Ali, A; Hasan, Z; McNerney, R; Mallard, K; Hill-Cawthorne, G; Coll, F; Nair, M; Pain, A; Clark, TG; Hasan, R (2015) Whole Genome Sequencing Based Characterization of Extensively Drug-Resistant Mycobacterium tuberculosis Isolates from Pakistan. PloS one, 10 (2). e0117771. ISSN 1932-6203 DOI: https://doi.org/10.1371/journal.pone.0117771

Downloaded from: http://researchonline.lshtm.ac.uk/2121482/

DOI: 10.1371/journal.pone.0117771

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
RESEARCH ARTICLE

Whole Genome Sequencing Based Characterization of Extensively Drug-Resistant Mycobacterium tuberculosis Isolates from Pakistan

Asho Ali1,2‡, Zahra Hasan1,*‡, Ruth McNerney3, Kim Mallard3, Grant Hill-Cawthorne4,5, Francesc Coll3, Mridul Nair4, Arnab Pain4, Taane G. Clark3, Rumina Hasan1

1 Department of Pathology and Microbiology, Aga Khan University Hospital, Karachi, Pakistan, 2 School of Nursing and Midwifery, The Aga Khan University Hospital, Karachi, Pakistan, 3 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, 4 Pathogen Genomics Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia, 5 Sydney Emerging Infections and Biosecurity Institute and School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia

‡ These authors are joint first authors on this work.
* zahra.hasan@aku.edu

Abstract

Improved molecular diagnostic methods for detection drug resistance in Mycobacterium tuberculosis (MTB) strains are required. Resistance to first- and second-line anti-tuberculous drugs has been associated with single nucleotide polymorphisms (SNPs) in particular genes. However, these SNPs can vary between MTB lineages therefore local data is required to describe different strain populations. We used whole genome sequencing (WGS) to characterize 37 extensively drug-resistant (XDR) MTB isolates from Pakistan and investigated 40 genes associated with drug resistance. Rifampicin resistance was attributable to SNPs in the \( rpoB \) hot-spot region. Isoniazid resistance was most commonly associated with the \( katG \) codon 315 (92%) mutation followed by \( inhA \) S94A (8%) however, one strain did not have SNPs in \( katG \), \( inhA \) or \( oxyR-ahpC \). All strains were pyraziamide resistant but only 43% had \( pncA \) SNPs. Ethambutol resistant strains predominantly had \( embB \) codon 306 (62%) mutations, but additional SNPs at \( embB \) codons 406, 378 and 328 were also present. Fluoroquinolone resistance was associated with \( gyrA \)9 1–94 codons in 81% of strains; four strains had only \( gyrA \) mutations, while others did not have SNPs in either \( gyrA \) or \( gyrB \). Streptomycin resistant strains had mutations in ribosomal RNA genes; \( rpsL \) codon 43 (42%); \( rrs \) 500 region (16%), and \( gidB \) (34%) while six strains did not have mutations in any of these genes. Amikacin/kanamycin/capreomycin resistance was associated with SNPs in \( rrs \) at nt1401 (78%) and nt1484 (3%), except in seven (19%) strains. We estimate that if only the common hot-spot region targets of current commercial assays were used, the concordance between phenotypic and genotypic testing for these XDR strains would vary between rifampicin (100%), isoniazid (92%), fluoroquinolones (81%), aminoglycoside (78%) and ethambutol (62%); while \( pncA \) sequencing would provide genotypic resistance...
in less than half the isolates. This work highlights the importance of expanded targets for drug resistance detection in MTB isolates.

Introduction

*Mycobacterium tuberculosis* (MTB) caused 9 million cases of tuberculosis (TB) and 1.5 million death worldwide in 2013 [1]. Drug resistance in MTB increases the burden due to the disease and complicates both diagnosis and treatment. Multidrug-resistant (MDR)-TB is caused by MTB resistant to at least the two first line drugs i.e. rifampicin and isoniazid. Globally, in 2013, an estimated 480,000 people developed TB while, there were about 210,000 deaths due to TB [1]. Extensively drug resistant (XDR)-TB is caused by MDR-TB strains resistant to fluoroquinolone and to any one of the injectable aminoglycosides; amikacin, kanamycin or capreomycin. In Pakistan, MDR-TB comprises 4–5% of all TB cases, and XDR-TB is 4–5% of the MDR-TB cases [2,3]. Appropriate MDR-TB management requires drug susceptibility testing to guide treatment regimens. The usual turnaround time for phenotypic MTB drug susceptibility testing (DST) is about 8–10 weeks, often leading to delays in initiation of appropriate therapy. In high TB burden low resource settings facilities for mycobacterial culture and DST are limited. Therefore, rapid genotypic methods for detection of drug resistant MTB strains are required for more effective management of TB.

The Xpert MTB/RIF assay (Cepheid, USA) assay was rolled out at a global level since 2012; it detects rifampicin resistance by targeting the ribosomal polymerase *rpoB* gene [4]. The LiPA MTBDRplus (Hain Lifescience GmbH, Germany) assay targets *rpoB* for rifampicin resistance and *katG* and *inhA* genes for isoniazid resistance. The LiPA MTBDRsl (Hain Lifescience GmbH, Germany) targets *gyrA*, *rrs* and *embB* for fluoroquinolone, aminoglycoside and ethambutol resistance respectively [5]. Whilst these assays have significant value and allow rapid diagnosis of MDR-TB it is apparent that further improvements in molecular DSTs are required. MTB is largely clonal [6,7]. The TBDReaM database (http://www.tbdreamdb.com), catalogues greater than 1400 SNPs associated with drug resistance [8]. Studies of drug resistant MTB have shown that mutations in hot-spot regions, and those occurring beyond hypervariable regions vary according to both strain type and geographical location [9,10,11]. Thus, it is important for all possible resistance conferring mutations or single nucleotide polymorphisms (SNPs) related to each anti-tuberculous drug to be identified. Whole genome sequencing (WGS) facilitates the identification of SNPs, insertions and deletions in genomic DNA [12]. Thus, WGS of MTB can reveal new SNPs in genes previously associated with drug resistance and also identify outbreak strains [13].

By providing information on multiple genes simultaneously WGS facilitates the understanding of relationships between genes and their interactions such as in the case of fitness conferring compensatory mutations which occur due to the presence of drug induced selection pressure. Compensatory mutations are believed to play a role in conferring fitness to drug resistant MTB strains by affecting the growth rate of the isolates [14,15,16]. Compensatory mutations in *rpoB* genes which aid fitness of rifampicin resistant strains are well documented [17]. Knowledge of additional resistance mutation and compensatory mutations in MTB will be important for determining appropriate treatment for MTB strains. Here we have used WGS analysis to describe XDR-TB strains from Pakistan and identify SNPs in genes known to be associated with drug resistance to particular drugs.
Materials and Methods

Ethics statement

This study was approved by the Ethical Review Committee, The Aga Khan University.

Strain selection

The *M. tuberculosis* strains studied were those collected from 2004–2009. Mycobacterial strains were obtained from the strain bank of Aga Khan University Clinical Microbiology laboratory. Demographic information on patients was obtained from clinical laboratory records and collected as part of good clinical practice. The hospital and its clinical laboratory are accredited by the Joint Commission of International Accreditation (JCIA, USA). MTB isolates were identified through a random selection method. Each strain had its unique identification number which at the time of testing was anonymised and unlinked with patient demographic information. The XDR-TB (n = 37) and drug susceptible (n = 5) isolates studied here were spoligotyped previously [3].

Spoligotyping

Spoligotype patterns for each strain were obtained as described previously [18], and also confirmed *in silico* from the sequencing reads using SpolPred software [19].

Culture and drug susceptibility testing

*M. tuberculosis* strains were isolated from Lowenstein-Jensen media and MGIT (Becton Dickinson, Franklin Lakes, NJ, and USA). *M. tuberculosis* was identified using BACTEC NAP TB differentiation test (Becton Dickinson), growth on para-nitrobenzoic acid containing media, nitrate reduction, and niacin accumulation [20]. DST of these isolates was performed using an agar proportion method on enriched Middlebrook 7H10 medium (BBL Microbiology Systems, Cockeysville, MD, USA) at the following concentrations: rifampicin 1 μg/mL, isoniazid 0.2 μg/mL, streptomycin 2 μg/mL, and ethambutol 5 μg/mL. Pyrazinamide sensitivity was determined by using BACTEC 7H12 medium, pH 6.0, at 100 μg/mL (BACTEC PZA test medium, Becton Dickinson). MDR TB strains were further tested with ciprofloxacin 2 μg/mL, capreomycin 10 μg/mL, amikacin 5μg/mL, kanamycin 6μg/mL and ethionamide 5 μg/mL. The reference strain *M. tuberculosis* H37Rv was used as a control with each susceptibility testing batch [21].

DNA extraction and Whole Genome Sequencing of MTB isolates

DNA was extracted by the cetyl-trimethyl ammonium bromide (CTAB) method [22]. One microgram of DNA was used for sequencing. All samples underwent WGS with 76-base paired end fragment sizes, using Illumina paired end HiSeq2000 technology, and the raw sequence data is available in the European nucleotide archive (http://www.ebi.ac.uk/ena/data/view/PRJEB7798). For each sample trimmomatic software (http://www.usadellab.org/cms/?page=trimmomatic) was used to remove low quality reads and trim low-quality 3’ ends of reads. Nucleotide positions in the reads with a quality score lower than Q20 were removed. High quality reads were then mapped to the H37Rv reference genome (Genbank accession: AL123456.3) using BWA-MEM software (http://bio-bwa.sourceforge.net). SAMtools (http://samtools.sourceforge.net) and GATK (https://www.broadinstitute.org/gatk) were used to call single nucleotide polymorphisms (SNPs) and small indels. Variants (with quality at least Q30) were then selected as the intersection dataset between those obtained from both programs. Sample genotypes were called using the majority allele (minimum frequency 75%) in positions supported by at least 20-fold total coverage; otherwise they were classified as missing. Samples
with a proportion of missing genotype calls greater than 15% were filtered out [21]. Similarly, we excluded positions in the genome with more than 15% missing genotypes across samples. Larger indels were called using a consensus from paired end mapping distance or split read approaches [23,24,25,26]. For more detail, the data processing pipeline has been described [27]. Spoligotypes were inferred from the sequencing reads SpolPred software [19]. Variation density maps were generated using Circos software (http://www.circos.com).

**Single nucleotide polymorphisms (SNPs)**

The SNP data obtained was analysed according to known resistance conferring genes for each anti-tuberculous drug. For rifampicin resistance, rpoB, rpoA and rpoC were analysed. For isoniazid resistance, katG, inhA, oxyR-ahpC, fabC, Rv1592C, Rv1772, fabD, kasA, accD, oxyR, ndh, fadE24, and nat were investigated. For pyrazinamide resistance, the pncA and rpsA genes were studied. For ethambutol resistance, embC, embA, embB, embR, iniA, iniB, iniC, Rv3124, manB, PPE49 and rmlD genes were studied. For streptomycin resistance, rpsL, rrs and gidB genes were investigated. For fluoroquinolone resistance, gyrA and gyrB genes were studied. For aminoglycoside resistance, rrs and tlyA, and for ethionamide resistance ethA and fabG1 genes were analysed. SNPs were divided into synonymous (sSNPs) and non-synonymous (ns-SNPs) mutations. Each SNP was compared with that listed in TBDReaMDB (http://www.tbdreamdb.com) database and other published reports of drug resistant MTB genotypes. For each strain, genotypic drug resistance findings were compared with phenotypic drug susceptibility results.

**Phylogenetic analysis**

Phylogenetic analysis was performed in PHYLP using a RAxML tree (Maximum likelihood phylogenetic software) with bootstrap to 1000 replicated [28]. Phylogenetic tree was visualized using Dendroscope software [29].

**Results**

**Demographic information, phenotypic drug susceptibility and genotypic profile**

The XDR strains comprised mainly CAS family isolates: CAS1-Delhi (n = 20), CAS (n = 2) and CAS2 (n = 1). Other strains belonged to EAT3-IND (n = 2), X3 (n = 1), T1 (n = 3) and Orphan (n = 8) lineages. Drug susceptible isolates were CAS1-Delhi (n = 2) and CAS (n = 3) strains.

The XDR strains were from the four provinces of Pakistan; Sindh (n = 21), Punjab (n = 16), Khyber Pakhtunkhwa Province (n = 4) and Baluchistan (n = 1), S1 Dataset. Sixty-two percent (n = 26) of MTB isolates were from males and 32% (n = 16) from females. The mean and median ages of patients was 34.2 and 31 years respectively (males: mean, 36.2, median, 34.5; female: mean 31.1, median, 26 years).

Isolates from Sindh comprised Central Asian Strain (CAS), T1, East African Indian (EAI), X and Orphan types and thus had more diverse lineages as compared to those from other provinces. However, within each province the CAS lineage was predominant, as described previously [30].

Shared type spoligotypes and unique MTB isolates were linked to their respective principal genetic group (PGG) based on the combinations of katG463-gyrA95 polymorphism [31], S1 Dataset and S2 Dataset. Most of the strain types (90.5%, 38/42) except, X3 (n = 1) and T1 (n = 3) belonged to the PGG1 lineage (based on katG codon 463 CTG (Leu) and gyrA codon 95 ACC (Thr) polymorphisms). PGG1 strains included CAS, EAI and Orphan strains. One,
X3 isolate was grouped in PGG2 as it showed katG 463 CGG (Arg) and gyrA codon 95 ACC (Thr) polymorphism. Three T1 isolates with katG 463 CGG (Arg) and gyrA codon 95 AGC (Ser) mutations were grouped in PGG3.

Polymorphisms in genes associated with drug resistances

A visual description of SNPs densities throughout the genome of MTB isolates was compiled by a circos plot (Fig. 1). The maximum likelihood analysis is demonstrated by the circular phylogenetic tree based on the polymorphism in the drug resistance genes using Dendroscope software version 3.2.8 [32], (Fig. 2). The clustering pattern of the strains demonstrated that T1 and X lineage strains clustered together (PGG3 and PGG2 groups respectively), and were distinct from CAS and EAI strains which clustered in a similar node. However, EAI remained distant from other PGG1 isolates in the phylogenetic tree.

By comparison with the H37Rv reference genome, SNPs were identified in 40 genes associated with drug resistance (Fig. 3). All synonymous SNPs (sSNPs) were removed from the analysis so that only non-synonymous SNPs (nsSNPs) are displayed in S2 Dataset. This nsSNP dataset is further described in Tables 1-5, which focus on the known resistance-determining regions in particular MTB genes.

Polymorphisms conferring rifampicin and isoniazid resistance

Resistance to rifampicin is associated with mutations in the RNA polymerase gene sub-unit B, rpoB. However, there is a fitness cost associated with the acquisition of rpoB mutation and rifampicin resistant MTB strains often acquire compensatory mutations in rpoC and rpoA genes to improve growth [33]. Similarly, there is a fitness cost associated with acquisition of isoniazid resistance, and while katG followed by the inhA gene mutations are most common, a number of other genes are thought to contribute to resistance [34]. We first evaluated SNPs in the context of MDR by studying the rifampicin-resistance determining region (RRDR) in rpoB, and compensatory mutations in rpoC and rpoA. For isoniazid resistance we analysed the katG and inhA genes, and also investigated sequences of oxyR-ahpC, fabC, Rv1592C, ndh, Rv2242, fabD, accD, fadE24 and nat. Combinations of mutations observed in the XDR strains related to rifampicin and isoniazid are illustrated in Table 1. All the XDR isolates exhibited mutations in the RRDR of rpoB gene; codons 531 (n = 24), 516 (n = 12) and 526 (n = 1). The rpoB codon 531 mutation was present in PGG1, PGG2 and PGG3 isolates whereas, a mutation at codon 526 was observed in one PGG1 group EAI3-IND isolate. Mutations at codons 515 and 516 were present concurrently in each case, in CAS and Orphan strains.

Twenty-two XDR-TB isolates with compensatory mutations in rpoC and rpoA genes had the low fitness cost mutation rpoB S531L. One isolate with the rpoB H526R mutation had two rpoC SNPs. In all, we observed 15 different rpoC SNPs and one rpoA SNP. The most common was rpoC A172V which present in 5 isolates and rpoA P601L which was present in 3 isolates.

Polymorphisms at codon 463 of katG were observed among all the PGG1 isolates. Resistance to isoniazid could be attributed to the katG codon in 34/37 of cases and most strains (86%) had low fitness cost mutation, katG S315T while, two strains had katG S315N. Two strains had an inhA S94A mutation in the absence of katG mutations. We did not observe any inhA promoter mutations amongst the XDR isolates studied. One EAI3-lineage strain had a katG Y413H change.

In summary, if rifampicin and isoniazid resistance were tested in these XDR strains using only MDR hot-spot region based targets (rpoB, katG and -15 region inhA), 100% of rifampicin and 89% of isoniazid resistance would be detected.
Fig 1. Comparative drug resistance genes information of *Mycobacterium tuberculosis* study isolates. Circos plot of *Mycobacterium tuberculosis* highlighting the genomic regions (Rv) where drug resistance associated SNPs were identified (colour coded in blue). The 5 inwards tracks represent SNP density, Non-synonymous/synonymous densities, small insertion/deletions densities, large deletion density and uniqueness track.

doi:10.1371/journal.pone.0117771.g001
When additional isoniazid resistance associated genes were analysed we found the oxyR g142a nucleotide substitution to be present in 4 strains together with a katG S315T mutation (S2 Dataset). The EAI3 lineage strain (with katG Y413H) had the mutations oxyR g38a, accD M439I and A142V; in addition to rpoB H526R, and rpoC A172V and P601L mutations. We also observed additional SNPs present only in resistant strains with a katG S315T mutation; Rv1592C (G9D, n = 4), ndh L104F and E360K, Rv2242 (A156V), fabD (A3T), accD (E224D and A243V), fadE24 (Q85R), nat (T175A).

Polymorphisms conferring resistance to pyrazinamide
All of the XDR strains were resistant to pyrazinamide treatment (Table 1). Non-synonymous SNPs in the pncA gene were observed in 14 (38%) strains, whereby 11 different SNPs were observed (S2 Dataset). Twenty-three pyrazinamide resistant isolates did not reveal any nsSNP in pncA; two such strains had the rpsA mutation Q1228R.

Both pyrazinamide and isoniazid are pro-drugs which are converted to their active forms by enzymatic activity within the mycobacterium. We explore an association between isoniazid and pyrazinamide resistance by comparing katG and pncA mutations within strains. Of the XDR strains, thirteen had pncA and katG S315T mutations, while one strain with a pncA mutation had inhA S94A (Table 1).

Polymorphisms in the genes for ethambutol resistance
All of the XDR strains were resistant to ethambutol. The ethambutol-resistance determining region (ERDR) is primarily in the embB gene and we found that 23 strains to have embB codon 306 mutations, where M306I was most common (Table 2). Based on this data, commercial
molecular diagnostic assays targeting \textit{embB} codon 306 would detect ethambutol resistance in 62\% of the XDR strains studied here.

Other ethambutol resistant strains had SNPs at \textit{embB} codon 1024 (n = 4), codon 497 (n = 3), codon 406 (n = 1), codon 378 (n = 3), codon 328 (n = 1) and codon 72 (n = 1), S2 Dataset.

We analyzed additional genes associated with ethambutol resistance; \textit{embC}, \textit{embA}, Rv3124, \textit{rmlD}, \textit{iniA}, \textit{iniB}, \textit{iniC} and PPE49. Different combinations of SNPs were observed in these ethambutol-resistance associated genes possibly, as compensatory mutations and or lineage associated mutations. Two EAI3-IND isolates (X5 and X32) had a distinct combination of mutations; \textit{embB} M306V, E378A and G406A, \textit{embA} P913S, \textit{embC} T270I, N394E, \textit{rmlD} S257P and \textit{iniA} H481Q (S2 Dataset). Two Orphan isolates (X17 and X57) also had a distinct combination of mutations; \textit{embB} E378A, \textit{embA} P913S, \textit{embC} T270I, N394E, \textit{rmlD} S257P, \textit{iniA} H481Q, \textit{iniC} R481C and PPE49 M67L. Another isolate had a SNP in Rv3124 (V273I) (S2 Dataset) while one strain displayed a single nucleotide (T) deletion at codon 173 of \textit{iniA} gene at 519 position.

Polymorphisms conferring resistance to streptomycin

The mechanism of streptomycin resistance in the XDR strains was investigated by analyzing sequences of the ribosomal genes \textit{rpsL}, \textit{rrs} (500 region) and \textit{gidB}. The \textit{rpsL} mutation K43R was found in 41\% of streptomycin resistant isolates; this SNP was found independently (n = 11) and also in combination with \textit{rrs} A514C (n = 2), Table 3. The \textit{rpsL} mutation K88R was found in combination with \textit{rrs} A906G and \textit{gidB} P75S. Three strains had \textit{rrs} A514C mutations only. There were eight different SNPs in \textit{gidB}. Eight streptomycin resistant strains had \textit{gidB}
mutations only, without rpsL and rrs (500 region) SNPs. However, six streptomycin resistant and five sensitive strains were without mutations in rrs, rpsL or gidB (S2 Dataset).

Polymorphisms in the genes for fluoroquinolone resistance

Mutations in gyrA and gyrB genes were analysed to investigate fluoroquinolone resistance in the XDR strains. The most common mutation in the quinolone-resistance determining region (QRDR) of gyrA gene covering codons 90–98 was A94G, followed by A90V and A94Y (Table 4). All CAS family strains had gyrA QRDR mutations. One X3 strain had gyrA D94Y together with gyrB T500N. In total, 81% of isolates had SNPs in the QRDR of gyrA, which is the target region in currently available commercial diagnostic assays for fluoroquinolone resistance.
Polymorphism at codon 95 of gyrA was observed among PGG1 and PGG2 isolates (S2 Dataset). The gyrA, G668D polymorphism was present in PGG3 (n = 3) isolates belonging to T1 lineage.

Of the eight strains which did not have gyrA QRDR SNPs, four had SNPs in the gyrB gene. One T1 strain only had the gyrB S447F mutation, two EAI3 strains only had the M291I mutation and an Orphan isolate had a gyrB M291I and A432V mutations. In one strain, a three nucleotide deletion was observed at codon 683 of gyrB gene at 7158 position. In all, four CAS1-Delhi XDR strains without gyrA or gyrB mutations.

Table 2. SNPs in genes conferring resistance ethambutol.

<table>
<thead>
<tr>
<th>Lineage of XDR isolate</th>
<th>Phenotype*</th>
<th>embB</th>
<th>No. of specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS1-Delhi (9), Orphan (3), T1 (1)</td>
<td>R</td>
<td>M306I</td>
<td>13</td>
</tr>
<tr>
<td>CAS1-Delhi (5), T1 (2)</td>
<td>R</td>
<td>M306V</td>
<td>7</td>
</tr>
<tr>
<td>Cas1-Delhi</td>
<td>R</td>
<td>M306L</td>
<td>1</td>
</tr>
<tr>
<td>CAS2</td>
<td>R</td>
<td>D354A, Q497P</td>
<td>1</td>
</tr>
<tr>
<td>X3</td>
<td>R</td>
<td>D328Y</td>
<td>1</td>
</tr>
<tr>
<td>Orphan</td>
<td>R</td>
<td>E378A</td>
<td>3</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>G406A</td>
<td>1</td>
</tr>
<tr>
<td>CAS</td>
<td>R</td>
<td>Q497K</td>
<td>2</td>
</tr>
<tr>
<td>Orphan</td>
<td>R</td>
<td>I72L #</td>
<td>1</td>
</tr>
<tr>
<td>CAS1-Delhi (3), Orphan (1)</td>
<td>R</td>
<td>D1024T #</td>
<td>4</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>wt #</td>
<td>1</td>
</tr>
</tbody>
</table>

*+* strains were tested against resistant to ethambutol (5 μg/ml).

‘#’ mutation outside the embB ethambutol resistance determining region (ERDR), the number in parenthesis indicates the number of MTB isolates from each lineage

doi:10.1371/journal.pone.0117771.t002

Table 3. Combination of SNPs in genes conferring resistance to streptomycin.

<table>
<thead>
<tr>
<th>Lineage of XDR isolates</th>
<th>Phenotype*</th>
<th>rpsL</th>
<th>rrs (500 region)</th>
<th>gidB</th>
<th>Number of specimens</th>
<th>Percentage of resistant isolates (n = 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS1-Delhi (5), T1 (3), Orphan (3)</td>
<td>R</td>
<td>K43R</td>
<td></td>
<td></td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>K43R</td>
<td>A514C</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Orphan</td>
<td>R</td>
<td>A514C</td>
<td></td>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>A514T</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>K43R</td>
<td>A514C</td>
<td>Q87 Stop</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>A906G</td>
<td>W148 Stop</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS</td>
<td>R</td>
<td>K88R</td>
<td>A906G</td>
<td>P75S</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>G71 Stop</td>
<td></td>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CAS (1), X3 (1)</td>
<td>R</td>
<td>P75S</td>
<td></td>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>R</td>
<td>W45 Stop</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Orphan</td>
<td>R</td>
<td>A82P</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS2</td>
<td>R</td>
<td>A119D</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CAS1-Delhi (6), EA13-IND (1)</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>CAS1-Delhi (2), Orphan (2), EA13-IND (1)</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*+* strains resistant to streptomycin (STR, 2 μg/ml), the number in parenthesis indicates the number of MTB isolates from each lineage

doi:10.1371/journal.pone.0117771.t003
To investigate the mechanism for aminoglycoside resistance, the codon 1400 region of the \textit{rrs} gene was analysed. Thirteen XDR strains were resistant to all three aminoglycosides (amikacin, kanamycin and capreomycin). Twenty-four XDR strains were resistant to amikacin and kanamycin but sensitive to capreomycin. The most common \textit{rrs} mutation across lineages was A1401G. Of strains resistant to amikacin, kanamycin and capreomycin, eleven had the SNP \textit{rrs} A1401G and one had G1484T, whilst 2 strains had no mutation in the 1400 region of \textit{rrs} (Table 5). Of the strains resistant to amikacin and kanamycin, eighteen had the SNP A1401G whilst five did not have a SNP in the \textit{rrs} 1400 region. As current commercial diagnostic assays for aminoglycoside resistance target \textit{rrs} nt 1401, 78\% of resistance would be detected in these XDR isolates.

In all, 6 aminoglycoside resistant isolates did not reveal any polymorphism in hyper-variable 1400 region of \textit{rrs} gene (S2 Dataset).

### Polymorphisms conferring resistance to aminoglycosides

To investigate the mechanism for aminoglycoside resistance, the codon 1400 region of the \textit{rrs} gene was analysed. Thirteen XDR strains were resistant to all three aminoglycosides (amikacin, kanamycin and capreomycin). Twenty-four XDR strains were resistant to amikacin and kanamycin but sensitive to capreomycin. The most common \textit{rrs} mutation across lineages was A1401G. Of strains resistant to amikacin, kanamycin and capreomycin, eleven had the SNP \textit{rrs} A1401G and one had G1484T, whilst 2 strains had no mutation in the 1400 region of \textit{rrs} (Table 5). Of the strains resistant to amikacin and kanamycin, eighteen had the SNP A1401G whilst five did not have a SNP in the \textit{rrs} 1400 region. As current commercial diagnostic assays for aminoglycoside resistance target \textit{rrs} nt 1401, 78\% of resistance would be detected in these XDR isolates.

In all, 6 aminoglycoside resistant isolates did not reveal any polymorphism in hyper-variable 1400 region of \textit{rrs} gene (S2 Dataset).

### Polymorphisms in the genes for ethionamide resistance

Five XDR isolates resistant to ethionamide but none showed any polymorphism in \textit{ethA} gene, but one ethionamide resistant isolate had the mutation \textit{ethA} R239G which was also present in an ethionamide sensitive isolate (S2 Dataset).

### Table 4. Combination of SNPs in genes conferring resistance to fluoroquinolone.

<table>
<thead>
<tr>
<th>Lineage of XDR isolates</th>
<th>\textit{gyrA}\textsuperscript{*}</th>
<th>\textit{gyrB}</th>
<th>Number of specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS1-Delhi (9), CAS2 (1)</td>
<td>D94G</td>
<td>M291I</td>
<td>10</td>
</tr>
<tr>
<td>Orphan</td>
<td>D94G</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Orphan (5), CAS1-Delhi (3), T1 (1)</td>
<td>A90V</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>CAS1-Delhi (2), CAS (2)</td>
<td>D94Y</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>A90V, D94Y</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td>S91P</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>T1</td>
<td>D94N</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>X3</td>
<td>D94Y</td>
<td>T500N</td>
<td>1</td>
</tr>
<tr>
<td>T1</td>
<td>S447F</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>EAI3-IND</td>
<td>M291I</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Orphan</td>
<td>M291I, A432V</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CAS1-Delhi</td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

All strains were resistant to ciprofloxacin (2\(\mu\)g/ml) and ofloxacin (2 \(\mu\)g/ml). \textsuperscript{*}mutations in the \textit{gyrA} quinolone- resistance determining region (QRDR), the number in parenthesis indicates the number of MTB isolates from each lineage.

### Table 5. SNPs present in XDR strains related to aminoglycoside resistance.

<table>
<thead>
<tr>
<th>Phenotypic data*</th>
<th>Genotypes</th>
<th>\textit{rrs} (1400 region)</th>
<th>Number of specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>R to AMK, KAN, CAP</td>
<td>CAS1-Delhi (6), T1 (2), Orphan (2), X3 (1)</td>
<td>A1401G</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Orphan</td>
<td>G1484T</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>EAI3-IND (1), CAS1-Delhi (1)</td>
<td>wt</td>
<td>2</td>
</tr>
<tr>
<td>R to AMK, KAN</td>
<td>CAS2(1), CAS1-Delhi (11), CAS (2), EAI3-IND (1), Orphan (3)</td>
<td>A1401G</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>CAS1-Delhi (3), Orphan (1), T1(1)</td>
<td>wt</td>
<td>5</td>
</tr>
</tbody>
</table>

\textsuperscript{*}resistance determined at amikacin (AMK, 6 \(\mu\)g/ml), kanamycin (KAN, 6 \(\mu\)g/ml) and capreomycin (CAP, 6 \(\mu\)g/ml). ‘wt’ wild type indicates absence of SNPs in the \textit{rrs} 1400 gene, the number in parenthesis indicates the number of MTB isolates from each lineage.
Discussion

Our results present the first WGS based characterization of XDR-TB strains from Pakistan. We focused on SNPs in genes known to confer resistance to first- and second-line drugs and present data relevant to molecular diagnostic assays for drug susceptibility testing.

Consistent with published reports all the rifampicin resistant MTB isolates in this study revealed at least one \textit{rpoB} gene codon mutation i.e. 531, 526, 516 or 515, in the RRDR region [35,36,37,38]. The most common mutation was the \textit{rpoB} S531L, corroborating with previous reports [33]. In 22/23 isolates, compensatory mutations in \textit{rpoC} and \textit{rpoA} genes were coincident with the low fitness cost \textit{rpoB}531 mutation, supporting previous reports which demonstrate increased fitness in rifampicin resistant isolates as a consequence of these compensatory mutations [39,40,41]. One EAI lineage isolate had the \textit{rpoB} H526R mutation together with two \textit{rpoC} mutations. As the \textit{rpoB} codon 526 mutation has been shown to have a higher fitness cost than the codon 531 mutation [17] therefore, it would be an advantage for such an isolate to accumulate the \textit{rpoC} compensatory mutations.

Based on the \textit{katG}463-gyrA95 polymorphism 91% of the MTB isolates belonged to PGG1 of which the majority were CAS family isolates, as previously reported [30]. PGG1 is considered the oldest of MTB lineages and has been previously been reported from Asia [42]. The predominance of CAS strains suggests that they are well adapted to human populations in this region [43,44,45,46].

The most frequently observed isoniazid resistance associated mutation observed in our strains was \textit{katG} S315T, which has previously been shown to be associated with a growth advantage in strains [47]. Three isolates with \textit{inhA} codon 94 mutations did not reveal \textit{katG}315 mutations, supporting a causal relationship of this mutation for isoniazid resistance [48,49].

Pyrazinamide is important as a first-line anti-tuberculous drug but phenotypic and genotypic discordance exists regarding testing its drug susceptibility. Consistent with previous reports, we found that more than 50% of resistant isolates did not reveal any mutation in \textit{pncA} gene [50]. Mutations in \textit{pncA} gene are thought to be the primary mechanism for pyrazinamide resistance, but our data consistent with others indicates that other resistance mechanisms may exist [51]. We found \textit{rpsA} mutations in two XDR isolates which did not have \textit{pncA} mutation, suggesting that this gene may play a role in resistance to pyrazinamide [52]. Additional pyrazinamide resistant strains had neither \textit{pncA} nor \textit{rpsA} mutations, and further supports the suggestion that mechanisms such an efflux pump for \textit{pncA} wild type resistant MTB isolates may be relevant [53].

The \textit{embB}306 SNP was found to be the most common mutation associated with ethambutol resistance in MTB strains, as reported previously [54]. The \textit{embB}306 mutation has been shown to be associated with high-level ethambutol resistance [55]. In addition, we also observed \textit{embB} codon 378, 406 and 497 mutations; which have reported previously [56,55,57]. Our observations of distinct combinations of \textit{embB} and compensatory mutations in XDR isolates of EAI-IND and Orphan types suggest that some of these may lineage associated mutations. Especially, we observed a particular pattern of \textit{embB}, \textit{embA}, \textit{embC}, \textit{rmlD} and \textit{ini} mutations was in four Orphan types, of which three were resistant and one was sensitive to ethambutol.

The most common streptomycin resistance associated mutation we observed was \textit{rpsL} K43R. This mutation has been shown to play a role in reduced fitness cost and enhanced growth rate of resistant MTB isolates [47]. The K43R mutation has also been associated with high-level streptomycin resistance [58]. In the \textit{rrs} gene we identified point mutations at positions 514 and 906, encoding for streptomycin resistance [58,59]. Previous studies have reported role of \textit{gidB} polymorphism for low level of streptomycin resistance in MTB [60]. Here we found eight different \textit{gidB} SNPs. Three strains had compensatory \textit{gidB} SNPs with \textit{rpsL} and/or \textit{rrs} mutations. Six streptomycin resistant isolates had \textit{gidB} SNPs but did not have either \textit{rpsL} or...
rrs gene mutations. The occurrence of gidB mutations in strains reduces drug pressure for selection of highly resistant isolates, therefore it has been suggested that high-level streptomycin resistant strains (e.g. rpsL mutants) arise more frequently in gidB mutants than in wild-type cells [61]. We observed six streptomycin resistant strains without mutations in rpsL, rrs or gidB. The aminoglycoside antibiotics inhibit protein translation therefore it is likely that in additional mutations may be relevant for inactivating the effect of streptomycin in inhibiting tRNA translocation in mycobacteria.

In the strains resistant to amikacin, kanamycin and capreomycin we found 16S rRNA gene rrs1401 and 1484 mutations to be the most common. These rrs mutations have previously been associated kanamycin and capreomycin [39,56,62]. In addition, it has been shown that the tlyA and rrs gene mutations jointly play a role for capreomycin resistance [63]. However, we did not observe any tlyA mutations amongst our strains.

We found that across MTB lineages, the most common fluoroquinolone resistance associated mutation was at gyrA codon 94, as shown previously [56,64]. Lineages other than CAS (X3, T1, EAI and Orphan) had gyrB SNPs in the 400 region. The gyrB gene has previously been shown to result in fluoroquinolone resistance [65]. Four XDR strains did not have gyrA or gyrB mutations. This may be explained by the fact that additional mechanisms such as efflux pumps may be involved in a generating resistance fluoroquinolone [66]. The predominant lineage amongst our XDR isolates was CAS and there is less representation of the other strain types. However, this data is suggestive of the association of fluoroquinolone resistance in CAS lineages with gyrA QRDR region.

In the case of ethionamide, we found that none of the isolates had ethA SNPs which were particular to ethionamide resistance, as one isolate had ethA R239G mutation which was also present in an ethionamide sensitive isolate. A low frequency of mutation in ethA gene of ethambutol resistant isolates have also been reported earlier and may suggest existence of some other mechanisms such as efflux pump for ethionamide resistance [56,67].

Overall, the most frequently seen drug resistance mutations observed for isoniazid, rifampicin, streptomycin, ethambutol, aminoglycosides and ofloxacin were katG315, rpoB531, rpsL43, embB306, rrs1401 and gyrA94 respectively as previously reported [35,36,37,38,55,56,68,69]. Currently available commercial assays for detection of drug resistance (MTBDRplus and MTBDRsl) target the rpoB531, 516 and 526, katG315, gyrA94, 91 and 90 and rrs1401 regions. Based on our data the frequency of drug resistance detection using such assays for these strains would have been 100% for rifampicin resistance, 92% for isoniazid resistance, 81% for fluoroquinolone resistance, 78% for aminoglycoside and 62% for ethambutol resistance. Our data indicates that SNPs in drug resistance genes vary with different MTB lineages. Therefore, it highlights the relevance for WGS data from prevalent global strain types in order to develop improved molecular diagnostic assays for resistance in MTB. Additional knowlegde of mutations that are compensatory to those that confer drug resistance will be important for prediction of levels of drug resistance and therefore, to guide treatment of drug resistance isolates.

Supporting Information

S1 Dataset. Drug susceptibility and demographic information of Mycobacterium tuberculosis isolates.

S2 Dataset. An overview of SNPs observed in genes associated with drug resistance in XDR TB isolates.
Acknowledgments

We thank the Aga Khan University Hospital Clinical Laboratory and its staff for providing samples used in the study. This work was made possible through the generous support for whole genome sequencing provided by Dr. Arnab Pain at KAUST, Saudi Arabia.

Author Contributions

Conceived and designed the experiments: RH RM ZH. Performed the experiments: AA GHC MN. Analyzed the data: FC KM TGC AA ZH. Contributed reagents/materials/analysis tools: RM AP FC KM TGC. Wrote the paper: RH ZH AA TGC KM.

References


