Identification of potential microRNAs in *Mycobacterium leprae* and their interaction with human genes

Durbba Nath[1], Supriyo Chakraborty[1*]

[1]Department of Biotechnology, Assam University, Silchar, Assam, India.

Email Id of authors: durbba.nath@gmail.com, supriyoch_2008@rediffmail.com

*Corresponding author, Email: supriyoch_2008@rediffmail.com

Mycobacterium leprae is a pathogenic species that is the primary cause behind the grievous Hansen's disease, generally termed Leprosy. Being one of the disparate species within its genus, *Mycobacterium leprae* stimulates peripheral nerve impairment. In this study, computational approach was adopted to screen the genome of *Mycobacterium leprae* for the detection of potential microRNAs. The microRNAs are non-protein coding RNAs that can act on multiple target genes and repress their activities. The validation of prokaryotic microRNAs or micro-size RNAs (msRNAs) was recently done on a few species like *Escherichia coli, Streptococcus mutans and Salmonella* etc. The computational screening of *Mycobacterium leprae* genome resulted in the recognition of potential microRNAs. The softwares CID-miRNA and MFOLD predicted precursor microRNAs based on the typical characteristics of their secondary structures and stochastic context-free grammar. It was reported that bacterial microRNA precursors disseminate into the host cytosol, where they exploit host cellular machinery to produce functional miRNAs. In the host cytosol, the bacterial miRNAs can target host mRNA molecules. The targets of predicted mycobacterial microRNAs were searched in human transcriptome (3' UTRs) with the webserver tool miRDB. The tool utilized the model 7mer-m8 seed match to search the possible targets. Further, STRING and Cytospace were employed to arrange the human target genes into a gene network and gene clusters. The tools MCODE ((Molecular Complex Detection) and Network Analyzer are integral tools of Cytoscape that categorized the essential regulator genes of the network. The Gene Expression Omnibus (GEO) available at National Centre for Biotechnology Information (NCBI) stores the list of human genes differentially expressed in Leprosy. A few target genes predicted in this study were also present in the list of differentially expressed genes reported in GEO, NCBI. Lastly, the genes were subjected to functional annotation analysis in the software BLAST2GO to understand their functional roles (biological process, molecular function and cellular compartment).

The results of the study were as follows:

1. The complete genome of *Mycobacterium leprae* was screened with CID-miRNA software that resulted in the prediction of precursor-microRNAs (pre-miRNAs). The resultant data was filtered and subjected to MFOLD to generate the possible secondary structures. In total, 15 pre-miRNAs were predicted and their genomic positions were ascertained.

2. The webtool MatureBayes predicts mature miRNAs from both the strands (5' and 3') of pre-miRNA. Thus, with the input of 15 putative pre-miRNAs, 30 mature miRNAs were predicted.

3. All the 30 mature miRNAs were selected for target prediction in the human transcriptome (3' UTRs) with the aid of the webtool miRDB. The webtool identified 297 human genes as possible targets.

4. The interactions among the target genes were studied in STRING and Cytoscape. The whole network connected 116 genes with 189 edges.

5. Gene clusters were identified with the application MCODE. The top 10 hub genes were identified with Network Analyzer.

6. The target genes (FN1, HECTD2, DACT1, MMP1, MCTP2) were amongst the few differentially expressed genes reported in GEO, NCBI.

7. The target genes were involved in crucial functions that were classified into biological processes (cell morphogenesis, ribosome biogenesis, developmental maturation and catabolic process etc) and molecular

functions (ubiquitin like protein binding, ion binding, RNA binding and lipid binding etc). The target genes were localized in plasma membrane, nucleoplasm and cytoplasm etc.

References

1. Dang, T.H.Y., Tyagi, S., D'Cunha, G., Bhave, M., Crawford, R., and Ivanova, E.P. (2019). Computational prediction of microRNAs in marine bacteria of the genus Thalassospira. PloS one *14*, e0212996.
2. Furuse, Y., Finethy, R., Saka, H.A., Xet-Mull, A.M., Sisk, D.M., Smith, K.L.J., Lee, S., Coers, J., Valdivia, R.H., and Tobin, D.M. (2014). Search for microRNAs expressed by intracellular bacterial pathogens in infected mammalian cells. PloS one *9*, e106434.
3. Shmaryahu, A., Carrasco, M., and Valenzuela, P.D. (2014). Prediction of bacterial microRNAs and possible targets in human cell transcriptome. Journal of Microbiology *52*, 482-489.
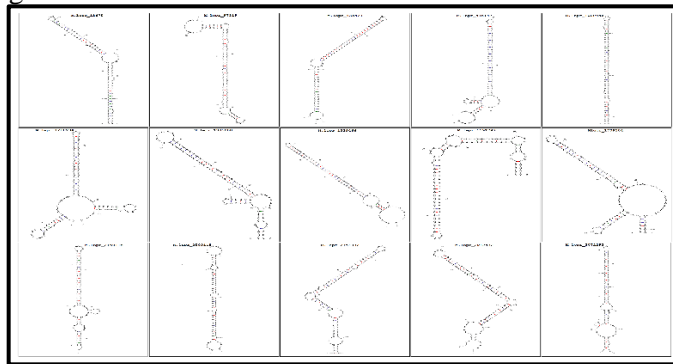
Figures

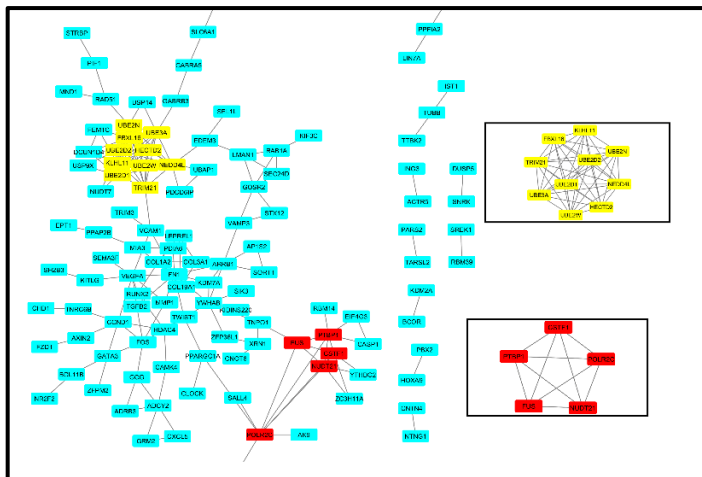Fig 1. Stem-loop structures of the precursor miRNAs predicted in *Mycobacterium leprae*.

Fig 2. Arrangement of target genes in a network. The genes (nodes) involved in Cluster 1 are depicted in yellow colour. The genes (nodes) involved in Cluster 2 are depicted in red colour.

*More figures and tables will be given in poster.

Table 1. Details on miRNA precursor sequences

Precursor miRNA name	Position in the genome		Length	MFE	(A+U) %
	start	end			
M.lepr_pre-mirna_13675	13675	13774	100	-29.80	65.00
M.lepr_pre-mirna_97319	97319	97425	107	-49.20	34.58
M.lepr_pre-mirna_435427	435427	435527	101	-64.60	44.58
M.lepr_pre-mirna_575147	575147	575259	113	-68.14	49.56
M.lepr_pre-mirna_1000056	1000056	1000158	103	-55.60	33.01
M.lepr_pre-mirna_1280294	1280294	1280399	106	51.10	45.28
M.lepr_pre-mirna_1305808	1305808	1305907	100	-27.83	58.00
M.lepr_pre-mirna_1339466	1339466	1339565	100	-69.50	48.00
M.lepr_pre-mirna_1391747	1391747	1391860	114	-70.30	40.35
M.lepr_pre-mirna_1709384	1709384	1709484	101	-36.82	58.42
M.lepr_pre-mirna_2160115	2160115	2160219	105	-50.33	47.62
M.lepr_pre-mirna_2509118	2509118	2509223	106	-49.80	36.79
M.lepr_pre-mirna_2754332	2754332	2754437	106	-40.10	36.79
M.lepr_pre-mirna_2942432	2942432	2942531	100	-47.30	37.00
M.lepr_pre-mirna_3072292	3072292	3072391	100	-59.04	40.00

Table 2. Mature miRNA sequences (5p & 3p)

Mature miRNA names	Sequence(5'-3')
M.lepr_5p_13675	AAATATATTCTTTGTGCAAATT
M.lepr_3p_13675	ACTTTTGCCGTAGATAGATCTA
M.lepr_5p_97319	CAGATCTTGCTGTTGCTCGGCG
M.lepr_3p_97319	CGAACCGACGAGCAGCTAGTTG
M.lepr_5p_435427	CTGGCGATGTTCGCCGACGACG
M.lepr_3p_435427	CGGCGAACATCGCCAGCCAGGT
M.lepr_5p_575147	TTGCGCTTGCGCAAACTAGTAC
M.lepr_3p_575147	ACCAGGAACTTATCGAATTGGT
M.lepr_5p_1000056	TGGGCACCACTGATCCGTGCCC
M.lepr_3p_1000056	CAGATCGGTGTCGCTGACGTCT
M.lepr_5p_1280294	AATATGCGGCGTCAGTACCGTC
M.lepr_3p_1280294	TACTGACGCCGCATATTAGCTT
M.lepr_5p_1305808	TATTGCCTTCAGGATTGTATTA
M.lepr_3p_1305808	CTACAATCCAGACGGCAATCTC
M.lepr_5p_1339466	CAACTAGCTTGGCTTCTCCTCA
M.lepr_3p_1339466	TGAGGAGAAGCCAAGCTAGTTG
M.lepr_5p_1391747	GATCACCATGGCCACGAAGTCC
M.lepr_3p_1391747	ACTTCGTGGCCATGGTCGACTG
M.lepr_5p_1709384	CGAGACATATCTATCTATCATA
M.lepr_3p_1709384	AGATAGATATGTTGCATGTTTG
M.lepr_5p_2160115	ATGTCCGACGCTCATGAACAGT
M.lepr_3p_2160115	CGTTGTTCATGAGCGTCGGACA
M.lepr_5p_2509118	TTTGGGCCAGGCAGCTTCAGTC
M.lepr_3p_2509118	CAAAGCTGGATGGCGACAAACC
M.lepr_5p_2754332	GTATCGGCTCGTCGGTGTGCGC
M.lepr_3p_2754332	GCCGACAACCAACGAGCGACTAT
M.lepr_5p_2942432	CAGGCCCTACTCGGCCTGGCGG
M.lepr_3p_2942432	AACCCACCGCACAGATCTAACT
M.lepr_5p_3072292	ACTACCAGCGGGATGCTGGTGG
M.lepr_3p_3072292	CCACCACCAGCATCCCGCTGCA

Table 3. Two identified clusters in the target gene network, their nodes, edges and scores.

Clusters	Human genes targeted by *M. leprae* miRNAs	Nodes	Edges	Scores
Cluster 1	UBE2N, UBE2W,UBE2D2, TRIM21, UBE3A, NEDD4L, KLHL11, FBXL16 (seed), HECTD2, UBE2D1	10	45	10
Cluster 2	PTBP1, CSTF1, NUDT21, POLR2C (seed), FUS	5	10	5