: NA	Chromosome: 3 Plasmid: NA
Transposase/integrase	53	68	56	98	Chromosome:110 Plasmid pMR01:1 Plasmid pMRLP01: 72	Chromosome:79 Plasmid pMK12478:3	Chromosome:41 Plasmid pRAW:4
Reference	[27]	[28]	[29]	[17]	This Study	[30]	[31]

175

176 The comparison of the annotated protein-coding genes from the *M. riyadhense*
177 MR226 strain to the genome assemblies of 152 *Mycobacterium* species (including 77
178 known mycobacterial plasmids) shows that *M. riyadhense* forms a tight cluster with
179 *M. marinum* and *M. angelicum*, while the MTBC members clusters together with *M.*
180 *shinjukuense*, *M. lacus* and *M. ulcerans* (S2(A) Fig, highlighted in box a). This clustering
181 is not unexpected as *M. lacus*, *M. shinjukuense* and *M. ulcerans* have smaller genomes and
182 have less predicted proteins as opposed to *M. riyadhense*, *M. marinum* and *M. angelicum*,
183 whose genomes are greater than 6 Mb. A total of 335 genes were identified as unique from
184 the orthologue group comparison, being present in only all sequenced *M. riyadhense*
185 isolates, including MR226. The vast majority of these genes belong to the PE/PPE family,
186 which are thought to be involved in antigen variation and are widely spread across the
187 slow-growing species within the *Mycobacterium* genus.

188 Linear plasmids were first described in 1989 in maize (which has a linear
189 mitochondrion)[32] and have also been found in *Actinomycetales*, including *Streptomyces*
190 [33], *Rhodococcus*[34] and *Mycobacterium* species, such as *Mycobacterium xenopi*,
191 *Mycobacterium branderi* and *Mycobacterium celatum*. They are often accompanied by a
192 circular plasmid in the same host[35]. The linear plasmid pMRLP01 in *M. riyadhense*
193 contains a pair of partitioning genes (*parA*/*parB*) that are involved in active segregation
194 and thus stabilize the inheritance of the plasmid[36]. The latter are known to contribute to
195 genome evolution by active DNA transfer and exchange[37]. As is often the case for both
196 circular and linear large plasmids, a relatively higher proportion of pMRLP01 genes (51%)
197 have no known function compared to those of the main chromosome (26%). This reinforces
198 the notion that plasmids are an important route by which new genes are introduced into the

199 genome in *Mycobacteria*. Of the 443 predicted protein-coding genes of pMRLP01, 118
200 have at least one orthologue on the main chromosome. Furthermore, we observed several
201 protein-coding genes in pMRLP01 that have orthologues in the genomes of *Mycobacterium*
202 *tusciae* JS617, *Mycobacterium aromaticivorans* JCM 16368, *Mycobacterium llatzerense*,
203 *Mycobacterium obuense*, *Mycobacterium novocastrense* and *Mycobacterium holsaticum*
204 (S2(B) Fig). This finding indicates that pMRLP01 can be readily exchanged with other
205 environmental mycobacterial species, with implications for horizontal gene transfer in
206 mycobacteria.

207 It is well known that plasmids are important "vehicles" for the exchange of genetic
208 material between bacteria or between chromosomes and extra-chromosomal plasmids. In
209 this study, we further identified a circular plasmid termed pMR01 in *M. riyadhense* (Figs
210 S2(C) and S3). When compared with the circular plasmids of other species, such as pRAW
211 in *M. marinum*[38], pMAH135[39] and pMA100[40] of *M. avium*, pMyong1 from
212 *Mycobacterium yongonense*[41], pMK12478[42] from *M. kansasii* and several plasmids
213 from *Mycobacterium chimaera*[43], a high similarity was observed. These plasmids all
214 harbor both a T4S and T7S, which are necessary for conjugation[38,44], and facilitate the
215 exchange of genetic material between different species of slow-growing mycobacteria[38].
216 We therefore speculate that pMR01 is a novel conjugative plasmid.

217 In mycobacteria, five type VII secretion systems have been described, named ESX-1
218 to ESX-5[45,46]. An ESX-P5 locus on pMR01, which shows high similarity to the ESX-5
219 loci on pMK12478, pRAW and pMAH135, is markedly different from the ESX-5 system
220 found on the main *M. riyadhense* chromosome. ESX-5 is involved in the secretion of PE
221 and PPE proteins in *M. tuberculosis* and is involved in modulating the host immune

responses to maintain a persistent infection[47]. The potential transmissibility of pMR01 and other pMR01-like plasmids may therefore play a role in the evolution of the ESX systems in mycobacteria.

The progressive alignments (S4(A) Fig) of the assembled chromosomes and plasmids (S4(B) Fig) of each *M. riyadhense* strain show that the chromosomes are relatively conserved; however, the linear plasmids present in all eight sequenced isolates are diverse from both structural and similarity perspectives, while the pRAW-like plasmids are present in only the MR226, MR193 and MR222 strains.

The SNP-based phylogeny of the sequenced *M. riyadhense* isolates based on 43,136 polymorphic sites (S2 Table, S5 Fig) indicates the presence of two different clades of *M. riyadhense* among the clinical isolates sequenced in this study. The nucleotide diversity between the MR222 clade is greater than the diversity between *M. tuberculosis* strains[48], while it is smaller than that seen between *M. canettii* strains[17], and the variation between the MR226 clade is comparable to the SNP variation in *M. tuberculosis* strains.

Regions of difference (RDs) in *M. riyadhense*

The RDs were originally described as genomic regions present in virulent *M. bovis* and *M. tuberculosis* but absent from the *M. bovis* BCG genome[49]. RD loci were subsequently described across the MTBC[50] and contain functions believed to contribute to pathogenicity[51–53] and evolution of MTBC species[54]. *M. riyadhense* was found to harbor most of the RD loci (RD1, RD3-R11, R13-RD16) that are also intact in *M. tuberculosis,* while 2 of the RDs show unique deletions, $RD2^{riyadh}$ (S6(A) Fig) and $RD12^{riyadh}$ (S6(B) Fig).

RD2 was originally described as deleted in BCG vaccine strains. Subsequently, it was shown that disruption of RD2 in *M. tuberculosis* leads to decreased proliferation *in vivo* and impaired modulation of the innate immune response[52]. The RD2 locus also has a deletion from the *M. riyadhense* genome. It is a larger deletion than the originally described $RD2^{BCG}$ as $RD2^{riyadh}$ contains 29 genes (*rv1971~rv2000*, location 2,216,498~2,246,766); eight genes within this locus (*rv1978*, *rv1979c*, *rv1980c*, *rv1981c*, *rv1983*, *rv1984*, *rv1987*, *rv1988*) have orthologues elsewhere in the *M. riyadhense* genome (*mr_05764*, *mr_05852*, *mr_02310*, *mr_02993*, *mr_00486*, *mr_02995*, *mr_02325*, *mr_02349*, *mr_01747*), suggesting possible functional redundancy.

The RD12 locus shows deletions across MTBC members, including *M. bovis*, *Mycobacterium caprae* and *M. orygis*[55], but it is present in other MTBC members. *M. canettii* isolates (except group B[56]) also show an independent deletion at the RD12 locus named $RD12^{can}$ (3,479,430~3,491,866, *rv3111~rv3126*), which is distinct from $RD12^{bovis}$ (3,484,740~3,487,515, *rv3117~rv3121*). We identified another unique deletion at the RD12 locus in *M. riyadhense*, designated $RD12^{riyadh}$, which encompasses a larger region than RD^{can} and $RD12^{bovis}$, encompassing *rv3108-rv3127* (3,477,171~3,492,150) (S6(B) Fig). It is intriguing that multiple mycobacteria show independent deletion events at the RD2 and RD12 loci, suggesting selective forces play a role in this variation.

Comparative phylogeny of *M. riyadhense* with other Mycobacteria

The phylogenetic tree shows that the slow-growers and rapid-growers are separated into two different clades and that fast-growers are ancestral compared to slow-growers (Fig 2). The overall topology of our tree is similar to that of previously published phylogenetic

trees[22]. In the tree, *M. riyadhense* is located within the same clade as the obligate and opportunistic causal organisms of mycobacterial diseases in humans that include the MTBC, *M. marinum*, *M. kansasii*, *M. leprae* and related host-restricted mycobacteria with reduced genomes and decreased survivability in the environment.

The PE/PPE family *mce* and *mce*-associated genes are known to be important for host adaptation[57] and pathogenicity[58]. The PE/PPE family genes are enriched in the MTBC members but also in *M. riyadhense* MR226 (278) and other pathogenic species, such as *M. kansasii* (228) and *Mycobacterium ulcerans* (200). The number of *mce*- or *mce*-associated genes has not significantly changed across mycobacterial genomes, indicating that this group of genes is under an evolutionary constraint and plays functional roles bridging both environmental and obligate pathogen lifestyles. Our results agree with previous findings that during their evolution, the ESX systems were derived from the ancestor ESX-4, as shown in Fig 2 at the root node, and then ESX-3, ESX-1, ESX-5 and ESX-2 evolved by horizontal transfer[59].

A comparative phylogenetic map based on 906 conserved proteins (S7 Fig) reveals this downsizing of the genome and the dynamic changes in genome components. Certain functional categories of genes are relatively enriched during the evolution of MTBC, including protein metabolism, regulation and cell signalling, cell division and cell cycle.

M. riyadhense shares a larger number of orthologues (3,122) with *M. tuberculosis* than with *M. kansasii* (2,978), *M. marinum* (2,962) and *M. szulgai* (2,724) among the environmental mycobacteria that are closely related to the MTBC (Fig 3(A)). A total of 134 orthologues are exclusively shared between *M. riyadhense* and *M. tuberculosis*, while the number of orthologues exclusively shared between *M. tuberculosis* and *M. kansasii*

(30), *M. marinum* (48) and *M. szulgai* (18) is less (Fig 3(A)). It is notable that genes from the phage-derived regions of RD3 and RD11 are shared exclusively between *M. riyadhense* and *M. tuberculosis*.

The comparative analysis of the orthologue groups of *M. riyadhense* and the MTBC is informative. Firstly, 385 orthologue groups present across all MTBC species are absent from *M. riyadhense*. Secondly, 221 orthologues uniquely present in *M. riyadhense* are not found amongst the MTBC (Fig 3(B)). This latter group of *M. riyadhense* unique orthologues likely depict the constraints imposed by the free-living biology of *M. riyadhense* which necessitates maintaining a broad functional repertoire to secure environmental survival. In contrast, the obligate MTBC species have lost genes involved in environmental survival but gained a large number of genes required for survival in the *in vivo* environment.

A hallmark of *M. tuberculosis* infection is the ability to survive long-term in host granulomas and develop a latent stage of infection. The molecular mechanisms and cellular components that are involved in the persistence of *M. tuberculosis* are still poorly understood, but several T/A systems have been implicated in the pathogenicity of *M. tuberculosis*[60]. T/A systems were first found on plasmids or plasmid-derived chromosomal loci where they promoted plasmid maintenance in bacterial populations[61], but when compared to other mycobacteria, the MTBC are remarkable for the extensive expansion of T/A systems. We thus compared the 79 pairs of T/A systems (belonging to the HigAB, MazEF, ParDE, RelEF, VapBC and UCAT families) in *M. tuberculosis* with the T/A pairs found in other members of the MTBC and NTMs. Based on the presence of 49 out of the 79 T/A orthologue pairs (Fig 4), *M. riyadhense* appears more closely related

314 to the MTBC than to other mycobacteria, including *M. lacus, M. shinjukuense* and
315 *M. decipiens*. It can be hypothesized that the shared component of T/A pairs play a
316 functional role in the pathogenicity or persistence of *M. riyadhense* infection in a way
317 similar to that described for the importance of T/A pairs in *M. tuberculosis in vivo* biology.
318

319 **_M. riyadhense_ strains produce a distinct pattern of LOSs**

320 The lipopolysaccharides (LOSs) are an important class of glycolipids that have
321 previously been linked to diverse mycobacterial phenotypes, such as colony morphology,
322 secretion of PE/PPE family proteins, and the pathogenicity of *M. marinum*[62]. It is
323 noteworthy that we observed both smooth (MR210, MR22, MR226, MR244, MR246, and
324 MR1023) (Fig 5(A)) and rough (MR193 and MR206) (Fig 5(B)) morphologies in *M.*
325 *riyadhense* strains. We therefore sought to first examine whether the genetic machinery for
326 the production of LOS is present in the *M. riyadhense* genome, and then to follow up on
327 the genome-level predictions with lipid analyses of rough and smooth variants using Thin-
328 layer Chromotagraphy (TLC).

329 We observed that the *wecE* and *galE6* LOS genes are absent from the *M. riyadhense*
330 genome (Table 2, S3 Table). These genes are linked with the removal of LOS II* and the
331 production of LOS IV[62]. Thus, their absence is likely to cause an accumulation of LOS
332 II* and the lack of fully formed LOS IV, which have previously been shown to increase
333 the pathogenicity of *M. marinum*[62]. Furthermore, both the *pks5* and *pap* genes in the
334 LOS locus are intact in *M. riyadhense*, as is the case in *M. canettii*, but not in
335 *M. tuberculosis,* where the former is truncated and the latter deleted[63]. This finding
336 indicates that *pks5* recombination and *pap* deletion occurred in a common ancestor of the

337 MTBC after its differentiation from both *M. riyadhense* and *M. canettii*. Remarkably, the
338 *M. riyadhense* LOS gene locus layout is dissimilar to that in *M. canettii*, *M. tuberculosis*,
339 *M. kansasii* and *M. marinum*: indeed, exclusive rearrangements of this locus in *M.*
340 *riyadhense* were observed (Fig 5(C)).

Table 2. Presence and absence of the *M. marinum* orthologs related to LOS synthesis in *M. riyadhense* and other species

	M. smegmatis	*M. kansasii*	*M. canettii*	*M. riyadhense*	*M. tuberculosis*
MMAR_1008	P	P	P	P	P
MMAR_2307	A	P	A	P	A
MMAR_2313	A	P	P	P	P
MMAR_2319	A	A	A	A	A
MMAR_2320 (*wecE*)	A	P	P	A	A
MMAR_2327	A	P	P	P	P
MMAR_2332	A	A	A	A	A
MMAR_2336 (*galE*)	A	P	A	A	A
MMAR_2340 (*pks*)	P	P	P	P	T
MMAR_2341	P	P	P	P	P
MMAR_2343 (*pap*)	P	P	P	P	A
MMAR_2353	P	P	A	P	A
MMAR_5170 (*whiB4*)	P	P	P	P	P

P: Present; A: Absent; T: Truncated

343 To correlate rough vs. smooth colony morphology with LOS production, we extracted
344 polar lipids from the strains and analyzed them by 2D-TLC using solvent system E[64],
345 which is designed to separate phospholipids and LOSs. Charring of the TLC plates with
346 alpha-napthol revealed glycolipids, including the accumulation of a species that migrated
347 at a position similar to that of LOS III. This lipid was seen in only smooth strains; species
348 with migration patterns similar to LOS I and LOS II were observed, while LOS IV was
349 absent. This result was not unexpected because all *M. riyadhense* strains lack a functional
350 *wecA* gene, which is required for the extension of LOS II to LOS IV (Fig 5). Additionally,
351 the relative levels of the predominant LOS species in *M. riyadhense* seem to be quite high
352 when compared to those seen in other LOS-producing mycobacteria (Figs 5(D)(E)).
353 Conversely, the rough strains did not produce any glycolipids that migrated in the positions
354 corresponding to LOSs (Figs 5(F)(H)).
355
356 ***PE-PGRS33* locus and type VII secretion system of *M. riyadhense***
357 The *pe-pgrs33* (*rv1818c*) locus encodes the exported protein PE_PGRS33 and plays
358 an important role in the pathogenesis of *M. tuberculosis*[65]. A previous study[66] showed
359 that *pe-pgrs33* is present in all MTBC members but not in *M. canettii*, which implies a
360 specific *pe-pgrs33* insertion event in the ancestor of MTBC strains. Genome comparison
361 of *M. riyadhense* with *M. tuberculosis*, *M. kansasii*, *M. marinum* and *M. canettii* provides
362 additional evidence that *M. riyadhense* is the missing link of the *pe-pgrs33*
363 deletion/insertion event. Our phylogeny strongly suggests that the deletion of *pe-pgrs33*
364 from *M. kansasii* and *M. marinum* occurred before the divergence of environmental
365 mycobacteria and the smooth tubercle bacilli (STB)/MTBC clade (S8 Fig).

366 All 5 ESX systems (ESX1-ESX5) were found in the *M. riyadhense* genome (S9 Fig)
367 but with minor modifications. Hence, the *eccA* and *eccB* genes are absent from the ESX-2
368 system, while the *espACD* operon, which is essential for secretion of virulence factors via
369 ESX1, is also missing in *M. riyadhense*. The overall gene arrangement of the ESX1-ESX5
370 loci is similar in both *M. riyadhense* and *M. tuberculosis*[45]. This conserved synteny
371 reinforces the previous results that phylogenetically *M. riyadhense* may represent an
372 ancestral state to MTBC. As noted before, the pMR01 plasmid also contains an extra ESX-
373 P5 locus, which could indicate a role for this plasmid in mediating pathogenicity (S3 Fig).
374

375 **Transcriptional response of murine macrophage cells upon *M. riyadhense***
376 **infection**

377 Our genomic analysis of *M. riyadhense* revealed a range of genes and potential
378 systems that could play a role in host-pathogen interactions. We therefore sought to assess
379 the initial interaction of *M. riyadhense* with macrophages, using the RAW264.7 cell line
380 of murine origin as our experimental model. As comparator strains in our analysis, we
381 performed parallel infections with *M. kansasii*, an opportunistic pathogen that also contains
382 an orthologous ESX-1 system (Fig 2), and *M. bovis* BCG, the live TB vaccine that is
383 attenuated through deletion of the ESX-1 system. These comparisons allowed us to explore
384 whether *M. riyadhense*-triggered innate immune responses were more similar to those
385 triggered by *M. kansasii* or BCG or were intermediate between the two.

386 To analyze the innate immune responses of the macrophages in the intracellular
387 presence or absence of these mycobacterial isolates, the transcriptional response was
388 analyzed using a 754 probe NanoString Murine Myeloid Innate Immunity panel V2[67]

389 at 3, 24 and 48 hours post infection. These analyses revealed an expected commonality in
390 the responses to infection with all mycobacteria, such as induction of proinflammatory
391 genes through TLR signalling (e.g., upregulation of *IL1B*, *TNFA*, *CCL4*, *PTGS2* and
392 *CXCL2*, albeit to different absolute levels, (Figs 6, S10(A) and S10(B)). Distinct responses
393 triggered by BCG infection compared to *M. riyadhense* and *M. kansasii* included, for
394 example, upregulation of *MARCO* by BCG infection (S10(B) Fig); *MARCO* is involved in
395 pathogen uptake via trehalose dimycolate, a lipid that is known to show variation in
396 structure between the MTBC and *M. kansasii*[68,69]. *CCL24* and *CXCL14*, which are
397 involved in the attraction of immune cells to the site of infection[70], were upregulated to
398 higher levels with *M. riyadhense* and *M. kansasii* than with BCG at 24 and 48 hours.
399 Overall, our analysis of the macrophage transcriptional profiles showed that the
400 response to infection with *M. riyadhense* and *M. kansasii* triggered more similar
401 macrophage transcriptional responses in comparison to the response produced by BCG
402 infection (S10 Fig).
403

404 **Developing a rapid PCR-based diagnostic marker for *M. riyadhense***

405 Due to the issues previously encountered in diagnosing *M. riyadhense* infections[2,8],
406 correct and prompt identification of cases upon presentation at healthcare units is of
407 paramount importance. We therefore sought to translate our knowledge on the genome
408 sequences into a PCR diagnostic test that could be used in a clinical microbiology setting
409 to distinguish *M. riyadhense* from other mycobacteria, including the members of the
410 MTBC.

411 By identifying unique K-mers ranging in size from 11 bp to 4,209 bp (S11 Fig) in the
412 assembled genome compared to the genomes of 152 other mycobacterial species, four
413 primer sets were developed targeting the *mr_00036*, *mr_00263*, *mr_00606* and *mr_01005*
414 genes. The MRDP primer pair MRDP-F/MRDP-R amplified a single product from each of
415 the eight isolates of *M. riyadhense* (Fig 7) but not from other mycobacterial species,
416 including *M. tuberculosis*, *M. bovis*, *M. kansasii*, *M. marinum*, *M. szulgai*, *M. avium* and
417 *Mycobacterium angelicum*. This result shows that the MRDP-F/MRDP-R primers are
418 highly specific to *M. riyadhense* and form the basis for a simple diagnostic PCR that can
419 inform appropriate treatment protocols.
420

421 ## Discussion

422 *M. riyadhense* has become a clinically relevant NTM species globally[8,71]. Contrary
423 to prior publications on *M. riyadhense* that have been based primarily on clinical case
424 reports, here we present the largest and most comprehensive genomic study undertaken to
425 date on clinical *M. riyadhense* isolates. The eight new *M. riyadhense* strains sequenced in
426 this study originated from pulmonary infections with some having additional extra-
427 pulmonary involvement, which fulfilled the American Thoracic Society/ Infectious
428 Diseases Society of America (ATS/IDSA) criteria for NTM infection[72].

429 Our comparative analysis of *M. riyadhense* genomes with the MTBC and a large
430 collection of NTMs provide unequivocal evidence that *M. riyadhense* is one of the closest
431 known environmental mycobacteria species to the MTBC and forms a phylotype with
432 *M. lacus*, *M. decipiens* and *M. shinjukuense*. Indeed, while our manuscript was in
433 preparation, independent work [73] also showed the close phylogenetic relationship of

434 *M. riyadhense* to the MTBC, suggesting that it forms part of an MTB-associated phylotype.
435 Our analyses of multiple *M. riyadhense* isolates complements and extends the findings of
436 Sapriel and Brosch by revealing that expansion of T/A pairs, modification of secretion
437 systems, alterations in cell wall lipids, and plasmid-mediated horizontal acquisition of new
438 functionality all played key roles in the evolution of the MTBC. Our work hence adds new
439 insight into the evolution of the MTBC from free-living environmental bacteria to obligate
440 pathogens.

441 Our study shows that *M. riyadhense* shares a larger number of orthologues with
442 *M. tuberculosis* than *M. kansasii* and *M. marinum*, notably in the T/A gene family (Figs 3
443 and 4). Forty-nine pairs of T/A gene orthologues were found in *M. riyadhense*, far greater
444 than the number of orthologues observed in any of the other NTMs. The expansion of T/A
445 genes among the MTBC offers additional evidence that suggested original acquisition of
446 T/A modules into mycobacteria through lateral gene transfer played a key role in the
447 development of pathogenicity[21,74].

448 *M. riyadhense* strains appeared as both smooth and rough colony forms when grown
449 on solid LJ media. The observation of smooth and rough colony variants is seen in other
450 mycobacterial pathogens where it is linked to presence or absence of LOS; for example,
451 the presence or absence of LOS from *M. canettii* causes a transition from smooth to rough
452 colony variants, respectively, with rough variants showing increased virulence; MTBC
453 strains lack LOS, and it has been suggested that the removal of LOS was a key event in the
454 evolution of the MTBC species towards their current obligate pathogen status [63]. In
455 *M. riyadhense* smooth colony variants we observed the presence of LOS I and LOS II but
456 the absence of LOS IV. These biochemical observations agree with the genomic prediction

457 that *M. riyadhense* strains lack a functional *wecA* gene, which is required for the extension
458 of LOS II to LOS IV. Overall, our results show that *M. riyadhense* exhibits a LOS
459 production phenotype distinct from that of other LOS-producing mycobacteria.

460 The ESX Type VII secretion systems are key elements of mycobacterial virulence.
461 All the 5 ESX type VII secretion systems present in *M. tuberculosis*, and known to be
462 involved in virulence and pathogenicity, were found in *M. riyadhense* with very similar
463 gene arrangement. An additional ESX-P5 system was also found on the circular plasmid
464 pMR01. pMR01-like plasmids have been shown in many different NTMs, including
465 *M. kansasii*, *M. marinum*, *M. chimaera* and *M. avium*, indicating that these groups of
466 plasmids containing both T4S and T7S are conjugative plasmids. The extensive presence
467 of these plasmids may also explain the origin of the type VII secretion system [38]. ESAT-
468 6 and CFP-10, which are secreted through ESX-1, are critical to phagosome perturbation
469 and manipulation of host macrophages induced by *M. tuberculosis*[75]. Linked to this, our
470 transcriptome analysis revealed that *M. riyadhense* and *M. kansasii* trigger similar overall
471 macrophage responses after infection when compared to *M. bovis* BCG (as a representative
472 of the MTBC). These data reinforce the phylogenetic relationship of *M. riyadhense* with
473 the MTBC species and its transitional status between an opportunistic and an obligate
474 pathogen.

475 The clinical presentation of our cases was by and large indistinguishable from disease
476 caused by *M. tuberculosis*, as reported earlier[7,8], but with a negative *M. tuberculosis*
477 PCR. Due to the relatively recent emergence of *M. riyadhense* as an important clinical
478 pathogen coupled with its misdiagnosis as *M. tuberculosis* by commercially available kits,
479 we developed an accurate set of diagnostic markers based on the genomic datasets

480 generated in this study. The primer sets accurately detect *M. riyadhense* in a mixed cocktail
481 of closely related mycobacteria, and can hence serve as part of an accurate and fast
482 diagnostic protocol in clinical settings thus reducing the need for strict isolation, laborious
483 contact tracing, and inappropriate use of TB antimicrobials. Going forward, these primers
484 could be used in a global epidemiological survey of cases of *M. riyadhense* infections in
485 humans, animals and the environment, providing a more complete picture of the
486 epidemiology of *M. riyadhense* following a broad One Health approach. It would, for
487 example, be of clinical interest to see if *M. riyadhense* infections occur in Africa and South
488 America, for which no reports are available, or if *M. riyadhense* is uniquely endemic in the
489 Arabian Peninsula. Indeed, a limitation of this and previous similar studies is that all
490 isolates available for genomic studies are clinical. *M. riyadhense* is an environmental
491 pathogen, and hence it may be found in animals sharing the same niche with humans that
492 were infected from the environment. As the environmental reservoir of *M. riyadhense*
493 remains unknown, we believe that systematic screening of relevant environmental samples
494 with the MRDP established in this study may help to establish the natural habitat of this
495 bacterium and hence allow improved infection control.

496 In conclusion, our study provides unprecedented insights into the ancestry and
497 adaptive evolution of the MTBC relative to extant NTM species and places *M. riyadhense*
498 as one of the closest relatives to the MTBC. Our work provides compelling data to support
499 the use of *M. riyadhense* as a novel mycobacterium for the study of virulence, evolution
500 and pathogenesis in the MTBC.

501

502 ## Materials and Methods

Ethics Statement

The research protocol was approved by the Institutional Review Board of King Fahad Medical City (Riyadh, Saudi Arabia; #16-345) and the Institutional Biosafety and Bioethics Committee of King Abdullah University of Science and Technology (Jeddah, Saudi Arabia; #18IBEC23). All adult subjects provided informed and written consent. A parent or guardian of any child participant provided informed consent on their behalf.

Clinical reports and bacterial strains

Eight *M. riyadhense* strains were collected in Riyadh, Saudi Arabia, between June 2011 and March 2016 (Fig 1) from patients with a positive culture for *M. riyadhense* isolated from the microbiology laboratory at the King Fahad Medical City (KFMC) in Riyadh, Saudi Arabia. The patient and sample data collected included demographic and clinical characteristics, age, sex, clinical features at presentation, presence of comorbidities, including HIV co-infection, antimicrobial susceptibility, initial and modified therapy, where applicable; and treatment outcome. Once an isolate was suspected to be an NTM, the samples were sent to the reference laboratory for full identification and antimicrobial susceptibility testing. Radiographic and pathologic data were also collected.

Culturing, DNA isolation and sequencing of bacteria

The *M. riyadhense* strains were grown on Lowenstein Jensen (LJ) slants at 37 °C for two weeks, DNA was extracted using a phenol-chloroform protocol [76], and the quality was measured by Qubit. Twenty micrograms of high-molecular-weight (HMW) DNA from the eight *M. riyadhense* strains was sequenced using a PacBio RSII sequencer (Pacific

526 Biosciences, Menlo Park, USA) with a 10 kb library. A NEBNext Ultra II DNA library
527 preparation kit (New England BioLabs, Massachusetts USA) was used to prepare libraries
528 according to the manufacturer's instructions, and sequences from each library were
529 generated for all *M. riyadhense* strains using the Illumina HiSeq 4000 platform (Illumina,
530 San Diego, United States).
531
532 **Genome assembly and annotation**
533 The Illumina short reads were trimmed, and low-quality reads were removed by
534 Trimmomatic [77]. Eight consensus genomes based on each strain were assembled into
535 contigs with the PacBio long reads using the Canu assembler [78]. After assembly, the draft
536 genome was then corrected with short Illumina reads using Pilon [79] software. The
537 circularity of assemblies was checked by Gepard [26], and assemblies were annotated by
538 Prokka [80]. A circular map of the chromosome was compared with that of *M. tuberculosis*
539 H37Rv and visualized with BRIG [81]. The genome of the *M. riyadhense* MR226 strain
540 was used as a high-quality reference in this study.
541
542 **Comparison of chromosomal and plasmid gene contents of *M. riyadhense* to those of**
543 **various *Mycobacterium* species**
544 DNA sequences of 152 mycobacterial species and 77 mycobacterial plasmids were
545 obtained from the NCBI genome database and independently annotated by Prokka [80].
546 The predicted protein sequences from the chromosome and each of the two plasmids
547 (pMRLP01 and pMR01) of the *M. riyadhense* MR226 strain were then compared with the
548 annotated genes from the rest of the mycobacterial species using Proteinortho[82] . The

549 obtained orthologues were visualized with the heatmap package in R [83]. A focused
550 OrthoMCL[84] comparison was performed between (1) *M. riyadhense*, *M. marinum*, *M.*
551 *kansasii*, *M. szulgai* and *M. tuberculosis* and (2) *M. riyadhense* and five species from the
552 MTBC, namely, *M. tuberculosis*, *M. bovis*, *M. canettii*, *M. mungi*, and *M. africanum*.
553
554 **SNP calling and phylogeny based on SNPs**
555 The corrected Illumina reads were mapped using BWA[85] on to the *M. riyadhense*
556 MR226 genome assembly. Picard tool[86] was used to clean SAM files, fix mate-pair
557 information and mark duplicates. SNPs were called for two iterations and filtered (QD<2.0,
558 FS>60.0, SOQ>4.0, ReadPosRankSum<-8.0) with Genome Analysis Toolkit (GATK)[87].
559 An alignment file was generated by SVAMP[88], and phylogeny was generated by
560 RaxML[89] with the TVM model.
561
562 **Phylogeny of *M. riyadhense***
563 The PhyoPhlAn2[90] pipeline was used to identify protein sequences from 400
564 conserved genes in the pangenome datasets and the *Mycobacterium* genomes available at
565 NCBI or JGI. A total of 149 species were used and *Nocardia abscessus* was used as the
566 out-group.
567 A whole-genome phylogenetic tree was constructed with the MTBC species (*M.*
568 *tuberculosis*, *M. bovis*, *M. canettii*, *M. mungi*, *Mycobacterium orygis*, *M. africanum*) and
569 with *M. kansasii*, *M. marinum*, *M. shinjukuense*, *Mycobacterium leprae*,
570 *Mycobacterium smegmatis*, *Mycobacterium parmense*, *Mycobacterium avium* and
571 *Mycobacterium abscessus*. The one-to-one orthologues of each species were obtained

572 using OrthoMCL and concatenated, then aligned with Muscle[91] and trimmed with
573 TrimAL[92]. The concatenated sequences were composed of 906 genes encoding 296,124
574 amino acids and were used to build a phylogenetic tree with the LG+G+F model, which
575 was selected by the ModelGenerator. The phylogenetic tree was generated by RaxML[89].
576
577 **Toxin/antitoxin systems, *mce*/*mce*-associated genes and ESX systems in *M. riyadhense***
578 **and other mycobacteria**
579 One hundred fifty-eight T/A proteins belonging to the VapBC, RelEF, HigBA,
580 MazEF, ParDE and UCAT families were downloaded from the NCBI protein database. We
581 identified the *M. tuberculosis* T/A orthologues from all of the 149 species by
582 Proteinortho[82], and the orthologue groups were also examined by Blast+ 2.4.0. The same
583 pipeline was also applied for the MCE/MCE-associate genes, PhoPR, DosRS, PE/PPE, PE-
584 PGRS, and ESX1-5.
585
586 **Infection of the RAW 246.7 cell line with *M. riyadhense*, *M. kansasii* and *M. bovis* BCG**
587 **Denmark**
588 The murine macrophage RAW264.7 cell line obtained from the American Type
589 Culture Collection (ATCC, Manassas, USA) was cultured in Dulbecco's modified Eagle's
590 medium (DMEM) (ThermoFisher Scientific, Waltham, USA) supplemented with 10% FCS,
591 streptomycin, and penicillin. *M. riyadhense*, *M. kansasii* (subtype I), and *M. bovis* BCG
592 Denmark strains were grown in Middlebrook 7H9 liquid medium after single-colony
593 isolation from LJ slants or 7H10 agar. 7H9 was supplemented with 10% albumin, dextrose
594 and catalase (ADC), while 7H10 was supplemented with oleic acid, albumin, dextrose and

595 catalase (OADC) in addition to 0.2% glycerol. Ready to use LJ slants were provided by
596 Saudi Prepared Media Laboratory (SPML, Riyadh, Saudi Arabia).

597 Before the infection, all bacterial cultures were centrifuged at 1,000xg for 10 minutes.
598 The supernatant was discarded, and 10-15 3 mm glass beads were added to the pellet and
599 then vortexed for 1 minute to break up clumps. Six ml of DMEM was added to the pellet
600 and left to rest for 5 minutes. The upper 5 ml of the suspension was removed to a fresh 15
601 ml falcon tube, which was then centrifuged for an additional 3 minutes at 200xg to remove
602 remaining bacterial clumps. The supernatant was then taken and passaged using a 26G
603 hypodermal syringe approximately 15 times to further break up any clumps of bacteria.
604 The optical density of the culture was measured again before the infection experiment.

605 RAW264.7 cells were seeded at 2 x 10^5 cells per well in 24-well flat-bottom tissue
606 culture plates 24 hours prior, reaching 5 x 10^5 cells per well. The DMEM over the cells was
607 removed, and the cells were washed once with phosphate-buffered saline (PBS). The cells
608 were infected by applying 1 ml of DMEM containing the prepared mycobacteria from the
609 former steps at an appropriate concentration to reach a MOI of 5:1, with DMEM alone used
610 for the control wells. The culture plates were then returned to the incubator at 37°C with
611 5% CO_2 for 3 hours to allow for bacterial uptake by the RAW264.7 cells. The supernatant
612 was removed after 3 hours, and the infected cells were washed with PBS to remove
613 extracellular bacteria. Subsequently, the cells were incubated in fresh DMEM with 10%
614 FCS for 24 hours and 48 hours. For harvesting, 400 µl of TRIzol (ThermoFisher Scientific,
615 Waltham, USA) was added to the wells at each time point, and the adherent cells were
616 scraped out and stored at -80°C for RNA extraction. Each bacterial infection was performed
617 in triplicate, in addition to the non-infected controls. RNA was isolated from the samples

618 using the Direct-zol™ RNA Miniprep kit (Zymo Research, Irvine, USA) according to the
619 manufacturer's instructions.

620 An Agilent RNA 6000 Nano kit was used to check the quality and quantity of the total
621 RNA. The NanoString murine nCounter Myeloid Innate Immunity Gene Expression Panel
622 (NanoString Technologies, USA) was used to assess transcript abundance across infections
623 and time points using the nCounter MAX Analysis System (NanoString Technologies,
624 Seattle, USA). The counts obtained were normalized using the nSolver™ Advanced
625 Analysis plugin (NanoString Technologies, Seattle, USA) using the geNorm algorithm,
626 and differential gene (DE) expression was analyzed using multivariate linear regression in
627 nSolver™ software, with 0.05 as the p-value cutoff.

628

629 **Thin-layer chromatography analysis of lipooligosaccharides in *M. riyadhense*, *M.***
630 ***kansasii* and *M. marinum***

631 For TLC analysis, mycobacterial strains were grown at 30°C (*M. marinum*) or 37°C
632 (*M. smegmatis*, *M. kansasii*, *M. riyadhense*) on LJ slants, and after sufficient incubation,
633 grown cells were collected and washed once with PBS. Apolar and polar lipids were
634 extracted from the cell pellets using methods described by Dobson *et al*[64]. Polar lipids
635 were analyzed by 2D-TLC using solvent system E, which is designed to separate
636 phospholipids and LOSs[64]. Glycolipids were visualized by charring following staining
637 with either molybdophophoric acid (MPA) or alpha-napthol (for glycolipids).

638

639 **Diagnostic PCR markers for *M. riyadhense***

To develop diagnostic markers for *M. riyadhense* for potential use in clinical diagnosis as well as environmental and animal studies, unique regions within the *M. riyadhense* reference genome compared to that of 152 other mycobacterial species were detected using Shustring[93]. These regions were also examined by Proteinortho[82] and Blastn[94]. *mr_00036*, *mr_00263*, *mr_00606*, and *mr_01005* were selected as the amplification targets. Two primers for each gene were designed in this study:

MRDP-MR_00036-F (5'-TTCGTTGTCGGTTTCGTCGC-3') and MRDP-MR_00036-R (5'-GCGTCAGCTCCACCGAAAAC-3');

MRDP-MR_00263-F (5'-CCACCGCTGTTGGCGA-3') and MRDP-MR_00263-R (5'-TTCGTCCCGTTGATCCCGTT-3');

MRDP-MR_00606-F (5'-AACCTGCCCGATACGCACTT-3') and MRDP-MR_00606-R (5'-ACTGTTCCTCCGTGGGGTTG-3');

MRDP-MR_01005-F (5'-GACTGTGGGGTAACGGTGGA-3') and MRDP-MR_01005-R (5'-CCGGTGATGTCGCCTACTCC-3').

PCR was performed in a 25 µl reaction volume with 12.5 µl of GoTaq ® Green Master Mix (Promega, USA), 1 µl of 100 ng/µl gDNA, 1 µl with 10 nmol of forward and reverse primers, 3 µl of dimethyl sulfoxide (DMSO) and 19 µl of nuclease-free water. The PCR mixture was denatured for 5 minutes at 94°C; followed by 35 cycles of amplification involving a denaturation step at 94°C for 30 seconds, a primer annealing step at 59°C for 45 seconds, and a primer extension step at 72°C for 45 seconds; and a final extension step at 72°C for 7 minutes. The ITS-F/mycom-2 primer set[95], which is a *Mycobacterium* genus-specific primer set, was used as a control, with amplification conditions as described

662 previously[95]. The products were electrophoresed in a 2% agarose gel for 60 minutes and
663 visualized.
664
665 ## Acknowledgements
666 We wish to thank members of the Biological Core Lab (BCL) of King Abdullah
667 University of Science and Technology for sequencing the bacterial genomic DNAs on the
668 Illumina HiSeq 4000 and PacBio RSII platforms.
669
670 ## Accession codes
671 The *M. riyadhense* dataset is available at European Nucleotide Archive (ENA) under
672 the study accession no. PRJEB32162. The assemblies are available with DOI
673 10.5281/zenodo.2873972.
674
675 **Financial Disclosure Statement**
676 Work in AP's laboratory is supported by the KAUST faculty baseline fund (BAS/1/1020-
677 01- 01). Work in SG and AB's laboratory was supported in part by BBSRC award
678 BB/N004574/1.
679
680 **Conflicts of interest**
681 The authors have no conflicts of interest to declare.
682
683 **Author contribution statement**

AP conceived the comparative genomics part of the study, obtained the funding and supervised the work; MG and AH conceived the clinical part of the study; MG and FA wrote the clinical part of the text; FA, MG, TA, SF, AH, MAR, AR, and TS collected the microbiological and clinical information; and AP, QG, SG, AB, and FA designed the experiments. QG performed the data analysis and prepared the initial draft of the manuscript, followed by edits from AP, SG, MG, SF, AB and CM. QG, SM, ASm and JB performed the transcriptome experiment. ASi and AB performed the TLC analysis. CN and YS provided materials for the diagnostic markers and intellectual advice. All authors have commented on various sections of the manuscript, which were finally curated and incorporated in the final version by QG and AP.

References

1. Falkinham JO. Ecology of nontuberculous mycobacteria-where do human infections come from? Semin Respir Crit Care Med. 2013;34: 95–102. doi:10.1055/s-0033-1333568

2. van Ingen J, Al-Haijoj SAM, Boeree M, Al-Rabiah F, Enaimi M, de Zwaan R, et al. Mycobacterium riyadhense sp. nov., a non-tuberculous species identified as Mycobacterium tuberculosis complex by a commercial line-probe assay. Int J Syst Evol Microbiol. 2009;59: 1049–1053. doi:10.1099/ijs.0.005629-0

703 3. Ding L-W, Lai C-C, Lee L-N, Huang L-M, Hsueh P-R. Lymphadenitis caused by
704 non-tuberculous mycobacteria in a university hospital in Taiwan: predominance
705 of rapidly growing mycobacteria and high recurrence rate. J Formos Med Assoc.
706 2005;104: 897–904.
707 4. Panwalker AP, Fuhse E. Nosocomial Mycobacterium gordonae pseudoinfection
708 from contaminated ice machines. Infect Control. 1986;7: 67–70.
709 doi:10.1017/S0195941700063918
710 5. Carbonne A, Brossier F, Arnaud I, Bougmiza I, Caumes E, Meningaud JP, et al.
711 Outbreak of nontuberculous mycobacterial subcutaneous infections related to
712 multiple mesotherapy injections. J Clin Microbiol. 2009;47: 1961–1964.
713 doi:10.1128/JCM.00196-09
714 6. Singh J, O'Donnell K, Ashouri N, Adler-Shohet FC, Nieves D, Tran MT, et al.
715 926. Outbreak of Invasive Nontuberculous Mycobacterium (NTM) Infections
716 Associated With a Pediatric Dental Practice . Open Forum Infect Dis. 2018;5:
717 S29–S29. doi:10.1093/ofid/ofy209.067

718 7. Garbati, Hakawi AM. Mycobacterium Riyadhense Lung Infection in a Patient with
719 HIV/AIDS [Internet]. Sub-Saharan African Journal of Medicine. 2014. Available:
720 http://www.ssajm.org/article.asp?issn=2384-
721 5147;year=2014;volume=1;issue=1;spage=56;epage=58;aulast=Garbati
722 8. Choi JI, Lim JH, Kim SR, Lee SH, Park JS, Seo KW, et al. Lung infection caused
723 by Mycobacterium riyadhense confused with Mycobacterium tuberculosis: The
724 first case in Korea. Ann Lab Med. 2012;32: 298–303.
725 doi:10.3343/alm.2012.32.4.298
726 9. King HC, Khera-Butler T, James P, Oakley BB, Erenso G, Aseffa A, et al.
727 Environmental reservoirs of pathogenic mycobacteria across the Ethiopian
728 biogeographical landscape. PLoS One. 2017;12.
729 doi:10.1371/journal.pone.0173811
730 10. Narendrula-Kotha R, Nkongolo KK. Microbial response to soil liming of damaged
731 ecosystems revealed by pyrosequencingand phospholipid fatty acid analyses. PLoS
732 One. 2017;12. doi:10.1371/journal.pone.0168497

733 11. Godreuil S, Marchandin H, Michon AL, Ponsada M, Chyderiotis G, Brisou P, et
734 al. Mycobacterium riyadhense pulmonary infection, France and Bahrain. Emerging
735 Infectious Diseases. 2012. pp. 176–178. doi:10.3201/eid1801.110751
736 12. Van der Werf, Marieke J. Ködmön, Csaba Katalinić-Janković, Vera Kummik,
737 Tiina Soini, Hanna Richter,E, et al. Inventory study of non-tuberculous
738 mycobacteria in the European Union. BMC Infect Dis. 2014;14. doi:10.1186/1471-
739 2334-14-62
740 13. Althawadi S, Johani S, Fernandez GM, Al-ghafli H, Enani RMA, Abdul K, et al.
741 Mycobacterium riyadhense. 2017;23: 2015–2017.
742 14. Tortoli E, Pecorari M, Fabio G, Messinò M, Fabio A. Commercial DNA probes for
743 mycobacteria incorrectly identify a number of less frequently encountered species.
744 Journal of Clinical Microbiology. 2010. pp. 307–310. doi:10.1128/JCM.01536-09
745 15. Murty DS PK. A Brief review on Ecology and Evolution of Mycobacteria.
746 Mycobact Dis. 2014;04. doi:10.4172/2161-1068.1000172

747 16. Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O. Horizontally acquired
748 genomic islands in the tubercle bacilli. Trends Microbiol. 2008;16: 303–308.
749 doi:10.1016/j.tim.2008.04.005

750 17. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al.
751 Genomic analysis of smooth tubercle bacilli provides insights into ancestry and
752 pathoadaptation of Mycobacterium tuberculosis. Nat Genet. 2013;45: 172–179.
753 doi:10.1038/ng.2517

754 18. Van Soolingen D, Hoogenboezem T, De Haas PEW, Hermans PWM, Koedam
755 MA, Teppema KS, et al. A Novel Pathogenic Taxon of the Mycobacterium
756 tuberculosis Complex, Canetti: Characterization of an Exceptional Isolate from
757 Africa. Int J Syst Bacteriol. 1997;47: 1236–1245. doi:10.1099/00207713-47-4-
758 1236

759 19. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, et al. Significance of the
760 Identification in the Horn of Africa of an Exceptionally Deep Branching

761 Mycobacterium tuberculosis Clade. PLoS One. 2012;7: e52841.

762 doi:10.1371/journal.pone.0052841

763 20. Wang J, McIntosh F, Radomski N, Dewar K, Simeone R, Enninga J, et al. Insights

764 on the emergence of Mycobacterium tuberculosis from the analysis of

765 Mycobacterium kansasii. Genome Biol Evol. 2015;7: 856–870.

766 doi:10.1093/gbe/evv035

767 21. Stinear TP, Seemann T, Harrison PF, Jenkin G a, Davies JK, Johnson PDR, et al.

768 Insights from the complete genome sequence of Mycobacterium marinum on the

769 evolution of Mycobacterium tuberculosis. Genome Res. 2008;18: 729–41.

770 doi:10.1101/gr.075069.107

771 22. Tortoli E, Fedrizzi T, Meehan CJ, Trovato A, Grottola A, Giacobazzi E, et al. The

772 new phylogeny of the genus Mycobacterium: The old and the news. Infect Genet

773 Evol. 2017;56: 19–25. doi:10.1016/j.meegid.2017.10.013

774 23. Brown-Elliott BA, Simmer PJ, Trovato A, Hyle EP, Droz S, Buckwalter SP, et al.

775 Mycobacterium decipiens sp. nov., a new species closely related to the

776 Mycobacterium tuberculosis complex. Int J Syst Evol Microbiol. 2018;68: 3557–
777 3562. doi:10.1099/ijsem.0.003031

778 24. Saito H, Iwamoto T, Ohkusu K, Otsuka Y, Akiyama Y, Sato S, et al.
779 Mycobacterium shinjukuense sp. nov., a slowly growing, non-chromogenic species
780 isolated from human clinical specimens. Int J Syst Evol Microbiol. 2011;61: 1927–
781 1932. doi:10.1099/ijs.0.025478-0

782 25. Fedrizzi T, Meehan CJ, Grottola A, Giacobazzi E, Fregni Serpini G, Tagliazucchi
783 S, et al. Genomic characterization of Nontuberculous Mycobacteria. Sci Rep.
784 2017;7. doi:10.1038/srep45258

785 26. Krumsiek J, Arnold R, Rattei T. Gepard: A rapid and sensitive tool for creating
786 dotplots on genome scale. Bioinformatics. 2007;23: 1026–1028.
787 doi:10.1093/bioinformatics/btm039

788 27. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering
789 the biology of mycobacterium tuberculosis from the complete genome sequence.
790 Nature. 1998. pp. 537–544. doi:10.1038/31159

791 28. Hurtado UA, Solano JS, Rodriguez A, Robledo J, Rouzaud F. Draft Genome
792 Sequence of a Mycobacterium africanum Clinical Isolate from Antioquia,
793 Colombia. Genome Announc. 2016;4: e00486--16. doi:10.1128/genomeA.00486-
794 16
795 29. de la Fuente J, Díez-Delgado I, Contreras M, Vicente J, Cabezas-Cruz A,
796 Manrique M, et al. Complete Genome Sequences of Field Isolates of
797 Mycobacterium bovis and Mycobacterium caprae. Genome Announc. 2015;3:
798 e00247-15. doi:10.1128/genomeA.00247-15
799 30. Wang J, McIntosh F, Radomski N, Dewar K, Simeone R, Enninga J, et al. Insights
800 on the emergence of Mycobacterium tuberculosis from the analysis of
801 Mycobacterium kansasii. Genome Biol Evol. 2015;7: 856–870.
802 doi:10.1093/gbe/evv035
803 31. Weerdenburg EM, Abdallah AM, Rangkuti F, El Ghany MA, Otto TD, Adroub
804 SA, et al. Genome-wide transposon mutagenesis indicates that Mycobacterium

805 marinum customizes its virulence mechanisms for survival and replication in
806 different hosts. Infect Immun. 2015;83: 1778–1788. doi:10.1128/IAI.03050-14
807 32. Leon P, Virginia W, Patricia B. Molecular analysis of the linear 2.3 kb plasmid of
808 maize mitochondria: Apparent capture of tRNA genes. Nucleic Acids Res.
809 1989;17: 4089–4099. doi:10.1093/nar/17.11.4089
810 33. Kinashi H. Giant linear plasmids in Streptomyces: A treasure trove of antibiotic
811 biosynthetic clusters. Journal of Antibiotics. 2011. pp. 19–25.
812 doi:10.1038/ja.2010.146
813 34. Crespi M, Messens E, Caplan AB, van Montagu M, Desomer J. Fasciation
814 induction by the phytopathogen Rhodococcus fascians depends upon a linear
815 plasmid encoding a cytokinin synthase gene. EMBO J. 1992;11: 795–804.
816 Available:
817 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC556518/%5Cnhttp://files/362/Cres
818 pi et al. - 1992 - Fasciation induction by the phytopathogen Rhodococ.pdf

819 35. Picardeau M, Vincent V. Characterization of large linear plasmids in
820 mycobacteria. J Bacteriol. 1997;179: 2753–2756. doi:10.1128/jb.179.8.2753-
821 2756.1997
822 36. Surtees JA, Funnell BE. Plasmid and Chromosome Traffic Control: How ParA and
823 ParB Drive Partition. Curr Top Dev Biol. 2003;56: 145–180. doi:10.1016/S0070-
824 2153(03)01010-X
825 37. Zrimec J, Lapanje A. DNA structure at the plasmid origin-of-Transfer indicates its
826 potential transfer range. Sci Rep. 2018;8. doi:10.1038/s41598-018-20157-y
827 38. Ummels R, Abdallah AM, Kuiper V, Aâjoud A, Sparrius M, Naeem R, et al.
828 Identification of a novel conjugative plasmid in mycobacteria that requires both
829 type IV and type VII secretion. MBio. 2014;5. doi:10.1128/mBio.01744-14
830 39. Uchiya KI, Takahashi H, Nakagawa T, Yagi T, Moriyama M, Inagaki T, et al.
831 Characterization of a novel plasmid, pMAH135, from Mycobacterium avium
832 subsp. hominissuis. PLoS One. 2015;10. doi:10.1371/journal.pone.0117797

833 40. da Silva Rabello MC, Matsumoto CK, de Almeida LGP, Menendez MC, de
834 Oliveira RS, Silva RM, et al. First description of natural and experimental
835 conjugation between mycobacteria mediated by a linear plasmid. PLoS One.
836 2012;7. doi:10.1371/journal.pone.0029884
837 41. Kim B-J, Kim B-R, Lee S-Y, Seok S-H, Kook Y-H, Kim B-J. Whole-Genome
838 Sequence of a Novel Species, Mycobacterium yongonense DSM 45126 T.
839 Genome Announc. 2013;1: 604–13. doi:10.1128/genomeA.00604-13
840 42. Veyrier F, Pletzer D, Turenne C, Behr MA. Phylogenetic detection of horizontal
841 gene transfer during the step-wise genesis of Mycobacterium tuberculosis. BMC
842 Evol Biol. 2009;9. doi:10.1186/1471-2148-9-196
843 43. van Ingen J, Kohl TA, Kranzer K, Hasse B, Keller PM, Katarzyna Szafrańska A, et
844 al. Global outbreak of severe Mycobacterium chimaera disease after cardiac
845 surgery: a molecular epidemiological study. Lancet Infect Dis. 2017;17: 1033–
846 1041. doi:10.1016/S1473-3099(17)30324-9

847 44. Morgado SM, Marín MA, Freitas FS, Fonseca EL, Vicente ACP. Complete
848 plasmid sequence carrying type IV-like and type VII secretion systems from an
849 atypical mycobacteria strain. Mem Inst Oswaldo Cruz. 2017;112: 514–516.
850 doi:10.1590/0074-02760160546
851 45. Houben ENG, Korotkov K V., Bitter W. Take five - Type VII secretion systems of
852 Mycobacteria. Biochim Biophys Acta - Mol Cell Res. 2014;1843: 1707–1716.
853 doi:10.1016/j.bbamcr.2013.11.003
854 46. Abdallah A, Pittius G, Nicolaas C, DiGiuseppe Champion P, Cox J, Luirink J, et
855 al. Type VII secretion - Mycobacteria show the way. Nat Rev Microbiol. 2007;5:
856 883–891. doi:10.1038/nrmicro1773
857 47. Weerdenburg EM, Abdallah AM, Mitra S, De Punder K, Van der Wel NN, Bird S,
858 et al. ESX-5-deficient Mycobacterium marinum is hypervirulent in adult zebrafish.
859 Cell Microbiol. 2012;14: 728–739. doi:10.1111/j.1462-5822.2012.01755.x
860 48. Jia X, Yang L, Dong M, Chen S, Lv L, Cao D, et al. The Bioinformatics Analysis
861 of Comparative Genomics of Mycobacterium tuberculosis Complex (MTBC)

862 Provides Insight into Dissimilarities between Intraspecific Groups Differing in
863 Host Association, Virulence, and Epitope Diversity. Front Cell Infect Microbiol.
864 2017;7. doi:10.3389/fcimb.2017.00088
865 49. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of
866 genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J
867 Bacteriol. 1996;178: 1274–1282. doi:10.1128/jb.178.5.1274-1282.1996
868 50. Gordon S V., Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST. Identification
869 of variable regions in the genomes of tubercle bacilli using bacterial artificial
870 chromosome arrays. Mol Microbiol. 1999;32: 643–655. doi:10.1046/j.1365-
871 2958.1999.01383.x
872 51. Van Ingen J, De Zwaan R, Dekhuijzen R, Boeree M, Van Soolingen D. Region of
873 difference 1 in nontuberculous Mycobacterium species adds a phylogenetic and
874 taxonomical character. J Bacteriol. 2009;191: 5865–5867. doi:10.1128/JB.00683-
875 09

876 52. Kozak RA, Alexander DC, Liao R, Sherman DR, Behr MA. Region of difference 2
877 contributes to virulence of mycobacterium tuberculosis. Infect Immun. 2011;79:
878 59–66. doi:10.1128/IAI.00824-10

879 53. Ru H, Liu X, Lin C, Yang J, Chen F, Sun R, et al. The Impact of Genome Region
880 of Difference 4 (RD4) on Mycobacterial Virulence and BCG Efficacy. Front Cell
881 Infect Microbiol. 2017;7. doi:10.3389/fcimb.2017.00239

882 54. Brosch R, Gordon S V., Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al.
883 A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc
884 Natl Acad Sci. 2002;99: 3684–3689. doi:10.1073/pnas.052548299

885 55. Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM,
886 et al. Novel mycobacterium tuberculosis complex pathogen, M. Mungi. Emerg
887 Infect Dis. 2010;16: 1296–1299. doi:10.3201/eid1608.100314

888 56. M Cristina G, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, et al. Ancient
889 origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS
890 Pathog. 2005;1: 0055–0061. doi:10.1371/journal.ppat.0010005

891 57. Delogu G, Brennan MJ. Comparative Immune Response to PE and PE _ PGRS
892 Antigens of Mycobacterium tuberculosis Comparative Immune Response to PE
893 and PE _ PGRS Antigens of Mycobacterium tuberculosis. Infect Immun. 2001;69:
894 5606–5611. doi:10.1128/IAI.69.9.5606
895 58. Isom GL, Davies NJ, Chong ZS, Bryant JA, Jamshad M, Sharif M, et al. MCE
896 domain proteins: Conserved inner membrane lipid-binding proteins required for
897 outer membrane homeostasis. Sci Rep. 2017;7. doi:10.1038/s41598-017-09111-6
898 59. Newton-Foot M, Warren RM, Sampson SL, Van Helden PD, Gey Van Pittius NC.
899 The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion
900 systems. BMC Evol Biol. 2016;16. doi:10.1186/s12862-016-0631-2
901 60. Slayden RA, Dawson CC, Cummings JE. Toxin-antitoxin systems and regulatory
902 mechanisms in Mycobacterium tuberculosis. Pathogens and Disease. 2018.
903 doi:10.1093/femspd/fty039

904 61. Cooper TF, Heinemann JA. Postsegregational killing does not increase plasmid
905 stability but acts to mediate the exclusion of competing plasmids. Proc Natl Acad
906 Sci. 2000;97: 12643–12648. doi:10.1073/pnas.220077897
907 62. Van Der Woude AD, Sarkar D, Bhatt A, Sparrius M, Raadsen SA, Boon L, et al.
908 Unexpected link between lipooligosaccharide biosynthesis and surface protein
909 release in Mycobacterium marinum. J Biol Chem. 2012;287: 20417–20429.
910 doi:10.1074/jbc.M111.336461
911 63. Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, Laval F, et al. Pks5-
912 recombination-mediated surface remodelling in Mycobacterium tuberculosis
913 emergence. Nat Microbiol. 2016;1. doi:10.1038/nmicrobiol.2015.19
914 64. G.Dobson., D.E.Minnikin., S.M.Minnikin., M.Parlett., M.Goodfellow., M.Ridell.,
915 et al. Systematic analysis of complex mycobacterial lipids. Chemical Methods in
916 Bacterial Systematics. Academic Press, London, United Kingdom.; 1985. pp. 237–
917 265.

65. Cohen I, Parada C, Acosta-Gío E, Espitia C. The PGRS domain from PE_PGRS33 of Mycobacterium tuberculosis is target of humoral immune response in mice and humans. Front Immunol. 2014;5: 1–9. doi:10.3389/fimmu.2014.00236

66. Boritsch EC, Supply P, Honoré N, Seeman T, Stinear TP, Brosch R. A glimpse into the past and predictions for the future: The molecular evolution of the tuberculosis agent. Molecular Microbiology. 2014. pp. 835–852. doi:10.1111/mmi.12720

67. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26: 317–325. doi:10.1038/nbt1385

68. Bowdish DME, Sakamoto K, Kim MJ, Kroos M, Mukhopadhyay S, Leifer CA, et al. MARCO, TLR2, and CD14 are required for macrophage cytokine responses to mycobacterial trehalose dimycolate and Mycobacterium tuberculosis. PLoS Pathog. 2009;5. doi:10.1371/journal.ppat.1000474

932 69. Fujita Y, Naka T, McNeil MR, Yano I. Intact molecular characterization of cord
933 factor (trehalose 6,6′-dimycolate) from nine species of mycobacteria by MALDI-
934 TOF mass spectrometry. Microbiology. 2005;151: 3403–3416.
935 doi:10.1099/mic.0.28158-0
936 70. Stavrum R, Stavrum AK, Valvatne H, Riley LW, Ulvestad E, Jonassen I, et al.
937 Modulation of transcriptional and inflammatory responses in murine macrophages
938 by the Mycobacterium tuberculosis Mammalian cell entry (Mce) 1 complex. PLoS
939 One. 2011;6. doi:10.1371/journal.pone.0026295
940 71. Katoch VM. Infections due to non-tuberculous mycobacteria (NTM). Indian
941 Journal of Medical Research. 2004. pp. 290–304. doi:10.1016/j.phrs.2009.12.004
942 72. Griffith DE, Aksamit T, Brown-Elliott B a., Catanzaro A, Daley C, Gordin F, et al.
943 An Official ATS/IDSA Statement: Diagnosis, Treatment, and Prevention of
944 Nontuberculous Mycobacterial Diseases. Am J Respir Crit Care Med. 2007;175:
945 367–416. doi:10.1164/rccm.200604-571ST

946 73. Sapriel G, Brosch R. Shared pathogenomic patterns characterize a new phylotype,
947 revealing transition towards host-adaptation long before speciation of
948 Mycobacterium tuberculosis. Genome Biol Evol. 2019;11: 2420–2438.
949 doi:10.1093/gbe/evz162
950 74. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, et
951 al. Contribution of horizontally acquired genomic islands to the evolution of the
952 tubercle bacilli. Mol Biol Evol. 2007;24: 1861–1871.
953 doi:10.1093/molbev/msm111
954 75. Jamwal S V., Mehrotra P, Singh A, Siddiqui Z, Basu A, Rao KVS. Mycobacterial
955 escape from macrophage phagosomes to the cytoplasm represents an alternate
956 adaptation mechanism. Sci Rep. 2016;6. doi:10.1038/srep23089
957 76. Belisle JT, Sonnenberg MG. Isolation of genomic DNA from mycobacteria.
958 Methods Mol Biol. 1998;101: 31–44. doi:10.1385/0-89603-471-2:31

959 77. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
960 sequence data. Bioinformatics. 2014;30: 2114–2120.
961 doi:10.1093/bioinformatics/btu170

962 78. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu:
963 Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat
964 separation. Genome Res. 2017;27: 722–736. doi:10.1101/gr.215087.116

965 79. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon:
966 An integrated tool for comprehensive microbial variant detection and genome
967 assembly improvement. PLoS One. 2014;9. doi:10.1371/journal.pone.0112963

968 80. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics.
969 2014;30: 2068–2069. doi:10.1093/bioinformatics/btu153

970 81. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image
971 Generator (BRIG): Simple prokaryote genome comparisons. BMC Genomics.
972 2011;12. doi:10.1186/1471-2164-12-402

973 82. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho:
974 Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;
975 doi:10.1186/1471-2105-12-124
976 83. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. J Comput
977 Graph Stat. 1996;5: 299–314. doi:10.1080/10618600.1996.10474713
978 84. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for
979 eukaryotic genomes. Genome Res. 2003;13: 2178–2189. doi:10.1101/gr.1224503
980 85. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
981 transform. Bioinformatics. 2009;25: 1754–1760.
982 doi:10.1093/bioinformatics/btp324
983 86. Broad Institute. Picard tools. In: https://broadinstitute.github.io/picard/ [Internet].
984 2016. Available:
985 https://broadinstitute.github.io/picard/%5Cnhttp://broadinstitute.github.io/picard/
986 87. Alkan C, Coe BP, Eichler EE. GATK toolkit. Nat Rev Genet. 2011;12: 363–76.
987 doi:10.1038/nrg2958

988 88. Naeem R, Hidayah L, Preston MD, Clark TG, Pain A. SVAMP: Sequence
989 variation analysis, maps and phylogeny. Bioinformatics. 2014;30: 2227–2229.
990 doi:10.1093/bioinformatics/btu176
991 89. Stamatakis A, Ludwig T, Meier H. RAxML-III: A fast program for maximum
992 likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005;21:
993 456–463. doi:10.1093/bioinformatics/bti191
994 90. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method
995 for improved phylogenetic and taxonomic placement of microbes. Nat Commun.
996 2013; doi:10.1038/ncomms3304
997 91. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time
998 and space complexity. BMC Bioinformatics. 2004;5: 113. doi:10.1186/1471-2105-
999 5-113
1000 92. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated
1001 alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:
1002 1972–1973. doi:10.1093/bioinformatics/btp348

1003 93. Haubold B, Pierstorff N, Möller F, Wiehe T. Genome comparison without
1004 alignment using shortest unique substrings. BMC Bioinformatics. 2005;6.
1005 doi:10.1186/1471-2105-6-123
1006 94. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
1007 BLAST+: Architecture and applications. BMC Bioinformatics. 2009;10.
1008 doi:10.1186/1471-2105-10-421
1009 95. Park H, Jang H, Kim C, Chung B, Chang CL, Soon Kew Park, et al. Detection and
1010 identification of mycobacteria by amplification of the internal transcribed spacer
1011 regions with genus- and species-specific PCR primers. J Clin Microbiol. 2000;38:
1012 4080–4085.

1013

1014

1015

1016

1017

Figures

									Therapy		Radiographic	Baseline		
	Age (Years)	Gender	HIV status	Other co-morbid conditions	Site(s) of the infection	Smooth/Rough	Auramine-rhodamine stain	Ziehl–Neelsen stain	Initial medication/duration	Continuous medication/duration	Location of lesion (Chest CT)	CD4 count ($X10^6$/L)	Viral load (Copies/mL)	Treatment Outcome
MR193	82	M	Unknown	Smoking; DM; HTN; heart disease	Pulmonary	R	4+	2+	NA	NA	Left upper lobe consolidation	NA	N/A	Died
MR206	32	M	Negative	Eisenmenger's syndrome; pulmonary HTN	Pulmonary	R	NA	1+	IMP/2M CLR/2M EMB/2M INH/1M PYZ/1M RIF/1M	CLR/4M EMB/4M	Bilateral air spice consolidation and nodular opacity	NA	N/A	Survived
MR210	66	M	Positive	PCP co-infection	Pulmonary	S	NA	Negtive	EMB/18M CLR/18M	NA	Bilateral ground glass opacities	0.7	415,927	Survived
MR222	37	M	Positive	Lymphoblastic lymphoma	Pulmonary; LN, Abdominal	S	NA	Negtive	CLR/1M EMB/1M IMP/1M	EMB/2M IMP/2M	Multiple lung lesions with central cavitation	0	838	Died
MR226	8	M	Negative	Free	Pulmonary; LA	S	NA	Negtive	NA	NA	Multiple bilateral nodular infiltrates, more in the left upper lobe	NA	N/A	Survived
MR244	28	M	Positive	Pulmonary disease; hematologic disease	Pulmonary, LN, Abdominal; Peritoneal;	S	NA	2+	CLR/1M RIF+INH/1M EMB/1M	MXF/4M	Multiple bilateral lung nodules; cavitation left lower lobe	0.02	399,652	Survived
MR246	17	M	Negative	Pulmonary disease; chronic diarrhea	Pulmonary; Mediastinal LN	S	NA	Negative	RIF/17M CLR/21M EMB/17M	NA	Left upper lobe consolidation	NA	N/A	Survived
MR1023	47	M	Negative	DM; subclinical hypothyroidism	Pulmonary	S	NA	Negative	NA	NA	Bilateral nodular infiltrate; cavity left upper lobe	NA	N/A	Survived

Fig 1. Clinical characteristics of the studies populations.

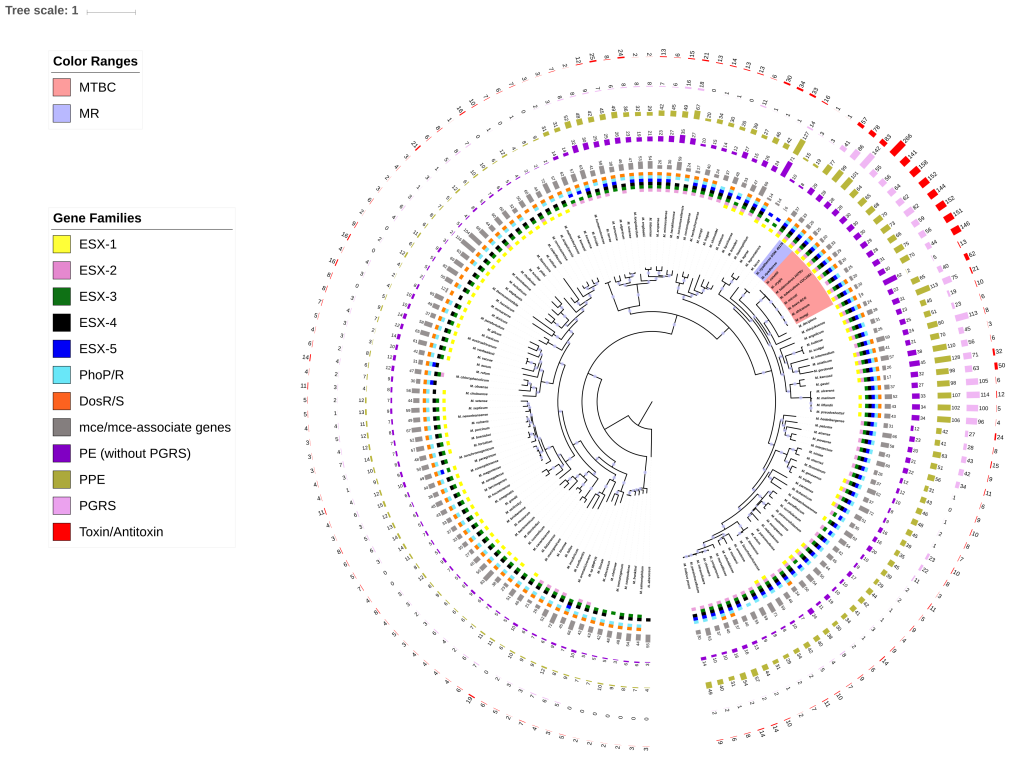
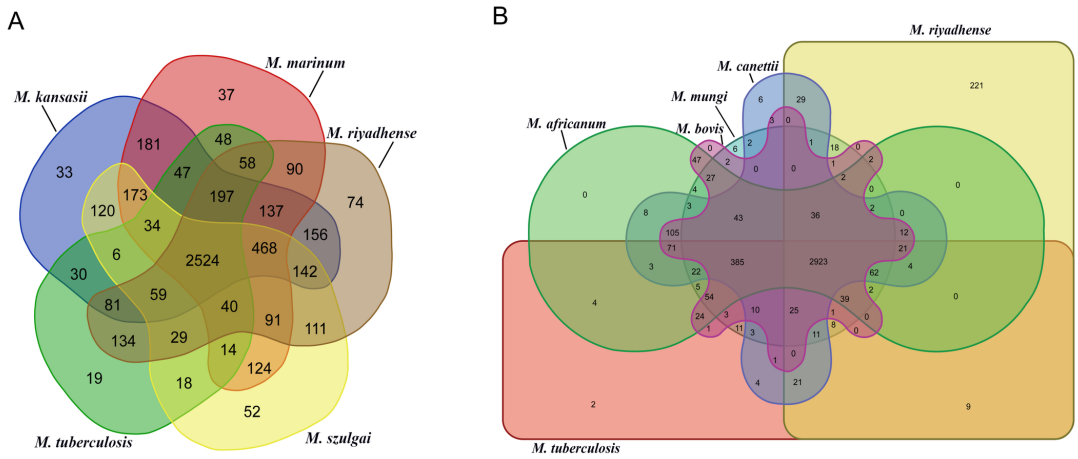


Fig 2. A phylogenetic tree of 149 species of mycobacteria showing close relationship of *M. riyadhense* and MTBC. The phylogeny is constructed using 149 available genomes by concatenating and aligning amino acid positions across 400 shared proteins automatically identified in the chosen genomes by PhyloPhlAn2. The description of the gene families is listed on the left side. For details please refer to the Materials and Methods. Circles on each branch indicate the bootstrap values above 95%.



1043

1044 **Fig 3. Comparison of the shared and unique gene orthologs amongst selected**

1045 **mycobacteria.** Venn diagrams represent the overlap of gene orthologs between the (A)

1046 *M. riyadhense*, *M. tuberculosis* H37Rv, *M. marinum* M, *M. kansasii* 12478 and *M.*

1047 *szulgai* and (B) *M. riyadhense* MR226 and five species within the MTBC.

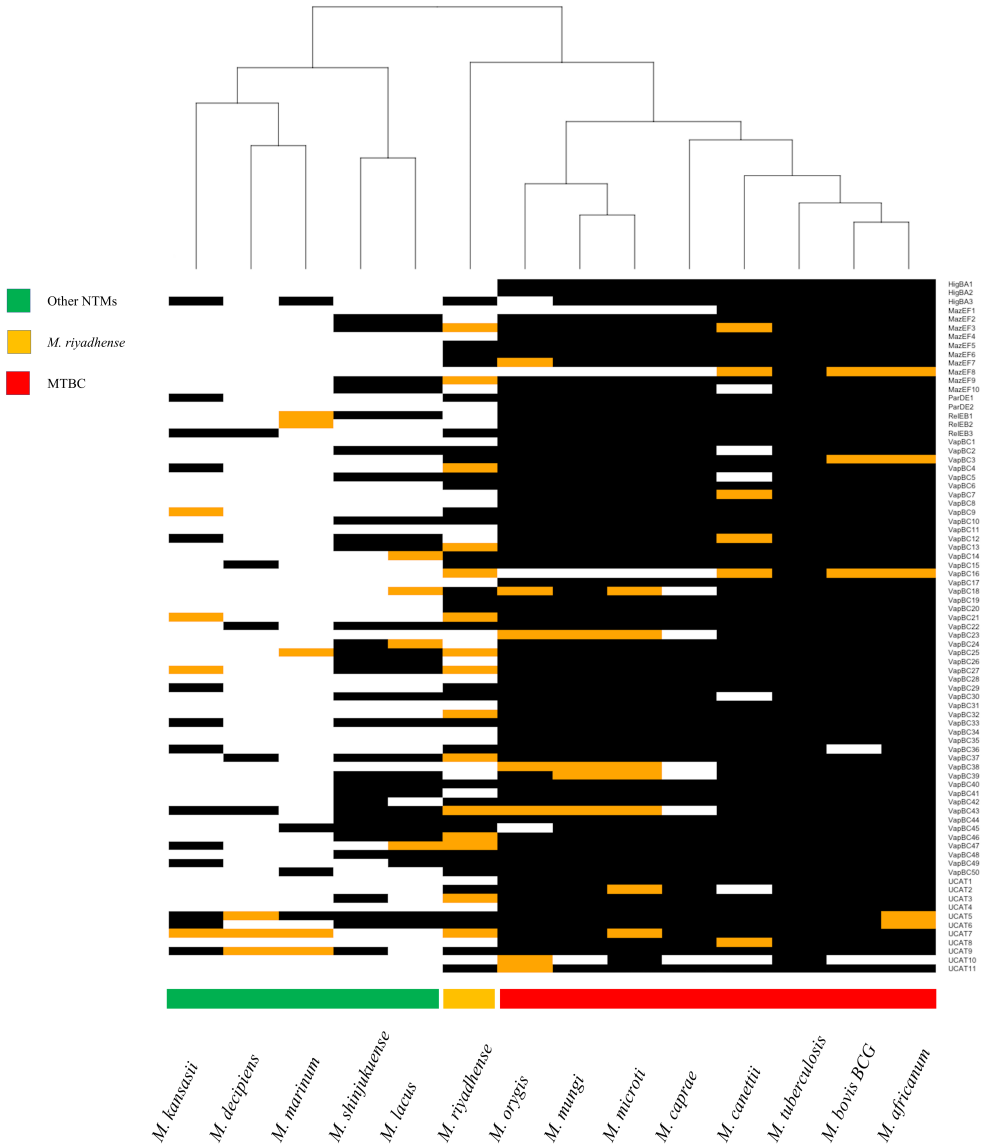


Fig 4. A hierarchical clustering of the presence (black) and absence (white) of *M. tuberculosis* H37Rv toxin/antitoxin orthologs in *M. riyadhense*, *M. marinum*, *M. kansasii*, *M. shinjukuense*, *M. lacus*, *M. decipiens* and MTBC species. The orange blocks denote the presence of either the toxin or antitoxin gene ortholog in a given pair of the T/A system. The black and white blocks represent presence and absence respectively. The name of the T/A system are shown for each row on the right.

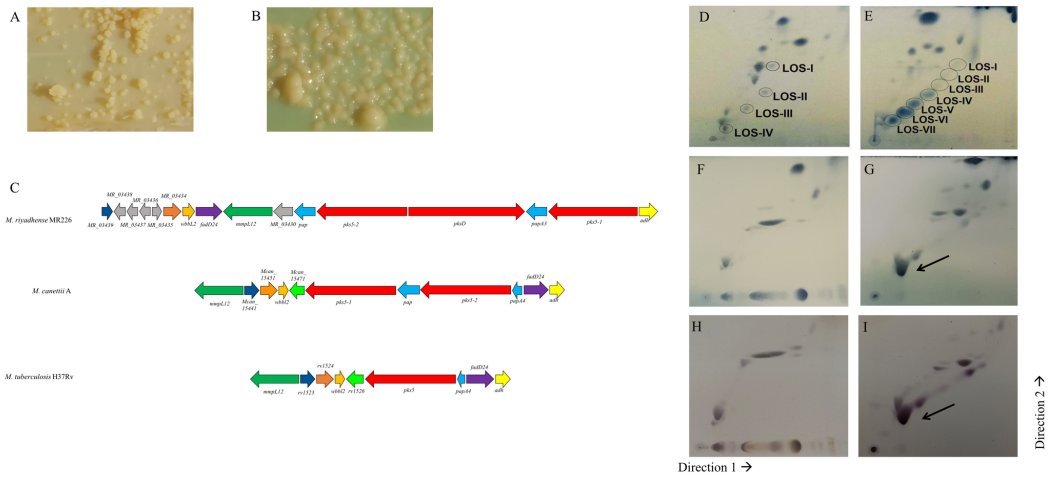


Fig 5. LOS systems in *M. riyadhense* and other related mycobacteria and 2D-TLC analysis of polar lipids extracted from selected *M. riyadhense* strains. (A) Rough-dry colony morphology (MR193) and (B) smooth morphology (MR226) of *M. riyadhense*. (C) Genetic locus map of the LOS biosynthesis gene cluster from *M. riyadhense*, *M. canettii* A and *M. tuberculosis* H37Rv (drawn to scale). The arrows show the direction of transcription and the genes are colored according to the orthologous relationships. Polar lipids from two known LOS producers, *M. marinum* (D) and *M. kansasii* (E), are included to illustrate the migration pattern of LOS species in System E. (F) 2D-TLC analysis of polar lipids extracted from select *M. riyadhense* rough strain or smooth (G) strain. A separate staining with alpha-napthol also confirmed that this was a glycolipid species from the same (H) rough and (I) smooth strain. (D), (E), (F) and (G) were charred after staining with MPA, while (H) and (I) were charred after staining with alpha napthol. LOS III from *M. riyadhense* is indicated by a solid arrow.

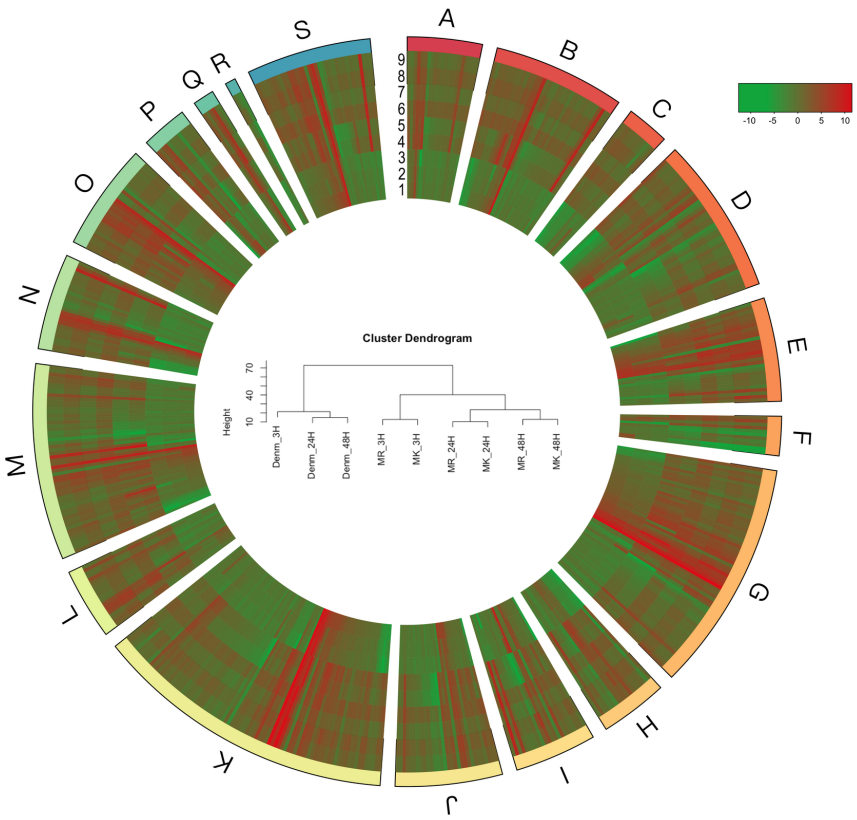
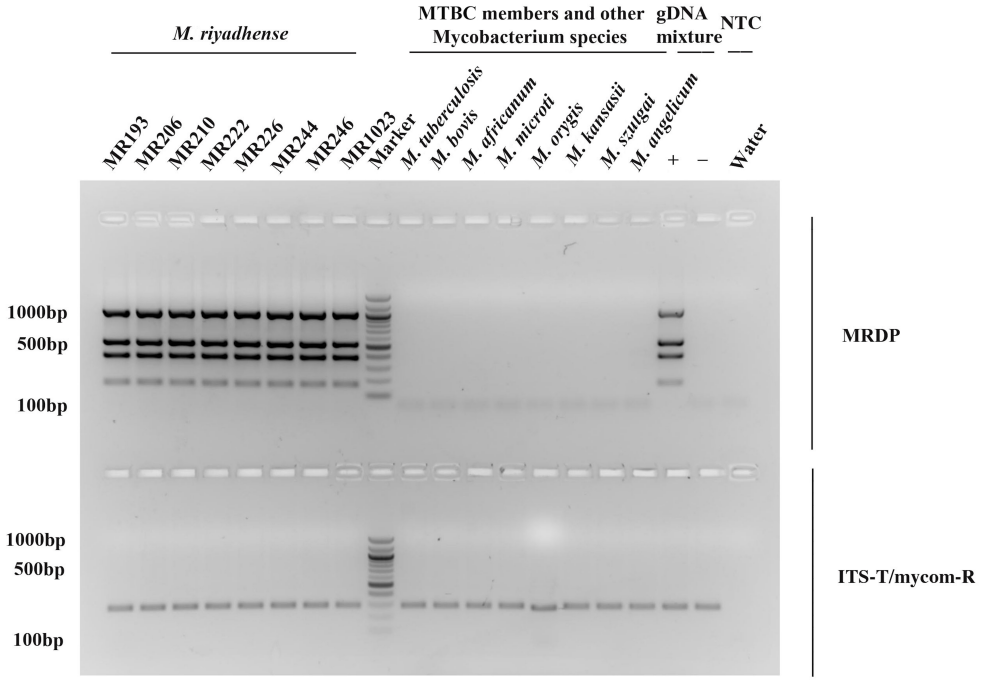


Fig 6. A circular heatmap demonstrating the host gene expression patterns from the RAW264.7 infections with *M. riyadhense* (MR), *M. kansasii* (MK) and *M. bovis* BCG Denmark (Denm) over 3, 24 and 48 hours as determined by the Myeloid Innate Immunity Gene Expression Panel (NanoString Technologies, USA). Data represent Log2FC expression values compared to the uninfected control from each time point. The gene sets of each clusters are: A: Angiogenesis; B: Antigen Presentation; C: Cell Cycle and Apoptosis; D: Cell Migration and Adhesion; E: Chemokine signaling; F: Complement Activation; G: Cytokine Signaling; H: Differentiation and Maintenance of Myeloid Cells; I: ECM remodeling; J: Fc Receptor Signaling; K: Growth Factor Signaling; L: Interferon Signaling; M: Lymphocyte activation; N: Metabolism; O: Pathogen Response; P: T-cell Activation and Checkpoint Signaling; Q: TH1 Activation;

R: TH2 Activation; S: TLR signaling. The circles 1-9 represent RAW264.7 transcriptome response upon infections of 1: Denm-3h; 2: Denm-24h; 3: Denm-48h; 4: MR-3h; 5: MK-3h; 6: MR-24h; 7: MK-24h; 8: MR-48h; 9: MK-48h. The cluster dendrogram in the middle show the hierarchical clustering of the individual gene expression datasets.



1107

1108 **Fig 7. Development of a rapid PCR-based diagnostic test for detection of *M.***

1109 ***riyadhense*.** Agarose gel (2%) electrophoresis patterns of the PCR products are shown as

1110 part of the diagnostic test for *M. riyadhense*. Lane M: DNA Marker, Lane 1-8: Varies *M.*

1111 *riyadhense* strains (From left to right: MR193, MR206, MR210, MR222, MR226,

1112 MR244, MR246, MR1023), Lane 9-16: Varies *Mycobacterium* species (From left to

1113 right: *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti*, *M. oryis*, *M. kansasii*, *M.*

1114 *szulgai* and *M. angelicum*) templates, Template cocktail, mycobacterium species (*M.*

tuberculosis, *M. bovis*, *M. kansasii*, *M. marinum*, *M. szulgai*, *M. avium* and *M. angelicum*) with (Lane 17, +) and without (Lane 18,-) *M. riyadhense* MR226 gDNA template. Lane 19: Non-template control (NTC). Upper Panel: MRDP (*M. riyadhense* diagnostic marker) set. Lower Panel B: *Mycobacterium* genus specific primer ITS-T and mycom-R amplified with mycobacterial gDNA.

Supporting Information

S1 Fig. Axial enhanced CT scan image of the chest from the anonymous patient infected with *M. riyadhense* MR193. Multifocal cavitating consolidation in both lungs predominantly involving the left upper lobe associated with ill-defined ground-glass centrilobular nodules and tree-in-bud appearance on both lungs were observed. The findings were suggestive of tuberculosis.

S2 Fig. A comparison heatmap of predicted protein-coding gene orthologs from *M. riyadhense* MR226 and other mycobacteria. The orthologs were determined by Proteinortho[82]**. The black and white spaces denote presence and absences of an ortholog in a given species respectively.** (A) Comparison of chromosome-encoded genes in 152 mycobacterial species. The box with red outline highlights MTBC. The red arrows indicate the genomic regions that are absent in the assembly GCA_002101845.1 while present in our MR226 assembly. (B) Comparison of the linear plasmid pMRLP01-encoded genes in 152 mycobacterial species genome assemblies. The box b with red outline highlights a region which shared orthologs with pMRLP01 in *Mycobacterium tusciae*, *Mycobacterium aromaticivorans*, *Mycobacterium llatzerense*, *Mycobacterium obuense*, *Mycobacterium novocastrense* and *Mycobacterium holsaticum*. (C) Comparison of the circular plasmid pMR01-encoded 88 genes in 152 mycobacterial species genome assemblies. The box c with red outline highlights the cluster of the pRAW-like plasmids.

S3 Fig. Circular map of pMR01. The circles show from outside to inside (1-12) BlastN results against the pMR01 of various *Mycobacterium* plasmids. The corresponding plasmids used are listed on the right panel. The innermost circle (13) represents pMR01 with the location of the predicted protein coding genes. Red colour indicates the location of the members of the *pe*/*ppe* gene family, blue indicates the *esx*-related genes and all the other predicted genes are colored in grey.

S4 Fig. Multiple alignment of 8 *M. riyadhense* assemblies using progressive Mauve. (A) The alignment of the 8 assemblies and (B) the alignment of the linear and circular plasmids from 8 strains. Circular plasmid is present in 3 strains and highlighted by boxes with black outlines.

S5 Fig. Phylogenetic tree of *M. riyadhense* clinical isolates used in this study. The *M. riyadhense* phylogenetic tree was constructed with SNP data from 8 datasets called by GATK pipeline by RaxML with the TVM model. The circles on each branch indicates the bootstrap values (above 95%, 1,000 replicates).

S6 Fig. Genome alignments comparing selected Region of Differences (RDs) in mycobacteria. The Artemis Comparison Tool (ACT) was used to compare and visualize the annotated genome sequences against the chosen mycobacteria. (A) The RD2 region of *M. riyadhense* MR226, *M. tuberculosis* H37Rv and *M. bovis* BCG Pasteur (B) The RD12 region of *M. tuberculosis*, *M. riyadhense* MR226 and *M. canettii* CIPT 140010059.

S7 Fig. The relative distribution of the functions of the predicted protein-encoding genes were normalized by the genome size of each genome and compared to the *M. riyadhense* MR226 strain.

The phylogenomic position of *M. riyadhense* within the *Mycobacterium* genus was determined by concatemer sequences of 906 shared single copy genes covering 296,124 amino acid positions. *Nocardia abscessus* was used as the out group. The concatenated sequences were used to build the phylogenetic tree with the LG+G+F model, which was selected by ModelGenerator. The phylogenetic tree was generated by RaxML with 1,000 iterations and the bootstrap values are shown above each branch.

S8 Fig. Genetic locus map of the *pe-pgrs33* gene cluster from *M. marinum*, *M. kansasii*, *M. riyadhense*, *M. canettii* and *M. tuberculosis* (drawn to scale).

The genes are shown with arrows and are colored according to the orthologs. The deletion event was highlighted with red arrows while the insertion event in green arrows.

S9 Fig. Comparison of different gene clusters that encode type VII secretion systems in the *M. riyadhense* MR226 strain.

The genes are shown with arrows and are colored according to the orthologs. The color codes for the figure are presented in the key. The black arrows indicate region-specific genes.

S10 Fig. TLR and NFkB pathway responses across the three infections.

Clustering of Nanostring data from infections with *M. riyadhense* (MR), *M. kansasii* (MK) and *M. bovis* BCG Denmark (Denm) over 3, 24 and 48 hrs showed commonality and variation in the transcriptional responses. Panel A and B show selected genes involved in the TLR (A) and NFkB (B) responses from the overall 754 gene panel used in this study.

S11 Fig. Basic statistics of unique K-mers of *M. riyadhense* across 152 mycobacteria genome assemblies.

S1 Table. Comparison of *M. riyadhense* strains' assemblies.

S2 Table. The SNPs analysis of different samples using MR226 as the reference.

S3 Table. *M. marinum* LOS locus gene orthologs shared with *M. riyadhense* strains.