Bioinformatics And Tuberculosis Databases- A Systematic Review

Arnika, Garima Verma[1], Jaspreet Singh[2], Anish Kumar[3]

School of Bioengineering and Biosciences, Lovely Professional University, Punjab, India
E-mail: [1]anish.20215@lpu.co.in

ABSTRACT

Tuberculosis (TB) is one of those diseases which shows greater impact on public health. It causes serious infections inside the human body, if it is present in the active state. Its treatment is available in the form of vaccines, drugs etc. It is caused by Mycobacterium tuberculosis. This bacterium has the tendency to become drug-resistant by its characteristic mechanism. For better treatment on this disease, we need to collect and analyse data properly during the drug designing. Currently, many databases like Mycobacterial Database, TB Database, TubercuList, TBrowse etc are available which helps the researcher in collecting the required data. These databases play a major role in a successful research. CADD (computer aided drug design) has been utilized in creating, modifying, analysing or optimizing a design. This software is used to raise the production of designer, improve quality design, enhance communications through documenting and for the crfeation of a database to manufacture.by using CADD researchers can know about the drug molecule which have potential with the help of databases.

KEYWORDS

Tuberculosis, bacterium, drug-resistant, database, CADD.

1. INTRODUCTION

Tuberculosis (TB) is one of the deadliest diseases present on the earth. It is caused by *Mycobacterium tuberculosis.* The primary target of this bacterium is lungs. It can also affect other organs like brain, liver, bones, intestines, lymph nodes and urinary or genital system [1]. This disease can be cured or treated.

 According to WHO's facts, worldwide approximately 10 million people got affected by TB in 2018 (5.7 were men, 3.2 were women and rest were children). Out of these 10 million people, it was estimated that 1.5 million died due to the illness. *Mycobacterium tuberculosis* can remain in latent phase for a longer period of time. The bacterium gets activated due to many factors like physiological changes, immune response, malnutrition, occurrence of other disease (HIV-AIDS) etc. In order to prevent this disease, vaccination is one of the most effective way. For example, BCG (Bacillus Calmette Guerin) vaccine is a live-attenuated vaccine which contains the weak form of the bacteria. It is usually given to infants to have better efficacy. According to various trials, it is observed that BCG vaccine is only 60-80% effective and the effect lasts for 10-15 years only as it does not stimulate the immune system (T-cells) effectively [2].

Earlier, TB vaccines can be divided into two major categories- therapeutic vaccine (used to treat TB) and preventive vaccine (used to prevent TB). But currently, there are four types of TB vaccines available which are developed by using advanced technology. Those are:

- Immunotherapeutic vaccine
- Immunopreventive vaccine
- Prime-boosting vaccine
- Priming vaccine

Many drugs are also commercially available for the treatment of tuberculosis. There are certain drugs which acts as antibacterial and some of them act as antibiotic. Rifampicin,3-formyl Rifamycin, isoniazid, clofazimine are some of the drugs for tuberculosis which shows antibacterial properties and route of delivery is oral. Amikacin and ofloxacin are some of the antibiotic drugs in which former is taken intravenously and the later one is taken orally. These are the drugs taken against tuberculosis and these all drugs are recommended by WHO (World health organization). The bacterium gets resistant to these all drugs which mentioned above and can hinder the treatment process. For survival *mycobacterium tuberculosis* requires various types of enzymes. There is also an urgent need to find the solution to cure drug-resistant TB. There are also certain enzymes required for the growth of mycobacterium tuberculosis. to study the tuberculosis disease in depth and to find the better solution of curing it we need to focus on both dry as well as wet lab[3]. There are various databases which can help the researcher in collecting and analysing the data systematically. The main aim of this study is to focus on the different type of databases which one can use during their research ton study about tuberculosis [4-5].

2. MATERIALS AND METHODS

 There are certain databases for tuberculosis which are extremely helpful in interpreting and presenting the data i.e. raw data into information. Primary function of databases is to manage the system by using software tools that help in organizing data in databases. As we are aware of the fact that 300000 compounds were tested against mycobacterium tuberculosis in lab alone, so several compounds were examined one by one. As several data occurred for different aspect of TB research. Following are the databases:

1. *BioHealthBase:*

It uses notes annotations subsystem technology for around 1,850 of the 2000 full and complete bacterial genomes as well as for mycobacterial tuberculosis and this system is rapid [6]. This is one of the social database and resources of bioinformatics to analyse specific biodefense and public health toxins like Flu virus, mycobacterium, microsporidia. It provides as a detailed fully integrated data storage facility which is taken from public databases and those data are extracted from various computational algorithm and that information taken from scientific literature are curated. The main aim of this database is to provide chemotherapeutics, diagnostics tools, and vaccinations via incorporating all the information on host-pathogen interactions, allowing researchers to understand the concept of toxicity, virulence factors, genomes and protein annotations are seen using either by formatted text or graphically using genome browser. 3D form visualization has the capability to check the proteins with essential structural and functional features (Fig.1). BioHelathBase has prepared biological pathways at the molecular level that represent host-pathogen interactions with the reactome database and influenza virus. [7]
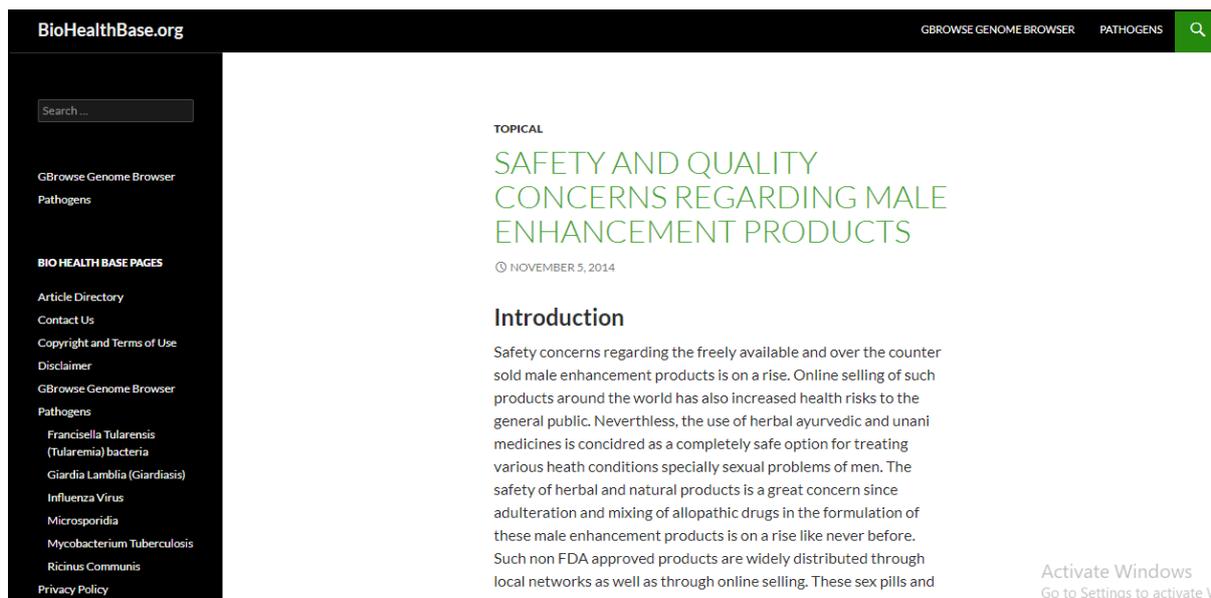
Fig1 - Represents the BioHealthBase.

2. *CDD TB:*

It is a software designed for small molecular compound libraries analysed towards mycobacterium tuberculosis (Fig 2). CDD have combined around 15 public databases on mycobacterium tuberculosis particular datasets that shows 300000 compounds which is originated from patent, literature, journals, etc [8]. it also has the facilities of storing and sharing data which are private. CDD database is helpful in finding similar compounds and molecular properties to known mycobacterium tuberculosis drugs, and to build unique computational machine learning and pharmacophore models for the identification of potential inhibitors [9-10].
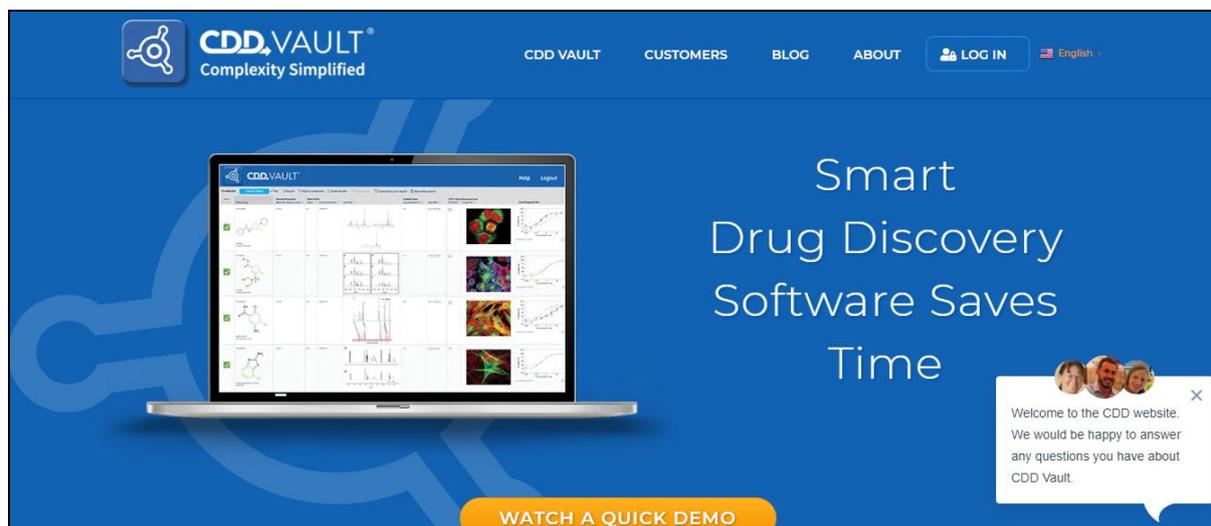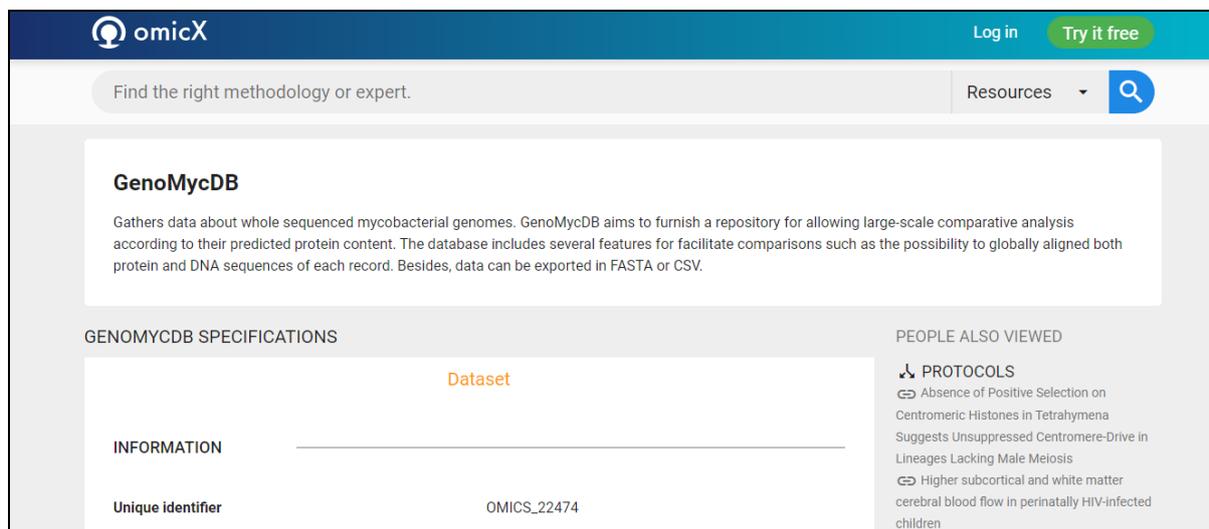

Fig2 - Represents CDD TB.

3.   *GenoMycDB:*

It is used for large scale comparative analysis for mycobacterial genomes which are completely sequenced. It helps in genome structure analysis and for classifying it. This database also prepared for the purpose of designing, incorporating and analysing

information's produced by wide-scale sequencing programmes of mycobacterial genomes present for tuberculosis bases on their protein content which may be predicted (Fig 3). This database inculcated the parameters which are similar with computed form of every aligned pair, which is permitted for every protein sequence the predicted subcellular localization, orthologous groups which are assigned with clusters, features of the particular gene and related to several databases which are useful. [11]

Fig3 - Represents GenoMycDB.

4.   TBrowse:

It is an integration type of analysis of TB genome, and is an open source for initiating drug discovery [12]. It also provides an Html browser for RNA-seq data. Tbrowse also known as module whose purpose is to facilitate the visualization of JSON files through a web browser, by using a web server. This application gives access to user for visualize sequence coverage, gene models, and splices as well as to count transcript abundance regions of interest. Tbrowse also provides a basic information about genome size and gene density etc. Detailed information on individual genes can be obtained by using TB database. Various targets can be downloaded which shows the data sequence, genes, and other genome data by TB database.

5.   Database for TDR targets:

The TDR targets database can be accessed with the help of the following URL link (http://tdrtargets.org). This database has brought together the genome sequencing and functional genomics project together. The main aspect of TDR target databases is to assess the target drug ability computationally and integrating large scale screening data in which raw data is manually curated and allow the assembling of candidate target to pursue [13].

6.   TubercuList:

It is one of the first databases for TB researchers (Fig 4). It gives DNAs complete dataset and sequence of proteins which is derived from MtbH37Rv, associated to annotation and function of assignments [14].
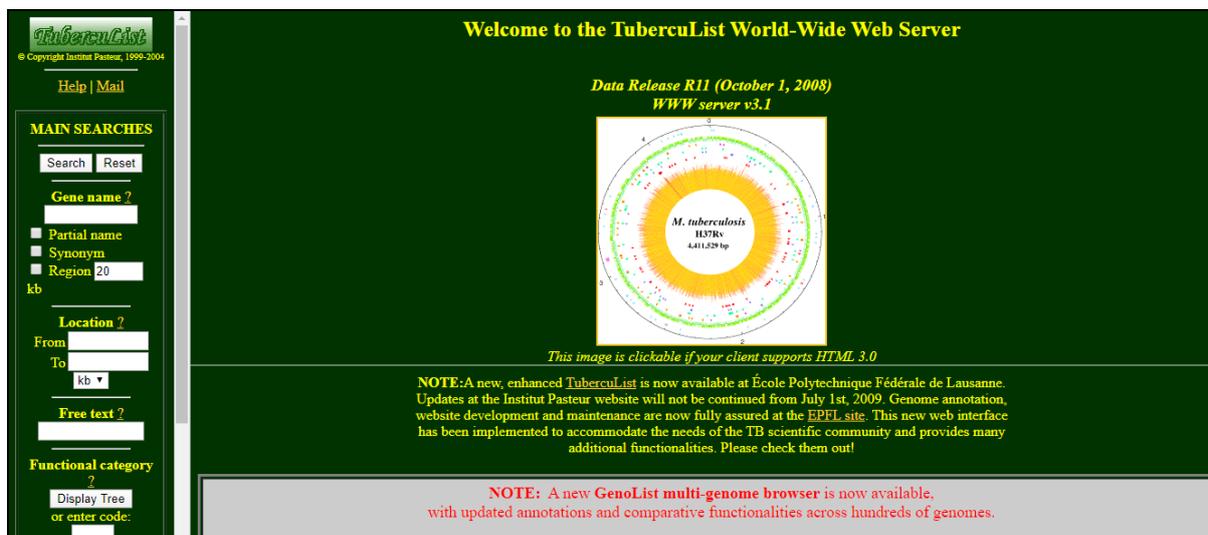
Fig4- Represents TubercuList.

7.   TBDB:

This offer genome and DNA data of an organism (for 28 genomes which are annotated) and tools that includes thousands of microarray datasets from Mtb- infected in vitro experiments and tissues. Here scholar can submit their data before publishing their paper, in which they browse for gene detail, genome visualization is performed and comparative analysis were done by using genome map methods, namely genomes and synteny map or operon map browser [15].

Fig- Represents TBDB.

8.   WebTB.org:
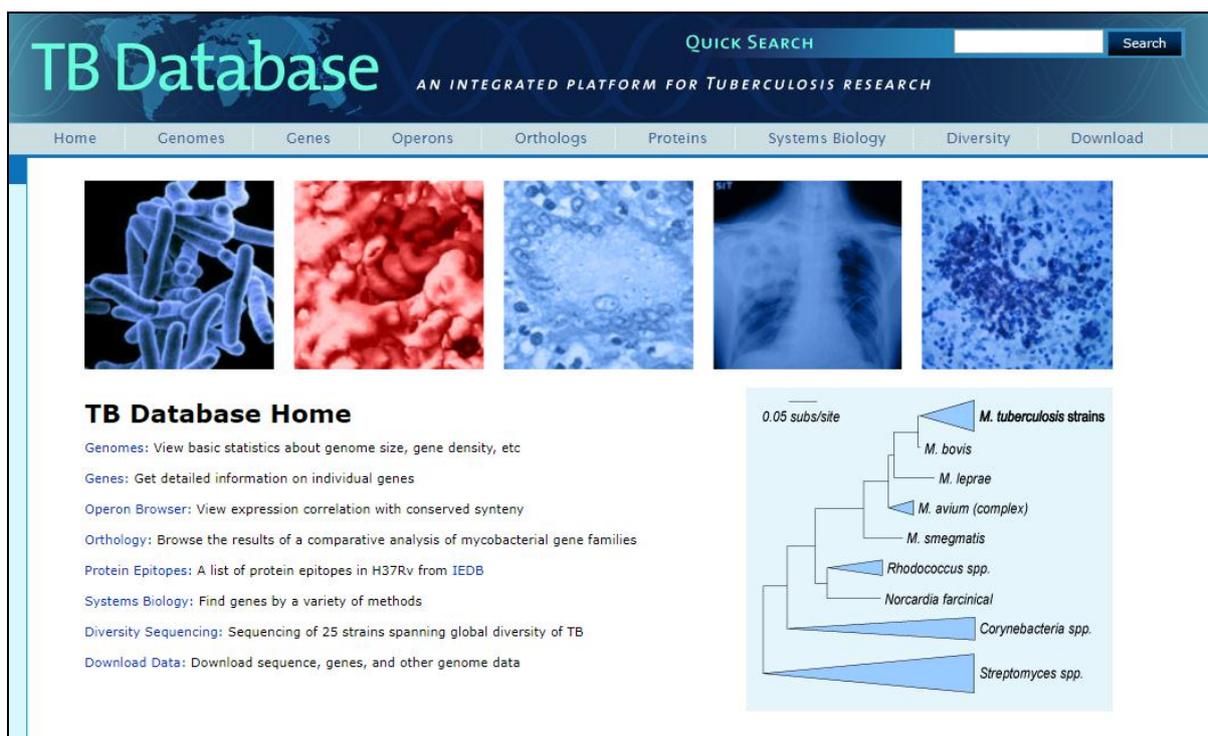
The TB structural genomics group offers this database. It mainly contains tools so that we can try to find out and go through the TB genomes in addition to summary pages for all recognized TB proteins, the MTBreg databases of proteins generally upregulated or downregulated in TB, around 100 targets which persists long in TB and various tools [16-18].

Various studies on tuberculosis have been performed and gathered many useful informations for successful preventive and therapeutic aspects of the disease [19-28].

S. No	Web Resource name	Description	Website/paper link
1.	Biohealthbase	Public bioinformatics database	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2238987/
2.	CDD TB	Web-Based database solution	https://www.collaborativedrug.com/cdd-receives-new-tb-funding/
3.	GenoMycDB	Comparative analysis of mycobacterial genes	https://www.ncbi.nlm.nih.gov/pubmed/16755503
4.	TBrowse	General purpose browsing mechanism for table-oriented data	https://www.itlnet.net/programming/program/Reference/c53g01c/ng15a315.html
5.	TDR	Rapid identification and prioritization of molecular *targets* for drug	https://tdrtargets.org/
6.	TBDB	comparative genomics and microarray analysis	https://www.hsls.pitt.edu/obrc/index.php?page=URL1234384552
7.	WebTB	Search and browse of TB genomes and structures	https://webtb.org.cutestat.com/

3. CONCLUSION:

In the present study, it mainly focuses on the different types of databases used in case of tuberculosis research work. As tuberculosis is one of the deadliest diseases on the earth which requires urgent action so that it can be cured properly. So, with the help of these databases we come to know about gene information, enzymes involved in tuberculosis metabolic activities. With the help of these databases the scientists can screen and study about drug molecules which are having potenticity to treat tuberculosis. In this work, we have noticed certain antibacterial drugs and antibiotic drugs which is useful in resistant against mycobacterial tuberculosis. Compounds present for tuberculosis are in huge number which cannot be studied manually so for that we need certain databases. Author sincerely hope that this work will be beneficial for those scholars who are working in antibacterial field and drug resistant tuberculosis.

4. REFERANCE

[1]   Reddy, T. B., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H., Koehrsen, M., Larson, L., Mao, M., Nitzberg, M., Sisk, P., Stolte, C., Weiner, B., White, J., Zachariah, Z. K., Sherlock, G. Schoolnik, G. K. (2009). TB database: an integrated platform for tuberculosis research. *Nucleic acids research*, *37*(Database issue), D499–D508. https://doi.org/10.1093/nar/gkn652

[2]   Gong, W., Liang, Y., & Wu, X. (2018). The current status, challenges, and future developments of new tuberculosis vaccines. *Human vaccines & immunotherapeutics*, *14*(7), 1697–1716. https://doi.org/10.1080/21645515.2018.1458806

[3]   Tiberi, S., Scardigli, A., Centis, R., D'Ambrosio, L., Muñoz-Torrico, M., Salazar-Lezama, M. Á. Caminero Luna, J. A. (2017). Classifying new anti-tuberculosis drugs: rationale and future perspectives. International Journal of Infectious Diseases, 56, 181–184. doi:10.1016/j.ijid.2016.10.026

[4]   Ehrt, S., Schnappinger, D., & Rhee, K. Y. (2018). Metabolic principles of persistence and pathogenicity in Mycobacterium tuberculosis. *Nature reviews. Microbiology*, *16*(8), 496–507. https://doi.org/10.1038/s41579-018-0013-4

[5]   Lougheed, K. E., Taylor, D. L., Osborne, S. A., Bryans, J. S., & Buxton, R. S. (2009). New anti-tuberculosis agents amongst known drugs. *Tuberculosis (Edinburgh, Scotland)*, *89*(5), 364–370. https://doi.org/10.1016/j.tube.2009.07.002

[6]   Ekins, S., Freundlich, J. S., Choi, I., Sarker, M., & Talcott, C. (2011). Computational databases, pathway and cheminformatics tools for tuberculosis drug discovery. *Trends in microbiology*, *19*(2), 65–74. https://doi.org/10.1016/j.tim.2010.10.005

[7]   Squires B, et al. BioHealthBase: informatics support in the elucidation of influenza virus host pathogen interactions and virulence. Nucleic Acids Res. 2008;36:D497–D503.

[8]   Ekins S, et al. A Collaborative Database And Computational Models For Tuberculosis Drug Discovery. Mol BioSystems. 2010;6:840–851.

[9]   Hohman M, et al. Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. Drug Disc Today. 2009;14:261–270.

[10]   Makarov V, et al. Benzothiazinones kill *Mycobacterium tuberculosis* by blocking arabinan synthesis. Science. 2009;324:801–804.

[11]   Catanho M, et al. GenoMycDB: a database for comparative analysis of mycobacterial genes and genomes. Genet Mol Res. 2006;5:115–126.

[12]   Bhardwaj A, et al. TBrowse: an integrative genomics map of *Mycobacterium tuberculosis*. Tuberculosis (Edinburgh, Scotland) 2009;89:386–387.

[13]   Aguero F, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. Nat Rev Drug Discov. 2008;7:900–907.

[14]   Cole ST. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. FEBS Lett. 1999;452:7–10.

[15]   Reddy TB, et al. TB database: an integrated platform for tuberculosis research. Nucleic Acids Res. 2009;37:D499–D508.

[16]   Terwilliger TC, et al. The TB structural genomics consortium: a resource for *Mycobacterium tuberculosis* biology. Tuberculosis (Edinburgh, Scotland) 2003;83:223–249.

[17]   Goulding CW, et al. The TB structural genomics consortium: providing a structural foundation for drug discovery. Curr Drug Targets Infect Disord. 2002;2:121–141.

[18]   Rupp B, et al. The TB structural genomics consortium crystallization facility: towards automation from protein to electron density. Acta crystallographica. 2002;58:1514–1518.

[19]   Sabareesh V., Singh G. Mass spectrometry based lipid(ome) analyzer and molecular platform: A new software to interpret and analyze electrospray and/or matrix-assisted laser desorption/ionization mass spectrometric data of lipids: A case study from Mycobacterium tuberculosis. Journal of Mass Spectrometry, 48(4), 2013.

[20]   Dua K., Rapalli V.K., Shukla S.D., Singhvi G., Shastri M.D., Chellappan D.K., Satija S., Mehta M., Gulati M., Pinto T.D.J.A., Gupta G., Hansbro P.M. Multi-drug resistant Mycobacterium tuberculosis & oxidative stress complexity: Emerging need for novel drug delivery approaches. Biomedicine and Pharmacotherapy, 107, 2018.

[21]   Vise E., Mawlong M., Garg A., Sen A., Shakuntala I., Das S. Recovery of Mycobacterium lentiflavum from bronchial lavage during follow-up of an extrapulmonary tuberculosis patient. International Journal of Mycobacteriology, 6(3), 2017.

[22]   Pandey R., Tamrakar D., Jaiswal S., Sharma A., Koju S., Duwal S.R., Sharma I., Jayaswal R.P., Pankaj P.P. Serum adenosine deaminase: A novel biomarker tool for the diagnosis of tuberculosis. Biosciences Biotechnology Research Asia, 13(1), 2016.

[23]   Garg A., Verma R. Cloning and molecular characterization of fibronectin attachment protein gene (ModD) of Mycobacterium avium subspecies paratuberculosis. Indian Journal of Animal Sciences, 87(1), 2017.

[24]   Singh H., Verma S., Marwah G.K. The new approach for medical enhancement in texture classification and feature extraction of lung MRI images by using Gabor filter with wavelet transform. Indian Journal of Science and Technology, 8(35), 2015.

[25]   Chaudhary A., Singh S.S. Lung cancer detection on CT images by using image processing. Proceedings: Turing 100 - International Conference on Computing Sciences, ICCS 2012.

[26]   Kaur P., Mishra V., Shunmugaperumal T., Goyal A.K., Ghosh G., Rath G. Inhalable spray dried lipidnanoparticles for the co-delivery of paclitaxel and doxorubicin in lung cancer. Journal of Drug Delivery Science and Technology, 56, 2020.

[27]   Shaheamlung G., Kaur H., Singla J. A Comprehensive Review of Medical Expert Systems for Diagnosis of Chronic Liver Diseases. Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019.

[28]   Sharma P., Mehta M., Dhanjal D.S., Kaur S., Gupta G., Singh H., Thangavelu L., Rajeshkumar S., Tambuwala M., Bakshi H.A., Chellappan D.K., Dua K., Satija S. Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. Chemico-Biological Interactions, 309, 2019.