Identification Of Taxa and Functional Pathway Information Of Mycobacterium tuberculosis Microbiome And High Throughput Simulation Studies With Mycobacteriophage

Divya Verma (divyaverma4451@gmail.com)
Padmashree Institute of Management and Sciences
Preenon Bagchi
Padmashree Institute of Management and Sciences
Shylesh Murthy IA
Junior Researcher, Vasishth Academy of Advanced Studies and Research

Research Article

Keywords: Mycobacterium tuberculosis, Microbiome, Mycobacteriophage lysine D29, next generation sequencing, global alignment, Operational taxonomic unit, Taxonomy, modeling, docking and molecular dynamics

Posted Date: February 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1253773/v1

Abstract

Tuberculosis caused by bacteria Mycobacterium tuberculosis. The bacteria usually attack the lungs, but TB bacteria can attack any part of the body such as the kidney, spine and brain. Mycobacterium tuberculosis in lungs microbiome can be studied by metamorphic mechanization specifically metatransriptomics sequencing. These sequencing allow us to investigate the DNA content, RNA content, bacterial taxa and functional pathways. Further by taking antibiotic resistant protein from bacteria, we establish that bacteriophage lysine B D29(PDB-3HC7)can lyse mycolylarabinogalactan bonds and releases free mycolic acid. It do not show action on peptidoglycan bond. Based on information a full surface docking was performed. To verify assignment, a molecular dynamics simulations was performed to assess the stability of the docked substrates. MD simulation suggested hydrolytic activity of amino acids residues on Mycobacterium tuberculosis. Docking and simulation of bacteriophage D29 lysin B gene protein 12 with multiple antibiotic resistant proteins which takes part in transcription process in bacteria. Here, mycobacteriophage D29 showed a potent inhibition on action of antibiotic resistant protein during transcription process. This action resulted by modification or deactivation of amino acid residues.

1. Introduction

Mycobacterium tuberculosis is potential reason for pulmonary complications include hemoptysis, pneomothorax, bronchiectasis, extensive pulmonary destruction and there is substantial interest in establishing the foremost effective treatment. It is the most typically reported as super bugs among the persons aged 25-65 years. Tuberculosis characterized by extracellular bacilli are ingested by macrophages and presented to other white blood cells. This triggers the immune response in which white blood cells kill or encapsulate most of the bacilli, leading to the formation of granuloma. Tuberculosis has been associated with serious health issues which includes extrapulmonary TB, miliary TB and central nervous system TB (which surrounds the brain or spinal cord) .Symptoms include a cough that last 3 weeks or longer, pain in chest, coughing up blood or sputum, weight loss, fever, sweating at night. *Mycobacterium tuberculosis* strain responsible for a large multidrug resistant TB. The outbreak started in the mid 2000 and in earlier 1990s.TB is still one leading causes of mortality worldwide. Factors contributing to the situation are HIV/AIDS pandemic [1–4].

Several genes and chromosomal regions have been found to be associated with mycobacterium in various linkages and analyses, case control studies, genome wide association studies, mapping studies. Studies shown the association of genetic variations with pathogenesis and drug resistance. Pathogenic variants in genes of high and moderate penetrance-Erm 37, TlyA, Rpob, Eis, InhA, KasB and MarA which confers resistant towards various types of antibiotics [1–4].

Due to arrival of Next generation sequencing the detection of these pathogenic variants genes became possible. Next generation sequencing technology initially was concerned with studying genomes that were tractable from the standpoint size and repetitive content and with characterization of multiple genes

associated with the disease. The technology used to determine the order of nucleotides or targeted regions of DNA or RNA. Here raw data generation is no longer a rate limiting factor in genome scale studies. Galaxy an open source for NGS data analysis. The pipeline worned here is metatranscriptomics analysis which allows us to understand how the microbiome retaliate to the host by studying effective analysis of genes expressed.Further using the tools of computer aided drug design, we have tried to establish the novel polypeptides lysine B D29 gene 12 from Mycobacteriophage [1–4].

Metatranscriptomics analysis-

Next-generation sequencing (NGS) is an advanced version of non-Sanger-based sequencing technology that offers ultra-high throughput, scalability, and speed. Galaxy is an open source, web-based platform for next generation computational biomedical research [5]. Metatranscriptomics analysis enables understanding of how the microbiome responds to the environment by studying the functional analysis of genes expressed by the microbiome [6]. The genes from the Metagenomic analysis were transcribed from functional data, active metabolic pathways can be identified in our selected microbiome community [7].

Computer aided Drug design-

Drug design is the whole process of taking a newly discovered compound or drug molecule. Structure based drug designing technique is used here to build, display, simulate and analyze the molecular structure. Here we have used SWISS-MODEL tool [8] for modelling the proteins (gene receptors) responsible to bind antibiotic resistant protein .Gene receptors are as follows CR3, Dectin 1, IRAK4 and CXCL8. Selected models from homology modelling output are docked with selected antibiotic resistant proteins from *Mycobacterium tuberculosis*. Selection of antibiotic resistant protein was done based on it's appropriate target sites for specific gene receptors. Molecular docking was done using Patchdock tool [9, 10] and best interacting antibiotic resistant bacterial proteins with the gene receptors was selected by identifying simulation prototype. SWISS-MODEL tool have also been used for modeling the bacterial proteins (antibiotic resistant protein) responsible to bind Mycobacteriophage Lysin B D29. Bacterial proteins are as follows InhA, KasB, Eis, WhiB7, Rpob, Erm37, TlyA, and Mar A. Selected model from homology modelling output are docked with selected Mycobacteriophage lysine B D29.Molecular docking was performed using patch dock tool and best interacting Mycobacteriophage lysine B D29 with the antibiotic resistant bacterial proteins was identified by Molecular dynamics simulation.

Mycolylarabinogalactan recognition by mycobacteriophage protein is an important part of biological system. However there is only one specific protein for recognition of mycolylarabinogalactan i.e., Mycobacteriophage lysine B gp12.Substrate recognition for enzyme is varied. The sites containing aromatic residues can be classified into 3 types depending on arrangement-parallel arrangement/ planar arrangement, juxtaposed arrangement, sandwich type arrangement. The interaction arrangement limits the type of substrates for catalytic mechanism. Bacteriophage receptors for bacteria divided into 3 major

groups depending on the nature of ligand. The first and the most predominant group recognize mycolylarabinogalactan in the cell wall and second group recognizes proteins and the third group recognizes mixed receptors (proteins or carbohydrate receptor).

It is shown that the Mycobacteriophage D29 lysin B gp 12 protein have hydrolytic activity towards mycolylarabinogalactan. The protein activity is unusual and there has been interest in elucidating potential mechanism of catalytic activity. Therefore insilico analysis of protein was performed to identify putative binding sites and catalytic residues. The docking was performed to analysed putative binding sites and catalytic residues and is verified by molecular dynamics simulation. The results were compared with other bacterial proteins involved in antibiotic resistant. Analysis of virtual mutants also suggested strong preference of enzyme for catalytic efficiency at the cost of stability and putative affinity.

MD Simulation by Vienna ptm 2.0-

MD simulation a whole process which undergo conformational changes of both ligand and proteins. Conformation is one of the factor taken into account on computer docking simulation. However side chain motion are generally coupled to back bone motion and the latter can be significant. The best way to explore relevant backbone and sidechain flexibility by Molecular dynamics simulation.MD simulation depicts distance dependent dielectric model, in which electrostatic screening expressed. It is conducted for the complexation of substrate with the inhibitor, complexation of protein domain with doubly phosphorylated peptide ligand.Here electrostatic interaction are important driving force for docking and protein undergo modest changes in conformation upon binding.

2. Materials And Methods

Mycobacterium tuberculosis fastq sequences SRR14690790.1.1 and SRR14690790.1.2 were retrieved from SRA database.

Sequence's quality was checked using FASTQC [11]. MultiQC [12] was done to aggregate results from FASTQC analyses into a single report. Sequences were trimmed using cutadapt.

FASTQC followed by MultiQC was re-run using the results of cutadapt.

Next, using SortMeRNA tool [13] any reads identified as rRNA in dataset was removed.

Next, using FASTQ INTERLACER tool [14] paired end FASTQ reads from two separate files were joined.

MetaPhlAn tool [15] was used for profiling the composition of microbial communities (Bacteria, Archaea and Eukaryotes) from our microbiota.

Krona tool [16] was used to visualize the results of a metagenomic profiling as a zoomable pie chart and GraPhlAn tool [17] for visualizing high-quality circular representations of taxonomic and phylogenetic

trees.

Further, HUMAnN [18] pipeline was used for efficiently and accurately profiling the presence/absence and abundance of microbial pathways in our microbiota.

Next, using the genes present in our microbiota, their 3d structure was modeled using SWISS-MODEL [8].

Mycobacteriophage lysin B D29 were downloaded from RCSB.

Further, docking was performed using patchdock [9, 10].

Vienna PTM using Molecular dynamics simulation. Here desired proteins are modified with one or supported Post translational modifications and obtain force field parameters(GROMOS, 45A3, 54A8) with the help of input files MD simulations were performed using GROMACS package.

3. Results And Discussion

3.1 Metagenome analysis

Metagenome, having accession number SRR14690790, for *Mycobacterium tuberculosis* was downloaded from SRA database.

As, per **Per base sequence quality** results (Fig. 1) of FASTQC and MultiQC, the sequence quality is not good hence we go ahead with trimming the sequence.

CUTADAPT tool [27] is used for trimming. It finds and removes adapter sequences, primers, poly-A tails, and other types of unwanted sequence from our data. It searches for the adapter in all reads and removes it when it finds it. Further, sequence quality of the cutadapt output is checked using FASTQC and MultiQC and it is found within the range.

SortMeRNA tool removes any reads identified as rRNA from our dataset. Fastq Interlace tool joins paired end FASTQ reads from two separate files. **Taxonomic profiling [28] was done using MetaPhlAn tool (Fig. 2). The output is visualized using** Krona and Graphlan (Table 1, Fig. 3).

After generation of taxonomy, we move to functional information of our microbiome. Functional information of the above microbiome community **[28]** was done using **HUMAnN pipeline.**

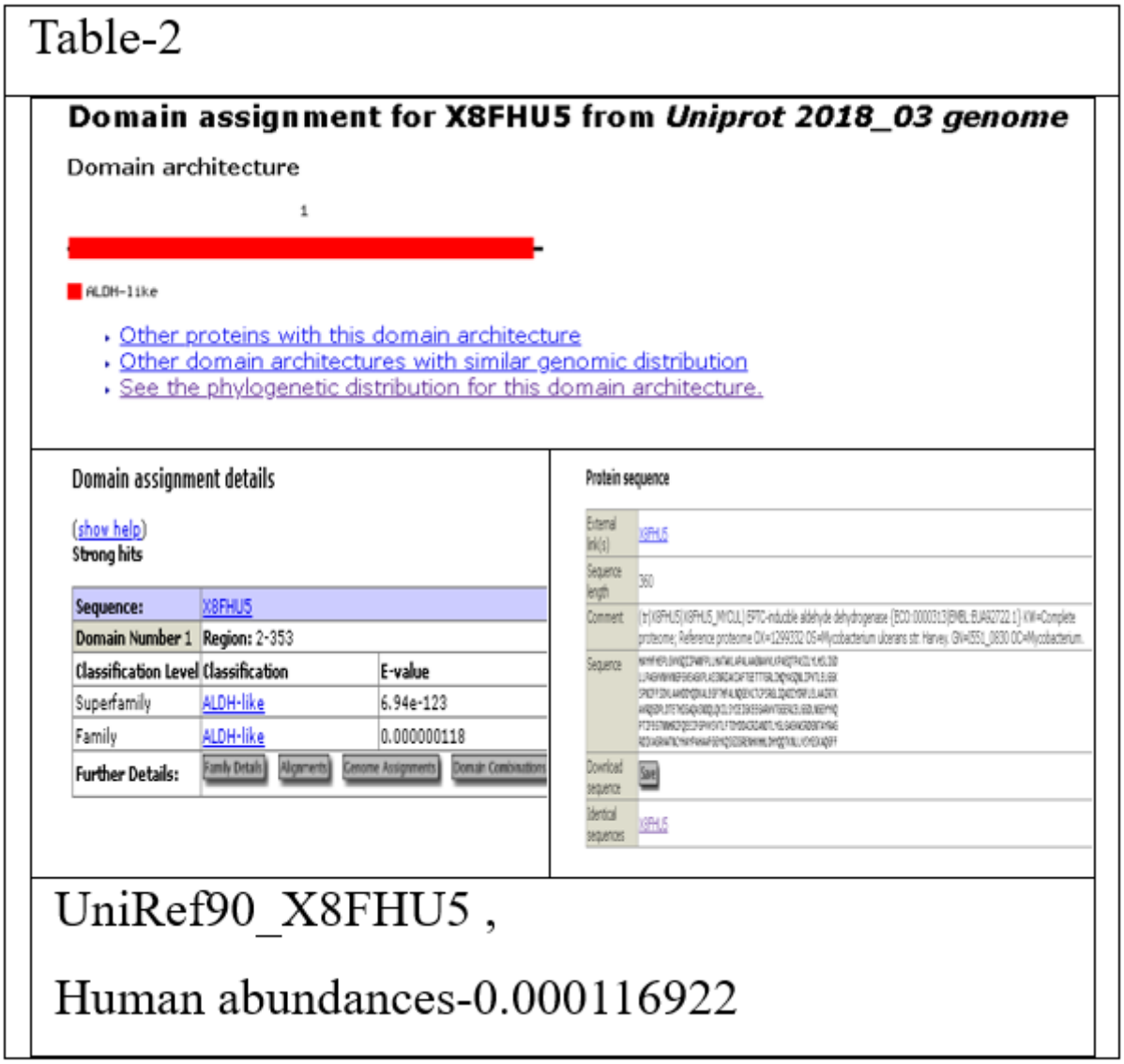
Table 1

Normalized gene families

# Gene Family	humann_Abundance-RELAB
UNMAPPED	0.999872
UniRef90_X8FHU5	0.000116922
UniRef90_X8FHU5|unclassified	0.000116922
UniRef90_Z9JRB3	1.08765e-05
UniRef90_Z9JRB3|unclassified	1.08765e-05

Next, from the gene family information, we obtain the functional information of our microbiome using Superfamily server. The Functional information of 1st five families from Normalized gene families as detected by Superfamily (HMM library and genome assignments server) is given in table 2 & 3 below.

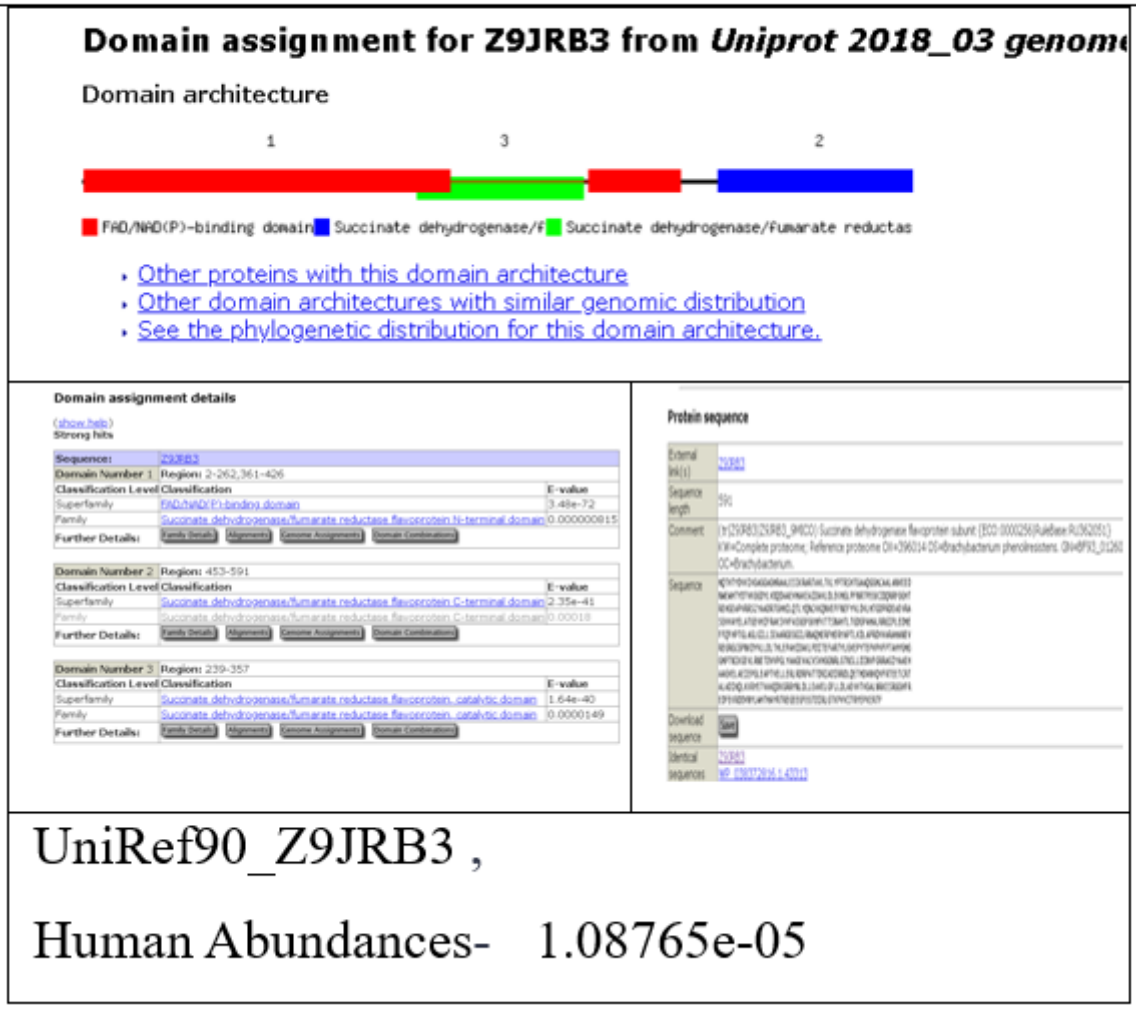
Table-2



UniRef90_X8FHU5 ,

Human abundances-0.000116922

Table-3



Relationships were determined –

By using this factor we selected the specific bacteriophage (Mycobacteriophage lysine B D29 gp12). It is determined by global alignment tool from NCBI (Table 4, Fig. 4).

Global alignment result:

Needleman-Wunsch algorithm

Source-https://blast.ncbi.nlm.nih.gov

Table 4

SL.NO	ORGANISMS	ACCESSION NO
1.	Mycobacteriophage D29 DNA coat protein	X70353.1
2.	*Mycobacterium tuberculosis* H37Rv	KY702779.1

3.2-Structure based Drug designing of Mycobacterium tuberculosis-

Since, Tuberculosis is bacterial disease; we further go ahead towards designing novel drug for the disease. From the MetaPhlAn: Bowtie2 output we get the gene ids. Corresponding gene receptors (macromolecules) are taken from NCBI for our work (Table 5).

Table 5

Genes with their NCBI Accession number

Sl. No	Gene Receptors	Number NCBI Accession	Homologous Template
1.	CR3	QRN45544.1	3K6SB
2	Dectin 1	AAH71746.1	1MPUA
3	IRAK4	NP_001338274.1	5UIUA
4	CXCL8	NP_001341769.1	6N2UA

Abbreviations of genes:

1. CR3 - Complementary Receptor

2. CLEC7A - C-type lectin domain family 7, member A

3.CXCL8 - C-X-C Motif chemokine ligand 8

4.IRAK4 - Interleukin 1 Receptor Associated kinase 4

Homology modeling

Homology modeling of the above receptors are done using SWISS-MODEL server. The receptor model and corresponding ramachandran plot results are given in Fig. 5. Template used for modeling is given in Table-5.

Table 6

Proteins with their NCBI Accession number

SL.NO	PROTEINS	ACCESSION NUMBER	HOMOLOGOUS TEMPLATE
1.	Eis	AVV29810.1	5EBV.1.F
2.	Erm37	KBG11004.1	6NVM.1.A
3.	Inh A	AVV29586.1	2PR.2.A
4.	kasB	CCE37716.1	2GP.6.A
5.	Mar A	OMH59859.1	3W6V.1.A
6.	Rpob	AEJ88322.1	6VW.0.1.C
7.	TlyA	CCP44459.1	5KS.2.1.A
8.	WhiB7	AJF05229.1	7KIM.1.K

Abbreviations of proteins:

1. Eis - Enhanced intracellular survival

2. Erm37 - Expression resistant macrolide

3. InhA - Inhibin alpha

4. KasB - Beta keto acyl carrier protein

5. Mar A -Multiple antibiotic Resistant

6. Rpob -Rifampin resistant gene (Beta subunit of bacterial RNA polymerase)

7. Tly A -Cytidine methyl transferase A.

8. WhiB7 - Probable transcription regulator

Homology modeling-

Homology modeling of the above receptors (micromolecules) are done using SWISS-MODEL server. The receptor model and corresponding ramachandran plot results are given in Fig. 6. Template used for modeling is given in Table 6.

Bacteriophage exhibit catalytic mechanism on binding with various proteins or carbohydrates motif. This study of phage-host interaction can inform small molecule drug discovery by revealing new drug targets and pinpointing their weakness. Mycobacteriophage lysine B D29 can hydolysed the mycolylarabinogalactan bonds and inactivates antibiotic resistant proteins (Table 7). The potential activity of Mycobacteriophage lysine B D29 against antibiotic resistant protein of *Mycobacterium tuberculosis* is studied here.

Table-7

Docking scores and RMSD value of Mycobacteriophage B D29 lysin with bacterial proteins

SL.NO	TEMPLATE.NO	PROTEINS	LIGANDS	DOCKING SCORE Kcal/mol	RMSD Value Angstrom
1.	2PR2.1.A	Inh A	Phage lysine B D29	-16652	4
2.	2GP6.1.A	Kas B	Phage lysine B D29	-16012	4
3.	5EBV.1.F	Eis	Phage lysine B D29	-19224	4
4.	7KIM.1.K	WhiB7	Phage lysine B D29	-13710	4
5.	6VW0.1.C	Rpob	Phage lysine B D29	-18264	4
6.	6NVM.1.A	Erm37	Phage lysin B D29	-13614	4
7.	5KS2.1.A	TlyA	Phage lysine B D29	-13792	4
8.	3W6V.1.A	Mar A	Phage lysine B D29	-12638	4

It is seen that Mycobacteriophage lysine B D29 has good docking scores with MarA (multidrug antibiotic resistant proteins), Erm37, whiB7. Patch dock server used to dock the proteins (Table 8). Protein protein

model interaction analysed. Gromacs minimization energy by Galaxy Europe server and structural charges , aminoacids identification were performed using Vienna-ptm server.

Further docking is performed with the receptors in Table 6 with the Mycobacteriophage B lysine D29.

Table 8

Docking scores and RMSD value of Human receptors with bacterial proteins.

SL. NO	RECEPTOS	LIGAND-1	LIGAND-2	DS-1 Kcal/mol	DS-2 Kcal/mol	RMSD value angst-rom
1.	CR3	Eis	InhA	-14726	-14260	4
2.	DECTIN	WhiB7	Rpob	-12610	-17644	4
3.	IRAK4	KasB	Erm37	-16198	-15514	4
4.	CXCL8	TlyA	MarA	-14418	-15096	4

Dectin 1 receptor with WhiB7 bacterial proteins (Transciptional regulators) has good docking scores. It exhibit good binding sites.

To select the putative site, an analogous experiment was performed with the bacteriophage protein and the sites were compared. The lysine B bacteriophage D29 had close overlap with Mar A gene binding sites and the binding energies were comparable (delta G bind=-12638 kcal/mol). Final verification of docking experiments performed with MD simulation which suggested stable binding sites. To help understand discrimination of different proteins in site 2, the docking of both protein protein was performed with a high precision Vienna-ptm server.

Molecular dynamics simulation of antibiotic resistant bacterial protein is selected sites for binding Mycobacteriophage lysine B D29 gp12 protein. MD simulation was performed in 150mM water at 300k for 100 ns (Table 8).

As per docking results and verification by molecular dynamics simulation it was found that the whib7 protein has good affinity in binding with Mycobacteriophage D29 lysin B. It depicts that Mycobacteriophage D29 Lysin B plays an important role at transcription process, it stop the transcription process of whib7 and do not allow the production of antibiotic resistant protein. Here we also characterize insilico the predicted interaction of gene protein 12 from Mycobacteriophage D29 with Mycobacterium tuberculosis antibiotic resistant protein (TlyA), Multidrug resistant protein(Mar A), Rifampin resistant protein(rpob), expression resistant macrolide (Erm 37). All these proteins plays an important role in the transcription process in bacterial cells and has been proposed (Table 9, Fig. 7).

Conclusion

The taxonomy and functional information of Mycobacterium tuberculosis microbiome are identified. As per docking studies and molecular dynamics simulation analysis it is seen that Mycobacteriophage lysin B D29 inhibits antibiotic resistant protein and hydrolysed mycolylarabinogalactan bonds. Further invitro or in vivo studies can be done on the Mycobacteriophage D29 lysin B to establish potential treatment against multi drug resistant strains i.e,. *Mycobacterium tuberculosis*. Taxonomic profiling obtained using krona pie chart and Graphlan which depicts the existence of Multidrug resistance *Mycobacterium tuberculosis* sequence. The phylogenetic tree obtained from gene Superfamily tool which depicts the evolutionary relationship between Mycobacteriophage and multi drug *Mycobacterium Tuberculosis* strain.To analyse the action of Mycobacteriophage protein on Multidrug resistant *Mycobacterium Tuberculosis* proteins Homology modeling, docking studies was carried out. It was observed that Mycobacteriophage lysine B D29 was very active and showed inhibitory action of aminoacid on WHIB7 protein which was validated by using tool Molecular dynamic Simulation.

References

1. Mostafa, M. M., Nassef, M., & Badr, A. (2016). Computational determination of the effects of virulent Escherichia coli and salmonella bacteriophages on human gut. *Computer methods and programs in biomedicine*, *135*, 27-35.
2. Nobrega, F. L., Costa, A. R., Kluskens, L. D., & Azeredo, J. (2015). Revisiting phage therapy: new applications for old resources. *Trends in microbiology*, *23*(4), 185-191.
3. Reindel R, Fiore CR, (2017) Phage Therapy: Considerations and Challenges for Development, Clinical Infectious Diseases, Volume 64(11): 1589–1590
4. Wittebole, X., De Roock, S., & Opal, S. M. (2014). A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*, *5*(1), 226-235.
5. Leite, D. M. C., Brochet, X., Resch, G., Que, Y. A., Neves, A., & Peña-Reyes, C. (2018). Computational prediction of inter-species relationships through omics data analysis and machine learning. *BMC bioinformatics*, *19*(14), 151-159.
6. Fernández-Ruiz, I., Coutinho, F. H., & Rodriguez-Valera, F. (2018). Thousands of novel endolysins discovered in uncultured phage genomes. *Frontiers in microbiology*, *9*, 1033.
7. Lenneman, B. R., Fernbach, J., Loessner, M. J., Lu, T. K., & Kilcher, S. (2021). Enhancing phage therapy through synthetic biology and genome engineering. *Current Opinion in Biotechnology*, *68*, 151-159.
8. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303.
9. Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002.

10. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res. 33: W363-367.
11. Andrews, S. (n.d.). FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
12. usegalaxy.eu
13. Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611
14. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., & and, A. N. (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics, 26(14), 1783–1785. https://doi.org/10.1093/bioinformatics/btq281
15. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., … Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature Methods, 12(10), 902–903.
16. Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-385
17. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., & Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ, 3, e1029. https://doi.org/10.7717/peerj.1029
18. Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., … Huttenhower, C. (2012). Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. PLoS Computational Biology, 8(6), e1002358. https://doi.org/10.1371/journal.pcbi.1002358

Tables

Table 8 and 9 are only available as a download in the Supplemental Files section.

Figures

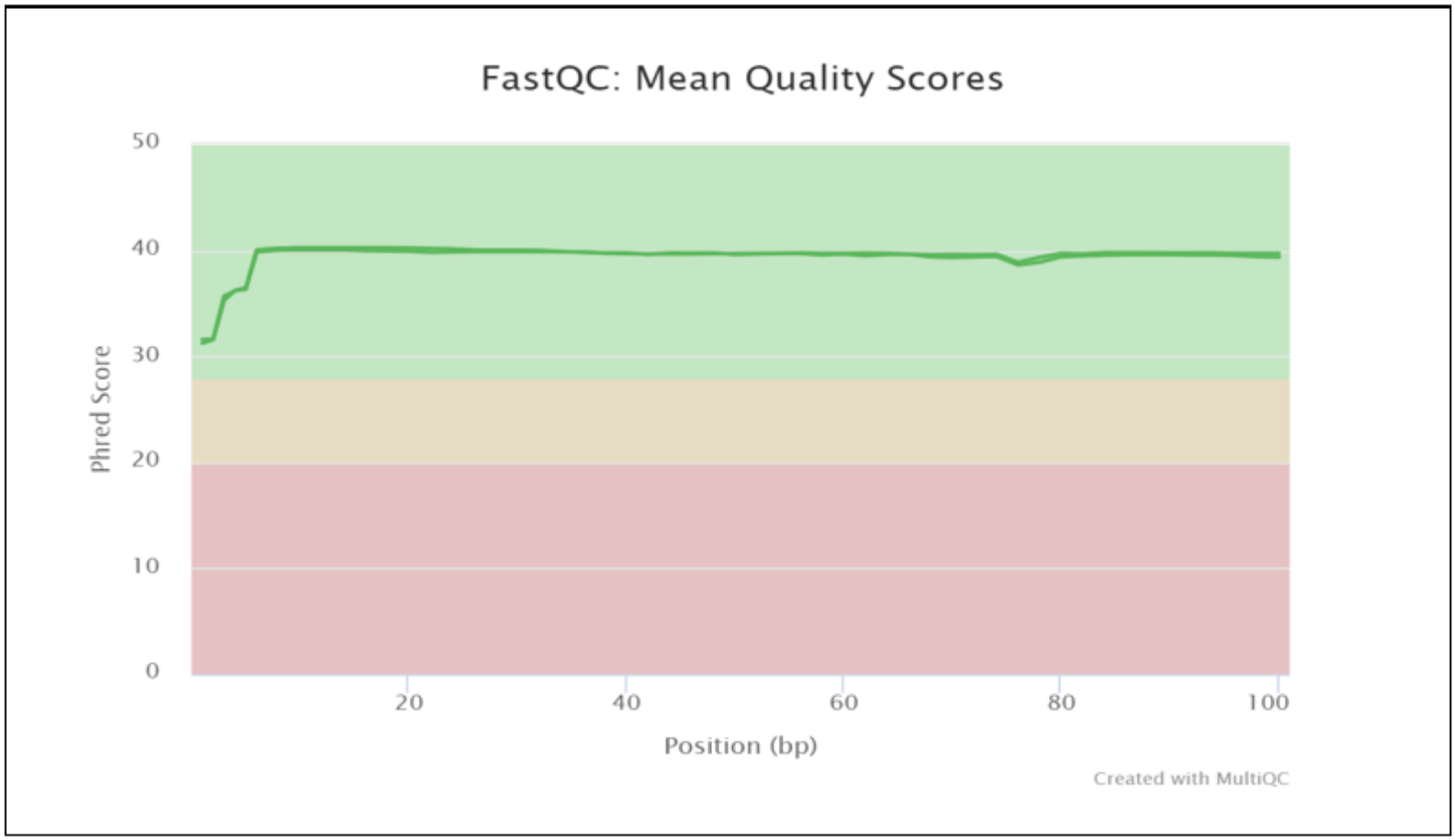


Figure 1

MultiQC result before Trimming-

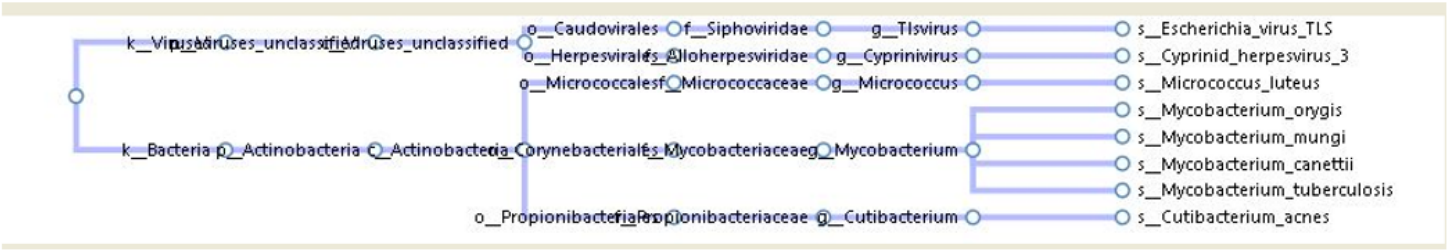


Figure 2

Generation, personalization and annotation of tree: Tree in PhyloXML

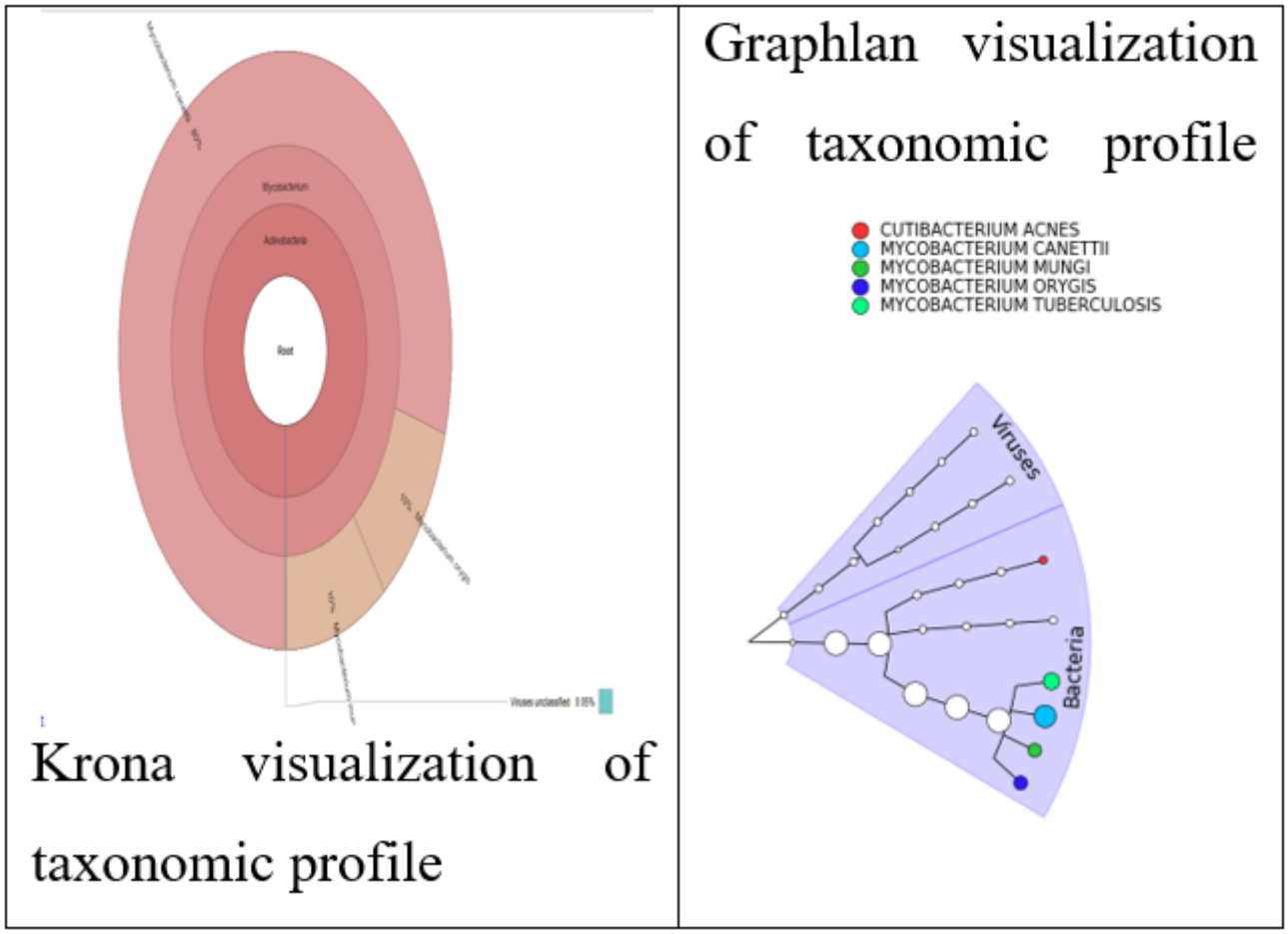


Figure 3

Visualization of **Taxonomic profile in Krona and** GraPhlAn

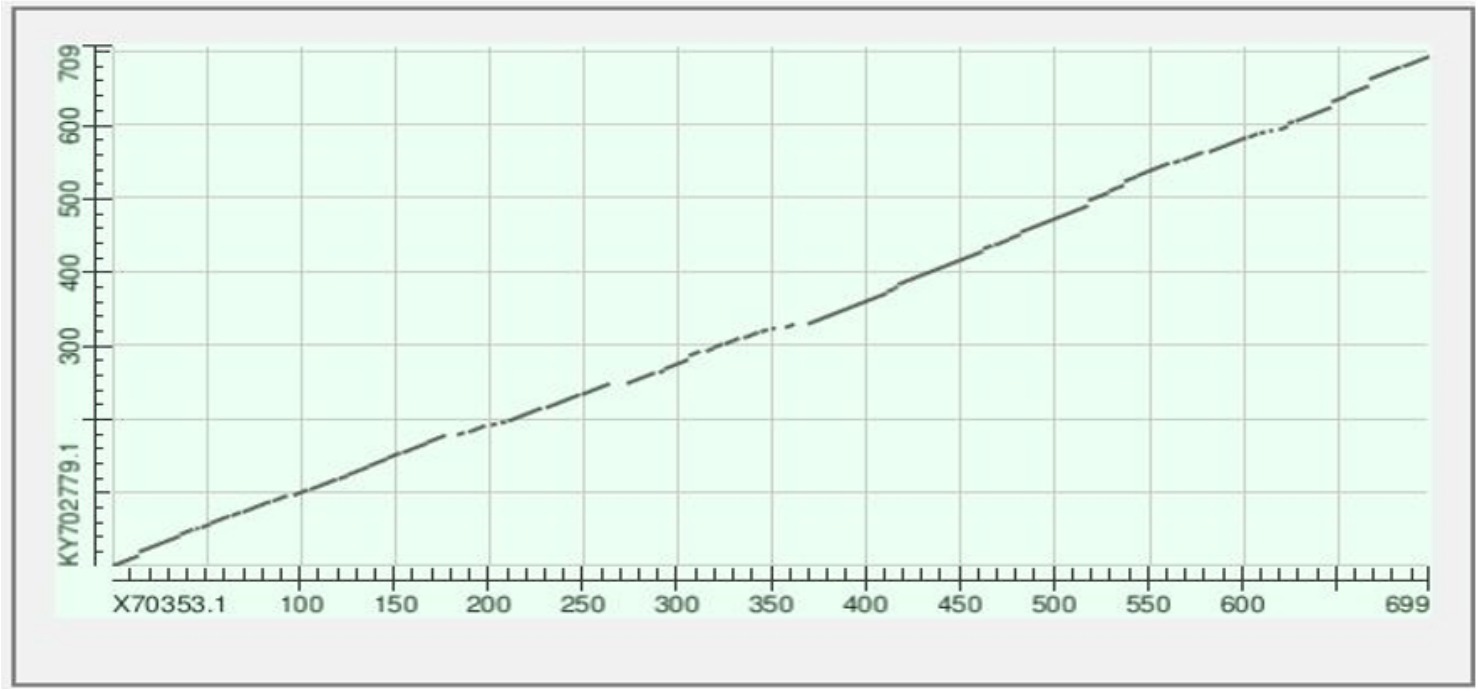


Figure 4

Dot-plot

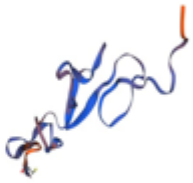
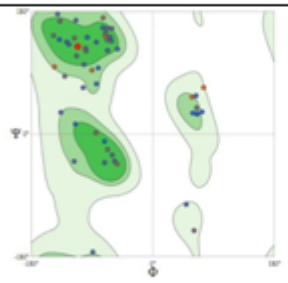
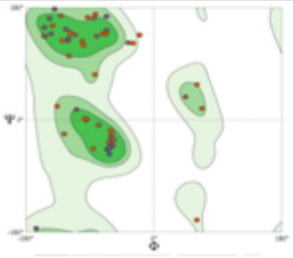
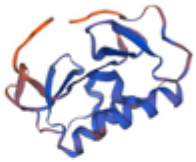
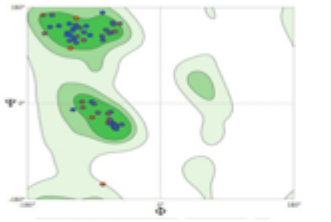
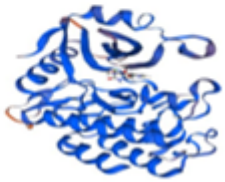
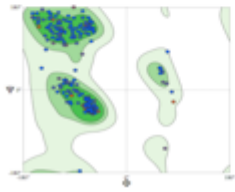
Sl. No.	Name of receptor	Modelled structure	Ramachandran plot of the modeled structure
1	CR3		
2.	DECTIN		
3.	CXCL8		
4.	IRAK4		

Figure 5

Swiss-model generated receptor models with their ramachandran plot

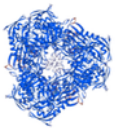
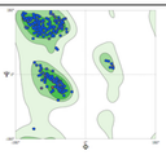
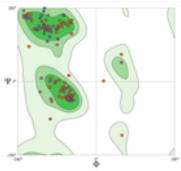
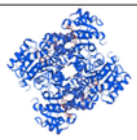
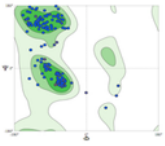
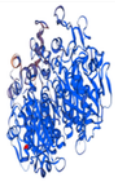
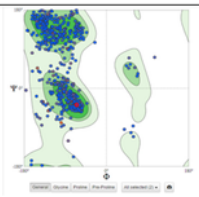
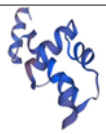
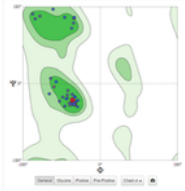
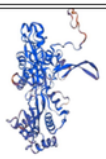
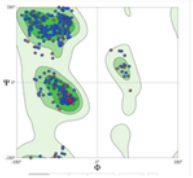
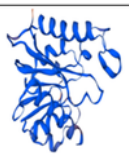
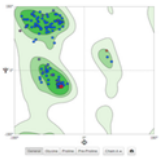
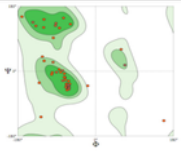
Sl. No.	Name of receptor	Modelled structure	Ramachandran plot of the modeled structure
1	EIS		
2.	ERM37		
3.	INHA		
4.	KASB		
5.	MARA		
6.	RPOB		
7.	TLYA		
8.	WHIB7		

Figure 6

Swiss-model generated receptor models with their ramachandran plot

Figure 7

Gromacs energy minimization by using galaxy Europe

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tab8.docx
- Tab9.docx