International Journal Of Public Mental Health And Neurosciences
ISSN: 2394-4668
(An Official publication of Sarvasumana Association)

Identification Of Taxa With Phylogenetic Profile And Functional Pathway Information And Novel Drug Leads Of Leishmania Microbiome

Darshna Kotharkar[1] , Milan Khandelwal[2] and Preenon Bagchi[3]

[1]Padmashree Institute of Management and Science, Bangalore,India

[2]Vasishth Academy of Advanced Studies and Research (Sarvasumana Association), Bangalore, India

[3]MGM Institute of Biosciences and Technology, Aurangabad, India

[*]Corresponding author Email: darshnakotharkar20@gmail.com.

Abstract:

Over 95% of cases of visceral leishmaniasis (VL), often known as kala-azar, are fatal if untreated. It is distinguished by irregular fever attacks, weight loss, spleen and liver enlargement, and anaemia. The majority of instances happen in India, East Africa, and Brazil. The second-deadliest parasitic illness and the most lethal neglected tropical disease is kala azar. Female sand flies are seen as the causal factor and it attacks the immune system. In this work, genome sequence of Leishmania microbiome was taken. After, pre-preprocessing of the sequence, the phylogeny was determined. Further, presence and absence with abundance of microbial pathways in our micro biota was efficiently and accurately profiling. Next, the expressed protein sequence was taken and novel drug lead was identified from the Indian traditional ayurvedic herbs.

Keywords: Leishmaniasis, metatranscriptomics, drug designing, Hex, molinspiration, pymol, 3D visualization.

I INTRODUCTION:

Any of the several species of flagellate protists in the genus Leishmania of the family Trypanosomatida. A parasite (protozoa) causes the ailment. These protists are vertebrate parasites that are transmitted by bloodsucking sand flies of the genera Phlebotomus and Lutzomyia. Visceral leishmaniasis (VL), often known as kala-azar, is lethal in many cases if left untreated.

Using a meta-transcriptomics technique, we attempted to identify the phylogeny of the microbiome in this study. We also applied HTS on the genomic sequence. Using drug-likeness software, molinspiration, we attempted to identify drug score values that indicate a compound's overall potential to be a drug candidate. Mol inspiration is a web-based application for predicting the bioactivity score of synthesised compounds against common human receptors such GPCRs, ion channels, kinases, nuclear receptors, proteases, and enzymes. Dave Ritchie created Hex, an interactive protein docking and molecular superposition programme. Hex recognises protein and DNA structures in PDB format, as well as small-molecule SDF files. PyMOL is a cross-platform molecular graphic programme that has been widely used for 3D macromolecule visualisation.

II. METHODS AND MATERIALS:

Metatranscriptomics is concerned with the study of how the microbiome responds to the environment, which is also explored by functional analysis of genes expressed by the microbiome. The science of virus discovery has been transformed by metatranscriptomic studies using whole RNA sequencing. It is more useful than microarrays in understanding the roles that host gene expression plays on a cellular or tissue level. Parallel sequencing and RNA-sequencing have opened up new and interesting possibilities for transcriptome study, offering insight and dynamic range. We may

investigate the gene expression of viral and bacterial communities in that environment, as well as their gene expression profiles. Metagenomics genes are observed to be transcribed from functional components, and active metabolic pathways can be discovered in bacterial and viral communities linked with that environmental situation. As a result, when compared to metagenomics, metatranscriptomics provides more detailed information regarding transcriptionally active .

Metatranscriptomics workflow: 1.Leishmania fastq sequences SRR12492456.1 and SRR12492456.2 were obtained from the SRA database.
FASTQC was used to assess the prominence of the sequence. Furthermore, MultiQC software was utilised to combine or aggregate quality checking data into a single report.

3.The trimomatic tool was used to slice the sequences.

4.FASTQC and ultiQC were re-run after cutting or slicing.

5.Any overlapping or unnecessary rRNA readings were then deleted.

6.FASTQ INTERLACE was used to connect paired end FASTQ reads from two different files.

7. The MetaPhlAn tool was used to profile the makeup of microbial communities (i.e., Bacteria, Archaea, and Eukaryotes).

The Krona tool was used to visualise metagenomic profiling [20, 21].

9.In addition, the HUMAnN pipeline was used to profile the presence/absence and abundance of microbial pathways.

1. Quality assurance
FastQC's goal is to make it simple to perform quality control checks on raw sequence data from high throughput sequencing workflows. MultiQC is a tool that allows you to aggregate numerous FastQC reports into a single file. The sequence quality is not as predicted, according to the sequence quality report. As a result, shortening the sequence becomes necessary. MULTIQC was used to combine the FASTQC results.
Trimomatic detects and removes unnecessary adapter sequences, primers, poly-A tails, and other sequences from our data.

Fig.1: It shows improved quality of sequence after trimming it using trimmotetric tool.

2.FILTER WITH SortMeRNA: SortMeRNA is a software that quickly filters ribosomal RNA fragments from metatranscriptomic data generated by next-generation sequencers. It can process massive RNA databases and sort out all fragments that match the database with high accuracy and specificity.
3. COMMUNITY PROFILE EXTRACTION a) MetaPhlAn MetaPhlAn is a computational tool for profiling the composition of microbial communities (Bacteria, Archaea, and Eukaryotes) from metagenomic shotgun sequencing data (not 16S) at the species level.
KronaThis programme displays hierarchical data from metagenomic profiling, in this case taxonomic levels, as a zoomable pie chart.

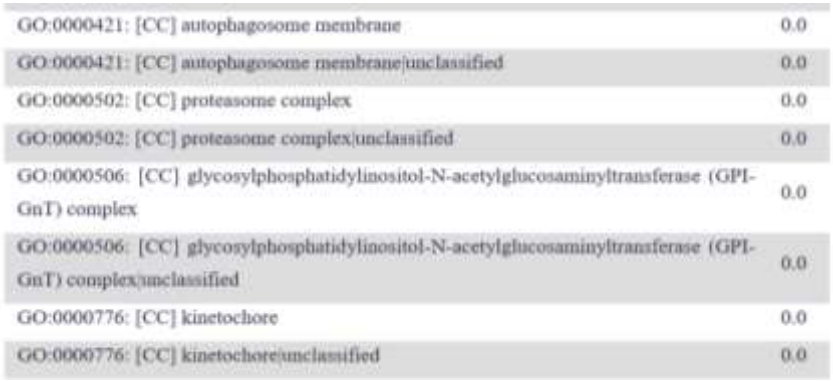
Fig. 2: Visualization of Taxonomic profile in Krona

4.Extract functional pathway information a) HUMAnN: HUMAnN is a pipeline for evaluating the presence/absence and abundance of microbial pathways in a community from metagenomic or metatranscriptomic sequencing data fast and reliably.

Sort gene families by GO terms.

The gene families can be a big list of ids, and reading through them one by one to find the interesting ones can be time consuming. Gene ontology analysis is commonly utilised in genome-wide expression investigations to minimise complexity and highlight biological processes. The GO keyword and id are extremely cryptic. They can be renamed and divided into three groups (molecular functions [MF], biological processes [BP], and cellular components [CC]).

GO:0000049: [MF] tRNA binding	0.0
GO:0000049: [MF] tRNA binding|unclassified	0.0
GO:0000287: [MF] magnesium ion binding	0.0
GO:0000287: [MF] magnesium ion binding|unclassified	0.0
GO:0000978: [MF] RNA polymerase II cis-regulatory region sequence-specific DNA binding	0.0
GO:0000978: [MF] RNA polymerase II cis-regulatory region sequence-specific DNA binding|unclassified	0.0
GO:0000979: [MF] RNA polymerase II core promoter sequence-specific DNA binding	0.0
GO:0002544: [BP] chronic inflammatory response	
GO:0002544: [BP] chronic inflammatory response|unclassified	0.0
GO:0002548: [BP] monocyte chemotaxis	0.0
GO:0002548: [BP] monocyte chemotaxis|unclassified	0.0
GO:0002553: [BP] histamine secretion by mast cell	0.0
GO:0002553: [BP] histamine secretion by mast cell|unclassified	0.0
GO:0002554: [BP] serotonin secretion by platelet	0.0



GO:0000421: [CC] autophagosome membrane	0.0
GO:0000421: [CC] autophagosome membrane|unclassified	0.0
GO:0000502: [CC] proteasome complex	0.0
GO:0000502: [CC] proteasome complex|unclassified	0.0
GO:0000506: [CC] glycosylphosphatidylinositol-N-acetylglucosaminyltransferase (GPI-GnT) complex	0.0
GO:0000506: [CC] glycosylphosphatidylinositol-N-acetylglucosaminyltransferase (GPI-GnT) complex|unclassified	0.0
GO:0000776: [CC] kinetochore	0.0
GO:0000776: [CC] kinetochore|unclassified	0.0

Fig.3: Gene ontology analysis is widely used to reduce complexity and highlight biological processes in genome-wide expression studies. There is a dedicated tool which groups and converts UniRef50 gene family abundances generated with HUMAnN into GO terms. The GO term with their id are quite cryptic. We can rename them and then split them in 3 groups (molecular functions [MF], biological processes [BP] and cellular components [CC]).

Structure based drug discovery for Leishmaniais

Drug discovery for fatal disease like visceral leishmaniasis is needed because the current drug used for its treatment is said to be toxic. Insilico drug design was done using softwares like Pubchem, Molinspiration, pymol and Hex which gave the following results.

Molinspiration:

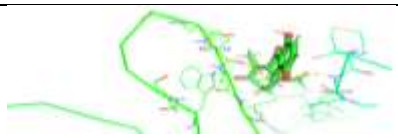
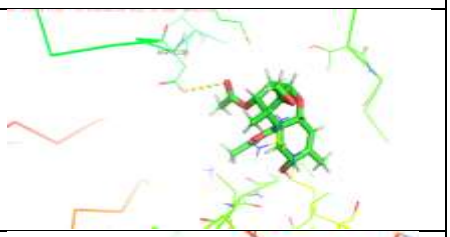
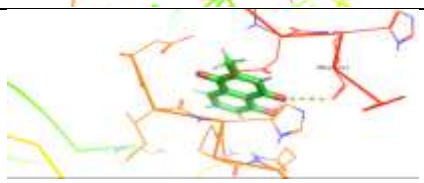
Table no. 1

Compound Name	plant	miLogP	TPSA	natoms	MW	nON	nOHNH	nrotb	nviolations	volume
sesquiterpene	Asteraceae	0.62	114.83	27	382.41	8	2	5	0	336.93
Acetylcordiaquinol	Cordiaima fragrantiss	3.94	52.61	24	328.41	4	0	0	6	317.78
Cordiachrome	Cordiaima fragrantiss	6	34.14	18	242.32	2	0	0	0	234.54
Cordiaquinol	Cordiaima fragrantiss	3.11	77.75	20	276.33	4	3	0	6	265.14
Cordiaquinol	Cordiaima fragrantis	3.39	74.60	20	272.30	4	2	0	0	243.81
Plumbagin	Cordiaima fragrantis	2.52	54.37	14	188.18	3	1	0	0	163.16
Ismailin	Cordiaima fragrantis	1.78	155.25	40	536.49	9	3	2	2	444.23

Betulin	Ocimum sanctum	5.47	40.46	32	442.73	2	2	1	2	469.86
Ursolic acid	Ocimum sanctum	7.16	57.53	33	456.71	3	2	1	1	471.49
Oleanolic acid	Ocimum sanctum	6.79	57.53	33	456.71	3	2	1	1	471.14
Stigmasterol	Ocimum sanctum	6.72	20.23	30	412.70	1	1	1	5	450.33

From the above table the compound with '0' nviolation value were selected to dock it with receptors that is the amino acids of organism. The value zero indicates the compound has drug-like property.

Table no. 2

Sr. No.	ligand	Interacting amino acid	Docking interaction
1	Cordia chrome	LEU-1219 ASP-1175 CYS-1174	
2	Cordia quinol	MET1156 LEU-1119	
3	Squisterpene	ASP-1130	
4	Plumbagin	PRO-1311	

Hex software was used to dock the ligand and receptor which showed interaction with different amino acid of the organism.

III. CONCLUSION

The genome of Leishmaniasis was phylogenetically determined. The genes are also found in the microbiome. Their phylogenetic relationship has been established. The functional data is further classified as BP (biological process), MF (molecular function), and CC (cellular component). The protein was extracted from the expressed genes, and the optimal ligand for the gene receptor was determined using a computer-aided drug design approach. According to the docking studies, Cordiachrome docks best with the receptor with the most three interactions, making it the best ligand for the receptor.

REFERENCES

1. Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., RodriguezMueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S. T., Methé, B., Schloss, P. D., Gevers, D., Mitreva, M., & 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
2. Andrews, S. (n.d.). FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
3. Barrett MP, Croft SL (2012). "Management of trypanosomiasis and leishmaniasis". British Medical Bulletin. 104: 175– 96. doi:10.1093/bmb/lds031. PMC 3530408. PMID 23137768.
4. Batut, B., K. Gravouil, C. Defois, S. Hiltemann, J.-F. Brugère et al., 2018 ASaiM: a Galaxy-based framework to analyze microbiota data. GigaScience 7: giy057.
5. Beghini, F., McIver, L. ., lanco-M gue , A., ubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. ELife, 10. https://doi.org/10.7554/elife.65088 6. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
6. Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., Eddy, S. R., Gardner, P. P., & Bateman, A. (2012). Rfam 11.0: 10 years of RNA families. Nucleic Acids Research, 41(D1), D226–D232. https://doi.org/10.1093/nar/gks1005
7. Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., Uva, P., Zanetti, G., & Fotia, G. (2014). Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics, 30(13),1928–1929. https://doi.org/10.1093/bioinformatics/btu135
8. 41
9. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19),2460–2461. https://doi.org/10.1093/bioinformatics/btq461
10. Franzosa, E. A., L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer et al., 2018 Species-level functional profiling of metagenomes and metatranscriptomes. Nature methods 15: 962.
11. Kopylova, E., L. Noé, and H. Touzet, 2012 SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28: 3211–3217.
12. 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
13. Kunath, B. J., F. Delogu, A. E. Naas, M. Ø. Arntzen, V. G. H. Eijsink et al., 2018 From proteins to polysaccharides: lifestyle and genetic evolution of Coprothermobacter proteolyticus. The ISME journal 1.
14. Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnology, 30(5), 434–439. https://doi.org/10.1038/nbt.219
15. Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17: 10–12.
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. Ondov, B. D., N. H. Bergman, and A. M. Phillippy, 2011 Interactive metagenomic visualization in a Web browser. BMC bioinformatics 12: 385.
18. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(D1), D590–D596. https://doi.org/10.1093/nar/gks1219
19. Roy, Mrinalini; Rawat, Aadish; Kaushik, Sanket; Jyoti, Anupam; Srivastava, Vijay Kumar (1 August 2022). "Endogenous cysteine protease inhibitors in upmost pathogenic parasitic 42 protozoa". MicrobiologicalResearch. 261:127061. doi:10.1016/j.micres.2022.127061. IS SN 0944-5013.
20. Truong, D. T., E. A. Franzosa, T. L. Tickle, M. Scholz, G. Weingart et al., 2015 MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature methods 12: 902.
21. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. (GBD 2015 Disease Injury Incidence Prevalence Collaborators) (October 2016). "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015". Lancet. 388 (10053):1545–1602. doi:10.1016/S0140-6736(16)316786. PMC 5055577. PMID 27733282