EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES

http://www.ejbps.com

ISSN 2349-8870
Volume: 5
Issue: 5
578-584
Year: 2018

COMPUTATIONAL ANALYSIS OF MOLECULAR SEQUENCES OF *MYCOBACTERIUM TUBERCULOSIS* AND *LEPRAE ON THE BASIS OF PATHWAY ANALYSIS* FOR DRUG TARGET IDENTIFICATION

Lakshmi Pillai* and Usha Chouhan

Research Scholar, Department of Computational and Integrated Science, MANIT, Bhopal.

***Corresponding Author: Lakshmi Pillai**

Research Scholar, Department of Computational and Integrated Science, MANIT, Bhopal.

Article Received on 22/03/2018 **Article Revised on 15/04/2018** **Article Accepted on 06/05/2018**

ABSTRACT

The genus *Mycobacterium* is best known for its two major pathogenic species, *M. tuberculosis* and *M. leprae*, the causative agents of two of the world's oldest diseases, tuberculosis and leprosy, respectively. *M. tuberculosis* kills approximately two million people each year and is thought to latently infect one-third of the world's population. Proteins of these two strains vary a lot and do share some similar characteristics too. Wet lab experiments usually used to classify the proteins of these strains are highly expensive, labor intensive and time consuming. Thus there arises a need for computational approach for classification of H37Rv and *Leprae*. These computational approaches are fast and economical as compared to wet lab techniques. Realizing their importance, in this paper an attempt has been made to correlate strains on the basis of literature review and pathway analysis and predict them with fair accuracy. A pathway analysis is done and pathway is drawn using bioinformatics tool on the basis of literature survey showing relation between *M Leprae* and *M.TB*.

KEYWORDS: Mycobacterium Tuberculosis (M.TB), M. Lepare, Pathway analysis, Fibrosis, Granuloma, Axenically.

INTRODUCTION

Mycobacterium is a genus of Actinobacteria, given its own family, the Mycobacteriaceae. The genus includes pathogens known to cause serious disease in mammals, including leprae and tuberculosis. The mycobacterial cell wall has unique characteristics and is impermeable to a number of compounds, a feature in part responsible for inherent resistance to numerous drugs. Tuberculosis[1,2] is the classical human mycobacterial disease, caused by *Mycobacterium Tuberculosis.*[3] The disease primarily affect the lung and causes pulmonary tuberculosis, as well as affect intestine, bone, joints, meninges, lymph nodes, skin and other tissue of the body, causing extrapulmonary tuberculosis. *Mycobacterium leprae* is the causative agent of the disease, leprosy, also known as Hanson's Disease. There are three forms of the disease, with lepromatous being the most severe, tuberculoid, and borderline leprosy.[5,6] The precise mode of transmission is not fully understood. The bacteria are likely to spread by direct contact and through air dispersement from coughing or sneezing. Virulence factors include a waxy exterior coating, formed by the production of mycolic acid unique to *Mycobacteria*.[7] Thus there arises the need to understand the relationships among various parameters of the proteins involved in these diseases for prediction of their classes, structures and functionality. And one of the ways to understand protein interaction is through pathway analysis. The computational approaches for prediction of their classes are fast and economical therefore can be used to complement the existing wet lab techniques. Realizing their importance, in this paper an attempt has been made to and predicts them with fair accuracy. This is a novel method where comparative analysis is made on the basis of pathway for *Mycobacterium Tuberculosis* and *Leprae.*[8]

The German scientist (**Robert Koch)** announced that he had cultured the causative agent from human TB lesions[9] and designated as "Bacillus of Tuberculosis". About 90% of those infected with *Mycobacterium tuberculosis*[4] have asymptomatic, latent TB infection [11] (sometimes called LTBI), with only a 10% lifetime chance that a latent infection will progress to TB disease. However, if untreated, the death rate for these active TB cases is more than 50%. TB infection begins when the *mycobacteria* reach the pulmonary alveoli, where they invade and replicate with the endosomes of alveolar macrophages. The primary site of infection in the lungs is called the Ghon focus, and is generally located in either the upper part of the lower lobe, or the lower part of the upper lobe. Bacteria are picked up by dendritic cells, which do not allow replication, although these cells

can transport the bacilli to local (mediastinal) lymph nodes. Further spread is through the bloodstream to other tissues and organs where secondary TB lesions can develop in other parts of the lung (particularly the apex of the upper lobes), peripheral lymph nodes, kidneys, brain, and bone. All parts of the body can be affected by the disease, though it rarely affects the heart, skeletal muscles, pancreas and thyroid. Tuberculosis is classified as one of the granulomatous inflammatory conditions. Macrophages, T- lymphocytes, B-lymphocytes and fibroblasts are among the cells that aggregate to form agranuloma, with lymphocytes surrounding the infected macrophages. The granuloma functions not only to prevent dissemination of the mycobacteria, but also provides a local environment for communication of cells of the immune system. Within the granuloma, T lymphocytes secrete cytokines such as interferon gamma, which activates macrophages to destroy the bacteria with which they are infected. Cytotoxic T cells can also directly kill infected cells, by secreting perforin and granulysin.[12]

Importantly, bacteria are not always eliminated within the granuloma, but can become dormant, resulting in a latent infection. Another feature of the granulomas of human tuberculosis is the development of cell death, also called necrosis, in the center of tubercles. To the naked eye this has the texture of soft white cheese and was termed caseous necrosis.

If TB bacteria gain entry to the bloodstream from an area of damaged tissue they spread through the body and set up many foci of infection, all appearing as tiny white tubercles in the tissues. This severe form of TB disease is most common in infants and the elderly and is called miliary tuberculosis. Patients with this disseminated TB have a fatality rate of approximately 20%, even with intensive treatment.[27]

In many patients the infection waxes and wanes. Tissue destruction and necrosis are balanced by healing and fibrosis. Affected tissue is replaced by scarring and cavities filled with cheese-like white necrotic material. During active disease, some of these cavities are joined to the air passages bronchi and this material can be coughed up. It contains living bacteria and can therefore pass on infection. Treatment with appropriate antibiotics kills bacteria and allows healing to take place. Upon cure, affected areas are eventually replaced by scar tissue.

The genome of *M.leprae*[14] has been sequenced; however, *M.leprae* still cannot be cultivated without contamination or association with other organisms. Before completion of the sequence of *M.leprae*, the only highly purified bacilli originated from infected armadillo tissues, which caused difficulty because certain secreted proteins were lost upon purification of the bacilli from the armadillos. New approaches to identifying genes from completed *M.leprae* genome sequences are being applied using standard bioinformatics tools that can identify proteins with special features like unique or shared amino acid sequence homologies with proteins of *M.tuberculosis* or other *mycobacteria*, and the presence of specialized peptide signatures suggesting their cellular location and possible secretion across the cell membrane (Scollard et. al). Proteins of interest can be prioritized based on potential B-cell and T-cell epitopes. Then the proteins selected for further study can be purified as recombinant proteins for an endless supply of the protein for vaccine research.

Currently, efforts are underway to develop new therapeutic agents and elucidation of metabolic pathway associated with diseases. Moreover, the mere understanding of *Mycobacterium* will assist in finding novel drug target with minimum side effects. The experimental attempts are reported in the literature for functional classification of *Mycobacterium Tuberculosis* and *Leprae*. But no computational technique is available in the literature for classification of *Mycobacterium tuberculosis* and *Leprae*[15] based on parameters like pathway analysis. Since the experimental identifications of them are labor and cost-intensive task, computational biology can provide a better alternative to develop a method for classifying *Mycobacterium Tuberculosis* and *Leprae on the basis of pathway*.

In view of the above an attempt has been made in this paper to develop a computational approach for pathway analysis and classifying two types of Mycobacterial pathogens i.e. *Mycobacterium Tuberculosis* and *M Leprae*. This paper is a step in the direction where computational biology techniques can be used to complement existing wet lab techniques[10]

The scientists are interested in identifying similarities and differences in composition of *Mycobacterium Tuberculosis* and *Leprae* and determining target proteins for further drug designing. Looking at the two *mycobacteria* on a stain, they appear very similar. They have similar cell walls and similar enzymatic capabilities. However, their chromosomes are sequenced quite differently. *M. Tuberculosis* has about 4000 gene capabilities, and therefore a greater capability of making and degrading compounds. *M.leprae* has about 1600 genes. The two *mycobacteria* have about the same number of molecules.[8]

About 1500 genes are common to both *M. leprae* and *M. tuberculosis*. The comparative analysis suggests both *mycobacteria* derived from a common ancestor and, at one stage, had gene pools of similar size. Downsizing from a genome of 4.42 Mbp, such as that of *M. tuberculosis*, to one of 3.27 Mbp would account for the loss of some 1200 protein-coding sequences. There is evidence that many of the genes that were present in the genome of the common ancestor of *M. leprae* and *M. tuberculosis* have been lost by recombination in the *M.leprae* genome.[17]

Wheeler enciphered the Biochemistry using genome analysis which showed that deletions and appearance of pseudogenes in pathways of carbon source utilisation and energy metabolism best explain the host-dependency and failure to culture *Mycobacterium leprae* axenically. This is because of the absence of some genes which are necessary for certain pathways like purine and pyrimidine synthesis is not present or is repressed in *Mycobacterium Leprae* but its present in *Mycobacterium Tuberculosis* which resulted in host dependency of *Leprae* for growth. Thus it depends on the culture media for growth.[14]

From a thorough examination of the genome of *M. leprae*, it was found that *M leprae* does have all the main metabolic and biosynthetic pathways, but there are some missing genes. Compared to the genome of *M. tuberculosis*, *M leprae* lacks the genes for vitamin B12 synthesis and metC, a gene that codes for the enzyme that converts cysteine into methionine.[24,10] This means that *M leprae* cannot synthesize the vitamin or methionine, an amino acid, and instead has to be taken from the environment.[24] To test the presence of methionine and vitamin B12 were enough for *M leprae* to grow; an attempt was made to culture the bacteria in media containing these molecules were not successful.[18,19]

The BCG vaccine is used as a vaccine against tuberculosis in many countries worldwide (not the U.S.) with variable efficacy. The effectiveness of the BCG vaccine in providing immunity against leprosy is quite controversial, however. Researchers have done complementary analysis of the BCG vaccine to see its effect on leprosy and levels of protection have varied from no effect to up to 80%. So we need a better vaccine for leprosy but in some cases it can be used.[7]

None of the characteristics described in the above sections allows by itself a direct classification of *Mycobacterium Tuberculosis* and *Leprae* as having any common pathogenic behavior.[16]

Classification (Fig.1) developed here is useful in evolutionary studies[20,21] where *Mycobacterium Tuberculosis* show conservation with *Mycobacterium Leprae* which can be clearly analysed by further clustering studies and have got applications in drug designing.[22,23] Both bacteria are non-opportunistic pathogen which is a common characteristic.

Till now various attempts have been made to discover similar proteins in *Mycobacterium Tuberculosis* and *Leprae* by using both in-vivo and in-silico techniques and elucidate their role in various regulatory processes, so as to find out a cure by targeting those proteins.[26,29] But from literature survey it appears that no attempt has been made to develop computational approaches for classification of *Leprae* and *Tuberculosis*. Thus in this paper an attempt is made to classify[19] *Mycobacterium Tuberculosis* and *Mycobacterium Leprae* on the basis of pathway analysis and specifically on the basis of carbon metabolism pathway.

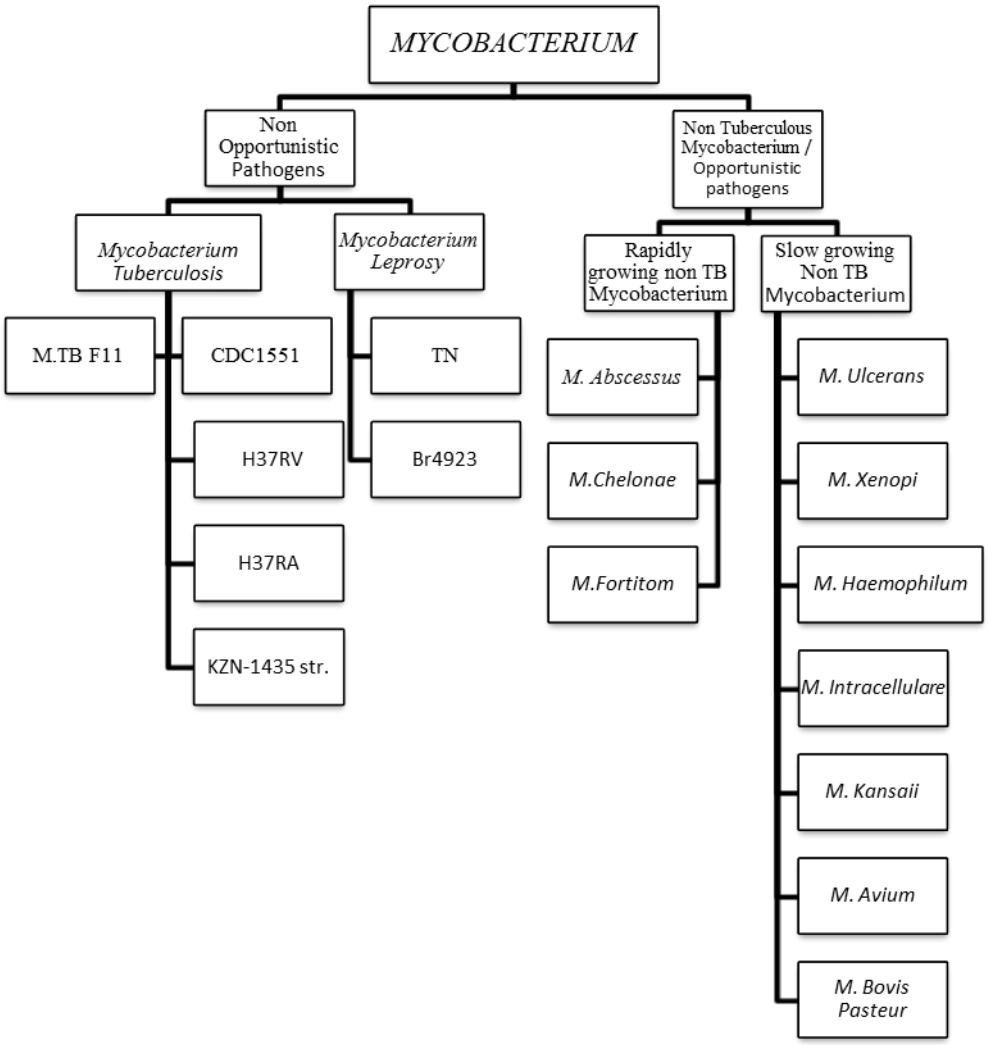

Fig. 1: Classification of *Mycobacteria.*

MATERIALS AND METHODS

Data set: To achieve our goal and develop our methodology we obtained the dataset from NCBI[30] and Swissprot/Uniprot databank of Expasy server. The following two data sets were used.

A pathway analysis was performed to show the relation between two species. For which several pathway from KEGG and literature are analysed and a target pathway i.e carbon metabolism pathway was selected and further analysed.

First analysis is performed using Biocyc.org.[25] In addition to reflecting differences in biology of different organisms, statistics differences are calculated i.e. in the levels of curation, data availability, and completeness of the PGDBs for these organisms.

Comparative analysis and statistics were computed for the following organism databases:

- *Mycobacterium leprae Br4923*[28]
- *Mycobacterium Tuberculosis H37Rv.*

CLASSIFICATION ON THE BASIS OF PATHWAY

Software: ACD/ChemSketch[1] Freeware is a drawing package; it is a chemically intelligent drawing interface that allows you to draw chemical structures including organics, organometallics, polymers, and Markush structures. It also includes features such as calculation of molecular properties (e.g., molecular weight, density, molar refractivity etc.), 2D and 3D structure cleaning and viewing, functionality for naming structures (fewer than 50 atoms and 3 rings), and prediction of log*P*. Chemketch is used here to produce professional looking pathway diagram for report.

RESULTS AND DISCUSSION

When comparative analysis is performed on the genome of Mycobacterium TB[31] and Leprae significant losses are found in the genome of *M. leprae* for carbon and energy metabolism. So a comparative pathway analysis is done on the basis of literature review and using bioinformatic tool chemsketch a pathway is drawn.

Evaluation

Pathway analysis was performed to show the relation between two species. For which carbon metabolism

pathway was analysed. In carbon metabolism pathway analysis we found similarity and dissimilarities between M. lepare and M.T.B. M. *leprae* metabolizes glucose, but other carbon sources cannot be metabolized for energy. There are pseudogenes present in *M. leprae* corresponding to genes in *M. tuberculosis* for proteins that enable the bacteria to breakdown other carbon sources such as acetate and galactose. Figure 2 shows carbon metabolism in *M. leprae* compared to *M. tuberculosis*. Pathways in black are present in both *M. leprae* and *M. Tuberculosis*;[25] grey pathways are pathways that exist in *M. tuberculosis* that *M. leprae* still has pseudogenes for. Bolded carbon sources are possible for use by *M. leprae*, grey carbon sources are unusable by *M. leprae*. The pathway with the red arrows is the Krebs cycle. Not being able to use these sources for energy means that *M. leprae* has to rely much more heavily on using glucose, which is not found in the environment as commonly as other carbon sources. Most microbes are able to convert other sources such as galactose into glucose, but *M. leprae* is not able to do this and must rely on another organism to make sure it has enough glucose. Thus this is a major comparison on the basis of pathway which says that both the disease are related and further study of this very important pathway may result in a drug target identification and finally we may get a cure for T.B.

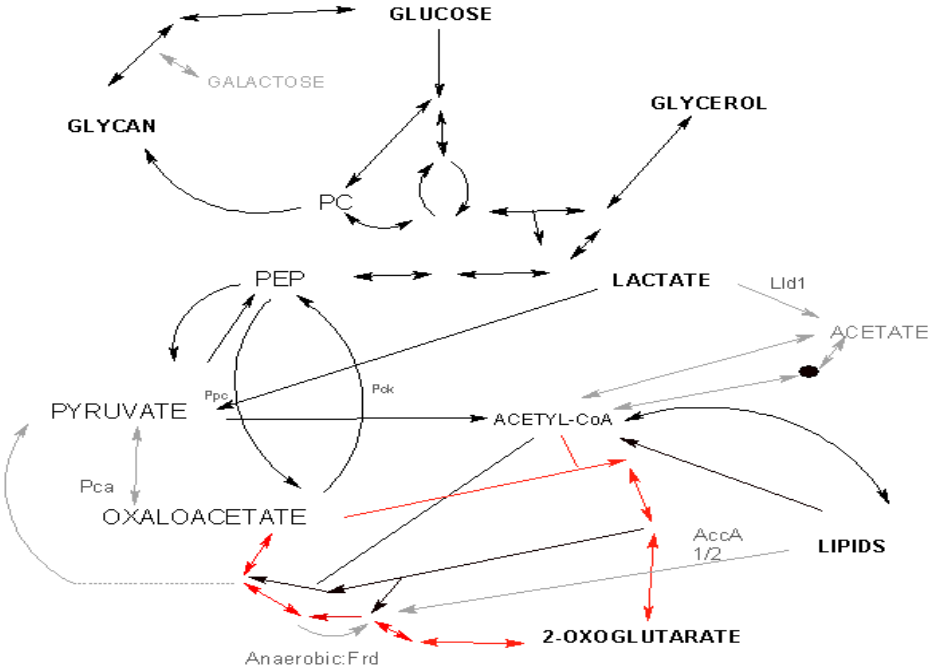

Figure 2: Carbon metabolism in *M. leprae* compared to *M.Tuberculosis*[24] Pathways are shown by arrows: black or red are present in M. leprae and M. tuberculosis, grey-blue in M. tuberculosis – here orthologues are deleted or are pseudogenes in *M. leprae*. Possible carbon sources for growth are shown in bold, regular script. Those in grey-blue are not available to M. leprae though they are to M.*tuberculosis*: to simplify the diagram keto acids are shown even if their more likely source is the corresponding amino acid. Cycles: the pentose cycle is shown as P.C, Krebs' cycle is shown with red arrows for clarity [with the kind permission of Springer Publications].[24]

On the basis of analysis of pathway and literature review it's found that certain genes are responsible for the ability of H37Rv to grow in macrophage like trp D, pro C and met C which are identified as the drug target genes as described in Table. 1.

Table 1: Drug Target genes.

Gene name	Gene Number ML	Predicted Function/ Comments
met B	2394	In methionine biosynthesis. Cystathionine c-synthase: catalyses O-succinylhomoserine + cysteine = cystathionine + succinate Cysteine is converted to methionine.
trpD	0883	Anthranilate phosphoribosyltransferase, essential for tryptophan biosynthesis.
proC	2430	Pyrroline-5-carboxylate reductase, Essential for proline biosynthesis

DISCUSSION

After comparative pathway analysis it's found that more number of pathways and pathway holes are there in *M.TB* compared to *leprae*. This means that genes without known function are more in mycobacterium tuberculosis H37Rv strain compared to leprae. These unknown genes may be ignored drug targets and can become the major drug target. After major and in-depth analysis it's found that there are some pathways in *leprae* without which it can't grow axenically like methionine synthesis pathway but mainly it depends upon the environment for growth. Similarly in Mycobacterium Tuberculosis tryptophan and proline pathway are important without which *M.TB* cannot grow in macrophage or mice. So targeting these pathways may lead us to drug for TB which may slow down the infection or may also abrogate the growth of bacteria causing TB. On the basis of analysis of pathway we found that certain genes are responsible for the ability of H37Rv to grow in macrophage and they can be targeted for further drug designing.

CONCLUSION

On the basis of literature survey carbon metabolic pathway analysis is done which concludes that there is a lot of relation between these two diseases and their respective pathways which may help in identifying the drug target through further analysis and may result in finding the cure which is our ultimate aim. In future work plan is to analyze the pathway further and to find out more target proteins or gene for cure through computational analysis. Also there is a need to take the work on experimental level for further gene analysis which may be difficult considering the time constraint which it poses on existing lab techniques.

ACKNOWLEDGMENT

The authors are highly thankful to the financial support in the form of Bioinformatics infrastructure facility, fully sponsored by Department of Mathematics, Computational and Integrated Science and Computer Applications, Maulana Azad National Institute of Technology, Bhopal to successfully carry out this research work. Also would like to thank D.B.T for infrastructure facility.

REFERENCES

1. ACD/Chemsketch (2009) version 12.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada. www.acdlabs.com.
2. Alamelu Raja (October 2004) Immunology of tuberculosis, Indian J Med Res 120, 213-232.
3. Alex Muller (2011) How TB infects the body: The Tubercle, TB Online. http://www.tbonline.info/posts/print/how-tb-infects-body-tubercle/.
4. Bloom, R. B., and C. J. L. Murray (1992) Tuberculosis: commentary on a reemergent killer. Science, 257: 1055-1064.
5. C.P. Blobel, Cell, 90: 589 (1999).
6. Case, C. L., Funke, B. R., & Tortora, G. J. , (2009) Chapter 7: The Control of Microbial Growth. Microbiology: An Introduction (10th Edition, 203), San Fransisco: Benjamin Cummings.
7. CDC TB Fact Sheets - BCG Vaccine, Centers for Disease Control and Prevention. (2008, Nov.) Retrieved May 19, 2010. http://www.cdc.gov/tb/publications/factsheets/vaccine/BCG.htm.
8. Cole ST (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence, Nature, 393(6685): 537-544.
9. Caspi R, Altman T et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, Nucleic Acids Research, 38: D473-9.
10. D.Barich (2010), Microbe wiki (online). http://microbewiki.kenyon.edu/index.php?title=Host_Dependency_of_Mycobacterium_leprae&oldid=54675.
11. Fletcher H and McShane H (2006 May) Tuberculosis vaccines: current status and future prospects, Expert Opin Emerg Drugs, 11(2): 207-15. Review.
12. Gormus (Sep 2002) Anti-leprosy protective vaccination of rhesus monkeys with BCG or BCG plus heat-killed Mycobacterium leprae: lepromin skin test results, Leprosy Reviews, 73(3): 254-61.
13. Hansen's Disease Research, National Hansen's Disease Programs, Bureau of Primary Health Care, (April 2006). http://www.bphc.hrsa.gov/nhdp/.
14. Hansen's Disease (Leprosy), Division of Bacterial and Mycotic Diseases, Centers for Disease Control and Prevention, (12 Oct 2005). http://www.cdc.gov/ncidod/dbmd/diseaseinfo/hansens_t.htm.
15. Issar Smith **(**July 2003) "Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence" Clin. Microbiology, 16(3): 463-496, Rev.
16. Jason T. Evans (March 2010) Global Origin of Mycobacterium tuberculosis in the Midlands, UK, Emerging Infectious Diseases, 16(3).
17. Kartikeyan (1991), Anti-leprosy vaccines: current status and future prospects, Journal of Postgraduate Medicine, 37(4): 198-204.
18. L. Barker (2006) Mycobacterium leprae interactions with the host cell: recent advances (Review article), Indian J Med Res, 123: 748-59.
19. Lakshmi pillai and Usha chouhan (2014) Comparative analysis of Machine learning algorithms for Mycobacterium Tuberculosis protein sequences on the basis of physicochemical parameters, Journal of Medical Imaging and Health Informatics, 4: 1-8. http://dx.doi.org/10.1166/jmihi.2014.1241.
20. L.Steyn, Mycobacterial Diseases. University of Cape Town / Welcome, (1999, March 12) Retrieved May 17, (2010). http://web.uct.ac.za/depts/mmi/lsteyn/lecture.html.

21. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11(1). http://www.cs.waikato.ac.nz/~ml/weka/.
22. Matsudaira P, Lodish H, Baltimore D, Berk A, Zipursky SL , Darnell J (1995) Molecular Cell Biology 3rd edition. Scientific American Books New York; 739-777.
23. Nelda B. Holmgren. (1954 October) Studies on the metabolism of Mycobacterium Tuberculosis, J Bacteriol, 68(4): 405–410.
24. Paul R. Wheeler (2003) Leprosy – clues about the biochemistry of Mycobacterium leprae and its host-dependency from the genome, World Journal of Microbiology & Biotechnology, Addlestone, Surrey KT15 3NB, UK, 19: 1–16.
25. Ron caspi, Altman T, Travers et al. (2013) A systematic comparison of the MetaCyc and KEGG pathway databases, BMC Bioinformatics, 14: 112.
26. Sasaki S (2001) Mycobacterium leprae and leprosy: a compendium, Microbiol Immunol, 45(11): 729–36. PMID 11791665.
27. Satchidanandam vijaya and Amararama rao (May 1996) An approach for studying the mediators of pathogenesis in, Mycobacterium Tuberculosis, J. Biosci., 21(3): 413-421.
28. Strydjswka Barbara (22 May 2006), Clinician at the National Hansen's Disease Programs, Personal Interview.
29. V.M.Katoch, A.K.Datta et al. (1987) Studies on metabolism of mycobacteria with special reference to M.leprae. of Indo-U.K. Symposium on leprosy held at Agra, April 7-10 , A CJIL Publication, Coronation Press, 182-194 .
30. NCBI (2015), Database resources of the National Center for Biotechnology Information, Nucl. Acids Res., 44(suppl 1): D7-D19, Nov 28 Available: http://www.ncbi.nlm.nih.gov.
31. Wiker.H, Spierings.E., Kolkman.M, Ottenhoff.T, Harboe. M (1999) The mammalian cell entrypoeron 1 (mce1) of Mycobacterium Leprae and Mycobacterium tuberculosis, Microbial Pathogenesis, 27: 173-177.