MICROBIAL GENOMICS

RESEARCH ARTICLE

Verma *et al.*, *Microbial Genomics* 2021;7:000542

DOI 10.1099/mgen.0.000542

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Genome analyses of 174 strains of *Mycobacterium tuberculosis* provide insight into the evolution of drug resistance and reveal potential drug targets

Helianthous Verma[1,2], Shekhar Nagar[3], Shivani Vohra[2,4], Shubhanshu Pandey[2,5], Devi Lal[2], Ram Krishan Negi[3], Rup Lal[6] and Charu Dogra Rawat[1,2,*]

Abstract

Mycobacterium tuberculosis is a known human pathogen that causes the airborne infectious disease tuberculosis (TB). Every year TB infects millions of people worldwide. The emergence of multi-drug resistant (MDR), extensively drug resistant (XDR) and totally drug resistant (TDR) *M. tuberculosis* strains against the first- and second-line anti-TB drugs has created an urgent need for the development and implementation of new drug strategies. In this study, the complete genomes of 174 strains of *M. tuberculosis* are analysed to understand the evolution of molecular drug target (MDT) genes. Phylogenomic placements of *M. tuberculosis* strains depicted close association and temporal clustering. Selection pressure analysis by deducing the ratio of non-synonymous to synonymous substitution rates (*dN/dS*) in 51 MDT genes of the 174 *M. tuberculosis* strains led to categorizing these genes into diversifying (D, *dN/dS*>0.70), moderately diversifying (MD, *dN/dS*=0.35–0.70) and stabilized (S, *dN/dS*<0.35) genes. The genes *rpsL, gidB, pncA* and *ahpC* were identified as diversifying, and *Rv0488, kasA, ndh, ethR, ethA*, *embR* and *ddn* were identified as stabilized genes. Furthermore, sequence similarity networks were drawn that supported these divisions. In the multiple sequence alignments of diversifying and stabilized proteins, previously reported resistance mutations were checked to predict sensitive and resistant strains of *M. tuberculosis*. Finally, to delineate the potential of stabilized or least diversified genes/proteins as anti-TB drug targets, protein–protein interactions of MDT proteins with human proteins were analysed. We predict that *kasA* (*dN/dS*=0.29), a stabilized gene that encodes the most host-interacting protein, KasA, should serve as a potential drug target for the treatment of TB.

DATA SUMMARY

All the bacterial genomes scanned in this study have been deposited previously in the National Center for Biotechnology Information (NCBI) Genome database and are listed in (Table S1) (available with the online version of this article). Non-synonymous to synonymous substitutions (*dN/dS*) versus synonymous substitutions (*dS*) in molecular drug target (MDT) genes of *Mycobacterium tuberculosis* are listed in (Table S2).

INTRODUCTION

Mycobacterium tuberculosis, the causative organism of the airborne infectious disease tuberculosis (TB), is responsible for a high incidence of mortality worldwide [1, 2]. The Global Tuberculosis Report 2019 of the World Health Organization (WHO) indicated that by the end of the year 2018, *M. tuberculosis* had infected 1.7 billion people and about 1.2 million had lost their lives due to TB. One fourth of the world's population is estimated to be latently infected with *M. tuberculosis* [3].

Received 01 June 2020; Accepted 09 February 2021; Published 22 March 2021

Author affiliations: [1]Molecular Biology and Genomics Research Laboratory, Ramjas College, University of Delhi, Delhi 110007, India; [2]Department of Zoology, Ramjas College, University of Delhi, Delhi 110007, India; [3]Department of Zoology, University of Delhi, Delhi 110007, India; [4]Faculty of Life Sciences and Biotechnology, South Asian University, New Delhi 110021, India; [5]Department of Biotechnology, Jamia Millia Islamia, Okhla, New Delhi 110025, India; [6]The Energy and Resources Institute, Darbari Seth Block, IHC Complex, Lodhi Road, New Delhi 110003, India.

***Correspondence:** Charu Dogra Rawat, cdrawat@ramjas.du.ac.in

Keywords: drug resistance; molecular drug target; *Mycobacterium tuberculosis*; protein–protein interaction; selection pressure; tuberculosis.

Abbreviations: ANI, average nucleotide identity; ETH, ethionamide; INH, isoniazid; MDT, molecular drug target; MSA, multiple sequence alignment; NCBI, National Center for Biotechnology Information; SSN, sequence similarity network; TB, tuberculosis; WHO, World Health Organization.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary tables and one supplementary figure are available with the online version of this article.

As *M. tuberculosis* acquired resistance to several first-line anti-TB drugs, such as isoniazid (INH), pyrazinamide and ethambutol, an effective second-line anti-TB drug regimen was developed. The strains that were found to be resistant to the first-line anti-TB drugs responded well to the second-line anti-TB drugs, but only for a brief period. This was soon followed by the emergence of strains that were found to be resistant to both the first- and the second-line anti-TB drugs. Various types of drug-resistant TB, multi-, extensively and totally drug resistant TB (MDR-, XDR- and TDR-TB, respectively), thus, emerged [4–7]. Newer drugs, bedaquiline and delamanid, and repurposed TB drugs, such as linezolid and clofazimine, are employed for the treatment of TB, but many recent studies have reported *M. tuberculosis* strains that have acquired resistance against these drugs [8–10]. The development of these antibiotic-resistant strains is due to the antibiotic pressure, which results in acquisition of resistance by chromosomal mutations or horizontal gene transfer mediated by phages, transposons and plasmids [11].

Since announcement of the first genome sequence of *M. tuberculosis* strain H37Rv in 1998 [12], more than 6000 genomes have already been sequenced and deposited in the National Center for Biotechnology Information (NCBI) Genome database. Most of the sequenced genomes are in the draft stage and less than 200 genomes are reported to be complete, i.e. fully sequenced with no sequencing gaps. The availability of these mycobacterial genomes led to several attempts of genomic analyses, many of which focused on phylogenetic analysis [13–15]. Genomic studies focusing on evolutionary changes, particularly on the selection of molecular drug targets (MDTs) in *M. tuberculosis* genomes, are lacking. A few significant studies expanded the panel of drug-resistant markers and reinforced proposals to employ high-throughput whole-genome sequencing approaches for routine diagnosis and drug-susceptibility testing [16].

In this study, we analysed complete genomes (as available from the NCBI Genome database in August 2019) of 174 *M. tuberculosis* strains to understand the evolutionary changes in them. We focused our analysis on the 51 MDT genes/proteins of *M. tuberculosis* reported in recent studies [7, 17]. These genes have been identified based on the genome architecture and mechanisms of resistance against anti-TB drugs. Mutations in these genes have been reported to play a significant role in the development of drug resistance, primarily by altering metabolic activities of *M. tuberculosis* against anti-TB drugs. Deciphering the genetic modifications and their association with drug resistance mechanisms may identify promising drug targets that may lead to the development of new anti-TB drugs [7, 17]. Towards this, we classified the selected drug target genes based on the ratio of non-synonymous to synonymous substitutions (*dN/dS*) into diversifying (D), moderately diversifying (MD) and stabilized genes (S). Sequence similarity network (SSN) analysis indicated that the diversified proteins possessed a greater number of isolated nodes, which reflects the diverse loci of these proteins as compared to the stabilizing proteins. Furthermore, analysis of point mutations in diversifying and stabilized drug target proteins elucidated the drug resistance across *Mycobacterium* strains. We then performed protein–protein interaction analysis of host proteins with diversifying (D) and stabilized (S) proteins to determine host interacting D and S proteins. Together *dN/dS*, SSN and protein–protein interaction analyses could pave the way to identify potential drug targets for the development of efficient therapeutic strategies to combat TB.

Impact Statement

Tuberculosis (TB), an airborne infectious disease caused by *Mycobacterium tuberculosis*, is a major cause of deaths worldwide. Treatment of TB remains a therapeutic challenge due to the frequent emergence of *M. tuberculosis* strains that are highly resistant to a wide range of antibiotics. The rapid evolution of *M. tuberculosis* strains to acquire multidrug resistance is attributed to the accumulation of point mutations in the drug target genes. This necessitates a deeper understanding of the mechanism that these molecular drug targets (MDTs) have evolved to confer resistance against anti-TB drugs. In this study, the whole genomes of 174 *M. tuberculosis* strains were analysed to deduce the selection pressure on 51 MDTs. Among these genes, those exhibiting high and low selection pressure were demarcated as diversifying and stabilized genes, respectively; the latter are projected as good MDTs for anti-TB drugs. Furthermore, based on the selection pressure analysis and the interaction between the MDTs and human proteins, we predict molecular targets that can be used for developing or testing anti-TB drugs. Our study lays a strong genome-based foundation for supplementing experimental studies to develop a robust therapeutic regimen against the challenging TB disease that impacts global health.

METHODS

Retrieval of genome sequence data

Complete genome sequences of 174 *M. tuberculosis* strains (as only 174 complete genomes of *M. tuberculosis* strains were available in August 2019) were retrieved from the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/) (Table S1).

Phylogenomic clustering

Pairwise average nucleotide identity (ANI) values were obtained using the ANIm method from the pyani master pipeline [18]. A two way-matrix obtained from these values was further processed on Multi experiment Viewer (MeV) [19]. The normalized data were processed with hierarchical clustering, which was performed using the Pearson correlation algorithm. The dendrogram was then visualized using

the Interactive Tree of Life (iTOL) server [20] (https://itol.embl.de/).

Genome annotation, calculation of selection and multiple sequence alignments (MSAs)

Genomes of *M. tuberculosis* strains included in the study were re-annotated using the Glimmer method [21] on the RAST server [22]. Fifty-one genes were analysed: *rpsL*, *atpE*, *gpsI*, *dprE1*, *dprE2*, *embC*, *embB*, *embA*, *rpsA*, *gidB*, *katG*, *inhA*, *ndh*, *ahpC*, *rpoB*, *pncA*, *embR*, *eis*, *whiB7*, *ethA*, *ethR*, *mshA*, *gyrA*, *gyrB*, *alrA*, *cycA*, *Rv0488*, *thyA*, *dfrA*, *folC*, *ribD*, *clpC*, *fabG1*, *panD*, *ddn*, *rpoA*, *rmlD*, *fbiAB*, *fbiC*, *fadE24*, *ald*, *pepQ*, *kasA*, *tlyA*, *rplC*, *Rv0678*, *rpoC*, *fgd1*, *iniA*, *mmpL3* and *Rv1979c* (Table 1) [7, 17]. These genes/proteins are reported as potential MDTs of anti-TB drugs as mutations in them play a significant role in the development of resistance against anti-TB drugs, primarily by altering the metabolic pathways in *M. tuberculosis* [7, 17]. All-versus-all BLASTN [23] was used to search for the presence and abundance of these genes in the genomes of all the *M. tuberculosis* strains. The gene sequences were extracted from all the genomes with a sequence identity and coverage cut-off of 80% aligned using the maximum-likelihood method [24] on Mega-X software version 10.0.5 [25]. To decipher the rate of evolution across the 174 strains of *M. tuberculosis*, the *dN/dS* calculation was performed using the single-likelihood ancestor counting (SLAC) method in DataMonkey software version 2.0 [26] (http://www.datamonkey.org/slac). The *dN/dS* values were plotted against *dS* in R [27]. MSAs of MDT protein sequences were performed by using Clustal-X software on BioEdit [28]. The *dN/dS* ratios ranged between 0.25 and 1.27 (Table S2). According to the general criteria, values of *dN/dS* ≥1.00 are indicative of undergoing positive diversification. We partitioned the *dN/dS* ratios from 0 to 1.00 into three parts and categorized the genes into most stabilized (S) (*dN/dS* <0.35), moderately diversifying (MD) (*dN/dS*=0.35–0.70) and diversifying (D) (*dN/dS* >0.70).These divisions are considered relative to one another as the one with the lowest value of *dN/dS* must possess more synonymous substitutions as compared to the one which has *dN/dS* values just below 1.0. The purpose of partitioning the genes based on *dN/dS* ratio was to identify the most stabilized genes among the ones having *dN/dS* values lower than 1.0, which can be further analysed for their potential as drug targets against TB.

SSN analysis

Sequences of the MDT proteins of the 174 strains of *M. tuberculosis* were retrieved using protein BLAST. An all-versus-all BLAST was performed to define the similarities/variations between sequence pairs of diversifying (AhpC, GidB, PncA and RpsL) and stabilized (Ddn, EmbR, EthR, EthA, KasA, Ndh and Rv0488) MDT proteins across the 174 mycobacterial genomes [23]. A user defined threshold was optimized according to the alignment score and the maximum length of BLAST results in diversifying and stabilized protein sequences. Clustering was performed using CD-HIT on the scores of BLASTP pairwise alignments at a threshold value, *E* value of 1×10^{-30}.

The pairwise alignment relationships of the sequences were analysed by constructing sequence-based similarity networks. The sequences were filtered for the removal of 100% identical sequences using CD-HIT [29] and SSNs were constructed with a threshold alignment score of 50%. This approach confers a better relationship between large datasets at defined cut-offs and is more advantageous than phylogenetic trees. The threshold cut-off value of 1×10^{-70} was used for the construction of both diversifying and stabilized MDT sequence networks upon analysing the trends of varying alignment lengths at different *E* values. The data were visualized in the complex network platform Cytoscape v3.7.2 [30]. Sequence and structural similarity was searched using BLASTP [23] and TM align [31].

Host–pathogen protein–protein interaction

The protein sequences of diversifying (D) and stabilized (S) MDTs were searched against various interaction databases, IntAct [32], MINT [33], BioGRID [34] and BIND [35], for predicting the maximum number of host and pathogen interactions. The interactions were visualized using Cytoscape v3.7.2 [30].

RESULTS AND DISCUSSION

General genome characteristics of *M. tuberculosis* strains

A total of 120 of the 174 *M. tuberculosis* strains were isolated from human sputum, lungs and cerebrospinal fluid samples (Table S1). The remaining 54 strains were isolates from non-human sources (Table S1). *M. tuberculosis* H37Rv was considered as a reference strain in the present study as most of the studies on *M. tuberculosis* genome analyses and antibiotic susceptibility have used it as a reference. Strain H37Rv, along with strain H37Ra, was derived from a clinical strain, H37, isolated from a pulmonary TB patient, resistant against rifampicin and ethambutol [36]. The genomic size of the *M. tuberculosis* strains ranges from 4.34 Mb (*M. tuberculosis* Moreau RDJ) to 4.44 Mb (*M. tuberculosis* Beijing-like/50148), with a mean genome size of 4.40 Mb. This suggests high uniformity in terms of genome size with a variation of only 0.1 Mb across the strains. *M. tuberculosis* var. *bovis* strain 26 [37] possessed the highest number of genes (4826) and proteins (4778), while strain *M. tuberculosis* RGTB423 [38] possessed the lowest number of genes (3670) and proteins (3622). The G+C content of the 174 strains was found to be high, i.e. 65.6 mol%, a characteristic of the phylum *Actinobacteria*.

Phylogenomic analysis

To define the evolutionary relatedness between the *M. tuberculosis* strains, a phylogenomic tree based on the pairwise ANI between the strains was generated. The ANI values of the genomes of *M. tuberculosis* strains were observed to be very high (99.5–99.9%), suggesting high genome similarity.

Table 1. List of 51 MDT genes of *M. tuberculosis*, their gene IDs, protein IDs (as obtained from the NCBI Genome database) and functions

Strain H37Rv was used as a reference.

Gene name	Gene ID	Protein ID	Function
rpsL	888259	NP_215196.1	30S ribosomal protein S12
gidB	886243	NP_218436.2	7-Methylguanosine methyltransferase
pncA	888260	NP_216559.1	Pyrazinamidase/nicotinamidase PncA
katG	885638	NP_216424.1	Catalase-peroxidase
ahpC	23491355	NP_216944.1	Alkylhydroperoxide reductase subunit AhpC
embB	886126	NP_218312.1	Arabinosyltransferase B
Rv0488	888415	NP_215002	Amino acid transporter
kasA	887269	NP_216761.1	3-Oxoacyl-ACP synthase 1
ddn	887496	NP_218064.1	Deazaflavin-dependent nitroreductase
ethR	886189	NP_218372.1	HTH-type transcriptional repressor EthR
ethA	886175	NP_218371.1	Monooxygenase EthA
inhA	886523	NP_216000.1	NADH-dependent enoyl-[ACP] reductase
mshA	887160	NP_215000.1	D-Inositol 3-phosphate glycosyltransferase
ndh	885746	NP_216370.1	NADH dehydrogenase
embR	887026	NP_215783.1	Transcriptional regulator EmbR
clpC	885104	YP_177995.1	ATP-dependent protease ATP-binding subunit
ald	888493	NP_217296.1	L-Alanine dehydrogenase
rpoB	888164	NP_215181.1	DNA-directed RNA polymerase subunit β
rpoC	888177	NP_215182.1	DNA-directed RNA polymerase subunit β
embC	886112	NP_218310.1	Arabinosyltransferase C
embA	886123	NP_218310.1	Arabinosyltransferase A
mmpL3	886752	NP_214720.1	Transmembrane transport protein
pepQ	888409	NP_217051.1	Cytoplasmic peptidase PepQ
gyrB	887081	NP_214519.2	DNA gyrase subunit B
fbiC	886061	NP_215689.1	2-Phospho-L-lactate transferase
fabG1	886551	NP_215999.1	3-Oxoacyl-ACP synthase 1
panD	885596	NP_218118.1	Aspartate 1-decarboxylase
Rv0678	888235	NP_215192.1	Hypothetical protein Rv0678
rpoA	887629	NP_217974.1	DNA-directed RNA polymerase subunit α
fadE24	887971	NP_217655.1	Acyl-CoA dehydrogenase
gpsI	888467	NP_217299.1	Bifunctional guanosine pentaphosphate synthetase/polyribonucleotide nucleotidyltransferase
cycA	888812	NP_216220.1	D-Serine/alanine/glycine transporter protein CycA
rplC	888343	NP_215215.1	50S ribosomal protein L3
rmlD	888704	NP_217783.1	dTDP-4-dehydrorhamnose reductase
fgd1	886418	NP_214921.1	F420-dependent glucose-6-phosphate dehydrogenase
atpE	886937	NP_215821.1	ATP synthase subunit C

Continued

Table 1. Continued

Gene name	Gene ID	Protein ID	Function
dfrA	887777	NP_217279.1	Dihydrofolate reductase
gyrA	887105	NP_214520.1	DNA gyrase subunit A
alrA	887634	NP_217940.1	Racemase
ribD	886721, 887389	NP_215925.1, NP_217187.1	Bifunctional riboflavin biosynthesis diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino) uracil reductase
whiB7	3205083	YP_177940.1	Transcriptional regulator WhiB7
rpsA	885188	NP_216146.1	30S ribosomal protein S1
fbiA, B	888701, 888693	NP_217778.1, NP_217779.1	Coenzyme F420:L-glutamate ligase
iniA	886510	NP_214856.1	Isoniazid inducible protein
folC	885902	NP_216963.1	Folylpolyglutamate synthase FolC
eis	885903	NP_216932.2	Enhanced intracellular survival protein
thyA	887728	NP_217280.1	Thymidylate synthase ThyA
Rv1979c	885819	NP_216495.2	Permease
tlyA	885396	NP_216210.1	16S/23S rRNA (cytidine-2′-*O*)-methyltransferase TlyA
dprE1	886125	NP_218307.1	Decaprenylphosphoryl-*β*-D-ribose oxidase
dprE2	886124	NP_218308.1	Decaprenylphosphoryl-D-2-keto erythropentose reductase

It indicates genome conservation in terms of loss or gain of genetic material during evolution. Difference in the ANI values is attributed to the point mutations that might have led to their adaptability against anti-TB drugs or other non-favourable conditions. Interestingly, the phylogenomic tree based on ANI values (Fig. 1) clustered the 174 strains into seven different clades, in which a major monophyletic clade covered nearly 1/4th of the genomes. The clade includes virulent *M. tuberculosis* H37Rv, which was used as the reference genome, and its non-virulent strain, H37Ra. Another group of virulent/non-virulent strains is strain F1 and strain F28, which were also clustered in the same clade. Strains Haarlem/NITR202 and CAS/NITR204 were the most diverged strains due to their differential clustering. A recent study by Saelens *et al.* (2018) has reported 16 strains of *M. tuberculosis* that were isolated from TB patients of Euro–American lineage with strain name header 'GG', such as GG-111-10, GG-5-10, GG-20-11, GG-27-11, GG-36-11, GG-37-11, GG-45-11 and GG-77-11 [39] (Table S1). Most of the GG strains clustered in a single clade, while GG-129-11, GG-27-11, GG-186-10 and GG-111-10 clustered in a different but closely placed clade. Similarly, seven strains with name headers 'Beijing-like' such as, Beijing-like/1104, Beijing-like, Beijing-like/50148, Beijing-like/36918, Beijing-like/38774, Beijing-like/35049 and Beijing/391 were clustered in a monophyletic clade. This close phylogenetic placement of the Beijing-like strains of *M. tuberculosis* was also reported in a previous study [13]. From this phylogenomic clustering analysis and corroboration from earlier reports of clustering, it can be predicted that *M. tuberculosis* strains isolated either from a similar environment or at a point of time, or both, are likely to be more similar genetically and tend to cluster temporally, i.e. based on their time of isolation.

Evolution of MDT genes and resistance to anti-TB drugs

To determine the evolutionarily accumulated variations in the 51 MDT genes (Table 1), we extracted their gene sequences from the 174 genomes of *M. tuberculosis* strains using *M. tuberculosis* H37Rv as the reference strain. We focused our analysis on diversifying (D) and stabilized (S) genes to deduce drug targeting strategies for treatment of drug-resistant mycobacterial strains.

MDT genes under diversifying selection (D) in *M. tuberculosis*

Under this category, we included four genes having *dN*/*dS* ratios between 0.70 and 1.27 (Table S2). The *rpsL* gene (D) that encodes 30S ribosomal protein S12 was found to be the most diversified with *dN*/*dS*=1.27. Streptomycin, the first drug administered against *M. tuberculosis* strains in the 1960s [40], acts on the *rpsL* gene and disrupts its functions leading to misreading of the genetic code, affecting initiation of translation and unusual proofreading [41]. Soon after, resistance against streptomycin was reported, and many other drugs were included more often in the treatment procedure [42, 43]. In *M. tuberculosis*, the reason for streptomycin resistance was reported as a change in the membrane permeability rather than mutations in the *rpsL*

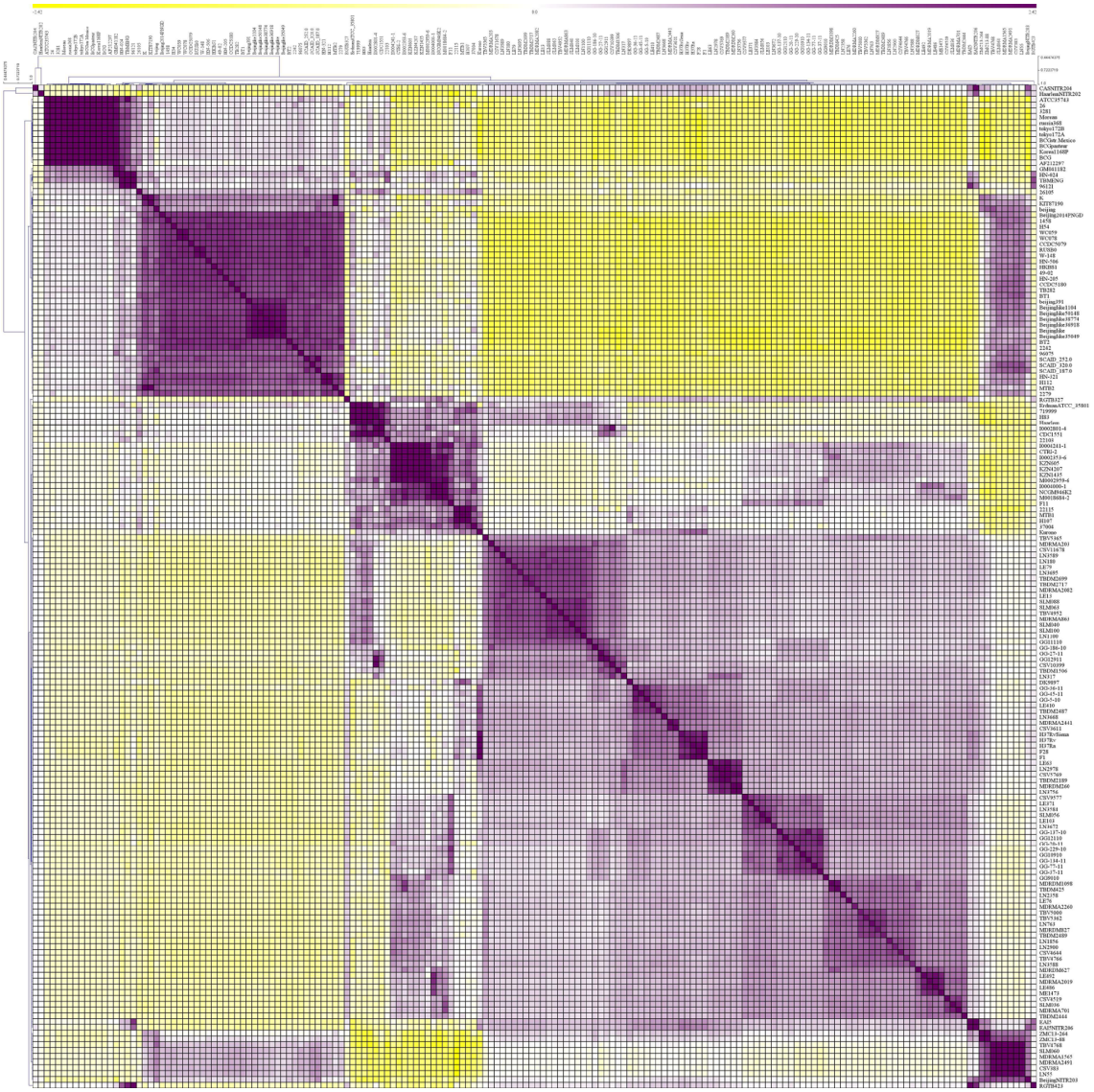

Fig. 1. Dendrogram on the matrix obtained using pairwise ANI between 174 strains of *M. tuberculosis* plotted with the heat map prepared on the values.

gene sequence [44], due to lesser number of *M. tuberculosis* genome sequences or *rpsL* gene sequences available at that time. In our study, the *rpsL* gene was identified as the most diversified (D) gene, indicating the possible reason for the development of early drug resistance against streptomycin.

Another gene that showed a high *dN*/*dS* ratio (*dN*/*dS*=0.99) was *gidB* (D) (encoding 7-methylguanosine methyltransferase). Mutations in the *gidB* gene were also linked with resistance against streptomycin as these methyltransferases are specific to 16S rRNA, and may affect the binding and proper functioning of streptomycin [45]. As *rpsL* and *gidB* are the most diversified genes and are linked to resistance against streptomycin, it suggests positive selection pressure on the gene targets of a drug may lead to the development of resistance with repeated encounters or relapse.

The gene *pncA* (D) (*dN*/*dS*=0.75) encodes pyrazinamidase/nicotinamidase, PncA, which catalyses the conversion of the prodrug pyrazinamide into the active form, pyrazinoic acid, in *M. tuberculosis* strains [46]. Pyrazinamide is part of the anti-TB first-line drug regime, along with INH, rifampicin and ethambutol, given during a WHO-recommended directly observed treatment, short course, DOTS [47]. From previous reports, a high rate of mutation in the *pncA* gene was observed that led to the development of resistance against pyrazinamide in *M. tuberculosis* [47, 48]. The resistance was due to the inability of mutated PncA to functionally activate the pyrazinamide drug. Development of resistance in *M. tuberculosis* strains against pyrazinamide correlates well with the high *dN*/*dS* ratio of *pncA,* suggesting

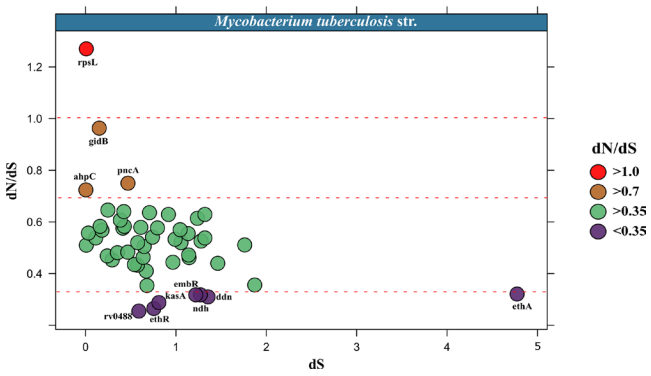

Fig. 2. *dN/dS* versus *dS* plot of the MDT genes of *M. tuberculosis* strains. The *x*-axis has *dS* values and *y*-axis has *dN/dS* values. Three grade lines have been drawn at 0.35, 0.70 and 1.0 *dN/dS* values. Genes with <0.35 *dN/dS* values are stabilized (S) and highlighted with purple. Genes with 0.35–0.70 *dN/dS* values are moderately diversifying (MD) and highlighted with green. Genes with 0.70–1.0 *dN/dS* values are diversified (D) and highlighted with brown, and the gene with a >1.0 *dN/dS* value is highlighted with red.

the diversifying genes are more likely to develop drug resistance.

As stated earlier, INH is a part of the first-line anti-TB drug treatment, which requires catalase-peroxidase KatG (*katG dN/dS*=0.54) to form active INH-NAD [49, 50]. In the active state, INH binds to enoyl-acyl carrier protein reductase (InhA), which is a component of fatty acid synthase II (FAS-II) required for the production of mycolic acids, the main cell wall component of mycobacteria [49]. Resistance against INH has been reported to arise possibly because of the development of efficient antioxidant systems by the pathogen [51], which affects the activation of the prodrug. Another gene, *ahpC* (D) (*dN/dS*=0.72), encodes alkylhydroperoxide reductase subunit C, which forms an alternative route to protect the mycobacteria from oxidative and nitrosative stress [52]. Mutations in *ahpC* can also compensate for the loss of catalase-peroxidase activity [53] and, thus, are able to confer resistance against the drug INH more effectively. The higher *dN/dS* value of *ahpC* represents positive selection pressure on this gene across the *M. tuberculosis* strains, which possibly promotes survival and propagation of the pathogen by protecting it against the toxic activity of organic peroxides. Another gene of the arabinan synthesis pathway, *embB* (D) (*dN/dS*=0.63), that encodes arabinosyltransferase B also has a noticeably higher value of *dN/dS*, suggesting the diversifying gene sequence accounts for the development of ethambutol drug resistance in *M. tuberculosis*.

We predict that these genes, *rpsL*, *gidB*, *pncA*, *ahpC* and *embB*, are the most diversifying (D) among the 51 MDT genes and, due to the copious variations, drug strategies targeting these genes may not stay effective for long. Also, these genes appear to have relatively more chances to develop drug resistance as compared to others due to their high non-synonymous diversification.

MDT genes under stabilizing selection (S) in *M. tuberculosis*

Seven genes (out of the 51 MDT genes) showed *dN/dS* values ≤0.35, indicating the most stabilized genes in the list (Fig. 2), (Table S2). The *Rv0488* (S) (*dN/dS*=0.26) gene that encodes a putative amino acid transporter was observed to have the lowest *dN/dS* value. The presence of a protective factor for INH resistance, ETRC, 210 bp upstream of *Rv0488* has been reported to be negatively linked with INH resistance [54]. Thus, our analysis suggests *Rv0488* is probably the most stabilized among MDT genes and can be considered as a suitable target. Another gene, *kasA* (S), which encodes 3-oxoacyl-ACP synthase 1 and catalyses the elongation of the fatty acid chain by adding an acyl group to malonyl-ACP, exhibited the low *dN/dS* value of 0.29 (Fig. 2). KasA has been reported to take part in mycolic acid synthesis, biofilm formation, cell wall synthesis and pathogenesis along with other such genes [55]. Based on the lower diversifying selection pressure on the *kasA* gene in *M. tuberculosis* strains, we suggest KasA as a potential drug target for new anti-TB drugs.

The gene *ddn* (S) (*dN/dS*=0.31) encodes deazaflavin-dependent nitroreductase and prevents the production of toxic semiquinones in the mycobacterial cells; thus, protecting them from oxidative stress. Ddn is also the only enzyme in *M. tuberculosis* that bioactivates bicyclic 4-nitroimidazole prodrugs like PA-824, pretomanid and delamanid for anti-TB therapy [56, 57]. Reports of resistance to nitroimidazole are limited to a certain population or community [8, 58]. This might be due to the recent inclusion of these drugs in anti-TB therapy and the lower *dN/dS* value of *ddn* suggests fewer accumulated non-synonymous to synonymous substitutions in the *ddn* gene in *M. tuberculosis* strains. Ddn has a dual role as it is required for mycobacterial cell survival as well as for the activation of nitroimidazole drugs.

Two genes, *ethR* (S) (*dN/dS*=0.26) and *ethA* (S) (*dN/dS*=0.32) (Fig. 2), encode HTH-type transcriptional repressor and monooxygenase, respectively. EthA is responsible for the bioactivation of an anti-TB drug called ethionamide (ETH), which when activated inhibits the production of InhA (NADH-dependent enoyl-[ACP] reductase) and synthesis of mycolic acid [59, 60]. The action of EthA is inhibited by EthR, which acts as a repressor of EthA and, thus, affects the subsequent activation of ETH [61]. This mechanism of repression and activation regulates the effectiveness of the ETH drug in anti-TB drug therapy. The lower *dN/dS* values of *ethA* and *ethR* suggest a stable bio-modulation mechanism for ETH activation. However, reports have shown the development of ETH resistance in many *M. tuberculosis* strains due to mutations in *inhA* (*dN/dS*=0.52), *mshA* (*dN/dS*=0.53), *ndh* (*dN/dS*=0.31) and *ethA* (*dN/dS*=0.32) genes (Table S2) [62–64]. Thus, for competent ETH targeted anti-TB therapeutic strategies, researchers may target

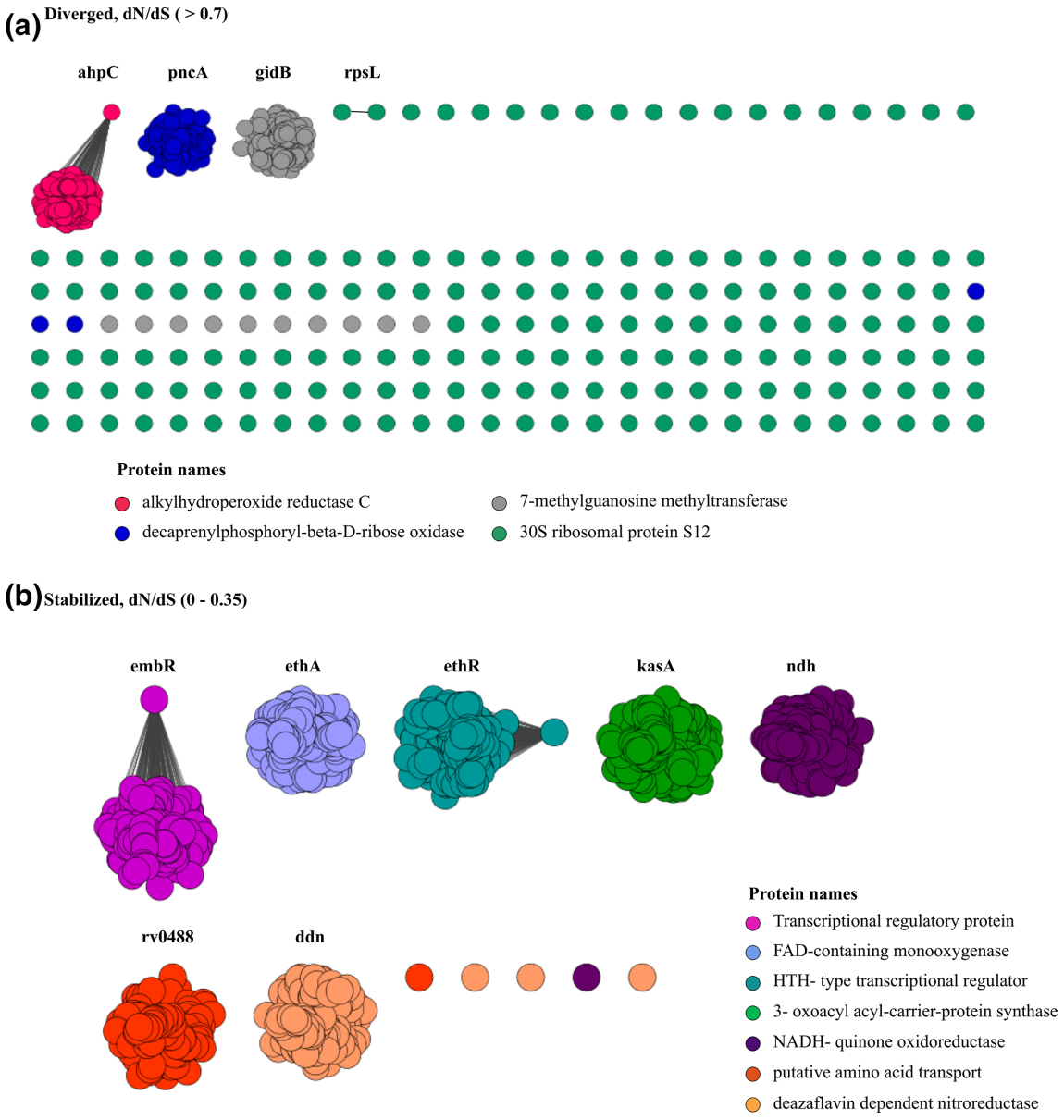

Fig. 3. The SSN was drawn across the 174 strains of *M. tuberculosis*. *M. tuberculosis* H37Rv was used as a reference. (a) Diversifying genes (>0.70 *dN/dS* values). (b) Stabilizing genes (<0.35 *dN/dS* values). Different colour codes are given for different genes, as shown in the keys.

the inactivation of EthR and production of its competent analogues.

Diversification of MDT gene loci

To determine the empirical measure of the sequence diversity among diversified (D) and stabilized (S) MDT genes, and to draw comparisons of their pairwise relationships, SSNs were prepared and analysed. Each protein sequence is represented by a node, and two nodes are connected by edges if they share more than the defined threshold similarity. Thus, each node in the resulting networks could not connect with all other nodes through a finite path. The analysis revealed that MDT proteins of the genes included under diversifying (D) selection consist of a total of 695 nodes (except AhpC from the Beijing strain, due to lower identity than the threshold set at an 80% cut-off) clustered into 189 connected components of the network with an average clustering coefficient of 0.73 (Fig. 3). There were 185/189 isolated nodes in the diversified (D) MDTs. These isolated nodes were distributed as follows: GidB, 10/11; AhpC, 0/1; RpsL, 172/173; and PncA, 3/4. Most of the proteins, GidB (90%), RpsL (99%) and PncA (75%), do not share homology across the genomes (Fig. 3). This reveals the extensive diversification due to host adaptability of individual genes. A recent study suggested the higher rate of drug resistance pertains to

a higher diversity of these loci [65]. In stabilized (S) proteins, the average clustering coefficient was 0.99 and a total of 1218 nodes partitioned into 12 connected components that include homologous clusters and isolated nodes. In total, 5/12 isolated nodes were identified, which partitioned into 0/1 of EmbR, 0/1 of EthA, 0/1 of KasA, 0/1 of EthR, 1/2 of Rv0488, 1/2 of Ndh and 3/4 of Ddn. When compared with diversifying proteins, stabilized proteins showed much lesser diversity, as maximum sequences were clustered into 12 components only and a very few were found isolated; diversified proteins were found more scattered with 189 components due to the maximum number of isolated nodes. Intuitively, sequences under stabilized selection that form robust clusters in SSNs may undergo less divergence throughout the evolution, while the sequences under diversified selection reveal a higher positive selection pressure to retain a function [66]. This supports the division of diversifying and stabilizing genes/proteins observed in *dN/dS* analysis, and our understanding that the stabilized genes/proteins owing to their lesser diversity should be kept under focus as potential drug targets against TB.

Prediction of sensitive and resistant strains of *M. tuberculosis*

Point mutations in drug targeted genes/proteins may or may not affect the affinity of the respective drug to its target; thus, these mutations can be categorized as resistant, sensitive or no effect mutations. We analysed previously reported point mutations in diversified (D) and stabilized (S) proteins of the 174 *M. tuberculosis* strains, and attempted to discover the ability of these strains to respond against the anti-TB drugs. The dendrogram based on the point mutations showed the temporal clustering of *Mycobacterium* strains, which is not identical but comparable to their whole-genome-based phylogenetic profile (Figs 1, 4 and S1). For instance, the GG strains clustered into two closely placed monophyletic clades. The Beijing strains also clustered in closely placed clades, and showed resistant point mutations in GidB, RpsL and EthA. Strains H37Rv, H37Ra and H37RvSiena were clustered in a close group along with the other strains that shared three common mutations, R70S and S100F in GidB and F110L in EthA. These mutations were, however, observed in all the strains included in the study. Similarity in the point-mutation-based and whole-genome-based clustering provides a glimpse of the evolution of antibiotic resistance and gene divergence in these strains.

As stated earlier, the anti-TB drug streptomycin was predominately given to patients in the 1960s and was later replaced with other drugs, INH and rifampicin [40]. It has been reported that streptomycin resistance is linked with mutations in RpsL (D) and GidB (D) proteins [67–70] . Three RpsL mutations, K43R, K88R and K88T, were identified as critical to demarcate a strain as resistant to streptomycin (Table 2) [69]. From MSAs of RpsL protein sequences from 174 *M. tuberculosis* strains, it was observed that 23 out of 174 strains have Arg at the 43rd position, three strains, TBDM425, MDRDM1098 and I0002353-6, have Thr at the 88th position, and one strain, LN2358, has Met at the 88th position, while no strain showed Arg at the 88th position (Figs 4 and S1a). These mutations have been previously reported to confer resistance against streptomycin. Thus, strains that possess K43R and K88R can be categorized as resistant to streptomycin. Previous studies have shown that many Beijing strains of *M. tuberculosis* harbour resistance against streptomycin [69, 70]. In this study, nine out of ten Beijing strains were observed to have a streptomycin resistance pattern. Streptomycin resistance in SCAID 187 and SCAID 320 has also been reported elsewhere [71, 72].

Another protein, GidB, contained many mutations at different codons and has been reported to confer low to intermediate level resistance against streptomycin [73, 74]. Surprisingly, a few mutational points, frequently reported in earlier studies, showed different substitutions or were missing in the MSA of GidB protein sequences (Fig. S1b, Table 2). For instance, L79W instead of L79Y, G164S instead of G164C, R47W instead of R47Q and D67H instead of D67G were observed. Two mutations, L16R and S100F, which were previously reported as non-determinant of resistance or spontaneous mutations [75], were also observed in the strains; S100F was found in every strain in the study (Fig. 4). Other noticeable spontaneous resistance mutations are E92D and A615G (A205A), which were observed in 44 and 62 strains, respectively. Previous studies have reported that A276C (E92D) and A615G (A205A) mutations are associated with the Beijing lineage of *M. tuberculosis* and their coexistence was proposed as the possible marker for the Beijing strains [76, 77]. Co-occurrence of A276C (E92D) and A615G (A205A) was also observed in many other strains, like zmc13-88, zmc13-264, wc078, wc059, SCAID320, SCAID252, SCAID187, SLM060 and so on (Fig. 4). This projects that the two mutations are not only limited to Beijing strains but also present in other lineages. The E92D mutation was referred to as conferring resistance in the SCAID187 strain [71]. Another notable mutation, R70S, previously reported to be a resistance mutation, was observed in all the mycobacterial strains [75]. This suggests the possible streptomycin resistance with varied degrees in the strains may depend upon other mutations in GidB and RpsL. A novel mutationV77G, which is not yet reported in the literature, was also observed. Since mutations in GidB are vast and varied, these mutations could not be used to demarcate the resistance or sensitivity to streptomycin.

Like GidB, PncA (D) sequences across the genomes showed high variability, in which many reported mutations were observed with similar or different substitutions and many mutations were not observed (Fig. S1c, Table 2). Mutations reported as the determinant of low or high resistance to pyrazinamide [78–80] were identified in the MSA of PncA protein sequences, K96T in strain SCAID320, I6T in strain SCAID 252, A171V in strain w-148, suggesting possible resistance against pyrazinamide. Two resistance mutations, H57D (in 12 strains) and W68R (in Beijing-like/35049) (Fig. S1c), which were previously reported as part of the iron centre and substrate-binding residues of PncA, respectively, were also observed [81]. In addition to this, A3E instead of A3P, L182W instead of L182S, T87A instead of T87M, T114A

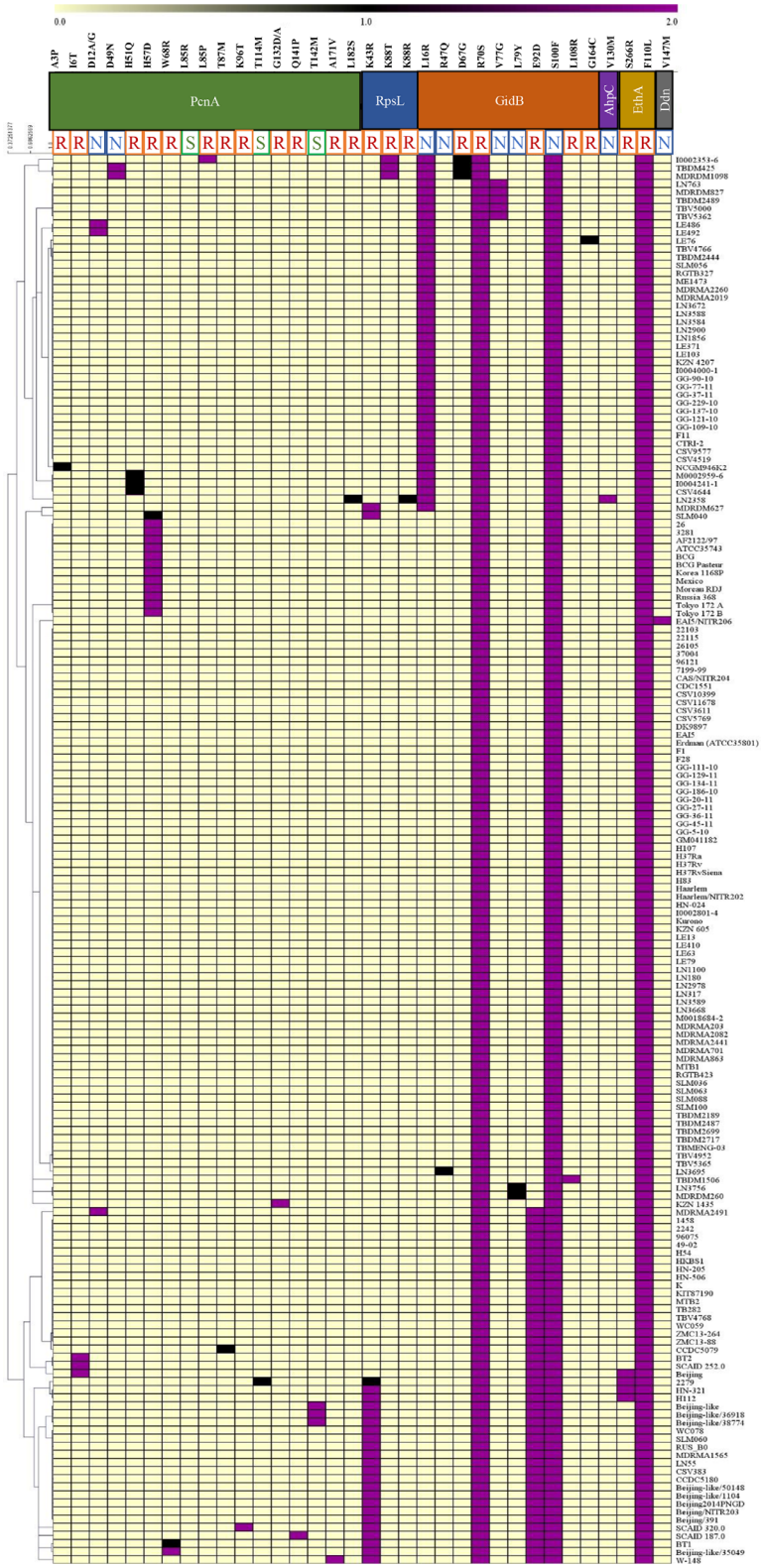

Fig. 4. Point mutations in PncA, GidB, RpsL, AhpC, EthA and Ddn across 174 *M. tuberculosis*. A matrix has been prepared based on the absence of mutation, '0', point mutation with different amino acids from reported ones, '1', and point mutation as mentioned in the columns, '2'. Colour codes depict this presence or absence of mutation (see the bar above the figure). Blocks: 'R', mutation reported to be resistant; 'S', reported to be sensitive; 'N', not known. The dendrogram has been reconstructed using antibiotic-resistance data.

Table 2. Amino acid mutations (previously reported) in the sequences of PncA, GidB, RpsL, AhpC, EthA and Ddn proteins

The reported effects of these mutations, as well as observed mutations in this study, are also listed.

MDT protein	Reported mutation	Reported effect of mutation	Observed mutation
PncA	A3P	Resistant	A3E
	I6T	Resistant	I6T
	D12A/G	No effect	D12A/G
	D49N	No effect	D49N
	H51Q	Resistant	H51R
	H57D	Resistant	H57D/R
	W68R	Resistant	W68L/R
	L85P	Resistant	L85P
	L85R	Sensitive	L85P
	T87M	Resistant	T87A
	K96T	Resistant	K96T
	T114M	Sensitive	T114A
	G132D/A	Resistant	G132A
	Q141P	Resistant	Q141P
	T142M	Sensitive	T142A
	A171V	Resistant	A171V
	L182S	Resistant	L182W
RpsL	K43R	Resistant	K43R
	K88R	Resistant	K88M
	K88T	Resistant	K88T
GidB	L16R	No effect	L16R
	R47Q	No effect	R47W
	D67G	Resistant	D67H
	R70S	Resistant	R70S
	V77G	No effect	V77G
	L79Y	No effect	L79W
	E92D	Resistant	E92D
	S100F	No effect	S100F
	L108R	Resistant	L108R
	G164C	Resistant	G164S
AhpC	V130M	No effect	V130M
EthA	F110L	Resistant	F110S
	S266R	Resistant	S266R
Ddn	V147M	No effect	V147M

instead of T114M and many others were determined as differential mutations at mutational points (Fig. 4). Among these, H51R instead of the H51Q mutation was observed at the 51st position, which belongs to the iron centre of the protein [81]. Strain SCAID 187 has been reported as being resistant against pyrazinamide with a mutation Q141P [71], which was also observed in the MSA of PncA sequences. Two resistance mutations, L85P and G132A, which have been reported after drug susceptibility tests against pyrazinamide [82], were also identified in strains I0002353-6 and KZN1435, respectively. Two observed novel substitutions, D49N and D12A/G, require experimental analysis to predict their effects on resistance or sensitivity towards the drug.

AhpC (D) mutations were mainly studied when present between *ahpC–oxyR* intergenic regions [83]. The MSA of AhpC showed a single substitution of V130M in strain LN2358 and a frame-shift in Haarlem/NITR202 after the 195th amino acid position (Figs 4 and S1d). Both mutations are novel and can be further studied to elucidate their role, if any, in determining resistant or sensitive *M. tuberculosis* strains.

As mentioned earlier, activation of the drug ETH is catalysed by EthA (S), which when mutated may confer resistance against the drug. As compared to diversifying genes/proteins, MSA of EthA protein sequences showed a lesser number of mutations (Fig. S1e). Only two mutations from the list of previously reported 14 mutational points were observed [84]. These are S266R in strains HN-321, H112, Beijing and 2279 strains, and F110S in all 174 strains (Fig. 4). Notably, S266R and F110L were included under high-level resistance mutations; thus, depicting the possible ETH resistance of strains HN-321, H112, Beijing and 2279, while the observed mutation of F110S instead of F110L requires further investigation.

Another stabilized protein, Ddn, showed a high-level of sequence conservation with a single substitution of V147M. Many reported mutations in the Ddn (S) sequence corresponding to the effectiveness of nitroimidazole and other drugs were checked, but none of them were observed in the MSA of the Ddn protein sequences of the 174 *M. tuberculosis* strains (Fig. S1f, Table 2) [85–87]. Thus, the analysis could draw a clear picture of drug resistance in a few strains based on already established mutations. However, the novel mutations and differential substitutions at known mutational points reported in this study need further investigation to discover their relevance in determining the resistance of *M. tuberculosis* strains against anti-TB drugs.

Host–pathogen interactions of stabilized and diversifying MDT proteins

It is evident from the analysis that among *M. tuberculosis* MDT genes, highly stabilized genes/proteins can serve as potential anti-TB drug targets for a longer duration. Thus, it becomes important to elucidate the interactions of MDT proteins and human host proteins to ascertain the role of MDT proteins in modulating the immune response. Here, we analysed the interactions among MDT proteins encoded by stabilized (S) and diversifying (D) genes with human host proteins.

KasA (3-oxoacyl-ACP synthase 1) was found to interact with 19 database proteins in which 15 were host targets and 4, AhpC (D), PncA (D), EthA (S) and Ndh (S), were MDT proteins (Fig. 5). Host targets of KasA perform different cell functions, like transcription modulation (SPRE), β-oxidation of fatty acids (HCD2, THIL and DHB4) and most importantly physiological regulation of programmed cell death (Fas and Hax1). Stromelysin-1 PDGF-responsive element (SPRE) is a stromelysin, a metalloprotease promoter [88]. It activates SPRE-binding protein (SPBP), which acts as a coactivator of many transcription factors like c-Jun, Sp1, Pax6 and Ets1 [89], critical for different physiological processes. For instance, c-Jun is well known for regulating cell proliferation, tumorigenesis, apoptosis and embryonic development [90]; Sp1 has been reported to activate or suppress the transcription of genes as a response to physiological or pathological stimuli [91]. Another important set of proteins to interact with KasA was Fas and Hax1. Fas is a receptor of the TNF (tumour necrosis factor) superfamily and has an important role in the regulation of programmed cell death by producing death-inducing signalling complex (DISC) for proteolytic cleavage of cascade 8, which induces subsequent cascade reactions [92]. A recent study has shown that Fas can activate NFκB [93], which we will discuss below. Hax 1 has been reported as an anti-apoptotic protein inhibiting CASP9 and CASP3 and, thus, promotes cell survival [94]. This suggests the significant role of KasA in controlling various host cellular processes and modulation of which may affect the host survival rate.

Another notable *M. tuberculosis* protein, AhpC, alkylhydroperoxide reductase subunit C, showed interactions with 22 proteins, of which 15 were host proteins and 7 were *M. tuberculosis* proteins (Fig. 5). Among these seven, four proteins, KasA (S), GidB (D), PncA (D) and RpsL (D), were encoded by MDT genes. Host proteins were mainly found to be involved in functions like innate and adaptive immune responses (LEG3, LEG9), molecular chaperones (HS71L, GRP78, GRP75), transcription factors (NFκB1) and antioxidants (PRDX1, PRDX2, PRDX3, PRDX4 and PRDX6). LEG3 produces acute inflammatory responses against stimuli such as neutrophil activation and adhesion, chemo-attraction of monocytes or macrophages, and activation of mast cells [95, 96], whereas LEG9 causes the release of cytokines and macrophage activation [97]. NFκB1, present in almost every cell, has several roles in different biological processes, like the immune response, cell growth, tumorigenesis, inflammation and apoptosis [98]. As mentioned above, FasA controls the activation of NFκB1 that provides a link between the interaction of KasA and AhpC. Another protein, IB42, that showed direct interaction with both KasA and AhpC is, however, not discussed in the literature. These results suggest the combined role of both KasA and AhpC in modulating the host immune response and programmed cell death.

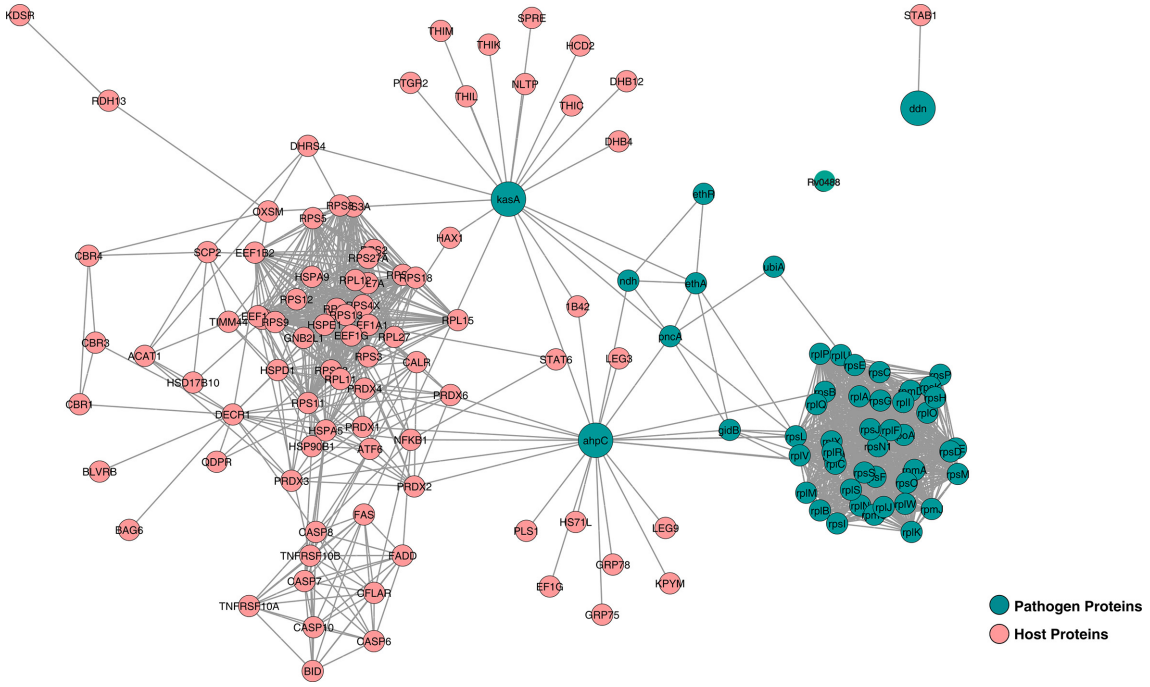

Fig. 5. Protein–protein interaction between the diversifying and stabilized MDT proteins of *M. tuberculosis* and human proteins. Proteins in light red are human and those in green belong to *Mycobacterium*.

Interestingly, Ddn, deazaflavin-dependent nitroreductase, was observed to interact with only one host protein, stabilin-1 (STAB1) (Fig. 5), and no interaction was observed with any other *M. tuberculosis* proteins. This might be due to the absence of sufficient literature on the protein. STAB1 receptor works as a scavenger for low-density acetylated lipoproteins, and is capable of binding both Gram-negative and Gram-positive bacteria; thus, providing defence against bacterial infections [98, 99]. So, STAB1 constitutes a defence mechanism that might be getting compromised by the action of Ddn. Finally, a close interacting network was identified between KasA (S), AhpC (D), PncA (D), EthR (S), EthA (S) and Ndh (S), suggesting their collective effort in host immune system modulation. Also, KasA and AhpC showed maximum interactions with host proteins, and KasA as compared to AhpC is a stabilized protein. We further checked the sequence and structural homology of KasA with human proteins and observed that it had 35% sequence identity and a 0.90 TM align score [31] to human mitochondrial β-ketoacyl synthase protein. This high structure homology questions KasA as a potential drug target. But a recent *in vitro* study by Abrahams *et al*. (2016) has identified an indazole sulfonamide (GSK3011724A) that was found to specifically act on KasA of *Mycobacterium* and was not found to interfere in the functioning of human cells [100]. This study opens the aspect of focusing on drug testing to discover suitable drug targets against KasA. Rv0488 does not show high sequence identity to any human protein; however, due to the absence of structural information for the protein, structural homology analysis could not be performed.

Conclusions

In this study, we analysed 51 MDT genes in the complete genomes of 174 *M. tuberculosis* strains to explore the development of resistance to antibiotics. Selection pressure on the drug target genes was estimated by the rate of selection pressure (*dN/dS* ratios) and focus was drawn towards the most (*dN/dS* ≥0.70) and least (*dN/dS* <0.35) diversifying MDT genes. We found that two genes, *rpsL* and *gidB,* reported to be involved in the development of resistance against streptomycin, the first administered anti-TB drug, showed high *dN/dS* ratios (*rpsL*, *dN/dS*=1.27, and *gidB*, *dN/dS*=0.96), becoming the most positively selected genes among MDTs included in this study. The genes *ahpC*, *pncA*, *ethA*, *rpoB* and *dprE*, reported as being the drug targets of WHO-recommended first-line TB treatment drugs and involved in the development of resistance to these anti-TB drugs, were identified as most to moderately diversifying genes. The higher *dN/dS* values of *rpsL* and *gidB* genes, followed by those of *ahpC*, *pncA*, *ethA*, *rpoB* and *dprE* genes, reflects that the pressure of administered antibiotics may cause random mutations in the target genes that accumulate over time and result in positive selection pressure. The analysis also revealed stabilized genes as suitable targets for the development of effective anti-TB drugs. Protein–protein interaction analysis between the drug target proteins of *M. tuberculosis* and host proteins revealed that a stabilized protein, KasA, and a diversifying protein, AhpC, possessed

maximum interactions with the host proteins. In conclusion, this study suggests that *kasA*/KasA of *M. tuberculosis*, due to its lesser diversification and maximum host protein interactions, can act as a promising drug target for the development of effective therapeutic strategies and next-generation anti-TB drug regimens. Its structural homology with human mitochondrial β-ketoacyl synthase protein and the development of a new effective anti-TB drug targeting KasA require deeper investigation of its role as a potential drug target.

Funding information
The authors received no specific grant from any funding agency.

Acknowledgements
H.V., D.L. and C.D.R. acknowledge Ramjas College, University of Delhi, for providing infrastructural support. S.N. thanks the Council of Scientific and Industrial Research (CSIR) for providing doctoral fellowships. Part of the study was done under the Department of Biotechnology (DBT) – Star College Project to Ramjas College, University of Delhi. R. L. acknowledges the National Academy of Sciences, India (NASI), for financial support under the NASI-Senior Scientist Platinum Jubilee Fellowship Scheme.

Author contributions
C. D. R., H. V., S. N. and D. L. planned the study; S. V. and S. P. collected the genome and resistance gene data; H.V. and S.N. performed the analyses; C. D. R. supervised the work; C. D. R., H. V. and S. N. wrote the manuscript; R. L. and R. K. N. critically reviewed the manuscript and improved it. All authors read and approved the final manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M *et al*. Tuberculosis. *Nat Rev Dis Primers* 2016;2:16076.
2. Koch A, Mizrahi V. Mycobacterium tuberculosis. *Trends Microbiol* 2018;26:555–556.
3. WHO. *Global Tuberculosis Report*. Geneva: World Health Organization; 2019.
4. Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J *et al*. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* 2009;136:420–425.
5. Klopper M, Warren RM, Hayes C, Gey van Pittius NC, Streicher EM *et al*. Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. *Emerg Infect Dis* 2013;19:449–455.
6. Udwadia Z. Totally drug resistant-tuberculosis in India: the bad just got worse. *J Assoc Chest Physicians* 2016;4:41–42.
7. Hameed HMA, Islam MM, Chhotaray C, Wang C, Liu Y *et al*. Molecular targets related drug resistance mechanisms in MDR-, XDR-, and TDR-*Mycobacterium tuberculosis* strains. *Front Cell Infect Microbiol* 2018;8:114.
8. Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S *et al*. Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis. *N Engl J Med* 2015;373:1986–1988.
9. Singh P, Kumari R, Lal R. Bedaquiline: fallible hope against drug-resistant tuberculosis. *Indian J Microbiol* 2017;57:371–377.
10. Polsfuss S, Hofmann-Thiel S, Merker M, Krieger D, Niemann S *et al*. Emergence of low-level delamanid and bedaquiline resistance during extremely drug-resistant tuberculosis treatment. *Clin Infect Dis* 2019;69:1229–1231.
11. Simner PJ, Antar AAR, Hao S, Gurtowski J, Tamma PD *et al*. Antibiotic pressure on the acquisition and loss of antibiotic resistance genes in *Klebsiella pneumoniae*. *J Antimicrob Chemother* 2018;73:1796–1803.
12. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;393:537–544.
13. Rodriguez-Castillo JG, Pino C, Nino LF, Rozo JC, Llerena-Polo C *et al*. Comparative genomic analysis of *Mycobacterium tuberculosis* Beijing-like strains revealed specific genetic variations associated with virulence and drug resistance. *Infect Genet Evol* 2017;54:314–323.
14. Zimpel CK, Brandao PE, de Souza Filho AF, de Souza RF, Ikuta CY *et al*. Complete genome sequencing of *Mycobacterium bovis* SP38 and comparative genomics of *Mycobacterium bovis* and *M. tuberculosis* strains. *Front Microbiol* 2017;8:2389.
15. Palaniyandi K, Kumar N, Veerasamy M, Kabir Refaya A, Dolla C *et al*. Isolation and comparative genomics of *Mycobacterium tuberculosis* isolates from cattle and their attendants in South India. *Sci Rep* 2019;9:17892.
16. Zhang H, Li D, Zhao L, Fleming J, Lin N *et al*. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet* 2013;45:1255–1260.
17. Nguyen L. Antibiotic resistance mechanisms in *M. tuberculosis*: an update. *Arch Toxicol* 2016;90:1585–1604.
18. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* 2016;8:12–24.
19. Saeed AI, Sharov V, White J, Li J, Liang W *et al*. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–378.
20. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44 2016:W242–W245.
21. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 2007;23:673–679.
22. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T *et al*. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
24. Fukami K, Tateno Y. On the maximum likelihood method for estimating molecular trees: uniqueness of the likelihood point. *J Mol Evol* 1989;28:460–464.
25. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
26. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV *et al*. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 2018;35:773–777.
27. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015. https://www.R-project.org/
28. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl Acids Symp* 1999;41:95–98.
29. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–1659.
30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
31. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on TM-score. *Nucleic Acids Res* 2005;33:2302–2309.
32. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A *et al*. The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 2010;38:D525–D531.

33. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L *et al*. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res* 2010;38:D532–D539.
34. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R *et al*. The BioGRID interaction database: 2011 update. *Nucleic Acids Res* 2011;39:D698–D704.
35. Bader GD, Betel D, Hogue CW. BIND: the biomolecular interaction network database. *Nucleic Acids Res* 2003;31:248–250.
36. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList – 10 years after. *Tuberculosis* 2011;91:1–7.
37. Zhu L, Zhong J, Jia X, Liu G, Kang Y *et al*. Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. *Nucleic Acids Res* 2016;44:730–743.
38. Madhavilatha GK, Joseph BV, Paul LK, Kumar RA, Hariharan R *et al*. Whole-genome sequences of two clinical isolates of *Mycobacterium tuberculosis* from Kerala, South India. *J Bacteriol* 2012;194:4430.
39. Saelens JW, Lau-Bonilla D, Moller A, Xet-Mull AM, Medina N *et al*. Annotated genome sequences of 16 lineage 4 *Mycobacterium tuberculosis* strains from Guatemala. *Genome Announc* 2018;6:e00024-18.
40. Bloom BR, Murray CJ. Tuberculosis: commentary on a reemergent killer. *Science* 1992;257:1055–1064.
41. Moazed D, Noller HF. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* 1987;327:389–394.
42. Barnes PF, Bloch AB, Davidson PT, Snider DE. Tuberculosis in patients with human immunodeficiency virus infection. *N Engl J Med* 1991;324:1644–1650.
43. Snider DE, Roper WL. The new tuberculosis. *N Engl J Med* 1992;326:703–705.
44. Honore N, Cole ST. Streptomycin resistance in mycobacteria. *Antimicrob Agents Chemother* 1994;38:238–242.
45. Okamoto S, Tamaru A, Nakajima C, Nishimura K, Tanaka Y *et al*. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol* 2007;63:1096–1106.
46. Zhang Y, Mitchison D. The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis* 2003;7:6–21.
47. Cheng SJ, Thibert L, Sanchez T, Heifets L, Zhang Y. pncA mutations as a major mechanism of pyrazinamide resistance in *Mycobacterium tuberculosis*: spread of a monoresistant strain in Quebec, Canada. *Antimicrob Agents Chemother* 2000;44:528–532.
48. Khan MT, Malik SI, Ali S, Masood N, Nadeem T *et al*. Pyrazinamide resistance and mutations in pncA among isolates of *Mycobacterium tuberculosis* from Khyber Pakhtunkhwa, Pakistan. *BMC Infect Dis* 2019;19:116.
49. Rawat R, Whitty A, Tonge PJ. The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. *Proc Natl Acad Sci USA* 2003;100:13881–13886.
50. Wong CF, Shin J, Subramanian Manimekalai MS, Saw WG, Yin Z *et al*. AhpC of the mycobacterial antioxidant defense system and its interaction with its reducing partner thioredoxin-C. *Sci Rep* 2017;7:5159.
51. Wengenack NL, Rusnak F. Evidence for isoniazid-dependent free radical generation catalyzed by *Mycobacterium tuberculosis* KatG and the isoniazid-resistant mutant KatG(S315T). *Biochemistry* 2001;40:8990–8996.
52. Wang H-W, Chung C-H, Ma T-Y, Wong H. Roles of alkyl hydroperoxide reductase subunit C (AhpC) in viable but nonculturable *Vibrio parahaemolyticus*. *Appl Environ Microbiol* 2013;79:3734–3743.
53. Springer B, Master S, Sander P, Zahrt T, McFalone M *et al*. Silencing of oxidative stress response in *Mycobacterium tuberculosis*: expression patterns of ahpC in virulent and avirulent strains and effect of ahpC inactivation. *Infect Immun* 2001;69:5967–5973.
54. Cheng XF, Jiang C, Zhang M, Xia D, Chu LL *et al*. Mycobacterial interspersed repetitive unit can predict drug resistance of *Mycobacterium tuberculosis* in China. *Front Microbiol* 2016;7:378.
55. Singh V, Mani I, Chaudhary DK, Somvanshi P. The β-ketoacyl-ACP synthase from *Mycobacterium tuberculosis* as potential drug targets. *Curr Med Chem* 2011;18:1318–1324.
56. Gurumurthy M, Mukherjee T, Dowd CS, Singh R, Niyomrattanakit P *et al*. Substrate specificity of the deazaflavin-dependent nitroreductase from *Mycobacterium tuberculosis* responsible for the bioreductive activation of bicyclic nitroimidazoles. *FEBS J* 2012;279:113–125.
57. Lee BM, Harold LK, Almeida DV, Afriat-Jurnou L, Aung HL *et al*. Predicting nitroimidazole antibiotic resistance mutations in *Mycobacterium tuberculosis* with protein engineering. *PLoS Pathog* 2020;16:e1008287.
58. Hoffmann H, Kohl TA, Hofmann-Thiel S, Merker M, Beckert P *et al*. Delamanid and bedaquiline resistance in *Mycobacterium tuberculosis* ancestral Beijing genotype causing extensively drug-resistant tuberculosis in a Tibetan refugee. *Am J Respir Crit Care Med* 2016;193:337–340.
59. Baulard AR, Betts JC, Engohang-Ndong J, Quan S, McAdam RA *et al*. Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem* 2000;275:28326–28331.
60. Donald PR, McIlleron H. Antibuerculosis drugs. *Tuberculosis: a Comprehensive Clinincal Reference*; 2009. pp. 608–617.
61. Willand N, Dirie B, Carette X, Bifani P, Singhal A *et al*. Synthetic EthR inhibitors boost antituberculous activity of ethionamide. *Nat Med* 2009;15:537–544.
62. Morlock GP, Metchock B, Sikes D, Crawford JT, Cooksey RC. ethA, inhA, and katG loci of ethionamide-resistant clinical *Mycobacterium tuberculosis* isolates. *Antimicrob Agents Chemother* 2003;47:3799–3805.
63. Vilchèze C, Jacobs WR. Resistance to isoniazid and ethionamide in *Mycobacterium tuberculosis*: genes, mutations, and causalities. *Microbiol Spectr* 2014;2:MGM2-0014-2013.
64. Dookie N, Rambaran S, Padayatchi N, Mahomed S, Naidoo K. Evolution of drug resistance in *Mycobacterium tuberculosis*: a review on the molecular determinants of resistance and implications for personalized care. *J Antimicrob Chemother* 2018;73:1138–1151.
65. Mortimer TD, Weber AM, Pepperell CS. Signatures of selection at drug resistance loci in *Mycobacterium tuberculosis*. *mSystems* 2018;3:e00108-17.
66. Copp JN, Akiva E, Babbitt PC, Tokuriki N. Revealing unexplored sequence-function space using sequence similarity networks. *Biochemistry* 2018;57:4651–4662.
67. Finken M, Krischner P, Meier A, Wrede A, Bottger EC. Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol* 1993;9:1239–1246.
68. Hlaing YM, Tongtawe P, Tapchaisri P, Thanongsaksrikul J, Thawornwan U *et al*. Mutations in streptomycin resistance genes and their relationship to streptomycin resistance and lineage of *Mycobacterium tuberculosis* Thai isolates. *Tuberc Respir Dis* 2017;80:159–168.
69. Sun Y-J, Luo J-T, Wong S-Y, Lee ASG. Analysis of *rpsL* and *rrs* mutations in Beijing and non-Beijing streptomycin-resistant *Mycobacterium tuberculosis* isolates from Singapore. *Clin Microbiol Infect* 2010;16:287–289.
70. Tudó G, Rey E, Borrell S, Alcaide F, Codina G *et al*. Characterization of mutations in streptomycin-resistant *Mycobacterium tuberculosis* clinical isolates in the area of Barcelona. *J. Antimicrob* 2010;65:2341–2346.
71. Ilin AI, Kulmanov ME, Korotetskiy IS, Islamov RA, Akhmetova GK *et al*. Genomic insight into mechanisms of reversion of antibiotic resistance in multidrug resistant *Mycobacterium tuberculosis* induced by a nanomolecular iodine-containing complex FS-1. *Front Cell Infect Microbiol* 2017;7:151.

72. Ilin AI, Kulmanov ME, Korotetskiy IS, Lankina MV, Akhmetova GK *et al*. Constraints of drug resistance in *Mycobacterium tuberculosis* - prospects for pharmacological reversion of susceptibility to antibiotics. *Open Conf Proc J* 2017;8:33–43.
73. Wong SY, Lee JS, Kwak HK, Via LE, Boshoff HIM *et al*. Mutations in gidB confer low-level streptomycin resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2011;55:2515–2522.
74. Perdigão J, Macedo R, Machado D, Silva C, Jordão L *et al*. GidB mutation as a phylogenetic marker for Q1 cluster *Mycobacterium tuberculosis* isolates and intermediate-level streptomycin resistance determinant in Lisbon, Portugal. *Clin Microbiol Infect* 2014;20:O278–O284.
75. Verma JS, Gupta Y, Nair D, Manzoor N, Rautela RS *et al*. Evaluation of gidB alterations responsible for streptomycin resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* 2014;69:2935–2941.
76. Jagielski T, Ignatowska H, Bakuła Z, Dziewit L, Napiorkowska A *et al*. Screening for streptomycin resistance-conferring mutations in *Mycobacterium tuberculosis* clinical isolates from Poland. *PLoS One* 2014;9:e100078.
77. Sun H, Zhang C, Xiang L, Pi R, Guo Z *et al*. Characterization of mutations in streptomycin-resistant *Mycobacterium tuberculosis* isolates in Sichuan, China and the association between Beijing-lineage and dual-mutation in gidB. *Tuberculosis* 2016;96:102–106.
78. Mestdagh M, Fonteyne PA, Realini L, Rossau R, Jannes G *et al*. Relationship between pyrazinamide resistance, loss of pyrazinamidase activity, and mutations in the *pncA* locus in multidrug-resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrobial Agents Chemother* 1999;43:2317–2319.
79. Juréen P, Werngren J, Toro JC, Hoffner S. Pyrazinamide resistance and pncA gene mutations in *Mycobacterium tuberculosis*. *Antimicrobial Agents Chemother* 2008;52:1852–1854.
80. Karmakar M, Rodrigues CHM, Horan K, Denholm JT, Ascher DB. Structure guided prediction of pyrazinamide resistance mutations in pncA. *Sci Rep* 2020;10:1875.
81. Karmakar M, Globan M, Fyfe JAM, Stinear TP, Johnson PDR *et al*. Analysis of a novel pncA mutation for susceptibility to pyrazinamide therapy. *Am J Respir Crit Care Med* 2018;198:541–544.
82. Chiu Y-C, Huang S-F, Yu K-W, Lee Y-C, Feng J-Y *et al*. Characteristics of pncA mutations in multidrug-resistant tuberculosis in Taiwan. *BMC Infect Dis* 2011;11:240.
83. Sreevatsan S, Pan X, Zhang Y, Deretic V, Musser JM. Analysis of the oxyR-ahpC region in isoniazid-resistant and -susceptible *Mycobacterium tuberculosis* complex organisms recovered from diseased humans and animals in diverse localities. *Antimicrob Agents Chemother* 1997;41:600–606.
84. Brossier F, Veziris N, Truffot-Pernot C, Jarlier V, Sougakoff W. Molecular investigation of resistance to the antituberculous drug ethionamide in multidrug-resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrobial Chemother* 2011;55:355–360.
85. Lee BM, Harold LK, Almeida DV, Afriat-Jurnou L, Aung HL *et al*. Predicting nitroimidazole antibiotic resistance mutations in *Mycobacterium tuberculosis* with protein engineering. *PLoS Pathog* 2020;16:e1008287.
86. Ramirez LMN, Vargas KQ, Diaz G. Whole genome sequencing for the analysis of drug resistant strains of *Mycobacterium tuberculosis*: a systematic review for bedaquiline and delamanid. *Antibiotics* 2020;9:133.
87. Zhang F, Li S, Wen S, Zhang T, Shang Y *et al*. Comparison of *in vitro* susceptibility of mycobacteria against PA-824 to identify key residues of Ddn, the deazoflavin-dependent nitroreductase from *Mycobacterium tuberculosis*. *Infect Drug Resist* 2020;13:815–822.
88. Kirstein M, Sanz L, Quiñones S, Moscat J, Diaz-Meco MT *et al*. Cross-talk between different enhancer elements during mitogenic induction of the human stromelysin-1 gene. *J Biol Chem* 1996;271:18231–18236.
89. Rekdal C, Sjottem E, Johansen T. The nuclear factor SPBP contains different functional domains and stimulates the activity of various transcriptional activators. *J Biol Chem* 2000;275:40288–40300.
90. Dunn C, Wiltshire C, MacLaren A, Gillespie DAF. Molecular mechanism and biological functions of c-Jun N-terminal kinase signalling via the c-Jun transcription factor. *Cellular Signalling* 2002;14:585–593.
91. Tan NY, Khachigian LM. Sp1 phosphorylation and its regulation of gene transcription. *Mol Cell Biol* 2009;29:2483–2488.
92. Scott FL, Stec B, Pop C, Dobaczewska MK, Lee JJ *et al*. The Fas-FADD death domain complex structure unravels signalling by receptor clustering. *Nature* 2009;457:1019–1022.
93. Shen Y, Song Z, Lu X, Ma Z, Lu C *et al*. Fas signaling-mediated TH9 cell differentiation favors bowel inflammation and antitumor functions. *Nat Commun* 2019;10:2924.
94. Vafiadaki E, Arvanitis DA, Pagakis SN, Papalouka V, Sanoudou D *et al*. The anti-apoptotic protein HAX-1 interacts with SERCA2 and regulates its protein levels to promote cell survival. *Mol Biol Cell* 2009;20:306–318.
95. Chen HY, Sharma BB, Yu L, Zuberi R, Weng IC *et al*. Role of galectin-3 in mast cell functions: galectin-3-deficient mast cells exhibit impaired mediator release and defective JNK expression. *J Immunol* 2006;177:4991–4997.
96. Haudek KC, Spronk KJ, Voss PG, Patterson RJ, Wang JL *et al*. Dynamics of galectin-3 in the nucleus and cytoplasm. *Biochim Biophys Acta* 2010;1800:181–189.
97. Dai SY, Nakagawa R, Itoh A, Murakami H, Kashio Y *et al*. Galectin-9 induces maturation of human monocyte-derived dendritic cells. *J Immunol* 2005;175:2974–2981.
98. Beinke S, Robinson MJ, Hugunin M, Ley SC. Lipopolysaccharide activation of the TPL-2/MEK/extracellular signal-regulated kinase mitogen-activated protein kinase cascade is regulated by IκB kinase-induced proteolysis of NF-κB1 p105. *Mol Cell Biol* 2004;24:9658–9667.
99. Adachi H, Tsujimoto M. FEEL-1, a novel scavenger receptor with *in vitro* bacteria-binding and angiogenesis-modulating activities. *J Biol Chem* 2002;277:34264–34270.
100. Abrahams KA, Chung C, Ghidelli-Disse S, Rullas J, Rebollo-López MJ *et al*. Identification of KasA as the cellular target of an anti-tubercular scaffold. *Nat Commun* 2016;7:12581.