



Identification of Taxa and Functional Pathway Information of *Mycobacterium tuberculosis* Microbiome and High Throughput Simulation Studies with Mycobacteriophage

Divya Verma^{1,2(✉)}, I. A. Shylesh Murthy², and Preenon Bagchi^{1,2,3}

¹ Padmashree Institute of Management and Sciences, Bengaluru, India
prithish.bagchi@gmail.com

² Vasishth Academy of Advanced Studies and Research, Sarvasumana Association, Bengaluru, India

³ MGM Institute of Biosciences and Technology, Aurangabad, India

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 metatranscriptomics 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 BD29 (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.

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

1 Introduction

Mycobacterium tuberculosis is potential reason for pulmonary complications include hemoptysis, pneumothorax, 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 worked 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.

Mycolyrarabinogalactan 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

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 (Table 1). The output is visualized using Krona and Graphlan (Fig. 2, 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.

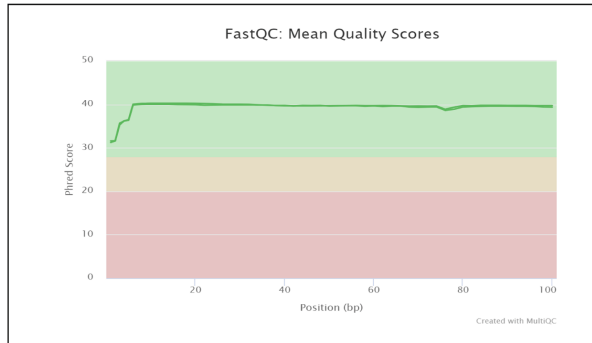


Fig. 1. MultiQC result before Trimming-

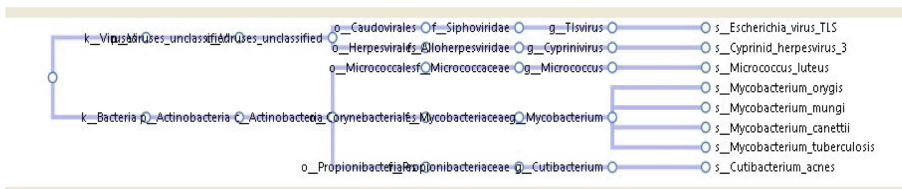


Fig. 2. Generation, personalization and annotation of tree: Tree in PhyloXML

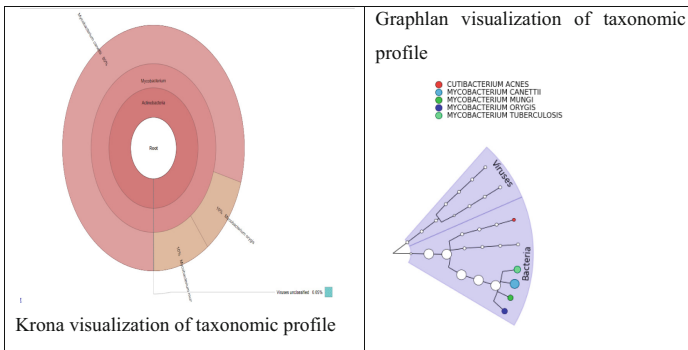


Fig. 3. Visualization of **Taxonomic profile** in Krona and GraPhlAn

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 and 3.

Table 1. Normalized gene families

# Gene Family	humann_Abundance-RELAB
UNMAPPED	0.999872
UniRef90_X8FHU5	0.000116922
UniRef90_X8FHU5unclassified	0.000116922
UniRef90_Z9JRB3	1.08765e-05
UniRef90_Z9JRB3unclassified	1.08765e-05

Table 2. Superfamily information

Domain assignment for X8FHU5 from Uniprot 2018_03 genome

Domain architecture

- Pfam-like
- Other proteins with this domain architecture
- Other domain architectures with similar genomic distribution
- See the phylogenetic distribution for this domain architecture.

<p>Domain assignment details</p> <p>View help</p> <p>Interpro</p> <p>Interpro ID IP00141</p> <p>Domain Number 1 Region 1-293</p> <p>Classification level Classification E-value</p> <p>Superfamily A02018 0.54e-113</p> <p>Family A02018 0.000000118</p> <p>Further details: View help View help View help View help</p>	<p>Protein sequence</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p>
---	--

UniRef90_X8FHU5 ,
Human abundances-0.000116922

Table 3. Superfamily information

Domain assignment for Z9JRB3 from Uniprot 2018_03 g

Domain architecture

- F10/NCF1-binding domain
- Succinate dehydrogenase/F
- Succinate dehydrogenase/ubiquinone reductase
- Other proteins with this domain architecture
- Other domain architectures with similar genomic distribution
- See the phylogenetic distribution for this domain architecture.

<p>Domain assignment details</p> <p>View help</p> <p>Interpro</p> <p>Interpro ID IP00141</p> <p>Domain Number 1 Region 1-293</p> <p>Classification level Classification E-value</p> <p>Superfamily A02018 0.54e-113</p> <p>Family A02018 0.000000118</p> <p>Further details: View help View help View help View help</p>	<p>Protein sequence</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p> <p>Start 1</p> <p>End 293</p>
---	--

UniRef90_Z9JRB3 ,
Human Abundances- 1.08765e-05

Relationships were determined

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

Global alignment result

Needleman-Wunsch algorithm.

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

Table 4. NCBI accession number of organisms

SL.NO	ORGANISMS	ACCESSION NO
1.	Mycobacteriophage D29 DNA coat protein	X70353.1
2.	<i>Mycobacterium tuberculosis</i> H37Rv	KY702779.1

See Fig. 4

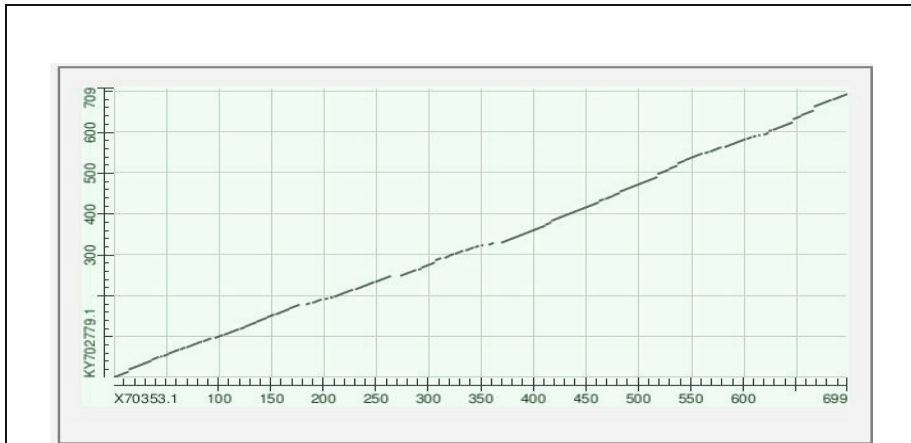


Fig. 4. Dot-plot

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

3.1 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).

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 6.

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.

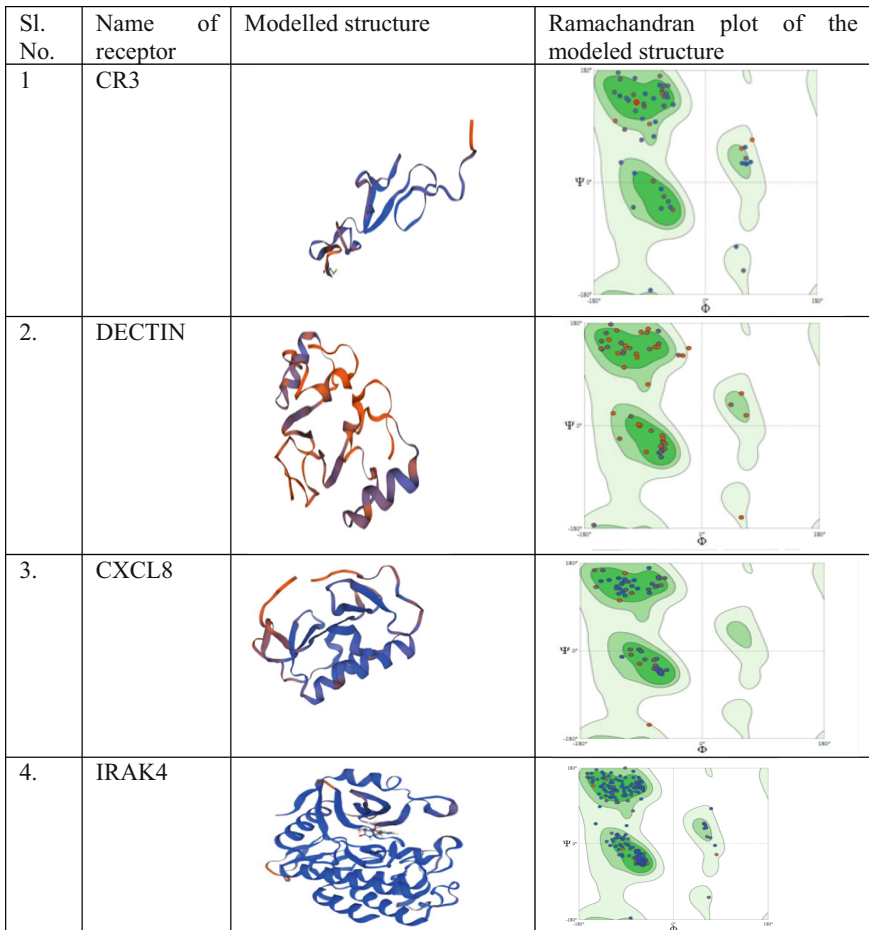


Fig. 5. Swiss-model generated receptor models with their ramachandran plot

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

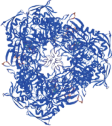
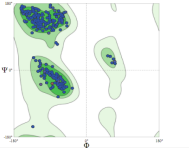

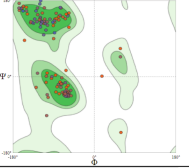
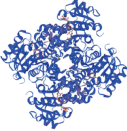
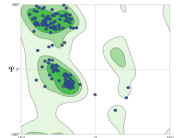
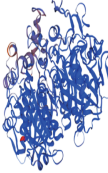
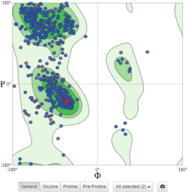
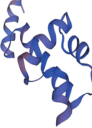
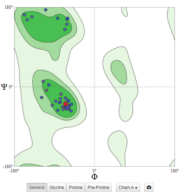
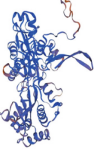
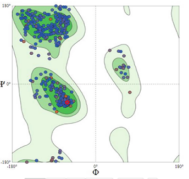
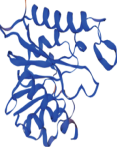
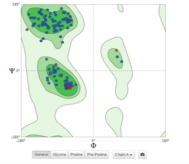

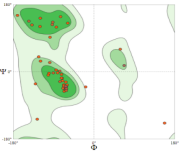
Sl. No.	Name of receptor	Modelled structure	Ramachandran plot of the modeled structure
1	EIS		
2.	ERM37		
3.	INHA		
4.	KASB		
5.	MARA		
6.	RPOB		
7.	TLYA		
8.	WHIB7		

Fig. 6. Swiss-model generated receptor models with their ramachandran plot

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 hydrolyse 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.

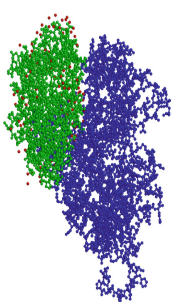
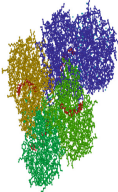
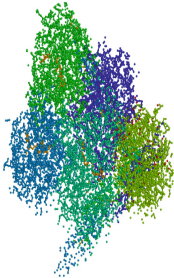
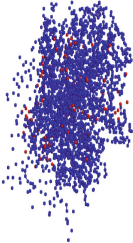
Table 7. Docking scores and RMSD value of Mycobacteriophage B D29 lysin with bacterial proteins

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

Table 8. Docking scores and RMSD value of Human receptors with bacterial proteins.

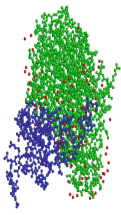
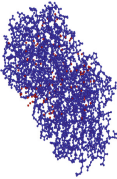
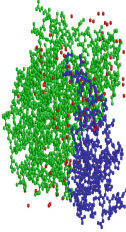
SL. NO	RECEPTOS	LIGAND-1	LIGAND-2	DS-1 Kcal/mol	DS-2 Kcal/mol	RMSD value angstrom
1.	CR3	Eis	InhA	-14726	-14260	4
2.	DECTIN	WhiB7	Rpob	-12610	-17644	4
3.	IRAK4	KasB	Erm37	-16198	-15514	4
4.	CXCL8	TlyA	MarA	-14418	-15096	4

Table 9. MD simulation results

SL .N O	Protein name	Mode-I	Residue-s (chainA)	Residue-s (chainA)	Charge-s
2.	Erm37-Expression resistant macrolide docked with Mycobacteriophage D29 lysin B gp12		1 14 24 35 38 45 53 55 59 84 91 106	Asp Ser Val Pro Gly Ala Pro Asn Val Ser Phe	0.129 -0.31 0.31 0.45 0.31 0 0.45 -0.31 0.29 0.31 -0.31
3.	InhA-Inhibin alpha docked with Mycobacteriophage D29 lysin B gp12		1 8 26 35 40 53 71 79 87 89 95 101	Gly Leu Asp Gly Lys Arg Ile Leu Leu Val Ser Ser	0.129 -0.31 -0.31 -0.31 -0.31 -0.31 -0.31 -0.31 -0.45 0 -0.45 0.31
4.	KasB-beta keto acyl carrier protein docked with Mycobacteriophage D29 lysin B gp12(Chain C)		1 9 27 37 47 53 75 85 91 97 105 113	Gly Leu Asp Gly Lys Arg Ile Leu Val Ser Gly Ile	0.129 0.31 0.31 0 0.129 -0.31 0 0 0 -0.31 0.31 0
5.	MarA-Multiple antibiotic resistant protein docked with Mycobacteriophage D29 lysin B gp12		1 13 22 30 46 60 77 86 93 98 99 100	Met Ser Pro Arg His Phe Thr Arg Arg Arg Arg Phe	-0.15 -0.31 0 0 0.31 0.31 0.31 0.31 0.34 0.24 0.35 0.31

(continued)

Table 9. (continued)

6.	Rprob-Rifampin resistant protein docked with Mycobacteriophage D29 lysin B gp12		1	Gln	0.129
			15	Thr	-0.31
			25	Pro	0
			32	Arg	0.31
			50	Gln	-0.31
			64	Arg	-0.31
			80	Leu	0
			89	Pro	0
			94	Val	0.31
			104	Leu	0
			110	Pro	0
7.	TlyA-Cytidine methyl transferase A docked with Mycobacteriophage D29 Lysin B gp12		1	Arg	0.129
			20	Ala	-0.31
			28	Trp	0
			42	Trp	0.14
			48	Val	0.31
			55	Ser	-0.31
			65	Arg	-0.31
			81	Gly	-0.31
			90	Ala	-0.45
			108	Lys	0
			122	Leu	-0.31
8.	WhiB7-Transcription regulator docked with Mycobacteriophage D29 lysin B gp12		1	Gln	0.129
			15	Thr	-0.31
			24	Pro	0
			32	Arg	0.31
			48	Gln	-0.31
			60	Arg	-0.31
			77	Leu	-0.31
			90	Pro	0
			100	Val	-0.45
			105	Leu	0
			114	Pro	-0.45

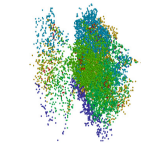
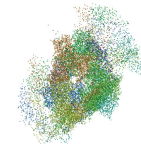
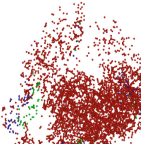
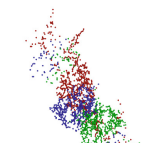
It is seen that Mycobacteriophage lysine B D29 has good docking scores with MarA (multidrug antibiotic resistant proteins), Erm37, whiB7. Patch dock server used to dock the proteins (Table 8). Protein protein model interaction analysed. Gromacs minimization energy by Galaxy Europe server and structural 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.

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

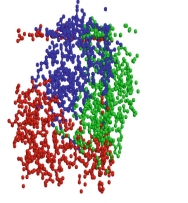
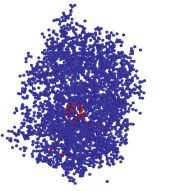
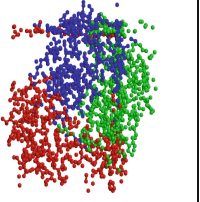
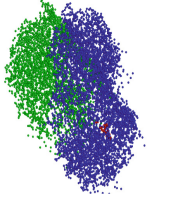
To select the putative site, an analogous experiment was performed with the bacteriophage protein and the sites were compared. The lysine B bacteriophage D29 had close overlap with Mar A gene binding sites and the binding energies were comparable ($\Delta G_{bind} = -12638$ kcal/mol). Final verification of docking experiments performed with MD simulation which suggested stable binding sites. To help understand discrimination

Table 10. Simulation model of human receptors and bacterial protein.

SL.NO	Human Receptors	Model	Numeral – residues	Amino –acids	Charges
1.	Complementary receptors docked with Eis protein.		1	Val	0.129
			11	Thr	-0.31
			20	Leu	-0.31
			30	Cys	0.31
			38	Ser	-0.31
			46	Pro	0
			53	Thr	0.31
			62	Glu	0.31
			72	Asp	0.31
			89	Trp	-0.31
			110	Pro	
2.	Complementary protein docked with InhA protein		1	Gly	0.129
			8	Leu	-0.31
			26	Asp	-0.31
			35	Gly	-0.31
			40	Lys	-0.31
			53	Arg	-0.31
			70	Ile	-0.31
			80	Leu	-0.31
			88	Val	0.31
			104	Gly	-0.31
			120	Gly	
3.	C-type lectin domain family 7, member A docked with RPOB		1	Val	0.05
			11	Leu	-0.31
			20	Ser	-0.31
			36	Pro	0
			44	Cys	0.31
			55	Pro	0.45
			65	Asn	0.31
			75	Trp	-0.31
			98	Ile	0
			115	Tyr	0.31
			135	Glu	0.16
4.	C-type lectin domain family 7, member A docked with whiB7		1	Val	0.05
			11	Leu	-0.31
			20	Ser	-0.31
			36	Pro	0
			45	Cys	0
			53	Pro	0
			65	Asn	-0.31
			75	Trp	-0.31
			96	Ile	-0.31
			117	Trp	0
			133	Glu	-0.31

(continued)

Table 10. (continued)

5.	C-X-C motif chemokine ligand 8 docked with Cytidine methyl transferase(tlyA)		1	Lys	0.05
			16	Glu	-0.31
			26	Leu	-0.31
			36	Arg	0.31
			52	Cys	-0.31
			60	Gln	0.31
			65	Gln	-0.45
			75	Cys	0
			82	Ile	0
			94	Lys	0.05
105	Thr	0.408			
6.	IRAK4 docked with kasB		1	Arg	0.129
			20	Phe	-0.31
			38	His	0.31
			52	Ser	-0.31
			60	Phe	0.31
			77	Ser	0.31
			88	Phe	0
			102	Tyr	0.31
			120	Glu	0.31
130	Leu				
7.	C-X-C motif chemokine ligand 8 docked with Mutidrug antibiotic resistant protein		1	Lys	0.05
			16	Glu	-0.31
			26	Leu	-0.31
			36	Arg	0.31
			52	Cys	-0.31
			60	Gln	0.31
			65	Gln	-0.45
			75	Cys	0
			82	Ile	0
			94	Lys	0.05
105	Thr	0.408			
8.	IRAK4 docked with expression resistant macroloide(erm 37)		1	Arg	0.129
			20	Phe	-0.31
			38	His	0.31
			52	Ser	0.31
			60	Phe	0.31
			78	Ser	0
			88	Phe	0
			103	Tyr	0
			120	Glu	0.31
			130	Leu	0.31
138	Lys	-0.31			

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 9).

<p>2-DO Structure (w/water)</p> <p>3088B</p> <p>JVAL N 1 9.250 3.702 5.015</p> <p>JVAL JI 2 9.218 3.684 5.071</p> <p>JVAL JI 3 9.231 3.687 4.916</p> <p>JVAL HA 9 9.320 3.632 5.040</p> <p>JVAL CA 5 9.304 3.638 5.029</p> <p>JVAL HA 6 9.263 3.605 4.965</p> <p>JVAL CB 7 9.258 3.686 5.175</p> <p>JVAL HB 9 9.395 3.818 5.236</p> <p>JVAL CI 9 9.356 4.027 5.195</p> <p>JVALHEH1 10 9.353 4.027 5.200</p> <p>JVALHG12 11 9.460 4.035 5.163</p> <p>JVALHDI12 12 9.298 4.100 5.139</p> <p>JVAL CO 13 9.152 3.983 5.232</p> <p>JVAL HD31 14 9.153 3.921 5.335</p> <p>JVAL HD22 15 9.087 3.945 5.418</p> <p>JVAL HD23 16 9.113 3.783 5.230</p> <p>JVAL C 17 9.448 3.836 4.972</p> <p>SLEU N 33 9.781 4.136 4.839</p> <p>SLEUJ 34 9.833 4.003 4.934</p> <p>SLEU CA 35 9.837 4.271 4.840</p> <p>SLEU HA 36 9.739 3.843 4.847</p> <p>SLEU CB 37 9.532 4.277 4.963</p> <p>SLEUHEH1 38 9.879 4.231 4.965</p> <p>SLEU HD 39 10.015 4.212 4.945</p> <p>SLEU HD40 5 9.880 4.410 5.013</p> <p>SLEU HD 11 42 10.377 4.402 5.023</p> <p>SLEU CD 43 10.043 4.401 5.152</p> <p>SLEU HD 143 10.377 4.402 5.181</p> <p>SLEUHD144 9.573 4.251 5.225</p> <p>SLEU HD1 145 10.137 4.231 5.150</p> <p>SLEU HD 46 10.397 4.486 5.210</p> <p>SLEUHD21 47 10.123 4.574 4.967</p> <p>SLEUHD148 10.193 4.420 4.892</p> <p>SLEUHD25 49 10.031 4.520 4.830</p> <p>SLEUHD50 9.978 4.286 4.769</p> <p>SLEU O 51 9.992 4.193 4.675</p> <p>BRPO CA 75 9.740 4.910 4.109</p> <p>BRPO HA 76 9.679 4.820 4.166</p> <p>BRPO CB 77 9.670 5.015 4.258</p> <p>BRPO HB 78 9.607 5.087 4.160</p> <p>BRPO HD 79 9.434 4.846 4.280</p> <p>BRPO CO 80 9.782 5.110 4.290</p> <p>BRPO H1 81 9.921 5.180 4.215</p> <p>BRPO H2 82 9.765 4.816 4.376</p> <p>BRPO CD 83 9.889 5.097 4.204</p> <p>BRPO HD 84 9.989 5.097 4.410</p> <p>BRPO HD 85 9.878 4.977 4.420</p> <p>BRPO C 86 9.763 4.846 4.310</p> <p>BRPO O 87 9.860 5.015 3.977</p> <p>9THR CA 89 9.608 4.905 3.929</p> <p>9THR H1 89 9.590 4.852 3.962</p> <p>9THR CA 90 9.608 4.905 3.986</p> <p>9THR HA 91 9.750 4.960 3.740</p> <p>9THR CB 92 9.679 4.844 3.732</p> <p>9THR HD 93 9.757 4.860 3.606</p> <p>9THR HD1 94 9.645 4.822 3.759</p> <p>9THR HD1 95 9.467 4.744 3.531</p>	<p>4THR N 19 9.484 3.930 4.883</p> <p>4THR H1 20 9.437 4.000 4.854</p> <p>4THR CA 21 9.633 4.050 4.830</p> <p>4THR HA 22 9.691 3.908 4.901</p> <p>4THR CB 23 9.644 3.875 4.096</p> <p>4THR HB 24 9.638 3.888 4.050</p> <p>4THR CO 25 9.545 3.813 4.598</p> <p>4THR HD 26 9.639 3.866 4.028</p> <p>4THR CG 27 9.641 3.724 4.708</p> <p>4THR HD21 28 9.661 3.678 4.612</p> <p>4THR HD 22 29 9.718 3.688 4.718</p> <p>4THR HD23 30 9.546 3.688 4.745</p> <p>4THR C 31 9.652 4.101 4.829</p> <p>4THR O 32 9.561 4.182 4.815</p> <p>SLEU N 33 9.781 4.136 4.839</p> <p>SLEU H1 34 9.801 4.003 4.830</p> <p>SLEU CA 35 9.837 4.271 4.840</p> <p>SLEU HA 36 9.759 4.346 4.847</p> <p>6CVS HB 66 9.598 4.388 4.265</p> <p>6CVS H1 67 9.654 4.374 4.260</p> <p>6CVS HB 68 9.593 4.281 4.409</p> <p>6CVS CO 69 9.721 4.384 4.382</p> <p>6CVS HD 69 9.698 4.298 4.281</p> <p>6CVS CA 1 9.598 4.574 4.386</p> <p>6CVS O 62 9.920 4.666 4.559</p> <p>ZSER CA 63 10.078 4.663 4.386</p> <p>ZSER H1 64 10.133 4.558 4.331</p> <p>ZSER CA 65 10.074 4.761 4.324</p> <p>ZSER HA 66 10.051 4.820 4.410</p> <p>ZSER CB 67 10.188 4.787 4.256</p> <p>ZSER HB 68 10.201 4.731 4.376</p> <p>ZSER HD 69 10.180 4.845 4.210</p> <p>ZSER CD 70 10.302 4.784 4.341</p> <p>ZSER HD 1 10.380 4.710 4.303</p> <p>ZSER O 72 9.836 4.777 4.297</p> <p>ZSER C 73 9.939 4.881 4.154</p> <p>BRPO N 75 9.684 4.981 4.127</p> <p>BRPO CA 75 9.740 4.910 4.150</p> <p>BRPO HA 76 9.679 4.820 4.166</p> <p>10GUL HA 105 9.495 5.350 3.712</p> <p>10GUL CB 106 9.495 5.311 3.698</p> <p>10GUL HB 107 9.530 5.305 3.671</p> <p>10GUL HD 108 9.435 5.247 3.627</p> <p>10GUL CD 109 9.436 5.458 3.482</p> <p>10GUL H01 110 9.432 5.245 3.377</p> <p>10GUL H02 111 9.330 5.470 3.520</p> <p>10GUL HD 112 9.585 5.450 3.588</p> <p>10GUL O 113 9.494 5.478 3.570</p> <p>10GUL HD 114 9.648 5.248 3.527</p> <p>10GUL C 115 9.310 5.214 3.660</p> <p>10GUL HD 116 9.285 5.297 3.697</p> <p>11ASP N 117 9.289 5.088 3.637</p> <p>11ASP CD 121 9.333 5.026 3.609</p> <p>11ASP CA 119 9.152 5.028 3.645</p> <p>11ASP HA 121 9.145 4.989 3.576</p> <p>11ASP CB 121 9.145 4.889 3.565</p> <p>11ASPH1 122 9.988 4.814 3.668</p> <p>11ASPH 122 9.911 4.859 3.567</p> <p>11ASPD 124 9.214 4.868 3.664</p> <p>11ASPD 125 9.274 4.749 3.628</p> <p>10AAL CA 158 11.460 5.160 5.137</p> <p>10AAL O 159 11.600 5.225 5.123</p> <p>10ASP N 160 11.440 5.074 5.046</p> <p>10ASP H 160 11.364 5.032 5.025</p> <p>10ASP CA 160 11.440 5.074 5.025</p> <p>10ASP HA 160 11.411 4.977 4.852</p> <p>10ASP CB 160 11.362 4.872 4.822</p> <p>10ASP H1 160 11.335 4.675 4.826</p> <p>10ASP HD 160 11.390 4.919 4.795</p> <p>10ASP CD 160 11.413 4.790 4.895</p> <p>10ASP HD 160 11.496 4.742 4.973</p> <p>10ASP CD 160 11.317 4.728 4.845</p> <p>10ASP C 160 11.559 5.144 4.832</p> <p>10ASP O 161 11.676 5.108 4.798</p> <p>11SER N 161 11.422 5.229 4.794</p> <p>11SER H 161 11.350 5.207 4.821</p> <p>11SER CA 161 11.462 5.260 4.701</p> <p>11SER HA 161 11.544 5.308 4.618</p> <p>11SER CB 161 11.355 5.388 4.646</p> <p>11SER HB 161 11.287 5.424 4.729</p> <p>11SER HD 161 11.357 5.485 4.574</p> <p>11SAL CA 167 11.315 5.954 5.090</p> <p>11SAL HA 168 11.311 5.846 5.795</p> <p>11SAL CB 169 11.324 6.058 5.783</p> <p>11SAL HB 167 11.261 6.056 5.854</p> <p>11SAL CO 167 11.405 6.016 5.861</p> <p>11SAL HD 167 11.271 6.013 5.844</p> <p>11SAL HD 167 11.480 5.960 5.911</p> <p>11SAL HD 167 11.521 6.078 5.902</p> <p>11SAL CO 167 11.329 6.192 5.710</p> <p>11SAL HD 167 11.375 6.077 5.860</p> <p>11SAL HD 167 11.420 6.209 5.628</p> <p>11SAL HD 167 11.289 6.197 5.927</p> <p>11SAL C 167 11.615 5.927 6.088</p> <p>11SAL O 168 11.386 5.950 5.847</p> <p>11SAL CA 168 11.072 5.893 5.675</p> <p>11SAL HA 168 11.088 5.880 5.776</p> <p>11SAL CA 168 11.088 5.890 5.624</p> <p>11SAL HA 168 10.935 5.924 5.224</p> <p>11SAL CO 168 10.889 5.743 5.564</p> <p>11SAL H1 168 10.791 5.730 5.529</p> <p>11SAL HD 168 10.958 5.683 5.561</p> <p>13LEU HD 13 1977 6.726 7.101 6.234</p> <p>13LEU CD 1978 9.985 6.198 6.815</p> <p>13LEU HD 1979 10.048 7.148 6.224</p> <p>13LEU HD 1980 9.816 7.278 6.286</p> <p>13LEU HD 1981 10.078 7.245 6.247</p> <p>13LEU C 1982 9.957 7.916 6.792</p> <p>13LEU O 1982 9.964 7.926 6.892</p> <p>13BRIT N 1983 9.920 7.466 6.738</p> <p>13BRIT H 1983 9.837 7.490 6.647</p> <p>13BRIT CA 1984 9.875 7.538 6.843</p> <p>13BRIT HA 1983 9.874 7.492 6.925</p> <p>13BRIT CB 1984 10.077 7.611 6.893</p> <p>13BRIT HD 1984 10.091 7.544 6.897</p> <p>13BRIT HD 1984 10.098 7.646 6.821</p> <p>13BRIT CO 1981 9.997 7.678 6.927</p> <p>13BRIT HD 1984 10.098 7.710 6.826</p> <p>13BRIT CD 1982 9.905 7.770 7.071</p> <p>13BRIT HD 1984 9.818 7.814 7.023</p> <p>13BRIT C 1995 10.048 7.728 7.232</p> <p>13BRIT H1 1996 10.095 7.729 7.330</p> <p>13BRIT HD 1997 9.941 7.789 7.202</p> <p>20BRPO HB 2093 6.990 7.834 4.568</p> <p>20BRPO CO 2094 7.981 7.839 4.539</p> <p>20BRPO H01 2095 7.703 7.819 4.563</p> <p>20BRPO HB 2096 7.770 7.653 4.431</p> <p>20BRPO CD 2097 8.024 7.653 4.589</p> <p>20BRPO HD 2101 7.668 7.570 4.502</p> <p>20BRPO HD 2099 7.696 7.732 4.521</p> <p>20BRPO C 2100 7.725 7.653 4.821</p> <p>20BRPO O 2101 7.709 7.534 4.835</p> <p>20BRPO N 2102 7.723 7.791 4.886</p> <p>20POLY H 2103 7.733 7.799 4.802</p> <p>20SERV CA 2104 7.929 7.620 4.811</p> <p>20SERV HA 2105 7.910 7.682 4.815</p> <p>20SERV HD 2106 7.613 7.635 4.871</p> <p>20SERV C 2107 7.611 7.649 5.185</p> <p>20SERV O 2108 7.788 7.695 5.283</p> <p>21GOLY N 2109 7.910 7.598 5.144</p> <p>21GOLY HD 2109 7.825 7.696 5.096</p> <p>21GOLY CA 2111 8.037 7.537 5.197</p> <p>21GOLY HA 2112 8.035 7.495 5.265</p> <p>21GOLY HA 2113 8.035 7.495 5.265</p> <p>21GOLY HA 2113 8.035 7.495 5.265</p>
---	---

Fig. 7. Gromacs energy minimization by using gmxolx Europe

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 10, Fig. 7).

4 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., & Rodríguez-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>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

