

Whole genome sequencing and global metabolome profiling of clinical *Mycobacterium tuberculosis* isolates provide insights to their drug resistance status.

Ashutosh Sahoo,^{1,#} Amit Kumar Mohapatra,^{2,3,#} Haripriya Priyadarsini,^{2,#} Raghuv eer Varma Pemmadi,² Anjan Das,^{4,*} Tenzin Choedon,⁵ Chaitali Nikam,⁶ Rajendra Kumar Behera,³ Shyam K. Masakapalli,^{1,*} Ranjan K. Nanda^{2,*}

¹School of Biosciences and Bioengineering, Indian Institute of Technology Mandi, Kamand, Himachal Pradesh, 175005, India.

²Translational Health Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, 110067, India.

³School of Life Sciences, Sambalpur University, Sambalpur, Odisha, 768019, India.

⁴Department of Respiratory Medicine, Agartala Government Medical College, Kunjaban, Agartala, Tripura, 799006, India.

⁵Structural Biology Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, 110067, India

⁶Thyrocare Technologies, Mumbai, Maharashtra, 110067, India

#: Equal contribution and author list is arranged based on seniority.

*Corresponding author(s)

Dr Anjan Das, Associate Professor, Department of Respiratory Medicine, Agartala Government Medical College, Kunjaban, Agartala- 799006, Tripura, India. E. mail: anjan8870@gmail.com and/or

Dr. Shyam K. Masakapalli, Associate Professor, School of Biosciences and Bioengineering, Indian Institute of Technology Mandi, Kamand 175005, Himachal Pradesh, India. E. mail: shyam@iitmandi.ac.in and/or

Dr. Ranjan K. Nanda, Group Leader, Translational Health Group, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067, India. E. mail: ranjan@icgeb.res.in

Abstract

Whole genome sequence analysis of the *Mycobacterium tuberculosis* (Mtb) isolates show correlation to their drug resistance phenotype which may also reflect in their global metabolome. In this report, clinical Mtb isolates (S1, S4, S5, S6, S7, S10) harvested from the sputum of tuberculosis patients were characterized using drug sensitive test (DST), electron microscope, whole genome sequencing (WGS) and metabolomics. Majority of these Mtb isolates showed similar size (length: 1.0–3.2 μm ; width: 0.32–0.52 μm) to the H37Rv Mtb strain whereas significant variations were observed in their growth kinetics, WGS and metabolome profiles. In-silico drug resistance prediction, from the WGS data (single-nucleotide polymorphisms (SNP) pattern) of these Mtb isolates showed resistance to tuberculosis drugs and matched with DST results. Differences in the genes involved in stress response, pathogenicity, drug efflux pumps were observed between isolates but genes of the central carbon metabolic pathways and amino acid metabolism were conserved. Gas chromatography and mass spectrometry (GC-MS) based metabolite profiling of these clinical isolates identified 291 metabolites involved in various metabolic pathways and a sub set of these metabolites (glutamic acid, aspartic acid and serine) contributed to the drug resistance patterns. These clinical Mtb isolates could be useful as alternate reagent for understanding host pathogen interaction and the pipeline used for WGS analysis could be used to predict drug resistance pattern of new Mtb isolates.

Key words: Tuberculosis, *Mycobacterium tuberculosis*, drug resistance, whole genome sequencing, metabolome

Introduction

Tuberculosis disease, caused by the infection of *Mycobacterium tuberculosis* (Mtb), affects millions of lives and is a major global issue.^{1,2} Increase in the number of multi-drug resistant tuberculosis cases in certain geographical settings is a cause of concern.^{2,3} Characterizing the clinical Mtb isolates and their drug resistance pattern, at the time of diagnosis, is critical to initiate appropriate therapeutic interventions and reduce community transmission.

However, current method of drug resistance phenotyping is carried out by culture method and it has a more than three weeks of long result waiting time.⁴ Recent advancement in whole genome sequencing (WGS) of Mtb isolates provide an alternate method for predicting drug resistance status, with certain refinement, it may reduce the result waiting period to initiate appropriate and effective personalized therapy.^{5,6}

In this study, Mtb isolates from the tuberculosis patient sputum samples were characterized by their microbiological phenotyping, morphology, WGS and metabolome data. In silico analysis of the WGS data of the Mtb isolates were used to identify single nucleotide polymorphisms (SNPs) and predict drug resistance patterns. Mass spectrometry was employed to monitor the global metabolite levels of these clinical isolates and compared with reference strain like H37Rv Mtb. The adopted informatics pipeline could be useful for predicting drug resistance phenotype and the clinical Mtb isolates could be useful reagents for better understanding host pathogen interaction.

Methodology

Study subject recruitment and isolation of clinical Mycobacterial isolates: This study was approved by the human ethics committees of the Agartala Government Medical College, Agartala (F.4[6-9]/AGMC/Academic/IEC Committee/2015/8965, dated 25 April 2018) and International Centre for Genetic Engineering and Biotechnology, New Delhi (ICGEB/IEC/2017/07).⁷ Subjects with symptoms of tuberculosis disease like cough for 2 weeks or more, weight loss, fevere, night sweat reporting to the outpatient department of the hospital were recruited after receiving their signed informed consent forms. Sputum samples (day 0 and day 1) of these subjects were taken for sputum microscopy at hospital and culture test for drug sensitivity test using Mycobacteria Growth Indicator Tube (MGIT) in BACTEC MGIT 960 System at a National Accreditation Board for Testing and Calibration Laboratories (NABL) accredited laboratory following WHO guidelines. Clinical Mtb strains (S1, S4, S5, S6, S7, and S10) were isolated from the colonies of culture positive MGIT tubes and expand using 7H9 media supplemented with Oleic acid-albumin-dextrose-catalase (OADC, 10%), Tween-80 (0.05%, vol/vol) and glycerol (0.2%, vol/vol) at 37°C in a shaker incubator at 180 rpm in BSL-III laboratory.^{8,9} Equal number (~10⁷) of each clinical isolates and laboratory Mtb strain (H37Rv) were cultured for 28 days using similar conditions and their OD was monitored at 600 nm using spectrophotometer (Biorad, USA) on alternate days for growth kinetics.^{8,9} Aliquots of these Mtb samples were pre-processed and used for further characterization using transmission electron microscope (TEM), whole genome sequencing (WGS) and mass spectrometry (MS) (**Figure 1**).

Transmission electron microscopy of clinical isolates and laboratory Mycobacterial strain: Clinical isolates and laboratory Mtb strain (S1, S4, S5, S6, S7, S10, H37Rv; ~5x10⁸ colony forming units:CFU/ml) were centrifuged at 3,000 g for 5 min. The resulting bacterial pellet was washed with phosphate buffer saline (20 mM, PBS), and incubated with Gluteraldehyde (2% in PBS:0.1 M) for 1 hour at room temperature for fixing in BSL-III laboratory.^{10,11} These bacterial preparation was washed using PBS buffer (20 mM) and twice with distilled water. After a short centrifugation, the resulting bacteria were reconstituted in distilled water (10-20 µl) before loading on a copper grid covered by carbon formvar (300 mesh, Ted Pella, USA) and stained with Uranyl Acetate (1%). The Mtb loaded grids were examined under a transmission electron microscopy (TEM, Tecnai G2 spirit, HT 120 kV) and images were captured. Using ImageJ software (vs 1.53e), the length and width of 15 individual bacteria of each samples were measured for comparative analysis.

Global metabolite profiling of of clinical isolates and laboratory Mycobacterial strain: Five biological replicates of each Mtb samples (S1, S4, S5, S6, S7, and H37Rv), except S10 (4 replicates) were harvested at OD=1.0 at 600 nm and centrifuged at 1,500 g for 10 min at 4°C¹². After washing the bacterial pellets twice with NaCl solution (0.9%) were centrifuged at 1,500 g for 10 min at 4°C and pre-chilled (at -80°C) Acetonitrile: Methanol: Water (40: 40: 20, 2 mL) was added.¹³ Spike in standard (Ribitol, 4 µg) was added to it and metabolites were extracted using bead beating method. Briefly, after adding Zirconium beads (0.1 mm, Biospec, CatNo. 11079101z) to the bacterial pellets, 4 cycles (on/off: 40 sec/40 sec) of bead beating was carried out. The bacterial lysates were centrifuged at 2,000 g for 3 min at 4°C and the supernatant was

filtered (0.22 μm , nylon filters) before completely drying at 50°C using a speed vac evaporator (Labconco, USA). Toluene (100 μL) was added to each sample, again dried for at least 20 mins using speed vac at 50°C. The extracted metabolites were incubated with O-Methoxiamine HCl (40 μL , 20 mg/mL) at 60°C for 2 h at 400 rpm in a thermomixer (Eppendorf, USA). N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA, 70 μL) was added to the reaction mixture and incubated at 60°C for 1 h at 400 rpm in a thermomixer. Within 48 hours of derivatization, the GC-MS data acquisition using a Pegasus 4D time of flight (TOF) mass spectrometer (Leco, USA) coupled to a GC system (7890A GC from Agilent Technologies, USA) was completed.¹² Derivatized samples (1 μL) were injected to the GC column (HP-5MS, Agilent) in splitless mode using an automatic liquid sampler (Rail system, Gestrel) and helium at a constant flow of 1 mL/min was used as the carrier gas. Separation of the metabolites were achieved using a GC oven temperature programme, from 50 to 200°C at a ramp of 8.5°C/min, from 200 to 280°C at a ramp of 6°C/min and a hold time of 5 min at final temperature. A mass range of 33 to 600 m/z at an EM voltage of 70 eV with a 20 spectra/sec scan rate and 600 sec solvent delay was selected during data acquisition. Using statistical compare feature of the ChromaTOF, all GC-MS data were analysed using parameters, minimum presence in >50% of the samples within group, signal to noise ratio (S/N) > 20, and peak width of 1.35. Analytes qualifying these criteria were aligned. Out of the identified 413 analytes, after merging peaks of similar analyte, removing silane related molecules resulted a set of 291 analytes which was selected for chemometric data analysis using Metaboanalyst 5.0.^{14,15} Data normalised using the spike in standard as the reference feature and auto-scaling was used for principal component analysis (PCA) and partial least square discriminate analysis (PLS-DA) model was built. Analytes qualifying a variable important project parameter (VIP) score ≥ 1.8 of the PLS-DA model were selected as important. Identity of the important analytes were validated by running commercial standards using similar derivatization and GC-MS method.

Whole genome sequencing analysis of the clinical Mycobacterial isolates and in silico prediction of drug resistance status: Genomic DNA from the clinical Mtb isolates were extracted from the exponentially growing cultures (OD=1.0 at 600 nm) following earlier reported method and quantified using Qubit 3.0 (ThermoFischer, USA)⁷. Quality of the extracted genetic material was monitored by running in agarose gel (0.8%) and TapeStation (Agilent, USA). The extracted genetic materials were sequenced using a HiSeq 4000 system (Illumina, Inc., San Diego, CA, USA) and the TruSeq SBS v3 reagent kit (2 \times 150 bp) was used for library preparation. The WGS data of these clinical strains (S1, S4, S5, S6, S7, S10) are available in the NCBI database (BioProject: PRJNA595470).⁷ Additionally, sequence data of laboratory reference strains (H37Rv: NC_000962.3 and H37Ra: NC_009525.1), lineage-specific clinical isolates from India and abroad were selected from NCBI for comparative informatics analysis (**Supplementary Table 1**). Informatics analyses included genome comparison (alignment, Average Nucleotide Identity: ANI, Single Nucleotide Polymorphisms: SNPs based phylogenetic closeness analysis and lineage classification), pathway comparison (reconstruction of metabolic pathways using RAST and BlastKOALA) and drug resistance prediction. ANI calculator was carried out employing the OrthoANIu algorithm with a minimum identity of 70%, minimum alignment length of 50 bps was used to calculate the GC contents and nucleotide identity was determined by considering the reference strain (H37Rv Mtb).¹⁶ Mauve-multiple Genome alignment and progressive-Mauve that uses both local alignment HODX scoring matrix to determine potential anchors from the ungapped genome sequence of selected strains was considered for complete genome alignment using H37Rv as the reference strain.^{17,18} The alignment was based on the default parameter settings that was optimised for bacterial whole genome alignment. A phylogenetic tree of the selected strains was constructed using CSI Phylogeny v1.4¹⁹ and visualized using FastTree tool.²⁰ And SNPs qualifying with a minimum depth of SNP position to be 70 \times , quality >90, relative depth >10%, the minimum distance of 10 bp between SNPs and Z-score >1.96 were considered. Employing Jonny 2.1, the common SNPs in the studied clinical isolates and from the reported lineage specific strains were monitored.²¹ Using Res Finder (v 4.0) the susceptibility pattern of these clinical isolates to anti-TB drugs (rifampicin, isoniazid, macrolides, fluoroquinolone, kanamycin, and clarithromycin etc.) and probable gene mutations conferring this resistance were predicted.^{22,23} A chromosomal point mutation with >60-90% and antimicrobial resistance (AMR) genes >60-90% threshold were selected and reconfirmed using Resistance gene Identifier (RGI) from CARD and compared with the microbiology drug phenotyping data.²⁴

Genome annotation and unique gene identification: The WGS Mtb sequence data were analysed employing Rapid Annotation using Subsystem Technology (RAST) with tRNAscan-SE and RNAmmer to annotate tRNA and rRNAs.^{25,26} GLIMMER2 was used for annotating coding protein sequence.²⁷ The putative genes identified were verified by establishing phylogenetic context to proteins available in the

FIGfam database using the BLAST against the corresponding FIGfam.²⁸ RAST tool was used to classify the genes into subsystems. Orthovenn2 was used for clustering of genes among new clinical strains to segregate orthologous genes of Mycobacterium species based on similarities with reference H37Rv strain.²⁹ An E-value > 1e-5 and inflations > 1.5 were selected for identification of homogeneous clusters and unique genes present in these clinical isolates were also identified.

Metabolic pathway reconstruction by assigned KEGG orthology (KO numbers) to newly annotated genes: Annotated genes of all these six new clinical isolates were assigned with KEGG orthology (KO) numbers using BlastKOALA.³⁰ These KO numbers were compared using G2KO with the KO numbers of the Mtb strains available in the KEGG database.³¹ Comparative pathway analysis between isolates were carried out using the KEGG mapper pathway reconstruction tool with particular focus to glycolysis, gluconeogenesis, tricarboxylic acid (TCA) cycle, pentose phosphate pathway and amino acid metabolic pathways. Using KEGG mapper, presence of the transporters including efflux pumps, conferring drug resistance and unique transporters were identified. Morpheus tool was used to identify the presence of absence of the enzymes.³²

Statistical Analysis: All data points were presented as mean ± stdev and t-test with a P-value < 0.05 was considered to be statistically significant.³³ All statistical procedures were performed using GraphPad Prism 9.3.1.

Results

Characteristics of the clinical Mycobacterial strains: From the microbiology data, the clinical Mtb isolates S1, S4, S5 and S10 were found to be susceptible to rifampicin, isoniazid, ethambutol, ofloxacin, capreomycin and amikacin drugs. Whereas S6, S7 and S10 were resistant to rifampicin, isoniazid, streptomycin, clofazimine. Upon culture in 7H9 media, these clinical isolates (S1, S4, S5, S6, S7, and S10) showed slower growth kinetics compared to the virulent laboratory strains (H37Rv) (**Figure 2A**). The clinical isolate S1 showed slower growth kinetics with a growth rate (μ) of 0.13/day, whereas the rest was in between of 0.14–0.2/day. The cells had a standard rod shape (1.0–3.2 μ m in length and 0.32–0.52 μ m in width), except S4, rest of the clinical Mtb isolates showed similar length and width to the laboratory H37Rv Mtb strain (**Figure 2B, 2C, 2D**).

Whole-genome sequence of the clinical Mycobacterial strains: All the raw sequences of the clinical Mtb isolates showed different genome sizes and based on fastp v0.20.0 analysis showed a coverage of 100× and pipeline employed for informatics analysis presented in **Figure 3A**.³⁴ Among them S5 had a larger genome size (4,456,211 bp) followed by S10 (4,405,137 bp) and S6 (4,401,104 bp) with contig size varied from 306, 210, 124 between isolates respectively. The sequence length and contig distribution of clinical isolates (S1:4,399,078 bp:111, S4:4,396,813bp: 202, S7:4,391,229 bp: 145) showed certain difference from the other isolates (**Supplementary Table 2**). The N50 value, ANI and SNPs of individual isolates tabulated in **Supplementary Table 2**. The GC% of all the clinical isolates was 65.6% except for S10 (65.5%). The ANI score of the S7 isolate was higher (99.79%) than other isolates compared to H37Rv reference strain. The genome sequence alignment of all clinical isolates with H37Rv showed a high degree of similarity (**Figure 3B**). In the S5 isolate, 3148 SNPs were identified and similarly, 2494, 2519, 1950, 1280 and 2870 SNPs were observed in S1, S4, S6, S7 and S10 isolates respectively (**Supplementary Table 2**). A set of 414 SNPs were found to be common to all study isolates and the rest were unique (**Supplementary Figure 1A**). Using high-quality SNPs (quality score > 90), a maximum-likelihood phylogenetic tree was constructed (**Figure 3C**). The S5 and S1, S4 and S10 isolates have a distance of 0.093 and 0.0858 respectively with RGTB423 that belongs to East-African Indian lineage. S4 and S1, S10 clade shared internal nodes with a branch distance of 0.103. Larger branch length of 0.1749 between S6 and CDC5079 isolates from the Beijing lineage was observed. The S7 isolate was grouped closer to the reported H37Rv and H37Ra reference strains with a branch length of 0.283. The SNPs distribution between isolates showed a common set of 1558 in clinical isolates (S1, S4, S5, S10) and a comparison of S6 with reported strains belonging to Beijing lineage showed 1387 SNPs (**Supplementary Figure 1B, 1C**).

Prediction of drug resistance status of the clinical Mycobacterial isolates: Reported SNPs that confer resistance to TB drugs and identified in Mtb clinical isolates presented in **Supplementary Table 3 and Supplementary figure 2**. All the mutated SNPs that become the base for predicting resistance or susceptible status in clinical isolates to known drugs like ofloxacin, capreomycin, kanamycin, amikacin, moxifloxacin, clofazimine and ethanamide were tabulated (**Supplementary Table 4**). Based on these mutation patterns, in-silico analysis predicted resistance to rifampicin, isoniazid and streptomycin drugs in S1 and S6 isolates (**Table 1**). The S7 isolate showed resistance to isoniazid. The predicted drug resistance status of the clinical isolates matched with the microbiological laboratory results (**Table 1**). The total coding DNA sequences (CDS) of these clinical isolates showed variation and the annotated genes contributes to various functions (**Supplementary Table 5**). Approximately ~23-24% of CDS were categorised into the virulence, disease, and defence categories with additional numbers involved in the invasion and intracellular resistance categories (**Supplementary Table 6**). Except in S6 isolate, amino acid metabolism associated genes were annotated in all the study isolates. A set of copper resistance genes and transporters were predicted to be present in S4, S5 and S6 isolates. Additional genes involved in regulation, signalling, DNA, protein, carbohydrates, fatty acids, lipids, and isoprenoids metabolism were identified in these clinical isolates (**Supplementary Table 6**).

Clustering of orthologous genes and comparative metabolic pathways in the clinical Mycobacterial isolates: A set of 3967 gene clusters consisting of 27,915 proteins were shared between clinical isolates and laboratory Mtb strains (**Figure 4A, 4B**) and isolate-specific unique gene clusters (S1:4, S4:5, S5:2 and S10:4) were observed. Metabolic pathway construction showed conserved Glycolysis, Gluconeogenesis and TCA cycles in these clinical isolates and reference Mtb strains (**Supplementary Figure 3A**). Genes of clinical isolates involved in the Pentose phosphate pathway as well as pentose and glucuronate interconversion were found to be conserved (**Supplementary Figure 3C**). These clinical isolates had all the enzymes required for biosynthesis and degradation of all proteogenic amino acids (**Supplementary Figure 3D**). Certain unique enzymes or gene products were predicted to be present in these clinical isolates. Interestingly, with respect to the laboratory reference strains (H37Ra and H37Rv), many transporters were missing with few unique transporters identified in these clinical isolates which may have a role in their pathogenicity (**Figure 4C**).

Global metabolite profile of these clinical ioslates showed significant differences: Metabolome of the mycobacterial isolates and laboratory strain were analyzed using GC-MS after derivatization (**Figure 5A**). Important to mention that the global metabolite analysis of the laboratory virulent Mtb H37Rv strain clustered it away from all the clinical isolates. PCA plot showed that metabolites of S1 isolate cluster separately from the other clinical isolates and laboratory reference H37Rv strain (**Supplementary Figure 4**). PLS-DA plot showed overlap between S6 and S7 as well as S4 and S5 (**Figure 5B**). S4 and S5 of East-African Indian lineage showed lower abundance of aspartic acid and glutamic acid with respect to S1 and S10. S4 showed a low glutamic acid level with lower aspartic acid level in S5. In the S6 isolate of Beijing lineage, serine abundance was the highest among all the clinical isolates. S7 of Euro American lineages showed lower glutamic acid and aspartic acid levels with respect to H37Rv. Glutamic acid level was found to be lower in S4 with highest in H37Rv strain. Aspratic acid level was higher in S1 isolate with a minimum level in S7. Serine levels in S6 clinical isolate was higher with very low level in S10 (**Figure 5C**).

Discussion

Post COVID19, tuberculosis disease related cases are increasing including the number of drug resistant cases. Appropriate personalized treatment to these patients is the need of the hour and providing the drug resistance pattern in shorter time span will introduce personalized therapeutic interventions to cure the patients. Mtb the pathogen causing tuberculosis is divided into multiple lineages and show varied drug resistance patterns, virulence and transmission dynamics.³⁵ Variation in virulence of these Mtb isolates from clinical materials showed dependence to patient immunity, degree of exposure and other factors including environment. It is critical to characterize clinical Mtb isolates at multiple molecular levels like WGS and their metabolic components to rationally predict their drug resistant patterns.

The difference in the growth kinetics observed in these clinical isolates could be partially explained to the lack of adoption to the available carbon and nitrogen components present in the 7H9 media. These clinical isolates, irrespective of their drug resistance and susceptible status, showed similar size with the laboratory

H37Rv strain. One of the clinical isolate S4 was found to be smaller in size than the rest of the Mtb isolates. Similar morphological parameters were reported from clinical Mtb isolates from African, South east Asian and Russian tuberculosis patients. As expected the average whole genome size of the Mtb isolates was ~4.4 Mbps with a high GC content (65.5-65.6%). The genome sequence of these clinical strains were compared with available drug susceptible and resistant Mtb strains reported from East Asia, Africa, Russia, North America and East Europe. A set of 414 SNPs were found to be common among the Mtb clinical isolates. These SNPs distribution provided insights into the evolutionary pattern of these new clinical isolates. In the study clinical site, circulating Mtb isolates of East African-Indian (S1, S4, S5, S10), Beijing (S6) and Euro-American lineages (S7) were observed. Hence identifying the genomic markers, in clinical isolates, which could predict the drug resistance pattern might rationalise the therapeutic modalities.³⁶

SNP clustering analysis of all clinical Mtb isolates (S1, S4, S5 and S10), of Beijing lineage (S6, CCDC5079, CCDC5180, BT1, BT2) and clinical isolates of EAI lineage strains showed similarity of ~30% (1387) and ~34% (1558) SNPs respectively. These clustered SNPs among lineages of Mtb strains shows their adaptation, evolutionary changes and mutation conferring resistance to tuberculosis drugs. In silico prediction of the drug resistance patterns of the S1 and S6 clinical isolates showed resistance to rifampicin, isoniazid, and streptomycin and susceptible to capromycine and moxifloxacin corroborating culture test results. However, SNP profile of S6 isolate showed resistance to ethambutol and did not match to the microbiology data. These observations indicate that the algorithms for predicting drug resistance patterns of clinical Mtb isolates need further refinement and provide useful clinically relevant information.³⁷

In the clinical drug resistance isolates few unique proteins were identified which may be partially involved in conferring drug resistance status. For example, EsxL (Rv1198) protein down regulating MHC-II by suppressing the nitric oxide production and p38 MAPK pathway in the macrophages and helps drug resistance Mtb to remain undetected for longer time for a delayed immune response.³⁸ Genome annotation in these clinical isolates, predicted some genes with increased copy numbers, belonging to fatty acid synthesis and utilization, DNA metabolism and stress responses. Such alterations reported to help drug resistant Mtb to survive in the limited nutrients and extreme stress condition for long.^{39,40} Interestingly, the proteins involved in the central carbon metabolism (glycolysis, gluconeogenesis, TCA cycle and Pentose phosphate pathways) were conserved in these clinical isolates irrespective of their drug resistance status. Also the genes involved in amino acid metabolism in these clinical isolates were found to be conserved. This suggests that possibly all these pathogens survive in the host system by maximum utilization of different carbon sources like glucose and fatty acids.⁴¹ Previous studies showed that fatty acids taken from the host gets utilized via TCA cycle for further gluconeogenesis to fulfil energy demand and production of glucogenic amino acids.⁴²

The growth and replication of the pathogen depend on the nutritional niche of the host cell and the transporters across the membrane play a significant role. ABC transporters act as efflux pumps expelling antimicrobial agents to confer drug resistance to the pathogen.^{43,44} Comparative genomic analysis revealed presence of ABC transporters (Rv0194, Rv2686c, Rv2687c, Rv2688c) and unique ones (OmpA, Glutathione-S-transferase, opp transporters) in these clinical isolates. Rv0194, is an ATP binding cassette transporter provide resistance to ampicillin, streptomycin, and chloramphenicol and the others (Rv2686c, Rv2687c, Rv2688c) provide resistance to fluoroquinolones (**Figure 4D**).^{45,46} Similarly, peptidoglycan binding protein transporters (ArfA and OmpA) which act as porins to maintain pH by secreting ammonia was identified in these clinical isolates.⁴⁷ OmpA gene was annotated in S1, S4, S5, S6 and S10 whereas arfA was present in the S7 strain. The oppA, oppB and oppC transporters, import oligopeptides in the cell and function as oligopeptide permease and were identified in all these clinical isolates (**Figure 4D**).^{48,49} Interestingly, S4 isolate has glutathione-S-transferase, a known detoxification enzyme involved in biodegradation of xenobiotics, protection against chemical agents and possibly provide antimicrobial.^{40,50}

These clinical isolates, with varied drug resistance status and lineages (Beijing, East African Indian and Euro American) may reflect in their metabolite phenotypes. S1 strain exhibited distinct metabolic phenotype from the rest. The clinical isolates clustered together based on their drug sensitive (S4 and S5) and resistance status (S6 and S7) in the PLS-DA plot. S4 and S5 belongs to the East-African Indian lineage showed lower aspartic acid and glutamic acid respectively than its closely linked isolates (S1 and S10). We observed slow growth rate in S1 isolate and their intracellular serine level was low which may partially explain isolate specific observed phenotype. S6, belongs to the Beijing lineage and showed high serine levels whereas S7 of Euro American lineages showed lower glutamic acid and aspartic acid levels with respect to H37Rv. Serine biosynthesis follows a multistep reaction using 3-phosphoglyceric acid (PGA) as a precursor produced during glycolysis and is essential for the survival of Mtb. Phosphoserine transferase serC (Rv0884c)

catalyses the addition of nitrogen to the backbone of pyruvate.⁵¹ Reduced serine production contributes to the slow growth phenotype of Mtb and serine supplementation accelerates the growth of Mtb.^{12,51,52} Earlier reports showed lower aspartic acid level and a higher serine levels in the hypervirulent Mtb strain of Beijing lineage than a hypovirulent strain.⁵³ This may indicate that S7 might show hyper-virulence than H37Rv which needs experimental validation . Glutamic acid content of H37Rv was reported to be much higher than nine Mtb strains and we also observed similar findings.⁵⁴

In this study, we characterized clinical Mtb isolates and employed whole genome sequencing and in silico analysis to annotate presence of SNPs to predict and compare drug resistance status with microbiology findings. Interestingly, these Mtb isolates, irrespective of their drug resistance or lineage status showed varied metabolic phenotypes. This adopted informatics pipeline and metabolite data could be useful as a resource to predict drug resistance phenotype of the new clinical isolates.

Data availability:

The raw reads of the whole-genome sequence of the clinical *M. tuberculosis* isolates (S1, S4, S5, S6, S7, S10) are available in the NCBI database (BioProject: PRJNA595470).

Acknowledgment:

This work was supported by the Department of Biotechnology, Government of India (grant BT/PR23238/NER/95/636/2017 to A. Das, and R. K. Nanda).

References

1. *Global Tuberculosis Report 2020*. Vol 7. CC BY-NC-SA 3.0 IGO; 2020.
2. Central TB Division - Directorate General of Health Services. India Tb Report National Tuberculosis. Published online 2020:292.
3. Huddart S, Svadzian A, Nafade V, Satyanarayana S, Pai M. Tuberculosis case fatality in India: A systematic review and meta-analysis. *BMJ Glob Heal*. 2020;5(1):1-9. doi:10.1136/bmjgh-2019-002080
4. Migliori GB, Matteelli A, Cirillo D, Pai M. Diagnosis of Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis: Current Standards and Challenges. *Can J Infect Dis Med Microbiol*. 2008;19:857901. doi:10.1155/2008/857901
5. Coll F, McNerney R, Preston MD, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med*. 2015;7(1):51. doi:10.1186/s13073-015-0164-0
6. Walker TM, Kohl TA, Omar S V, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis*. 2015;15(10):1193-1202. doi:https://doi.org/10.1016/S1473-3099(15)00062-6
7. Aparna M, Hritusree G, Varma PR, et al. Whole-Genome Sequence of Drug-Resistant Mycobacterium tuberculosis Strain S7, Isolated from a Patient with Pulmonary Tuberculosis. *Microbiol Resour Announc*. 2020;9(17):e01567-19. doi:10.1128/MRA.01567-19
8. Beste DJ V, Peters J, Hooper T, Avignone-Rossa C, Bushell ME, McFadden J. Compiling a molecular inventory for Mycobacterium bovis BCG at two growth rates: evidence for growth rate-mediated regulation of ribosome biosynthesis and lipid metabolism. *J Bacteriol*. 2005;187(5):1677-1684. doi:10.1128/jb.187.5.1677-1684.2005
9. Baker JJ, Abramovitch RB. Genetic and metabolic regulation of Mycobacterium tuberculosis acid growth arrest. *Sci Rep*. 2018;8(1):4168. doi:10.1038/s41598-018-22343-4
10. Yamada H, Chikamatsu K, Aono A, Mitarai S. Pre-fixation of virulent Mycobacterium tuberculosis with glutaraldehyde preserves exquisite ultrastructure on transmission electron microscopy through cryofixation and freeze-substitution with osmium-acetone at ultralow temperature. *J Microbiol Methods*. 2014;96:50-55. doi:10.1016/j.mimet.2013.10.019
11. Hosseini H, Fooladi AAI, Arjomandzadegan M, Emami N, Bornasi H. Genetics study and transmission electron microscopy of pili in susceptible and resistant clinical isolates of Mycobacterium tuberculosis. *Asian Pac J Trop Med*. 2014;7S1:S199-203. doi:10.1016/S1995-7645(14)60232-7
12. Borah K, Mendum TA, Hawkins ND, et al. Metabolic fluxes for nutritional flexibility of Mycobacterium tuberculosis. *Mol Syst Biol*. 2021;17(5):e10280. doi:10.15252/msb.202110280
13. de Carvalho LPS, Fischer SM, Marrero J, Nathan C, Ehrt S, Rhee KY. Metabolomics of Mycobacterium tuberculosis reveals compartmentalized co-catabolism of carbon substrates. *Chem & Biol*. 2010;17(10):1122-1131. doi:10.1016/j.chembiol.2010.08.009
14. Pang Z, Chong J, Zhou G, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(W1):W388-W396. doi:10.1093/nar/gkab382
15. Díaz C, Pérez del Palacio J, Valero-Guillén PL, et al. Comparative Metabolomics between Mycobacterium tuberculosis and the MTBVAC Vaccine Candidate. *ACS Infect Dis*. 2019;5(8):1317-1326. doi:10.1021/acsinfecdis.9b00008
16. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average

- nucleotide identity. *Antonie Van Leeuwenhoek*. 2017;110(10):1281-1286. doi:10.1007/s10482-017-0844-4
17. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14(7):1394-1403. doi:10.1101/gr.2289704
 18. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One*. 2010;5(6):e11147. <https://doi.org/10.1371/journal.pone.0011147>
 19. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the Problem of Comparing Whole Bacterial Genomes across Different Sequencing Platforms. *PLoS One*. 2014;9(8):e104984. <https://doi.org/10.1371/journal.pone.0104984>
 20. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490. doi:10.1371/journal.pone.0009490
 21. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvarkit: an interactive Venn diagram viewer. *BMC Bioinformatics*. 2014;15(1):293. doi:10.1186/1471-2105-15-293
 22. Bortolaia V, Kaas RS, Ruppe E, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75(12):3491-3500. doi:10.1093/jac/dkaa345
 23. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother*. 2017;72(10):2764-2768. doi:10.1093/jac/dkx217
 24. Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2020;48(D1):D517-D525. doi:10.1093/nar/gkz935
 25. Aziz RK, Bartels D, Best AA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*. 2008;9(1):75. doi:10.1186/1471-2164-9-75
 26. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol*. 2019;1962:1-14. doi:10.1007/978-1-4939-9173-0_1
 27. Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL. Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res*. 2012;40(1):e9. doi:10.1093/nar/gkr1067
 28. Meyer F, Overbeek R, Rodriguez A. FIGfams: yet another set of protein families. *Nucleic Acids Res*. 2009;37(20):6643-6654. doi:10.1093/nar/gkp698
 29. Xu L, Dong Z, Fang L, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019;47(W1):W52-W58. doi:10.1093/nar/gkz333
 30. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. 2016;428(4):726-731. doi:<https://doi.org/10.1016/j.jmb.2015.11.006>
 31. Joshi C, Sharma S, MacKinnon N, Masakapalli SK. Efficient System Wide Metabolic Pathway Comparisons in Multiple Microbes Using Genome to KEGG Orthology (G2KO) Pipeline Tool. *Interdiscip Sci Comput Life Sci*. 2020;12(3):311-322. doi:10.1007/s12539-020-00375-7
 32. Niehr F, Eder T, Pilz T, et al. Multilayered Omics-Based Analysis of a Head and Neck Cancer Model of Cisplatin Resistance Reveals Intratumoral Heterogeneity and Treatment-Induced Clonal Selection. *Clin Cancer Res*. 2018;24(1):158-168. doi:10.1158/1078-0432.CCR-17-2410
 33. Khandazhinskaya AL, Matyugina ES, Alexandrova LA, et al. Interaction of 5-substituted pyrimidine nucleoside analogues and M.Tuberculosis: A view through an electron microscope. *Biochimie*. 2020;171-172:170-177. doi:10.1016/j.biochi.2020.03.004
 34. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560
 35. Arjomandzadegan M, Sadmia M, Titov L, et al. Transmission Electron Microscopy of XDR Mycobacterium tuberculosis Isolates Grown on High Dose of Ofloxacin. *Sci Pharm*. 2017;85(1). doi:10.3390/scipharm85010003
 36. Liu F, Hu Y, Wang Q, et al. Comparative genomic analysis of Mycobacterium tuberculosis clinical isolates. *BMC Genomics*. 2014;15(1). doi:10.1186/1471-2164-15-469
 37. Witney AA, Cosgrove CA, Arnold A, Hinds J, Stoker NG, Butcher PD. Clinical use of whole genome sequencing for Mycobacterium tuberculosis. *BMC Med*. 2016;14(1):46. doi:10.1186/s12916-016-0598-2
 38. Sengupta S, Naz S, Das I, et al. Mycobacterium tuberculosis EsxL inhibits MHC-II expression by promoting hypermethylation in class-II transactivator loci in macrophages. *J Biol Chem*. 2017;292(17):6855-6868. doi:10.1074/jbc.M117.775205
 39. Daniel J, Sirakova T, Kolattukudy P. An Acyl-CoA Synthetase in Mycobacterium tuberculosis Involved in Triacylglycerol Accumulation during Dormancy. *PLoS One*. 2014;9(12):e114877. <https://doi.org/10.1371/journal.pone.0114877>
 40. Singh M, Vaughn C, Sasaninia K, et al. Understanding the Relationship between Glutathione, TGF- β , and

- Vitamin D in Combating Mycobacterium tuberculosis Infections. *J Clin Med*. 2020;9(9):2757. doi:10.3390/jcm9092757
41. Nazarova E V, Montague CR, Huang L, La T, Russell D, VanderVen BC. The genetic requirements of fatty acid import by Mycobacterium tuberculosis within macrophages. Kana BD, Storz G, Jacobs WR, Boshoff H, eds. *Elife*. 2019;8:e43621. doi:10.7554/eLife.43621
 42. Serafini A, Tan L, Horswell S, et al. Mycobacterium tuberculosis requires glyoxylate shunt and reverse methylcitrate cycle for lactate and pyruvate metabolism. *Mol Microbiol*. 2019;112(4):1284-1307. doi:https://doi.org/10.1111/mmi.14362
 43. da Silva PEA, von Groll A, Martin A, Palomino JC. Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. *FEMS Immunol Med Microbiol*. 2011;63(1):1-9. doi:10.1111/j.1574-695X.2011.00831.x
 44. Sarathy JP, Dartois V, Lee EJD. The role of transport mechanisms in mycobacterium tuberculosis drug resistance and tolerance. *Pharmaceuticals (Basel)*. 2012;5(11):1210-1235. doi:10.3390/ph5111210
 45. Danilchanka O, Mailaender C, Niederweis M. Identification of a novel multidrug efflux pump of Mycobacterium tuberculosis. *Antimicrob Agents Chemother*. 2008;52(7):2503-2511. doi:10.1128/AAC.00298-08
 46. Pasca MR, Guglielame P, Arcesi F, Bellinzoni M, De Rossi E, Riccardi G. Rv2686c-Rv2687c-Rv2688c, an ABC fluoroquinolone efflux pump in Mycobacterium tuberculosis. *Antimicrob Agents Chemother*. 2004;48(8):3175-3178. doi:10.1128/AAC.48.8.3175-3178.2004
 47. Yao Y, Barghava N, Kim J, Niederweis M, Marassi FM. Molecular structure and peptidoglycan recognition of mycobacterium tuberculosis ArfA (Rv0899). *J Mol Biol*. 2012;416(2):208-220. doi:10.1016/j.jmb.2011.12.030
 48. Dasgupta A, Sureka K, Mitra D, et al. An Oligopeptide Transporter of Mycobacterium tuberculosis Regulates Cytokine Release and Apoptosis of Infected Macrophages. *PLoS One*. 2010;5(8):e12225. <https://doi.org/10.1371/journal.pone.0012225>
 49. Danelishvili L, Stang B, Bermudez LE. Identification of Mycobacterium avium genes expressed during in vivo infection and the role of the oligopeptide transporter OppA in virulence. *Microb Pathog*. 2014;76:67-76. doi:https://doi.org/10.1016/j.micpath.2014.09.010
 50. Semianiv I, Todoriko L, Sprynsian T, Yeremenchuk I. Prevention of adverse reactions due to pharmacotherapy in MRTB considering polymorphism of glutathione-S-transferase M1 and T1 genes. *Eur Respir J*. 2019;54(suppl 63):PA3008. doi:10.1183/13993003.congress-2019.PA3008
 51. Borah K, Beyß M, Theorell A, et al. Intracellular Mycobacterium tuberculosis Exploits Multiple Host Nitrogen Sources during Growth in Human Macrophages. *Cell Rep*. 2019;29(11):3580-3591.e4. doi:10.1016/j.celrep.2019.11.037
 52. Bai G, Schaak DD, Smith EA, McDonough KA. Dysregulation of serine biosynthesis contributes to the growth defect of a Mycobacterium tuberculosis crp mutant. *Mol Microbiol*. 2011;82(1):180-198. doi:10.1111/j.1365-2958.2011.07806.x
 53. Meissner-Roloff R, Koekemoer G, Warren R, Loots DT. A metabolomics investigation of a hyper- and hypovirulent phenotype of Beijing lineage M. tuberculosis. *Metabolomics*. 2012;8:1194-1203. doi:10.1007/s11306-012-0424-6
 54. Ben Ginsburg, Sarah L. Lovett MSD. Amino acid composition of Mycobacteria. *Arch Biochem Biophys*. 1956;60(1):164-170. doi:https://doi.org/10.1016/0003-9861(56)90408-8

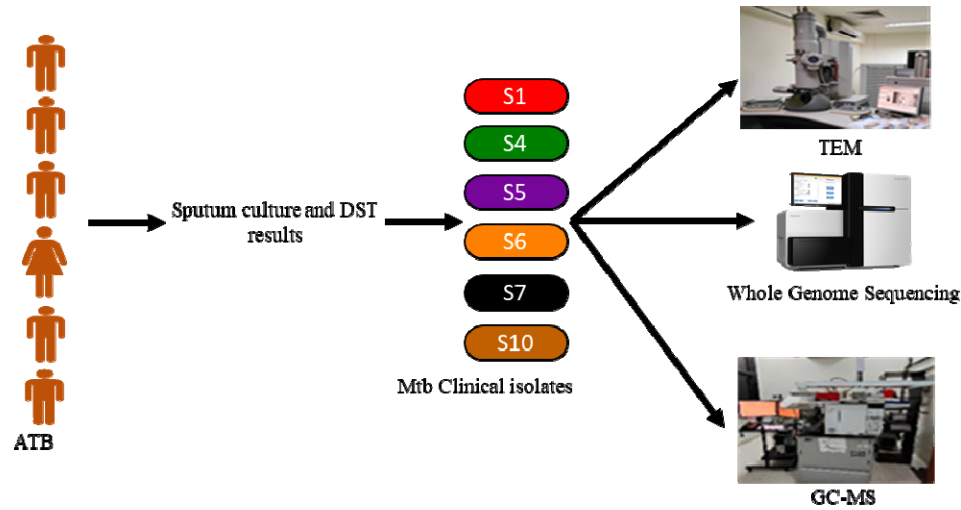


Figure 1: Schematics presenting the methods adopted for isolation and analysis of clinical *Mycobacterium tuberculosis* isolates. DST: drug sensitive test, TEM: transmission electron microscope, GC-MS: gas chromatography and mass spectrometry.

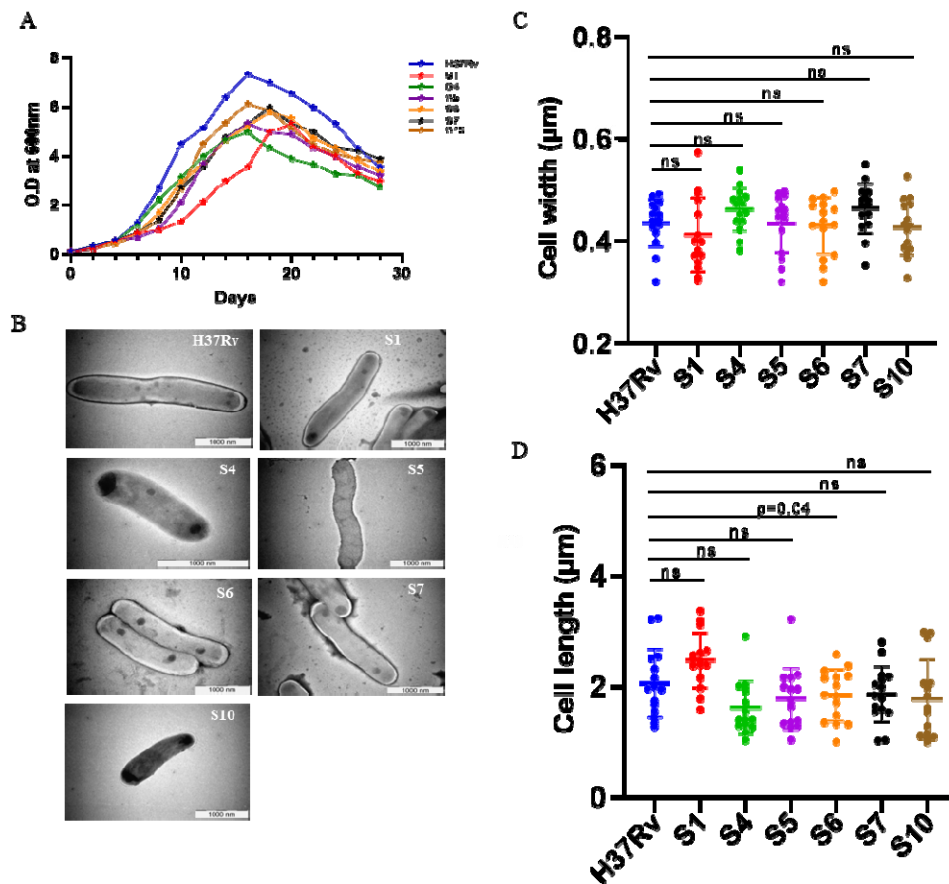


Figure 2: Growth kinetics and morphology of representative laboratory *Mycobacterium tuberculosis* H37Rv strain and clinical isolates showed variations. A. Growth kinetics under standard laboratory conditions. B. Transmission electron micrograph of laboratory H37Rv strain and clinical isolates (S1, S4, S5, S6, S7, S10). C. Bacterial length and D. width of individual isolates from 15 independent images. ns: not significant.

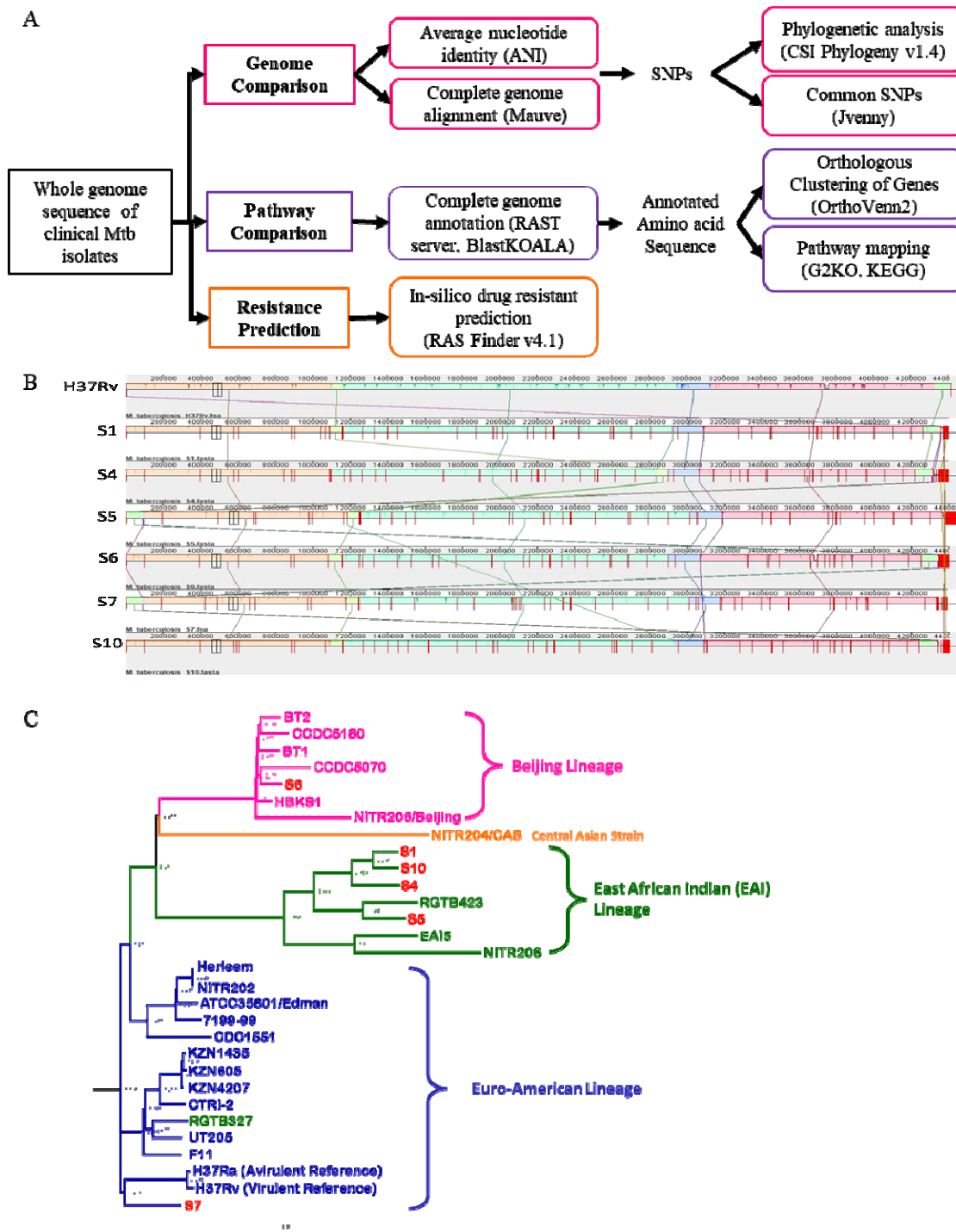


Figure 3: Comparative genome sequence analysis of the new clinical *Mycobacterium tuberculosis* isolates with 23 reported strains from major lineages. A. Pipeline used for comparative whole genome sequence analysis for functional characterization and drug resistance prediction. B. Complete genome alignment of new clinical isolates and H37Rv strain for nucleotide identity and single-nucleotide polymorphisms (SNPs). C. Maximum likelihood phylogenetic tree of the clinical isolates and strains of multiple lineages constructed using CSI Phylogeny v1.4.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

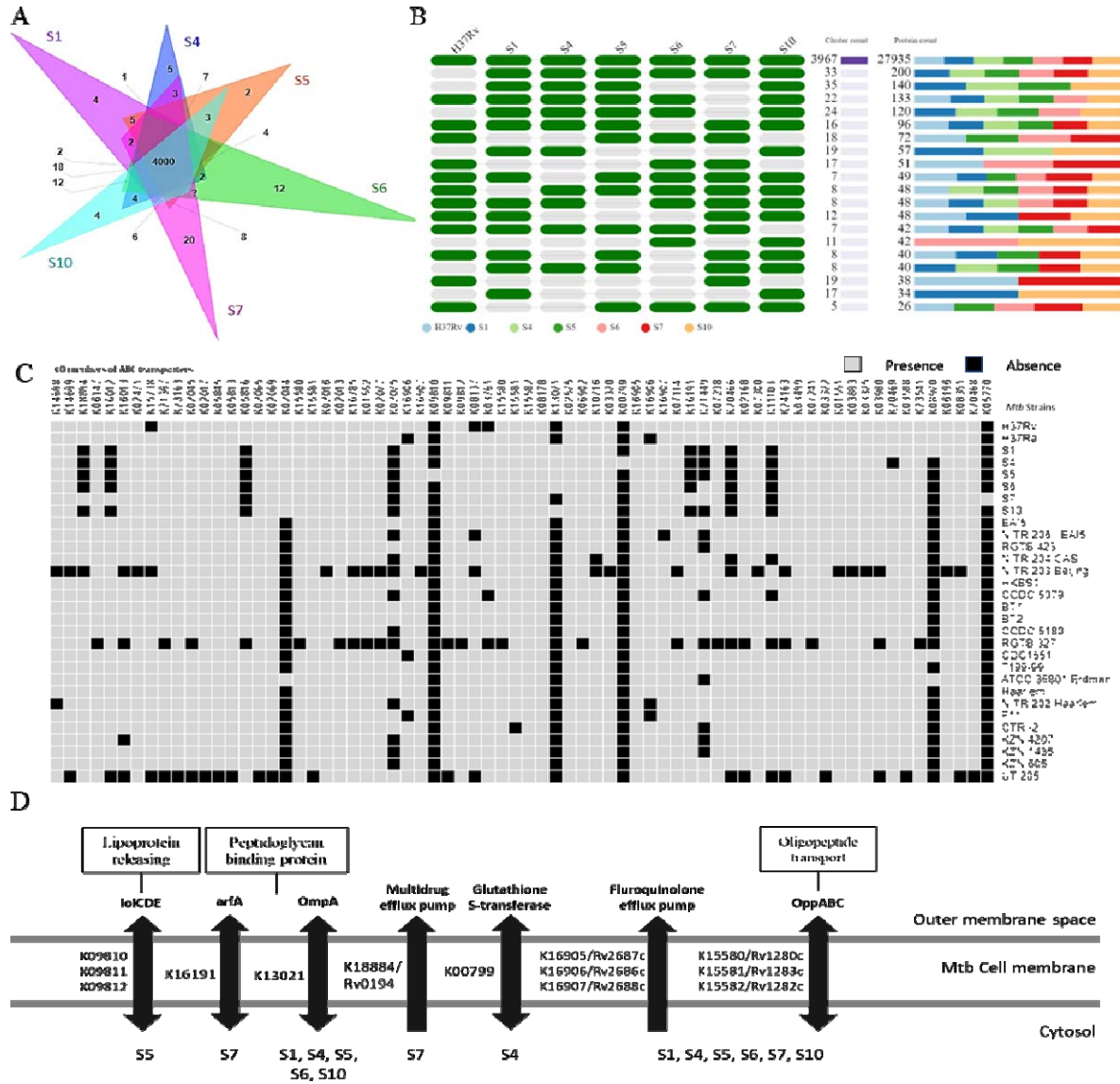


Figure 4: Representation of orthologous genes and unique pathways obtained from pathway comparison of the clinical *Mycobacterium tuberculosis* (Mtb) isolates. A. Venn diagram presenting orthologous clusters of annotated genes of the clinical Mtb isolates by OrthoVenn2. B. Heat map showing the proteins involved in different orthologous clusters between clinical isolates and reported Mtb strains. C. Presence and absence of the ABC transporters in clinical Mtb isolates and reported strains. D. Presence of important transporters in the newly isolated clinical Mtb isolates.

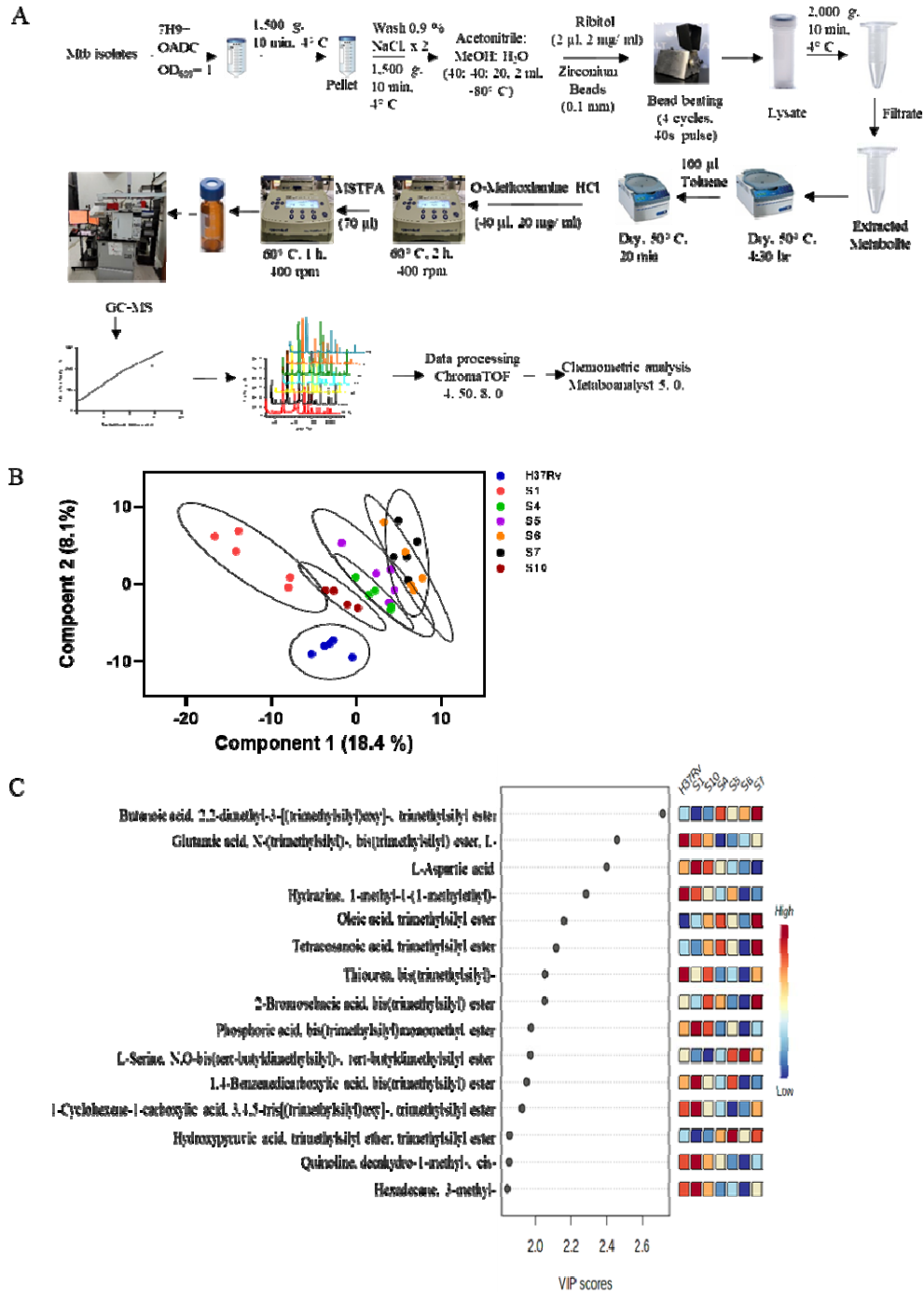


Figure 5: Global metabolite profiling of laboratory *Mycobacterial tuberculosis* strain (H37Rv) and clinical isolates (S1, S4, S5, S6, S7, S10) using Gas chromatography-Mass spectrometry showed significant differences. A. Pipeline used for metabolite extraction and analysis of *Mycobacterial* samples. B. Partial least squares discriminant analysis (PLS-DA) score plot showing differences in the metabolite patterns between clinical isolates and H37Rv strain. C. Variable importance in projection (VIP) score plot generated using MetaboAnalyst 5.0 showing the important deregulated metabolites.

Table 1: Predicted drug resistance status of the new clinical *Mycobacterium tuberculosis* isolates from in-silico analysis and the microbiology experimental data. R/S: Resistance/Susceptibility. The matched cases were colour coded as gray and the drug resistance status is presented in black.

Drug Names	Microbiology results						In-silico drug resistance prediction					
	S1	S4	S5	S6	S7	S10	S1	S4	S5	S6	S7	S10
Rifampicin	S	S	S	R	S	S	R	S	S	R	S	S
Isoniazid	S	S	S	R	R	S	R	S	S	R	R	S
Ethambutol	S	S	S	S	S	S	S	S	S	R	S	S
Ofloxacin	S	S	S	S	S	S	S	S	S	S	S	S
Capreomycin	S	S	S	S	S	S	S	S	S	S	S	S
Streptomycin	R	S	S	R	S	S	R	S	S	R	S	S
Kanamycin	S	S	S	S	S	S	S	S	S	S	S	S
Amikacin	S	S	S	S	S	S	S	S	S	S	S	S
Moxifloxacin	S	S	S	S	S	S	S	S	S	S	S	S
Clofazimine	S	S	S	S	R	R	S	S	S	S	S	S
Ethionamide	S	S	S	S	S	S	S	S	S	S	R	S
Levofloxacin	Not Determined						S	S	S	S	S	S
Clindamycin							R	R	R	R	R	R
Clarithromycin							S	S	S	S	S	S
Erythromycin							R	R	R	R	R	R
Telithromycin							R	R	R	R	R	R
Tobramycin							R	R	R	R	R	R
Gentamicin							R	R	R	R	R	R
Dibekacin							R	R	R	R	R	R
Netilmicin							R	R	R	R	R	R
Lincomycin							R	R	R	R	R	R