# Virulence factors of Mycobacterium tuberculosis and its analysis

**Submitted By.**

SIDDHARTH K SHAILENDRA 16BCB0129

Under the guidance of.

**Rishin Haldar**

PROFESSOR,

SCOPE

**Computer Science and Engineering Spec In Bioinformatics**

**School of Computer Science and Engineering**



1. **Problem Statement**

Mycobacterium tuberculosis (M. tb) is a species of [pathogenic](https://en.wikipedia.org/wiki/Pathogen) [bacteria](https://en.wikipedia.org/wiki/Bacteria) in the family [Mycobacteriaceae](https://en.wikipedia.org/wiki/Mycobacteriaceae" \o "Mycobacteriaceae) and the causative agent of [tuberculosis](https://en.wikipedia.org/wiki/Tuberculosis). According to WHO In 2018, an estimated 10 million people fell ill with tuberculosis(TB) worldwide. 5.7 million men, 3.2 million women and 1.1 million children. There were cases in all countries and age groups. Humans are the only known reservoirs of M. tuberculosis. A misconception is that M. tuberculosis can be spread by shaking hands, making contact with toilet seats, sharing food or drink, sharing toothbrushes, or kissing. It can only be spread through [air droplets](https://en.wikipedia.org/wiki/Air_droplets) originating from a person who has the disease either coughing, sneezing, speaking, or singing. When in the lungs, M. tuberculosis is [phagocytosed](https://en.wikipedia.org/wiki/Phagocytosis" \o "Phagocytosis) by alveolar [macrophages](https://en.wikipedia.org/wiki/Macrophages), but they are unable to kill and digest the bacterium. For a pathogen to setup an infection inside a host it requires several properties. These properties are called determinants of “Virulence” or the ability to cause disease. The factors which causes the infection to spread are as follows. 1) Transmissibilty 2) Adhision 3) Propagation through interaction with host receptors 4) Spreading factors 5) Evading host defense mechanism. Though TB is curable but there are genes are causing drug resistance in multiple drugs used to treat tuberculosis. The disease is only minimally preventable; the vaccine that is used in areas of high endemicity is about 20% effective

1. **Objective**

* To provides in-depth coverage major virulence factors of the best-characterized bacterial pathogens, with the structure features, functions and mechanisms used by these pathogens to allow them to conquer new niches and to circumvent host defence mechanisms, and cause diseases.
* To find out the genes that are resistant to particular drugs used for treatment of TB.
* An attempt to find out specific regions in the genes across all the genomes of MTb which are more prone to change, find some pattern of that change and hopefully predict the next mutation in a new genome.

1. **Literature survey**

* **vf analyzer for mycobacterium tuberculosis**

The virulence factor database (VFDB) is an integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogens. Instead of using simple BLAST searches, VFanalyzer first constructs orthologous groups within the query genome and pre-analyzed reference genomes from VFDB to avoid potential false positives due to paralogs. Then, it conducts iterative and exhaustive sequence similarity searches among the hierarchical pre-build datasets of VFDB to accurately identify potential untypical/strain-specific VFs. Finally, via a context-based data refinement process for VFs encoded by gene clusters, VFanalyzer can achieve relatively high specificity and sensitivity The conventional method for the identification of VFs in bacterial genomes usually relies on sequence similarity searches (such as BLAST) among datasets of known VFs.

* **Tuberculosis Drug Resistance Mutation Database**

Centralized databases, curated from the literature or high-throughput experiments. A new database devoted to drug resistance mutations in TB, called the TB Drug Resistance Mutation Database (TBDReaMDB). By providing a comprehensive, single resource of drug resistance mutations in TB, to accelerate and encourage new discoveries that will have applications ranging from diagnostics to drug discovery. To envision that this database will expand as additional mutations are identified in the coming years and will serve as a platform for diverse analyses.

* **PHIDIAS**

PHIDIAS is a web-based database and analysis system that aims to manually curate, computationally analyze, and address different scientific issues in the areas of pathogen-host interactions (PHI, or called host-pathogen interactions or HPI). PHIDIAS has emphasized the study of those pathogens that cause various infectious diseases in humans and animals.

One central theme in PHI research is the identification and understanding of virulence factors from various pathogens. We have generated a manually curated database, which is called [Victors](http://www.phidias.us/victors/index.php), to store and analyze various virulence factors from different pathogens that infect humans and animals. The Victors data has now been utilized by different groups and projects, including the [PATRIC Specialty Genes](http://enews.patricbrc.org/faqs/genome-feature-data-and-tools/specialty-genes-faqs/) resource.

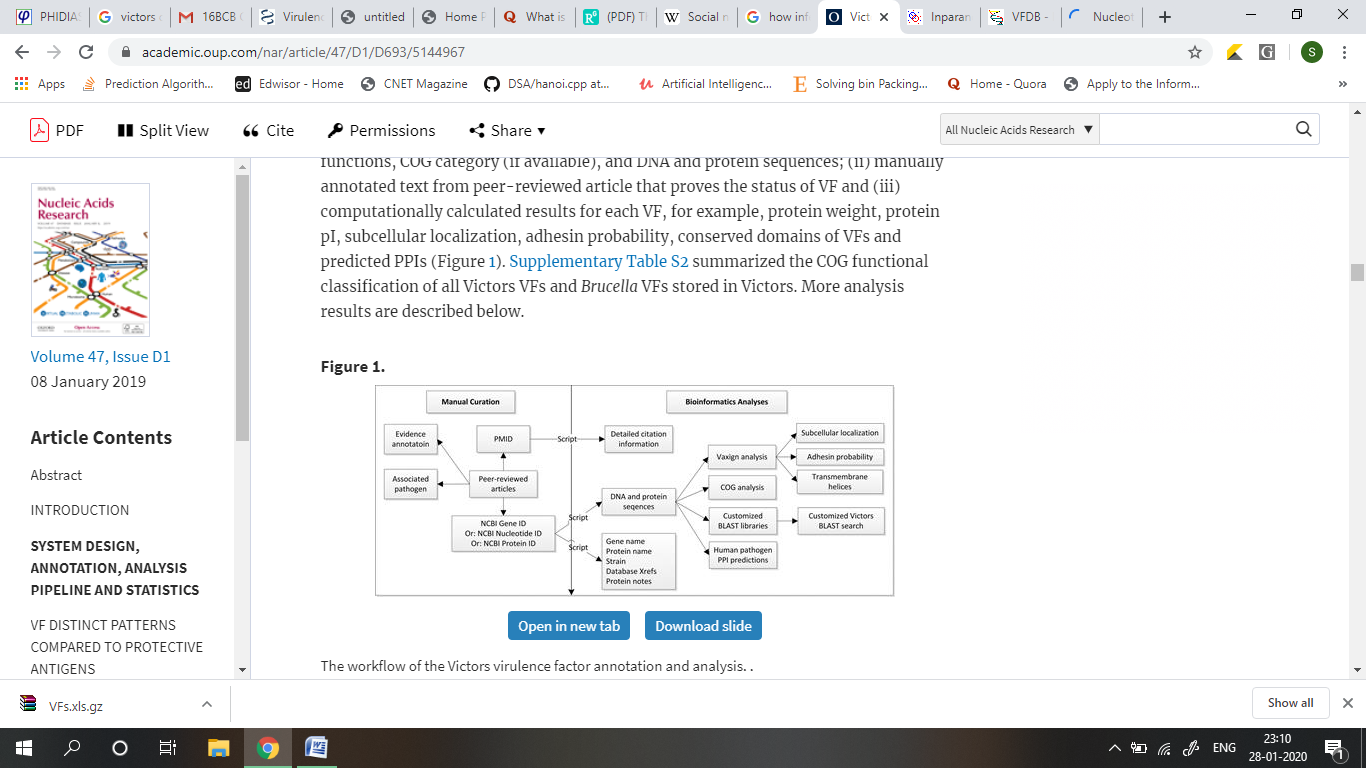
1. **Requirements**

Software- Python, BLAST, Uniprot, Victors,Swissprot.

Hardware- Quad core Intel Core i5, 4 GB of RAM, Premium graphics cards for GPU’s

1. **Methodology and Future works**

* For each specific VF, the Victors database contains the following information: (i) general information on each VF gene symbol, protein name, general gene/protein functions, COG category (if available), and DNA and protein sequences; (ii) manually annotated text from peer-reviewed article that proves the status of VF and (iii) computationally calculated results for each VF, for example, protein weight, protein pI, subcellular localization, adhesin probability, conserved domains of VFs and predicted PPIs. Customized BLAST was used for BLAST sequence similarity search. For the prediction of host–pathogen protein–protein interactions (PPIs).

****

* The genomes of M. tuberculosis strains were downloaded from NCBI. The annotation refinement was carried out through coding sequence (CDS) mapping by BLAST. Genome Topology Networks will be made in order to analyze clusters of orthologous group in which can be used as reference for orthology.
* Using information networks to create a genome network and applying algorithms like HITS and other algorithms to make out how MTB spreads from one genome to another and also to know about the genes resistant to particular drugs.

1. **References**

<http://www.mgc.ac.cn/VFs/main.htm>

<http://www.phidias.us/index.php>

<https://tbdreamdb.ki.se/Info/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3544749/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342819/pdf/12864_2015_Article_1259.pdf>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5712783/pdf/97320630013380.pdf>