

Drug-resistance profiling and transmission dynamics of multidrug-resistant *Mycobacterium tuberculosis* in Saudi Arabia revealed by whole genome sequencing

Hawra Al-Ghaffli^{1,*}
Thomas A Kohl^{2,3,*}
Matthias Merker^{2,3,*}
Bright Varghese^{1,*}
Anason Halees⁴
Stefan Niemann^{2,3}
Sahal Al-Hajoj¹

¹Mycobacteriology Research Section, Department of Infection and Immunity, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia; ²Molecular and Experimental Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Borstel 23845, Germany; ³German Centre for Infection Research (DZIF), Partner site Borstel, Borstel 38124, Germany; ⁴Data and Information Office, Anfas Medical Centre, Hittin District, Riyadh, Saudi Arabia

*These authors contributed equally to this work

Background: In Saudi Arabia, cross-border transmission of multidrug-resistant (MDR) *Mycobacterium tuberculosis* complex (MTBC) strains might be particularly fostered by high immigration rates. Herein, we aimed to elucidate the transmission dynamics of MDR-MTBC strains and reveal a detailed prediction of all resistance-conferring mutations for the first- and second-line drugs.

Methods: We investigated all MDR-MTBC strains collected between 2015 and 2017 from provincial mycobacteria referral laboratories and compared demographic and clinical parameters to a cohort of non-MDR-TB patients using a whole genome sequencing approach. Clusters were defined based on a maximum strain-to-strain genetic distance of five single-nucleotide polymorphisms (SNPs) as surrogate marker for recent transmission, and then investigated molecular drug-resistance markers (37 genes).

Results: Forty-eight (67.6%) MDR-MTBC strains were grouped in 14 different clusters, ranging in size from two to six strains; 22.5% (16/71) of all MDR-MTBC isolates were predicted to be fully resistant to all five first-line drugs, and five strains (7.0%) exhibited fluoroquinolone resistance. Moreover, we revealed the presence of 12 compensatory mutations as well as 26 non-synonymous SNPs in the *rpoC* gene and non-hotspot region in *rpoB*, respectively

Conclusion: Optimized TB molecular surveillance, diagnosis, and patient management are urgently needed to contain MDR-MTBC transmission and prevent the development of additional drug resistance.

Keywords: molecular profiling, drug resistance, genotyping, detailed prediction, compensatory mutations, drug-resistance mutations

Introduction

The global figures on tuberculosis (TB) urge for optimized control strategies and efficient allocation of resources to reduce TB mortality and morbidity rates. In 2016, the WHO reported 10.4 million new cases and approximately 1.8 million deaths caused by TB.¹ A particular concern in TB control is the worldwide emergence of resistant, multidrug-resistant (MDR), and extensively resistant (XDR) *Mycobacterium tuberculosis* complex (MTBC) strains.¹ M/XDR-TB is considerably more difficult to treat (longer duration, second-line drugs with severe side effects) and has much poorer treatment outcomes.²

Saudi Arabia is classified by the WHO as a moderate TB-burden country;¹ however, several factors may contribute to TB transmission and the emergence of drug resistance

Correspondence: Sahal Al-Hajoj
Mycobacteriology Research Section,
Department of Infection and Immunity,
King Faisal Specialist Hospital and
Research Center, Post Box #3354,
Takasusi Street, Riyadh 11211, Saudi
Arabia
Tel/fax +966 11 464 7272 ext 24992
Email hajoj@kfshrc.edu.sa

among new cases. The country has a unique geographical location, situated in the heart of three different continents: Africa, Asia, and Europe. As such, it serves as a major migration hub. In addition, the country hosts millions of Islamic pilgrims annually, most of whom arrive from highly TB-endemic regions, and only 60% of the total population are Saudis.³ The resulting mobility increases the risk of the introduction and transmission of “imported” M/XDR strains from patients who have not been registered or identified by local TB-control measures.

The first molecular epidemiology study in the country, based on spoligotyping, revealed the presence of the geographically dominating Delhi/CAS genotype among others, such as Latin American Mediterranean (LAM), East African Indian (EAI), Uganda, Ghana, Cameroon, Haarlem, and Beijing, which may reflect transmission from immigrated TB patients.⁴ Recently, Varghese et al applied 24 loci MIRU-VNTR typing for the analysis of drug resistant (mono, poly, and multidrug resistant) MTBC strains, and showed the predominance of Delhi/CAS, EAI, Beijing, Ghana, and Haarlem genotypes. In addition, the study revealed a high cluster rate between local and immigrant population with no association between patient origin and strain genotype.⁵ However, both, spoligotyping and MIRU-VNTR typing have a low discriminatory power to precisely analyze the genetic relationship of MTBC strains as a surrogate marker for the extent of recent patient-to-patient MTBC transmission, especially in high-incidence populations.^{2,6,7} Here, whole genome sequencing (WGS)-based approaches have a clear advantage by allowing the discrimination of clinical MTBC strains at the highest resolution and simultaneously enabling detailed resistance predictions for, virtually, all drugs.^{2,8} More recently, the first national surveillance study conducted on 1,902 MTBC strains revealed an alarming proportion of 4.5% MDR-TB cases in the country.⁹ However, high-resolution data on transmission patterns or molecular drug-resistance profiles using WGS in Saudi Arabia is lacking, which is a critical step to define the underlying epidemiology of MDR-TB and its transmission dynamics with high precision.

To address these questions, we utilized a WGS approach to characterize all MDR-MTBC strains sampled from different provinces of the country between 2015 and 2017. In addition, we included a cohort of non-MDR MTBC strains to point out putative associations between patients infected with MDR and non-MDR MTBC strains, which were also stratified on the basis of their demographic and clinical parameters.

Materials and methods

Study design

The Mycobacteriology Research Section of King Faisal Specialist Hospital and Research Center maintains a national-level collection of clinical isolates as part of the National TB Surveillance Study during August 2015 to July 2017. The collection of isolates applied the WHO’s 100% collection policy during the execution time; thus, a fair and extensive coverage rate was acquired. During the period, 2,156 isolates were collected; 394/2156 were either not successfully cultured or recognized as non-MTBC strains. The remaining 1,762 isolates were subjected to phenotypic drug-susceptibility testing (DST) and mutation-detection via Line Probe Assay (LPA) as an initial screening method. A total of 71 isolates were phenotypically defined as MDR strains and were accordingly considered in the study. Another 134 non-MDR-TB strains were included as background samples to obtain a final sample size of 205 isolates.

Demographic and clinical data were gathered using data collection forms from the provincial laboratories on the basis of national TB codes, and the data were verified at the National TB Registry. The data collection was strictly anonymous and no patient identifiers were used throughout the collection and analysis period. The study has been reviewed and approved by the Office of Research Affairs at King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

Laboratory methods and data analysis

All the isolates were subjected to DST of four first-line drugs (isoniazid, rifampicin, streptomycin, and ethambutol) using the MGIT SIRE kit (Becton Dickinson, USA) as an initial screening method. Genomic DNA was extracted with a standard spin-column technique as per the manufacturer’s recommendations (QIAamp DNA mini Kit, Qiagen, Germany). Primary rapid screening of mutations conferring resistance to rifampicin and isoniazid was carried out using the LPA Genotype MTBDR *Plus* (Hain Life Science, Nehren, Germany).

From the extracted genomic DNA, WGS was carried out with Illumina next-generation sequencing technology, using the Nextera XT library preparation kit and the NextSeq-500 sequencing platform as instructed by the manufacturer (Illumina, San Diego, CA, USA). Reads were mapped to the genome of *M. tuberculosis* H37Rv (GenBank ID: NC_000962.3) with the alignment program SARUMAN.¹⁰ All datasets reached a minimum average coverage depth of

54.8-fold, with at least 98.6% of the reference genome covered by at least one read for all isolates. For variant detection in the mapped reads, we employed minimum thresholds of 10× coverage and 75% allele frequency. Variant positions were then combined, supplementing the joint list with the respective information from the original mappings where necessary. SNP positions with a reliable base call (10× coverage and 75% allele frequency) in at least 95% of the isolates were concatenated to a sequence alignment, excluding SNPs occurring within a window of 12 bp from each other in the same isolate. Genome clusters were defined by grouping all isolates within a maximum distance of five SNPs to the nearest group member – an empirically defined cutoff for epidemiologically related cases (ie, household contacts).¹¹ Polymorphisms in 37 drug-resistance and bacterial fitness-associated target genes were investigated and used to construct molecular drug-resistance profiles of all MDR-MTBC strains (Table S1). SNPs in 37 resistance-related target genes that were found in at least two isolates of the same MTBC genotype but were not associated to phenotypic resistance were excluded as putative phylogenetic variants (Table S1).

Data analysis was conducted using the software package SPSS version 19.0 (IBM Corporation, Armonk, NY, USA). Pearson's χ^2 statistics or Fisher's two-tailed test was used to compare categorical variables. All variables were studied in a univariate logistic regression analysis. In all analyses, 95% CIs and *P*-values were corrected for the definite population of genomic clusters.

Results

Overall, 71 genotypically confirmed MDR-TB and 134 non-MDR-TB cases were enrolled for final data analysis, including the mainly pulmonary specimens (88.3%) and extra-pulmonary specimens (12%; Table 1). All enrolled isolates were culture-positive and 72% were smear-positive (Table 1). MTBC strains were derived from 62.9% male and 37.1% female patients; the dominating age groups were 16–29 years (32.2%) and 30–45 years (37.1%). Similar proportions of Saudi and non-Saudi patients were included, mainly originating from the Central (27.3%) and Western provinces (34.1%). On comparing MDR and non-MDR-TB cases in a univariate analysis, we found no differences in the general patient characteristics described earlier. However, MDR-TB patients were more often observed with smear-positive acid-fast bacilli specimens ($P=0.001$), higher treatment-failure rates ($P=0.002$), deportation ($P=0.006$), and patient death ($P=0.017$; Table 1).

All strains were phylogenetically classified in accordance with a recently reported SNP barcode nomenclature.¹² Overall, four major MTBC lineages were identified: 32/205 (15.6%) of isolates were assigned to Lineage 1 (EAI), 16/205 (7.8%) to Lineage 2 (Beijing), 43/205 (21.0%) to Lineage 3 (Delhi/CAS), 113/205 (55.1%) to Lineage 4 (Euro-American super-lineage), and one isolate was classified as *Mycobacterium bovis* (Figure 1, Figure S1). With regard to the MDR-MTBC population structure in particular, we assigned 11/71 (15.5%) isolates to Lineage 1, 11/71 (15.5%) to Lineage 2, 17/71 (23.9%) to Lineage 3, and 32/71 (45.0%) to Lineage 4 (Table 2).

On a sub-lineage level, the most prevalent MDR-MTBC groups that could be further differentiated by the applied SNP barcode were Lineage 4.2.2 (associated with the TUR genotype; 13/71, 18.3%), Lineage 1.1.2 (associated with the EAI genotype; 10/71, 14.1%), Lineage 2.2.1 (associated with the Beijing genotype; 10/71, 14.1%), and Lineage 4.6.2.2 (associated with the Cameroon genotype; 7/71, 9.9%; Table 2). Compared to the non-MDR-MTBC strains ($n=134$), strains belonging to Lineage 2 were enriched in the MDR-TB cohort ($P=0.006$) whereas all other identified lineages and sub-lineages were observed with similar proportions in both datasets (Table 2).

Primary screening of phenotypically and genotypically confirmed MDR-MTBC strains with LPA identified 43/71 (60.6%) S531L (ie, S450L in the *M. tuberculosis* reference genome, see “Methods” section), 4/71 (5.6%) H526Y (ie, H445Y) and 4/71 (5.6%) D516V (ie, D435V) mutations in the *rpoB* gene, and 20/71 (28.2%) had a missing wild-type band that indicated another rare mutation in the *rpoB* hotspot region (Tables S1 and S2). With regard to isoniazid-resistance-mediating mutations, 59/71 (83.8%) of the MDR-MTBC isolates showed the mutation *katG* S315T, (8/71, 10.3%) *inhA* mutation in Position –15 (C>T) and four missing wild-type bands for the *katG* gene (Table S2).

Using WGS, we analyzed insertions, deletions, and SNPs in 37 target genes associated with bacterial fitness and antibiotic resistance. Isoniazid resistance was dominant with the mutation (S315T, 61/71, 85.9%) in *katG* gene. The second common mutation was found in the *inhA* gene (–15 C>T, 8/71 (11.3%; Table S2). Rifampicin resistance was associated with S450L mutation (43/71, 60.6%) followed by H445Y (9/71, 12.7%) of the *rpoB* gene. All strains with undefined mutations in the *rpoB* gene on LPA have mutations at S450W, L452P, L430P, H445R, H544N, and S428G codon positions (Tables S1 and S2).

We analyzed the *rpsL* and *rrs* genes for the mutations conferring streptomycin resistance. The K43R mutation in

Table 1 Demographic and clinical summary of study subjects

Parameters	n/% (N=205)	MDR, n/% (N=71)	Non-MDR, n/% (N=134)	OR (95% CI)	P-value
Gender					
Male	129 (62.9)	48 (67.6)	81 (60.4)	REF	
Female	76 (37.1)	23 (32.4)	53 (39.6)	1.365 (0.745–2.502)	0.313
Age group (years)					
<15	9 (4.4)	3 (4.2)	6 (4.5)	REF	
16–29	66 (32.2)	19 (26.8)	47 (35.1)	1.236 (0.280–5.459)	0.779
30–45	76 (37.1)	24 (33.8)	52 (38.8)	1.083 (0.249–4.701)	0.914
46–59	33 (16.1)	17 (23.9)	16 (11.9)	0.470 (0.100–2.206)	0.338
>60	21 (10.2)	8 (11.3)	13 (9.3)	0.812 (0.157–4.197)	0.805
Nationality					
Saudi	109 (53.2)	40 (56.3)	69 (51.5)	REF	
Non-Saudi	96 (46.8)	31 (43.7)	65 (48.5)	1.215 (0.681–2.168)	0.508
-Indian subcontinent	36 (37.5)	11 (15.5)	25 (18.6)	1.317 (0.587–2.959)	0.504
-African	29 (30.2)	12 (16.9)	17 (12.7)	0.821 (0.356–1.893)	0.644
-Southeast Asian	20 (20.8)	6 (8.5)	14 (10.4)	1.352 (0.482–3.799)	0.566
-Middle Eastern	11 (11.4)	2 (2.8)	9 (6.7)	2.609 (0.536–12.67)	0.234
Residential province					
South	18 (8.9)	5 (7.1)	13 (9.7)	REF	
West	70 (34.1)	24 (33.8)	46 (34.3)	0.737 (0.235–2.312)	0.601
East	48 (23.4)	17 (23.9)	31 (23.1)	0.701 (0.213–2.303)	0.558
Central	56 (27.3)	19 (26.8)	37 (27.6)	0.749 (0.232–2.414)	0.628
North	13 (6.3)	6 (8.4)	7 (5.2)	0.448 (0.100–2.013)	0.295
Previous treatment					
Yes	66 (32.2)	23 (32.4)	43 (32.1)	REF	
No	139 (67.8)	48 (67.6)	91 (67.9)	1.014 (0.548–1.876)	0.964
Positive AFB smear					
Yes	148 (72.2)	62 (87.3)	86 (64.2)	REF	
No	57 (27.8)	9 (12.7)	48 (35.8)	3.845 (1.756–8.415)	0.001
Site of infection					
Pulmonary	181 (88.3)	67 (94.4)	114 (85.1)	REF	
Extra-pulmonary	24 (11.7)	4 (5.6)	20 (14.9)	2.938 (0.963–8.962)	0.058
Treatment outcome					
Completed	81 (39.5)	21 (29.6)	60 (44.8)	REF	
Failed	12 (5.8)	9 (12.6)	3 (2.2)	0.116 (0.028–0.472)	0.002
Deported	9 (4.4)	7 (9.9)	2 (1.5)	0.100 (0.019–0.519)	0.006
Died	11 (5.4)	7 (9.9)	4 (3.0)	0.200 (0.053–0.752)	0.017
Unknown	92 (44.9)	27 (38.0)	65 (48.5)	0.842 (0.431–1.645)	0.616

Notes: Indian subcontinent – Afghanistan, India, Pakistan, Bangladesh, Nepal, and Myanmar; African – Somalia, Sudan, Nigeria, Eritrea, Ethiopia, and Egypt; Middle Eastern – Yemen and Syria; South East Asian – Philippines and Indonesia. Provincial cities: South – Jizan and Asir; West – Jeddah, Baha, Taif, and Medina; East – Dammam; Central – Riyadh; North – Hail. Unknown: After diagnosis, the patient's follow-up was lost from the records.

Abbreviations: AFB, acid-fast bacilli; MDR, multidrug-resistant; *M. bovis*, *Mycobacterium bovis*; REF, reference.

the *rpsL* gene was found to be the most common mutation (16/71, 22.5%) followed by the 514 A>C mutation in the *rrs* gene (12/71, 16.9%). Ethambutol-resistant isolates were mainly associated with the mutation *embB* M306V (28/71, 39.4%; Table S2). Furthermore, we identified variants in the upstream region of the gene *pncA*, associated with pyrazinamide resistance, as well as non-synonymous and frame-shift mutations in 52.1% (37/71) of isolates. Overall, 22.5% (16/71) of MDR-MTBC isolates were identified with molecular resistance markers to all five first-line drugs; 5/71

(7.0%) of the MDR-MTBC strains were found with a fluoroquinolone-resistance-mediating mutation (ie, *gyrA* A90V and *gyrA* D94G, respectively) and were defined accordingly as pre-XDR (Table S2). Overall 18/205 isolates (including 13 MDR-MTBC isolates) were observed with prothionamide- and ethionamide-resistance-mediating mutations – that is, *fabG1/inhA* –15 c/t, *ethA* 672 insG, and *ethA* 377 delT (Tables S1 and S2). No resistance marker was identified for second-line injectable drugs or WHO Group Five anti-TB drugs (ie, bedaquiline, delamanid, clofazimine, linezolid, and

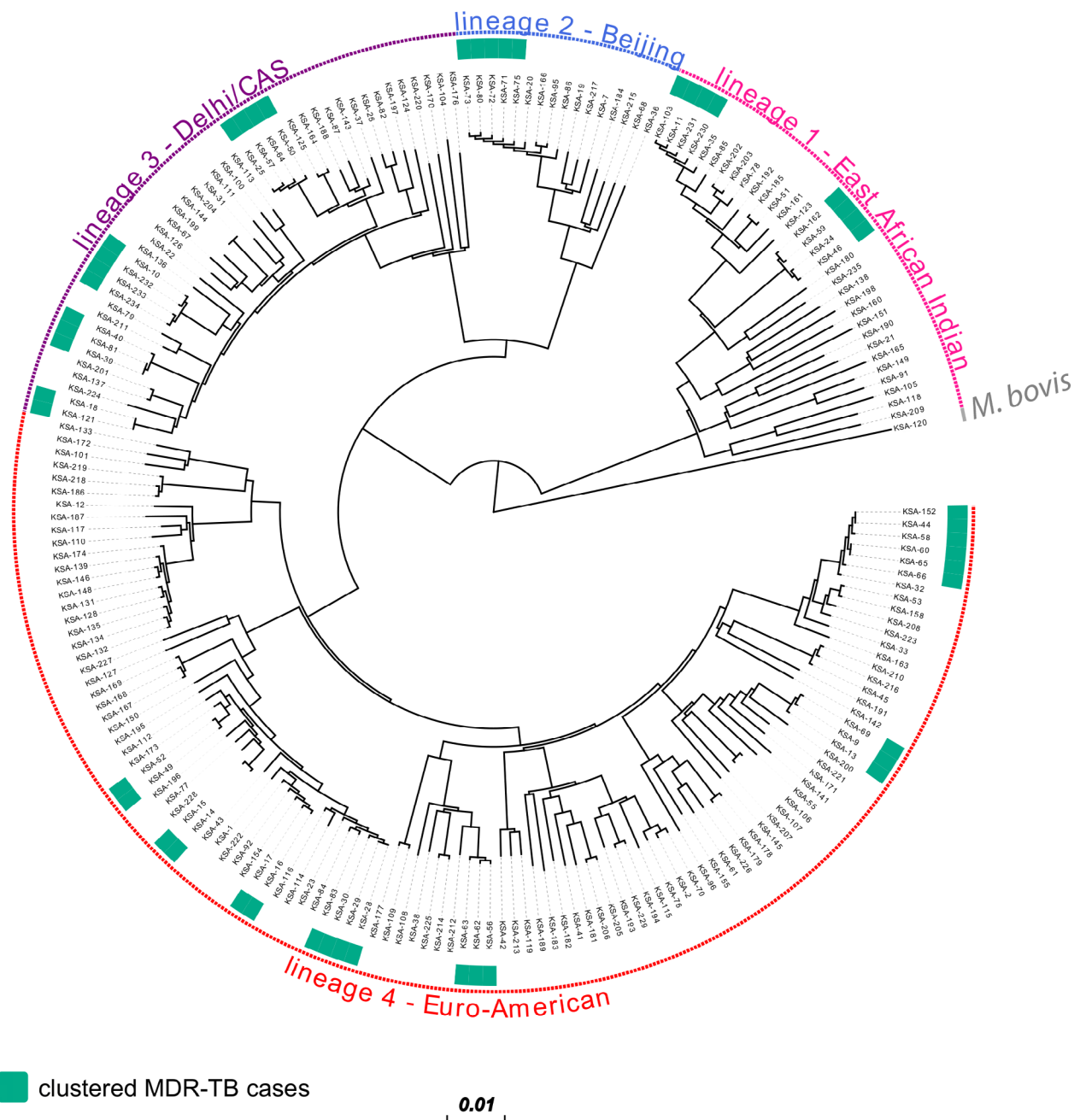


Figure 1 Phylogeny of 71 MDR and 134 non-MDR-MTBC strains from Saudi Arabia.

Notes: Maximum likelihood phylogeny of 205 isolates with annotated genomic MDR-MTBC clusters using 21,745 variable genome positions and 1,000 bootstrap resamples. Main MTBC lineages are annotated.

Abbreviations: MDR, multidrug-resistant; MTBC, *Mycobacterium tuberculosis* complex; *M. bovis*, *Mycobacterium bovis*; CAS, Central Asian.

paraminosalicylic acid). Among 47/71 (66.2%) MDR-MTBC isolates, we identified putative compensatory mutations in the RNA-polymerase genes *rpoA* (1/71; 1.4%), *rpoB* (32/71; 45.1%), and *rpoC* (23/71; 32.4%). The mutation in the *rpoC* gene was found to dominate (17/23, 73.9%) among isolates as was the S450L mutation of the *rpoB* gene (Table S2).

In order to determine molecular strain clusters indicative of recent chains of MDR-MTBC transmission (within the

last 3 years), we defined groups of at least two strains that had a maximum distance of five SNPs to at least one other group member. This threshold was recently shown to be a reliable indicator to identify transmission among household contacts in the UK.¹¹ Overall, 48/71 (67.6%) MDR-MTBC isolates were assigned into one of 14 molecular clusters, each comprising two to six patients (Figure 1, Table 3). In contrast, non-MDR strains showed a clustering rate of only

Table 2 Major phylogenetic lineages and sub-lineages classified by whole genome sequencing approach

Phylogenetic lineages	n/% (N=205)	Clustered isolate/rate (N=98)	MDR, n/% (N=71)	Non-MDR, n/% (N=134)	95% CI	P-value
Lineage 1	32 (15.6)		10 (14.1)	22 (16.4)	REF	
Lineage 2	16 (7.8)		12 (16.9)	4 (2.9)	0.151 (0.039–0.588)	0.006
Lineage 3	43 (20.9)		17 (23.9)	26 (19.4)	0.695 (0.264–1.825)	0.460
Lineage 4	113 (55.1)		32 (45.1)	81 (60.4)	1.150 (0.490–2.697)	0.747
Sub-lineages						
3 Delhi-CAS	33 (16.1)	12 (12.2)	11 (15.5)	22 (16.4)	REF	
4.2.2 Euro-American	26 (12.7)	15 (15.4)	13 (18.3)	13 (9.7)	0.500 (0.174–1.437)	0.198
1.1.2 EAI	20 (9.7)	12 (12.2)	10 (14.1)	10 (7.5)	0.500 (0.160–1.558)	0.232
2.2.1 Beijing	14 (6.8)	7 (7.1)	9 (12.7)	5 (3.7)	0.277 (0.075–1.030)	0.055
4.1.2.1 Haarlem	13 (6.3)	9 (9.2)	1 (1.4)	12 (8.9)	6.00 (0.689–52.26)	0.104
4.6.2.2 Cameroon	13 (6.3)	9 (9.2)	7 (9.8)	6 (4.5)	0.429 (0.115–1.586)	0.204
4.4.1.1 S-type	8 (3.9)	5 (5.1)	3 (4.2)	5 (3.7)	0.833 (0.167–4.143)	0.823
4.8 Euro-American	8 (3.9)	3 (3.1)	3 (4.2)	5 (3.7)	0.833 (0.167–4.143)	0.823
4.3.3 LAM	7 (3.4)	2 (2.0)	3 (4.2)	4 (2.9)	0.667 (0.126–3.515)	0.632
1.2.1 EAI Manila	5 (2.4)	–	1 (1.4)	4 (2.9)	2.000 (0.199–20.104)	0.556
3.1.2.1 Delhi-CAS	5 (2.4)	4 (4.1)	2 (2.8)	3 (2.2)	0.750 (0.108–5.167)	0.770
4 Euro-American	4 (2.0)	–	1 (1.4)	3 (2.2)	1.50 (0.139–16.144)	0.738
3.1.1 Delhi-CAS	2 (1.0)	–	1 (1.4)	1 (0.7)	0.500 (0.028–8.772)	0.635
1.2.2 EAI	3 (1.5)	–	–	3 (2.2)	–	–
4.5 Euro-American	6 (2.9)	2 (2.0)	–	6 (4.5)	–	–
1.1.3 EAI	4 (2.0)	–	–	4 (2.9)	–	–
4.6 Euro-American	4 (2.0)	3 (3.1)	–	4 (2.9)	–	–
3.1.2 Delhi-CAS	3 (1.5)	3 (3.1)	3 (4.2)	–	–	–
4.1 Euro-American	3 (1.5)	3 (3.1)	–	3 (2.2)	–	–
4.1.1.3 X-type	3 (1.5)	–	–	3 (2.2)	–	–
4.2.2.1 TUR	3 (1.5)	3 (3.1)	–	3 (2.2)	–	–
4.7 Euro-American	3 (1.5)	2 (2.0)	–	3 (2.2)	–	–
4.2.1 Ural	2 (1.0)	–	–	2 (1.5)	–	–
4.3.4.2.1 LAM	2 (1.0)	2 (2.0)	–	2 (1.5)	–	–
4.4 Euro-American	2 (1.0)	2 (2.0)	–	2 (1.5)	–	–
2.2.1.1 Beijing	1 (0.5)	–	1 (1.4)	–	–	–
2.2.2 Beijing	1 (0.5)	–	1 (1.4)	–	–	–
4.3 LAM	1 (0.5)	–	–	1 (0.7)	–	–
4.3.1 LAM	1 (0.5)	–	–	1 (0.7)	–	–
4.3.2 LAM	1 (0.5)	–	–	1 (0.7)	–	–
4.3.4.1 LAM	1 (0.5)	–	1 (1.4)	–	–	–
4.3.4.2 LAM	1 (0.5)	–	–	1 (0.7)	–	–
4.6.1.1 Uganda	1 (0.5)	–	–	1 (0.7)	–	–
<i>M. bovis</i>	1 (0.5)	–	–	1 (0.7)	–	–

Abbreviations: MDR, multidrug-resistant; *M. bovis*, *Mycobacterium bovis*; EAI, East African Indian; LAM, Latin American Mediterranean; TUR, Turkish.

30.6%. Individual MDR-MTBC strains within genomic clusters each presented with identical resistance-conferring mutations in the genes *katG* or *inhA*, and *rpoB* mediating a MDR phenotype (Table 3 and S1). In addition, 29 MDR-TB patients in eight different transmission networks were infected with strains showing cluster-specific *pncA* mutations associated with pyrazinamide resistance – again, pointing toward ongoing patient-to-patient transmission. However, molecular markers associated with second-line injectable drug resistance were identical only in one cluster.

Demographic data from patients involved in the 14 suggested transmission networks revealed five clusters with only Saudi nationals and one cluster with only non-Saudi nationals. The remaining eight clusters comprised both Saudi and non-Saudi nationals. Ten out of the 14 clusters were linked to a single city, with the remaining four clusters linked to two cities. There was no significant difference in the MDR-MTBC cluster rate on comparing patient origin – Saudi nationals (57.1%) and non-Saudi nationals (41.2%; $P=0.629$; Table 4). However, lack of proper social information restricts further

Table 3 Characteristics of 14 MDR-TB genomic clusters

Clusters	No of isolates	Lineage	Gender	Age, (mean)	Nationality	City of residence	Treatment outcome	Pza (N/%)	Emb (N/%)	Str (N/%)	Flq (N/%)	Compensatory mutations N (%)	XDR/Pre XDR (N)
Cluster-6	4	Delhi-CAS	M (3), F (1)	49	Saudi	Dammam	Completed (1), Unknown (3)	-	4 (100)	4 (100)	-	-	-
Cluster-7	4	EAI	M (3), F (1)	43	Saudi (3), Non-Saudi (1)	Taif (3), Jeddah (1)	Completed (1), Failed (2), died (1)	3 (75)	3 (75)	4 (100)	-	-	-
Cluster-8	4	Euro-American	M	42.2	Non Saudi (3), Saudi (1)	Jizan (3), Jeddah (1)	Deported (2), Unknown (2)	4 (100)	3 (75)	-	-	rpoC: 4 (100); rpoB: 4 (100)	-
Cluster-9	6	Cameroon	M (4), F (2)	35.3	Saudi (5), Non-Saudi (1)	Taif (5), Medina (1)	Completed (2), Unknown (3)	6 (100)	5 (83.3)	6 (100)	-	rpoB:6 (100)	-
Cluster-11	4	Delhi-CAS	M (3), F (1)	60.5	Saudi (3), Non-Saudi (1)	Taif(3), Jeddah(1)	Died (1), Deported (1), completed (2), Unknown (1)	-	-	4 (100)	-	rpoB:4 (100)	-
Cluster-12	4	EAI	F	54.7	Saudi	Dammam	Completed (2), Died (2)	-	4 (100)	-	-	rpoC:4 (100)	-
Cluster-13	5	Beijing	F (3), M (2)	36.4	Saudi (3), Non-Saudi (2)	Riyadh	Failed (3), completed (1)	5 (100)	5 (100)	5 (100)	-	rpoB:5 (100)	-
Cluster-14	3	Delhi-CAS	M	40.7	Saudi (2), Non-Saudi (1)	Riyadh	Unknown (1), Unknown (2), Died (1)	-	3 (100)	3 (100)	-	-	-
Cluster-15	3	Euro American	M	27.3	Non-Saudi	Riyadh	Unknown (3)	3 (100)	3 (100)	3 (100)	-	-	-
Cluster-19	3	S-type	F	19.3	Saudi (2), Non-Saudi (1)	Taif	Completed (1), unknown (2)	3 (100)	3 (100)	-	-	-	-
Cluster-27	2	Euro-American	M	31	Saudi	Dammam	Completed (1), Failed (1)	2 (100)	2 (100)	-	-	rpoB:2 (100)	-
Cluster-30	2	Euro-American	M	35	Saudi	Hail	Completed (1), Failed (1)	-	-	-	-	rpoB:2 (100)	-
Cluster-31	2	Delhi-CAS	M	36.5	Saudi (1), Non-Saudi (1)	Hail	Deported (1), completed (1)	2 (100)	2 (100)	-	2 (100)	rpoC:2 (100)	2
Cluster-10	2	Euro-American	M/F	23.5	Saudi	Taif	Completed (1), Unknown (1)	-	2 (100)	2 (100)	-	-	-

Notes: Treatment regimens have been given to patients; no culture conversion was documented (failed), with confirmed culture conversion (completed).

Abbreviation: Emb, ethambutol; Flq, fluoroquinolone; PZA, pyrazinamide; Str, streptomycin; XDR, extensively resistant.

Table 4 Demographic and clinical risk

Parameters	Total (MDR)	Clustered; N=48 (n%)	Non-clustered; N=23 (n%)	OR (95% CI)	P-value
Gender					
Male	50	33 (68.7)	17 (73.9)	REF	
Female	21	15 (31.3)	6 (26.1)	0.776 (0.255–2.363)	0.597
Age group (years)					
<15	3	2 (4.2)	1 (4.3)	REF	
16–29	17	11 (22.9)	6 (26.1)	0.545 (0.028–10.368)	0.687
30–45	24	14 (29.2)	10 (43.5)	1.428 (0.113–18.00)	0.782
46–59	19	17 (35.4)	2 (8.7)	0.235 (0.014–3.917)	0.313
>60	8	4 (8.3)	4 (17.4)	2.00 (0.125–31.97)	0.624
Nationality					
Saudi	43	30 (62.5)	13 (56.5)	REF	
Non-Saudi	28	18 (37.5)	10 (43.5)	1.281 (0.467–3.521)	0.629
residential province					
South	5	3 (6.2)	2 (8.7)	REF	
West	25	19 (39.6)	6 (26.1)	0.474 (0.063–3.538)	0.467
East	17	12 (25.0)	5 (21.7)	0.625 (0.079–4.958)	0.656
Central	18	10 (20.8)	8 (34.8)	1.2 (0.159–90.131)	0.859
North	6	4 (8.3)	2 (8.7)	0.75 (0.064–8.833)	0.819
Previous treatment					
Yes	22	4 (8.3)	18 (90.0)	REF	
No	49	44 (91.7)	5 (21.7)	0.025 (0.006–0.105)	<0.001
Positive AFB smear					
Yes	62	41 (85.4)	21 (91.3)	REF	
No	9	7 (14.6)	2 (8.7)	0.557 (0.106–2.925)	0.490
Treatment outcome					
Completed	21	14 (29.2)	7 (30.4)	REF	
Failed	8	7 (14.6)	1 (4.3)	0.285 (0.029–2.802)	0.282
Deported	7	4 (8.3)	3 (13.1)	1.50 (0.260–8.635)	0.649
Died	6	5 (10.4)	1 (4.3)	0.400 (0.039–4.114)	0.441
Unknown	29	18 (37.5)	11 (47.9)	1.222 (0.377–3.967)	0.738
Strain lineage					
Lineage 1	10	8 (16.7)	2 (8.7)	REF	
Lineage 2	12	5 (10.4)	7 (30.4)	5.6 (0.814–38.51)	0.079
Lineage 3	17	13 (27.1)	4 (17.4)	1.231 (0.182–8.33)	0.831
Lineage 4	32	22 (45.8)	10 (43.5)	1.818 (0.325–10.157)	0.496

Abbreviations: AFB, acid-fast bacilli; MDR, multidrug-resistant.

analysis of epidemiological associations among the MDR-TB patients (Table 4). The two major clusters comprised MTBC strains belonging to the Cameroon sub-lineage (4.6.2.2; six cases) and the Beijing lineage (2.2.1; five cases), and included both Saudi and non-Saudi nationals (Table 3). The following mutations were shared among all strains from the Cameroon genotype cluster (*katG* S315T, *rpoB* S450L, *Rvnr01* 514A>C, *embB* M306V, and *pncA* T76P). Strains from the Beijing genotype cluster shared the mutations *katG* S315T, *rpoB* S450L, *rpsL* K43R, *embB* M306V, and *pncA* T47I. Overall, 12/14 (85.7%) of the clusters were observed with identical resistance-conferring mutations (Table 3).

Analysis on confounding factors for MDR-TB clustering showed no statistically significant association with other

variables tested, except for the treatment history. Clustered MDR-TB cases were more likely to be new cases ($P<0.001$), confirming the cluster-definition threshold of five SNPs as an indicator for patient-to-patient transmission (Table 4).

Discussion

To the best of our knowledge, this is the first study elucidating the transmission dynamics of MDR-MTBC strains using a WGS approach in Saudi Arabia – a country challenged by high immigration rates. Alarmingly, high cluster rates (67.6%) were determined among the MDR-MTBC strains, leading to a total of 14 putative recent transmission chains. Together with the fact that MDR-TB was not associated with patient origin, this points to ongoing, endemic transmission of

MDR-TB within Saudi Arabia. This is particularly important as several clustered strains have already acquired resistance-conferring mutations against all first-line drugs. At the same time, genomic markers for second-line drug resistance are still rare among our study cohort, urging for improved TB-control measures for complete resistance detection and tailored therapy.

An association of MDR-TB with recent transmission has been collectively reported from different geographical settings.^{13–15} Moreover, our findings are in concordance with high rates of MDR-MTBC transmission (45.1%) in countries such as Ethiopia.¹⁵ Other intermediate TB-burden countries, such as Iran, Iraq, and Turkey reported lower rates of transmission, ranging from 24% to 35.9%, based on classical genotyping methods.^{14,16,17} Nevertheless, the different methodological approaches employed herein – compared with studies conducted in nearby regions – make the process of contextualizing transmission dynamics very challenging. However, strikingly, our data suggest that, transmission of full first-line-resistant MTBC strains leading to extended primary MDR-TB cases is a major factor for the emergence of MDR-TB in Saudi Arabia. This has been further evidenced by identical resistance-conferring mutations in *katG*, *inhA*, *rpoB*, and *pncA* in clustered strains. Thus, transmission chains must be effectively interrupted by proper case finding (beyond conventional contact-tracing), isolation of MDR-TB cases until they are negative, and comprehensive diagnosis to allow for prompt and effective treatment. The latter are critical to significantly decrease the incidence of MDR-TB and avert the emergence of XDR-TB in Saudi Arabia.

A slightly increased cluster rate was evident among Saudi nationals (62.5%) compared to non-Saudi cases. Of the total clusters, 13 had Saudi cases whereas non-Saudis were represented in only nine clusters. There were four clusters exclusively composed of Saudi nationals, whereas only one cluster was exclusively composed of non-Saudi cases. Interestingly, eight clusters contained both local and foreign patients, which corroborate findings from previous studies;⁵ 71.4% of the clusters were found limited to a single city, which supports the possibility of an active ongoing transmission in the local area. However, no statistical evidence has been observed in the geographical origin (province of residence) of patients and strain clusters.

Among the MDR-TB cases found as part of any cluster, only 29.2% of the patients received a full course of therapy, whereas 30.4% patients received therapy among non-clustered cases. The majority of the MDR-TB cases found in clusters had an “unknown” status without follow-up

visits, according to the medical records. Thus, Saudi Arabia is still far (62%) from the WHO’s MDR-TB treatment-success target (85%). In addition, several previous studies showed a large scale non-compliance of treatment among Saudi nationals in particular.^{18,19} Unsuccessful treatment is, therefore, another likely factor for the ongoing transmission of MDR-TB strains.

As expected from the high immigration rate reported in Saudi Arabia, a high phylogenetic diversity of MTBC strains was identified, unlike what has been reported elsewhere.²⁰ Besides Lineage 3, MTBC strains that are associated with a South Asian evolutionary origin, and which are also frequently isolated from TB patients in East Africa and the Middle East, we found 55% of all patients being infected with Lineage 4 strains. Strains of the TUR/4.2.2 sub-lineage are most prevalent, followed by strains of the Haarlem/4.1.2.1, and Cameroon/4.6.2.2 lineages. Although Haarlem is one of the globally widespread sub-lineages, the sub-lineages Cameroon and TUR are geographically restricted and have been identified mainly in Asia and Africa, highlighting the impact of high immigration rates from these regions into Saudi Arabia.²¹ In fact, the large clusters of Cameroon and TUR strains identified in our study included non-Saudi nationals, particularly of African (Nigeria) and Middle Eastern (Syria) origin. Similarly, Lineage 3 is frequently identified in Asia and Africa and has been previously reported in Saudi Arabia and Oman with dominating rates.^{22,23} The high influx of immigrants coming from Asia and Africa is well reflected by the highly diverse patient population in the study (Figure S1). This might explain the predominance of these sub-lineages in our settings. Strains of lineages 1 and 2, on the other hand, accounted for less than 10% of the investigated strains in this study. This is in contrast to studies in Russia and China, which reported a dominance of Beijing strains and have elucidated an association between Lineage 2 strains and drug resistance.^{24,25} In addition, four strains from Saudi patients belonging to Lineage 4 failed to be identified to the sub-lineage level in this study. This might be attributed to a small “blind spot” in the employed SNP-based algorithm,¹² potentially representing minor novel sub-lineages. Analysis on mutations conferring resistance to isoniazid and rifampicin showed a clear domination of well-established mutations in the *katG* codon 315 and the *rpoB* hotspot region in concordance with previous Saudi Arabian studies by Varghese et al.^{13,26} However, there were six other mutations in the *rpoB* hotspot region and one mutation in *katG* which have never been reported in the country before. Likewise, we have identified variants in the upstream region of the gene *pncA* as well as non-synonymous and frame-shift

mutations that are associated with pyrazinamide resistance in 22.4% of the isolates. Similar to our results, many studies have reported a variety of scattered and variable set of pyrazinamide resistance-conferring mutations.^{27,28} Of note, pyrazinamide is not part of the routine first-line therapy in Saudi Arabia and phenotypic DST is not applied. Thus, our WGS results suggest that mutations in the *pncA* gene have been acquired elsewhere or that individual patients had non-programmatic access to this drug.

Compensatory mutations can be a major factor in restoring the fitness reduction of strains due to drug-resistance-conferring mutations. Compensation for rifampicin resistance has been shown to be associated with mutations in *rpoA*, *rpoC*, and the non-hotspot region of the *rpoB* gene.²⁹ Our analysis detected mutations in the *rpoC* gene and outside the rifampicin-resistance-determining region, with four mutations (one in MDR-MTBC strain and three in non-MDR-MTBC strains) found in the *rpoA* gene. Altogether, 12 and 26 non-synonymous SNPs were detected in the *rpoC* gene and non-hotspot *rpoB* region, respectively. The *rpoC* mutation was more common in strains with an *rpoB* gene S450L mutation, in concordance with a recent study.³⁰ The rate of probable compensatory mutations found in MDR-MTBC isolates was relatively high (66.2%) compared to previous studies from other parts of the world (27–66%).^{29,30}

A larger scale population-based research is, of course, required in the near future through employing WGS to better understand the nature and route of MTB transmission in Saudi Arabia. Although the WGS technique is powerful, lack of accessibility to other demographic and medical indicators has prevented us from epidemiologically tracing the transmission chains in individual clusters. Due to the restrictions of the study design and time frame, we could possibly have missed MDR-MTBC isolates. To conclude, for the first time our results showed that Saudi Arabia is indeed suffering from ongoing MDR-MTBC transmissions indicated by high cluster rates and identical resistance-conferring mutations. As such, effective interruption of transmission chains requires improved TB surveillance, diagnosis, and treatment. In addition, incorporating WGS into the routine TB-management plan might be a crucial step in the future in Saudi Arabia. It facilitates an accurate diagnosis, optimized case management, and a detailed estimation of ongoing transmission dynamics, which is especially important in settings with high immigration and mobility rates.

Acknowledgments

This project has been funded by the National Science, Technology, and Innovation Plan (NSTIP) of King Abdulaziz of

Science and Technology (KACST, Saudi Arabia; grant no. 11-BIO-1436–20). This work was partly supported by the PathoNgenTrace (grant no. FP7-278864–2) and the German Center for Infection Research (DZIF). The sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank Julia Zallet, Vanessa Mohr, and Tanja Ubben (all from Research Institute Borstel) for their help with library preparation and sequencing, Ms. Tracy Lynn AlSarhani for language editing support, and Prof. Mohammed Shoukri for support with the statistical analysis.

Author contributions

HA-G contributed to literature search, data analysis, data interpretation, and manuscript writing, as well as creation of figures and tables. TAK and MM contributed to data analysis, data interpretation, creation of tables and figures, manuscript writing, and literature search. BV contributed to study design, data collection, data analysis, data interpretation, creation of tables and figures, manuscript writing, and literature search. AH contributed to study design, data analysis, and data interpretation. SN contributed to study design, data analysis, data interpretation, manuscript writing, literature search, and creation of tables and figures. SA-H contributed to study design, data collection, data analysis, manuscript writing, and creation of tables and figures. All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. WHO. Global Tuberculosis Report 2017. Geneva: World Health Organization. Available from: <http://apps.who.int/medicinedocs/documents/s23360en/s23360en.pdf>. Accessed June 20, 2018.
2. Walker TM, Merker M, Kohl TA, Crook DW, Niemann S, Peto TE. Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin Microbiol Infect*. 2017;23(3):161–166.
3. Statistical Year Book: Population by Gender, Age Groups and Nationality (Saudi/Non-Saudi). Saudi Arabia: General Authority for statistics in Saudi Arabia. Available from: <https://www.stats.gov.sa/en/5305>. Accessed June 20, 2018.
4. Al-Hajj SA, Zozio T, Al-Rabiah F, et al. First insight into the population structure of *Mycobacterium tuberculosis* in Saudi Arabia. *J Clin Microbiol*. 2007;45(8):2467–2473.
5. Varghese B, Supply P, Allix-Béguec C, et al. Admixed phylogenetic distribution of drug resistant *Mycobacterium tuberculosis* in Saudi Arabia. *PLoS One*. 2013;8(2):e55598.
6. Bjorn-Mortensen K, Lillebaek T, Koch A, et al. Extent of transmission captured by contact tracing in a tuberculosis high endemic setting. *Eur Respir J*. 2017;49(3):1601851.

7. Niemann S, Merker M, Kohl T, Supply P. Impact of Genetic Diversity on the Biology of *Mycobacterium tuberculosis* Complex Strains. *Microbiol Spectr*. 2016;4(6).
8. Walker TM, Kohl TA, Omar SV, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*. 2015;15(10):1193–1202.
9. Al-Hajjaj S, Varghese B, Shoukri MM, et al. Epidemiology of anti-tuberculosis drug resistance in Saudi Arabia: findings of the first national survey. *Antimicrob Agents Chemother*. 2013;57(5):2161–2166.
10. Blom J, Jakobi T, Doppmeier D, et al. Exact and complete short-read alignment to microbial genomes using Graphics Processing Unit programming. *Bioinformatics*. 2011;27(10):1351–1358.
11. Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*. 2013;13(2):137–146.
12. Coll F, McNERNEY R, Guerra-Assunção JA, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun*. 2014;5:4812.
13. Varghese B, Shoukri M, Memish Z, et al. Occurrence of diverse mutations in isoniazid- and rifampicin-resistant *Mycobacterium tuberculosis* isolates from autochthonous and immigrant populations of Saudi Arabia. *Microb Drug Resist*. 2014;20(6):623–631.
14. Durmaz R, Zozio T, Gunal S, et al. Genetic diversity and major spoligotype families of drug-resistant *Mycobacterium tuberculosis* clinical isolates from different regions of Turkey. *Infect Genet Evol*. 2007;7(4):513–519.
15. Yimer SA, Norheim G, Namouchi A, et al. *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *J Clin Microbiol*. 2015;53(4):1301–1309.
16. Torkaman MR, Nasiri MJ, Farnia P, Shahhosseiny MH, Mozafari M, Velayati AA. Estimation of Recent Transmission of *Mycobacterium Tuberculosis* Strains among Iranian and Afghan Immigrants: A Cluster-Based Study. *J Clin Diagn Res*. 2014;8(9):DC05–DC08.
17. Ahmed MM, Mohammed SH, Nasurallah HA, Ali MM, Couvin D, Rastogi N. Snapshot of the genetic diversity of *Mycobacterium tuberculosis* isolates in Iraq. *Int J Mycobacteriol*. 2014;3(3):184–196.
18. Samman Y, Krayem A, Haidar M, et al. Treatment outcome of tuberculosis among Saudi nationals: role of drug resistance and compliance. *Clin Microbiol Infect*. 2003;9(4):289–294.
19. Al-Hajjaj MS, Al-Khatim IM. High rate of non-compliance with anti-tuberculosis treatment despite a retrieval system: a call for implementation of directly observed therapy in Saudi Arabia. *Int J Tuberc Lung Dis*. 2000;4(4):345–349.
20. Ssengooba W, Meehan CJ, Lukoye D, et al. Whole genome sequencing to complement tuberculosis drug resistance surveys in Uganda. *Infect Genet Evol*. 2016;40:8–16.
21. Stucki D, Brites D, Jeljeli L, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet*. 2016;48(12):1535–1543.
22. Al-Hajjaj S, Varghese B, Al-Habobe F, Shoukri MM, Mulder A, van Soolingen D. Current trends of *Mycobacterium tuberculosis* molecular epidemiology in Saudi Arabia and associated demographical factors. *Infect Genet Evol*. 2013;16:362–368.
23. Al-Maniri A, Singh JP, Al-Rawo O, et al. A snapshot of the biodiversity and clustering of *Mycobacterium tuberculosis* in Oman using spoligotyping. *Int J Tuberc Lung Dis*. 2010;14(8):994–1000.
24. Drobniewski F, Balabanova Y, Nikolayevsky V, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA*. 2005;293(22):2726.
25. Guang Xue H, Hai Ying W, Martien WB. Multidrug-Resistant Tuberculosis, People's Republic of China, 2007–2009. *Emerg Infect Dis*. 2011;17(10):1831–1838.
26. Varghese B, Hillemann A, Wijayanti DR, et al. New insight into the molecular characterization of isoniazid and rifampicin resistant *Mycobacterium tuberculosis* strains from Saudi Arabia. *Infect Genet Evol*. 2012;12(3):549–556.
27. Kim HJ, Kwak HK, Lee J, et al. Patterns of *pncA* mutations in drug-resistant *Mycobacterium tuberculosis* isolated from patients in South Korea. *Int J Tuberc Lung Dis*. 2012;16(1):98–103.
28. Whitfield MG, Soeters HM, Warren RM, et al. A global perspective on pyrazinamide resistance: systematic review and meta-analysis. *PLoS One*. 2015;10(7):e0133869.
29. Comas I, Borrell S, Roetzer A, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet*. 2011;44(1):106–110.
30. Yang C, Luo T, Shen X, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect Dis*. 2017;17(3):275–284.

Infection and Drug Resistance

Publish your work in this journal

Infection and Drug Resistance is an international, peer-reviewed open-access journal that focuses on the optimal treatment of infection (bacterial, fungal and viral) and the development and institution of preventive strategies to minimize the development and spread of resistance. The journal is specifically concerned with the epidemiology of antibiotic

resistance and the mechanisms of resistance development and diffusion in both hospitals and the community. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/infection-and-drug-resistance-journal>

Dovepress