

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

## 4,800

Open access books available

## 122,000

International authors and editors

## 135M

Downloads

Our authors are among the

## 154

Countries delivered to

## TOP 1%

most cited scientists

## 12.2%

Contributors from top 500 universities

**WEB OF SCIENCE™**Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us? Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)

---

# Web Resources on Tuberculosis: Information, Research, and Data Analysis

---

Edson Machado, Camillo Cerdeira,  
Antonio Basílio de Miranda and Marcos Catanho

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.73549>

---

## Abstract

Since the first edition of this book in 2013, many new tools and databases have become publicly available, as well as several have been discontinued. Here, we present an updated version of web resources on tuberculosis, providing more detailed information on some key concepts. However, the purpose of this chapter is by no means to offer an exhaustive list of all the resources available on the Internet about TB, the topic of this book. This would be a massive and perhaps futile work since the evolution of the Internet occurs at a very fast pace. Rather, this chapter concentrates on a selection of the most important, relevant and stable websites with relevance to several aspects of TB, such as research, treatment, main institutions, funding, and specialized platforms. We think this should complement all the other information already presented in this book, offering the reader a more integrated view of the disease, as well as access to new platforms and systems specialized in the analysis of data generated by a series of new technologies such as DNA sequencing.

**Keywords:** tuberculosis, omics, bioinformatics

---

## 1. Introduction

The Internet has quickly become the universal exponent of the digital world as we know it. Today, billions of people around the world use the Internet as their primary source of information and benefit from electronic mail systems, information distribution, file sharing, multimedia streaming services, and online social networking, to cite just a few examples. With the popularization of smartphones, access to the Internet has become even more frequent, and many people remain connected all the time. This medium is an unprecedented form of

---

communication in the history of humanity, capable of bringing information from virtually anywhere almost instantaneously.

Scientists have been pioneers in benefiting from global interaction and globalized information. Nowadays, we deal with algorithms, data mining, terabytes and petabytes of data, teraFLOPS and petaFLOPS to measure computer performance, distributed computing, cloud computing, and terms and aspects of what we call big data, referring to data sets with huge sizes, exceeding the capability of commonly used software tools to manage and process these data within a feasible time frame. This kind of data stems from a variety of new (as well as old) high-throughput technologies, employed by researchers in different fields, such as astronomy and (recently) biology, as well as from information technology companies, such as Amazon, Facebook, and Google, among many other examples. For instance, the Large Hadron Collider, a particle accelerator at CERN, Switzerland, generates data on the order of terabytes per second.

However, the global connection is not restricted to the virtual world: with the ease of intercontinental travels, infectious agents can traverse the whole world in one single day. Tuberculosis (TB) is a disease considered to be at high risk for people who maintain prolonged and repeated contact with the host. Therefore, the risk of infection when traveling is reduced, but some conditions might present as a risk factor. Immunocompromised travelers (whether by the disease itself or by medication), smokers, children under 5 years old, and healthcare workers are particularly vulnerable to TB. Before traveling, the individual should consult a specialist to assess the risks of contracting TB in the area of stay, checking the incidence of TB and, especially, of resistant strains. Prophylactic measures should be taken when necessary, and medical follow-up should be provided when returning from the trip. These procedures are critical to containing the global circulation of TB strains [1].

TB is a pandemic disease, with an estimated one-third of the world's population contaminated by the bacillus *Mycobacterium tuberculosis* (Mtb). Although treatable, patients without proper follow-up usually abandon the therapy as soon as they feel better. This and the indiscriminate use of antibiotics are causing the emergence of new drug-resistant strains, with patterns of epidemic or outbreak resembling the dissemination of a new piece of information throughout the Internet.

On the other hand, there has also been a revolution in other areas. High-performance technologies, such as genomics, transcriptomics, proteomics, and metabolomics, for instance, offer a new and more integrated view of the genetics and metabolism of targeted organisms. Currently, thousands of mycobacterial genomes are already sequenced or are in progress. Hence, for example, comparing genomes of virulent and non-virulent strains of Mtb, scientists can pinpoint genes and/or polymorphisms possibly involved in pathogenesis; similarly, analyzing transcriptome data, researchers might gain an idea of the effects caused by a given drug in the bacillus' metabolism.

Since the Internet content is continually updated and changed, providing an exhaustive list of all the resources available online concerning TB would be a cumbersome and perhaps futile work. Hence, in this chapter, we present a selection of the most relevant and stable websites

exploring several aspects of TB, such as treatment, research, and funding, or providing access to analytical tools, databases, or specialized platforms for mycobacterial research. We hope this chapter can offer the reader an overview of online, publicly available, computational resources that could help us to fight TB.

## 2. Bioinformatics and omics

### 2.1. Bioinformatics

Bioinformatics is an interdisciplinary field that involves distinct areas, such as biology, computer science, mathematics, and statistics, comprising an extensive list of activities, such as research, development, or application of computational techniques or tools to acquire, store, organize, analyze, visualize, and integrate biological, medical, behavioral, or health data and information. Bioinformatics addresses scientific questions imposed by biological phenomena applying several analytical approaches such as image processing, computational simulations, network analysis, and data mining, among others, to perform comparative genomics studies, gene expression analysis, structural protein analysis, phylogenetics, and metabolic networks, just to cite a few examples [2].

However, bioinformatics preceded the term, as its origins can be traced back to the early 1960s when computers became essential tools in the field of molecular biology, as well as in all other research areas [3]. At that time, Margaret Oakley Dayhoff, a biochemist with a PhD degree in quantum chemistry, pioneered the development of computer methods for comparing protein sequences and for inferring evolutionary histories based on protein sequence alignments. Among many other significant scientific contributions made during her career, she cataloged all known protein sequences and made them available to the scientific community, publishing in 1965 the *Atlas of Protein Sequence and Structure*, containing sequence information on 65 proteins, considered the first molecular biology database [4].

Indeed, toward the end of the 1960s, several algorithms and computer programs were already available for analyzing structure, function, and evolution of nucleotide and protein sequences, as well as rudimentary protein databases [3, 5]. In the following decades, new computational methods and approaches were introduced such as algorithms for sequence alignments, public domain databases, efficient data search and retrieval systems, sophisticated methods for protein structure prediction, tools to automatically annotate genes and genomes, and systems for functional genome analysis, among others [6].

### 2.2. Omics

The cost of DNA and RNA sequencing has been gradually decreasing over the years, and the improvement achieved with new sequencing technologies allows sequencing the whole genome or transcriptome of an organism in few hours, creating new avenues for biological research. However, genome sequencing involves the generation of a massive amount of data,

in the order of hundreds or even thousands of gigabytes. In the same way, but with higher cost and some technological limitations, mass spectrometry and magnetic resonance imaging are also capable of identifying thousands of proteins and metabolites of an organism, respectively. In this context, new “omics” (an informal term that collectively refers to research fields of biology ending in -omics, such as genomics, proteomics, or metabolomics) approaches have arisen, involving not only the production of sequence data but also the analysis, treatment, and interpretation of biological data on a large scale [7]. The following sections briefly present some of the main “omics” approaches, exemplifying their application in TB research.

### 2.2.1. Genomics

Genomics is a field of biology that seeks to understand the structure, functions, and evolution of genes and genomes analyzing large-scale genomic data obtained with high-throughput sequencing technologies. For instance, whole genome sequencing (WGS) and comparative analyses of genomic segments (chromosomes or syntenic regions) or the entire genome content (protein-coding genes, RNA-coding genes, pseudogenes, non-translated regions, etc.) allow the investigation of complex biological phenomena, giving insights into genome evolution and adaptation over time or even helping to improve clinical diagnoses. In the last two decades, WGS has become a routine among researchers, and currently there are thousands of completely sequenced genomes available in public databases. In fact, the availability and accessibility of genomic sequences are increasing rapidly, with the results generated by teams of researchers in collaborative networks or even in independent laboratories, supporting many essential analyses in the field, such as comparison of biological sequences and searching for similarities between them, enabling the inference of functional and evolutionary relationships between genes, gene families, and genomes [8, 9].

For instance, comparative genomics, together with the combination of archeological data and DNA sequencing, has already established a plausible evolutionary scenario for the origin of the principal etiological agent of TB, the pathogen *Mtb*. Although it is a millennial disease, drugs for the treatment of TB have appeared less than 100 years ago. Since then, multidrug-resistant (MDR) strains have emerged, as well as extensively drug-resistant (XDR) strains, and more recently, several countries have reported cases of total drug-resistant (TDR) strains, frustrating the efforts to fight TB. A typical feature of bacterial agents that develop drug resistance is the occurrence of lateral gene transfer (LTG), but comparative genomic studies did not find evidence of LTG in *Mycobacterium tuberculosis* complex (MTBC) organisms. Thus, there is a pathogen adapted for living in human host cells, able to remain for a long time as a latent disease, surviving in the host and adapting to yield a persistent infection, often immune to treatments. The primary challenge is then to develop an effective vaccine against most *Mtb* strains, which would be possible by targeting conserved elements [10].

### 2.2.2. Transcriptomics

Transcriptomics is the analysis of gene expression through the sequencing of RNA in large scale (RNA-Seq). Transcriptome, the whole set of RNA transcripts of a given organism, organ, tissue, or cell lineage, contains different types of RNAs. Thus, transcriptomics provides



essential information of the biological sample under analysis, allowing both quantitative and qualitative approaches to gene expression, providing a profile of all coding and noncoding transcripts, in specific conditions. RNA-Seq technologies have been improved gradually, including novel techniques such as single-cell RNA-Seq, in which individual cells of interest obtained from culture, tissue, or dissociated cell suspensions are isolated, converting RNA into cDNA and sequencing of cDNA libraries. However, transcriptomics still faces too many challenges. RNA-Seq produces very short reads and presents a high error rate, yielding a tremendous amount of data from massively parallel sequencing, requiring significant computational resources, as well as specific algorithms and software, to analyze it [11–13].

Mtb has an extensively resistant cell wall, can adopt an opportunistic switching over to latency, and has many strategies that fool the host's immune system, compromising the effectiveness of therapeutic approaches available. The analysis of Mtb transcriptome signatures during infection, for instance, provided by the genome-wide expression profile, showed the expression of numerous genes used to evade the host immune responses, suitable to the intracellular lifestyle, and to respond to various antibiotic drugs [14].

### 2.2.3. Proteomics

Proteomics establishes a global analysis of a cell's proteins. Gathering information about the proteome and comparing it with genome and transcriptome data is the way to understand the functioning of the cell. The physicochemical properties of highly diversified amino acids, protein modifications, and degradations and the interconnectivity of proteins in complexes are examples of difficulties encountered in collecting data compared to the other "omics" sciences. Nowadays, mass spectrometry is known as the state-of-the-art proteomics and has been improved with the development of instruments, sample preparation, and new analytical software. Mass spectrometry is capable of characterizing almost a complete proteome, revealing the profiling of the expressed proteins at the cellular and subcellular levels, providing knowledge of the functional status of a cell in response to environmental stimuli [15, 16].

Among all sequenced Mtb strains so far, with different genotypes and phenotypes, only a few have a complete identification of protein-coding regions, lacking the knowledge of protein functions that play roles in the physiology of mycobacteria. This information would be essential to allow, for example, understanding the causes that lead to mechanisms of mycobacterial pathogenicity and drug resistance. Even then, proteomics has already enabled us to know some aspects of virulence and its mechanisms of action in Mtb strains. The next step for proteomics contribution in mycobacteria research, the analysis of the immune response of the host, is one of the ways to establish new treatment programs, especially in the current scenario of drug resistance [17].

### 2.2.4. Metabolomics

Metabolomics encompasses the study of metabolism within a living cell, that is, it deals with the identification and analysis of biochemical reactions products and their processes within the cell. The metabolome is the set of these products (metabolites), and the metabolism analysis can reveal the cell's organic response. Metabolic analysis is a valuable resource for identification of

specific metabolite biomarkers, which would help, for example, evaluating response to drugs or stress agents. Magnetic resonance spectrometry (NMR) and mass spectrometry (MS) are the appropriate technologies for accurate measurements of metabolites, providing a good phenotype representation of any cell [18, 19].

Being complementary to other “omics,” the metabolomics fills some gaps to a better understanding of diseases caused by mycobacteria. It is known that mycobacterial exosomes can be used as biomarkers, since they come from infected cells, containing mycobacterial proteins, lipoarabinomannan, and metabolites. They have been used for TB diagnosis and in the research of new vaccines. It is important to emphasize that the signatures of TB biomarkers must be validated in geographical and ethnical context, given the worldwide and diversified nature of the strains, besides concomitant infections such as malaria and HIV, as they affect metabolic biosignatures [19, 20].

### 2.3. One example of integration of omics

An effective vaccine for all strains of TB and the development of therapeutic approaches appropriate to the variations and stages of the disease are the main goals of current research in TB. The achievement of the genomic expression catalog of a global collection of BCG vaccine strains, comparing genomes and transcriptomes of 14 of the most widely used BCG strains, is one example of interaction between “omics” sciences in TB research, contributing to display evidences for highly diverged metabolic and cell-wall adaptations. Moreover, quantitative proteomics has identified the major differences in protein expression, when changes observed in the proteome confirmed the changes observed in the transcriptome, showing how the adaptation to the environment causes phenotypic differences between BCGs [21].

## 3. Tuberculosis facts, information, and treatment research

This section presents websites covering diverse information about TB, including history, pathogenesis, transmission, epidemiology, diagnosis, treatment, and infection control, listed in alphabetical order (**Table 1**). They also provide information about courses, tokens, and links

Agency	URL
Americas American Lung Association (ALA) Lung Disease Programs	<a href="http://www.lung.org/">http://www.lung.org/</a>
American Public Health Association	<a href="https://www.apha.org/">https://www.apha.org/</a>
Bill & Melinda Gates Foundation	<a href="https://www.gatesfoundation.org/">https://www.gatesfoundation.org/</a>
Centers for Disease Control and Prevention, Division of Tuberculosis Elimination (CDC-DTBE)	<a href="https://www.cdc.gov/tb/">https://www.cdc.gov/tb/</a>
Food and Drug Administration (FDA)	<a href="https://www.fda.gov/default.htm">https://www.fda.gov/default.htm</a>
Global Tuberculosis Institute	<a href="http://globaltb.njms.rutgers.edu/">http://globaltb.njms.rutgers.edu/</a>
Institute for Tuberculosis Research	<a href="http://www.tuberculosisdrugresearch.org/">http://www.tuberculosisdrugresearch.org/</a>

	Agency	URL
	National Institute of Allergy and Infectious Diseases (NIAID)	<a href="https://www.niaid.nih.gov/">https://www.niaid.nih.gov/</a>
	National Library of Medicine, PubMed	<a href="https://www.ncbi.nlm.nih.gov/pubmed/">https://www.ncbi.nlm.nih.gov/pubmed/</a>
	Pan American Health Organization (PAHO)	<a href="http://www.paho.org/">http://www.paho.org/</a>
Africa	AllAfrica.com: TB News from Africa	<a href="http://allafrica.com/tuberculosis/">http://allafrica.com/tuberculosis/</a>
	Desmond Tutu TB Centre	<a href="http://www0.sun.ac.za/dttc/">http://www0.sun.ac.za/dttc/</a>
	South African Tuberculosis Vaccine Initiative	<a href="http://www.satvi.uct.ac.za/">http://www.satvi.uct.ac.za/</a>
	Tb Alliance	<a href="https://www.tballiance.org/">https://www.tballiance.org/</a>
Asia and Oceania	JATA, Research Institute of Tuberculosis	<a href="http://www.jata.or.jp/eindex/home.html">http://www.jata.or.jp/eindex/home.html</a>
	National Institute for Research in Tuberculosis	<a href="http://www.trc-chennai.org/">http://www.trc-chennai.org/</a>
	Pakistan Anti-TB Association	<a href="http://www.patba.org/">http://www.patba.org/</a>
	TBC India	<a href="http://www.tbcindia.nic.in/">http://www.tbcindia.nic.in/</a>
Europe	European Tuberculosis Surveillance Network	<a href="https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/european-tuberculosis/">https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/european-tuberculosis/</a>
	International Union Against Tuberculosis and Lung Disease (UNION)	<a href="https://www.theunion.org/">https://www.theunion.org/</a>
	Max Planck Institute for Infection Biology	<a href="http://www.mpiib-berlin.mpg.de/">http://www.mpiib-berlin.mpg.de/</a>
	National Health Service in England	<a href="http://www.nhs.uk/conditions/Tuberculosis/Pages/Introduction.aspx">http://www.nhs.uk/conditions/Tuberculosis/Pages/Introduction.aspx</a>
	Pasteur Institute	<a href="https://www.pasteur.fr/">https://www.pasteur.fr/</a>
	StopTB Partnership	<a href="http://www.stoptb.org/">http://www.stoptb.org/</a>
	TBnet	<a href="http://www.tb-net.org/">http://www.tb-net.org/</a>
Global	World Health Organization (WHO)	<a href="http://who.int/tb/en/">http://who.int/tb/en/</a>

**Table 1.** Websites covering TB facts, information, and treatment research.

to other websites and general guidelines. We do not prioritize any of them since every effort made to combat this disease has become important because of its comprehensiveness.

#### 4. Tuberculosis databases and computational tools

Bioinformatics had its origins in the 1960s when computers became essential tools in research. Since then, numerous computational resources have been created, providing the scientific community different analytical tools to interpret a range of biological data. Collectively, these online resources are publicly available and are dedicated to acquire, store, organize, analyze, visualize, and/or integrate the ever-increasing amount of biological data originated from scientific experiments, scientific literature, high-throughput technologies, and computational analyses.



In this section, we provide a selection of online publicly available resources (entirely or partially) dedicated to mycobacteria causing tuberculosis, categorized according to its purpose and functionality (**Table 2**). Each category is quickly reviewed, presenting the reference to the original paper describing each computational tool, as well as its electronic address.

Category	Resource	URL	Ref.
Generic and multifunctional resources	Resources for functional and evolutionary genomic study of the genus <i>Mycobacterium</i> , comprising extensive literature review and data annotation on mycobacterial genome polymorphism, virulence factors, <i>in silico</i> generated and manually reviewed information on the complete genome sequence of these organisms, and essential genes	MyBASE	<a href="http://mybase.psych.ac.cn/">http://mybase.psych.ac.cn/</a> [22]
		The MycoBrowser portal	<a href="http://mycobrowser.epfl.ch/">http://mycobrowser.epfl.ch/</a> [23]
Genomic mapping and functional annotation	Systems supporting functional annotation, including protein analysis, subcellular localization prediction, and mycobacterial membrane protein identification and characterization	MycoMemSVM	<a href="http://lin.uestc.edu.cn/server/MycoMemSVM">http://lin.uestc.edu.cn/server/MycoMemSVM</a> [24]
		MycoSub	<a href="http://lin.uestc.edu.cn/server/MycoSub/">http://lin.uestc.edu.cn/server/MycoSub/</a> [25]
		Genolist (TubercuList, BoviList, BCGList)	<a href="http://genolist.pasteur.fr/">http://genolist.pasteur.fr/</a> [26, 27]
Comparative genomics	Collection of databases dedicated to mycobacterial comparative genomics, providing precomputed data of comparative genome analyses among selected mycobacterial genera, as well as inferred orthologous groups, functional annotations, and protein features	GenoMycDB	<a href="http://genomycdb.fiocruz.br/">http://genomycdb.fiocruz.br/</a> [28]
		MycoDB	<a href="http://xbase.warwick.ac.uk/mycodb/">http://xbase.warwick.ac.uk/mycodb/</a> [29]
		<i>Mycobacterium tuberculosis</i> Comparative Database	<a href="https://olive.broadinstitute.org/projects/Mycobacterium_tuberculosis_Comparative">https://olive.broadinstitute.org/projects/Mycobacterium tuberculosis Comparative</a> [30]
Genetic diversity and epidemiology	Focused on genetic diversity and epidemiology of MTBC, providing information of epidemiological data, strain lineage, genotyping, and phylogeny. They offer analyses of mycobacterial interspersed repetitive units (MIRU), single nucleotide polymorphism (SNP), long sequence polymorphism (LSP), spoligotyping patterns, IS6110-based restriction fragment length polymorphism (RFLP), and regions of difference (RD) profiles	CASTB	<a href="http://castb.ri.ncgm.go.jp/CASTB/">http://castb.ri.ncgm.go.jp/CASTB/</a> [31]
		MIRU-VNTRplus	<a href="http://www.miru-vntrplus.org/">http://www.miru-vntrplus.org/</a> [32, 33]
		PhyTB	<a href="http://pathogenseq.lshtm.ac.uk/phytblive/index.php">http://pathogenseq.lshtm.ac.uk/phytblive/index.php</a> [34]
		PolyTB	<a href="http://pathogenseq.lshtm.ac.uk/polytblive/browser.php">http://pathogenseq.lshtm.ac.uk/polytblive/browser.php</a> <a href="http://pathogenseq.lshtm.ac.uk/polytblive/map.php">http://pathogenseq.lshtm.ac.uk/polytblive/map.php</a> <a href="http://pathogenseq.lshtm.ac.uk/polytblive/phy.php">http://pathogenseq.lshtm.ac.uk/polytblive/phy.php</a> [35]
		SITVITWEB	<a href="http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/">http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/</a> [36]

Category		Resource	URL	Ref.
Gene expression and regulation	Provide data on mycobacterial transcription factors, predicted operons, predicted transcriptional units, gene expression, and regulatory networks	SpolSimilaritySearch	<a href="http://www.pasteur-guadeloupe.fr:8081/SpolSimilaritySearch/">http://www.pasteur-guadeloupe.fr:8081/SpolSimilaritySearch/</a>	[37]
		spolTools	<a href="http://spoltools.emi.unsw.edu.au/">http://spoltools.emi.unsw.edu.au/</a>	[38]
		TB-Insight	<a href="http://tbinsight.cs.rpi.edu/">http://tbinsight.cs.rpi.edu/</a>	[39]
		CMRegNet	<a href="http://lgcm.icb.ufmg.br/cmregnet/">http://lgcm.icb.ufmg.br/cmregnet/</a>	[40]
		GeNET	<a href="http://bengi.cs.mun.ca/genet/">http://bengi.cs.mun.ca/genet/</a>	[41]
		MTBRegList	<a href="http://www.usherbrooke.ca/vers/MtbRegList/">http://www.usherbrooke.ca/vers/MtbRegList/</a>	[42]
Structural biology	These tools use three-dimensional models of mycobacterial proteins to provide information of domain assignments, functional annotation protein-protein or protein-small molecule interactions, and structural analyses of mutations potentially associated with drug resistance	MycoperonDB	<a href="http://cdfd.org.in/mycoperondb/home.html">http://cdfd.org.in/mycoperondb/home.html</a>	[43]
		CHOPIN	<a href="http://mordred.bioc.cam.ac.uk/chopin/">http://mordred.bioc.cam.ac.uk/chopin/</a>	[44]
		SInCRE	<a href="http://proline.biochem.iisc.ernet.in/sincre/">http://proline.biochem.iisc.ernet.in/sincre/</a>	[45]
Drug targets and resistance	Tools for diagnosis of drug resistance in tuberculosis, new vaccines, and drug targets	AuTuMN TB-modeling	<a href="http://www.tb-modelling.com/mdr_tb_at_retreatment/">http://www.tb-modelling.com/mdr_tb_at_retreatment/</a>	[46]
		MtbVeb	<a href="http://crdd.osdd.net/raghava/mtbveb/">http://crdd.osdd.net/raghava/mtbveb/</a>	[47]
		MUBII-TB-DB	<a href="http://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi">http://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi</a>	[48]
		MycobacRV	<a href="http://mycobacteriarv.igib.res.in/">http://mycobacteriarv.igib.res.in/</a>	[49]
		PhyResSE	<a href="http://phyresse.org/">http://phyresse.org/</a>	[50]
		TB Drug Resistance Mutation Database	<a href="https://tbdreamdb.ki.se/Info/">https://tbdreamdb.ki.se/Info/</a>	[51]
		TB-Profiler	<a href="http://tldr.lshrm.ac.uk/">http://tldr.lshrm.ac.uk/</a>	[52]
		TDR Targets database	<a href="http://tdrtargets.org/">http://tdrtargets.org/</a>	[53]
		TIBLE	<a href="http://mordred.bioc.cam.ac.uk/tible/">http://mordred.bioc.cam.ac.uk/tible/</a>	[54]
TuberQ	<a href="http://tuberq.proteinq.com.ar/">http://tuberq.proteinq.com.ar/</a>	[55]		

**Table 2.** Online publicly available resources dedicated to mycobacteria causing tuberculosis.

## Author details

Edson Machado<sup>1†</sup>, Camillo Cerdeira<sup>1,2†</sup>, Antonio Basílio de Miranda<sup>2</sup> and Marcos Catanho<sup>1\*</sup>

\*Address all correspondence to: [mcatanho@fiocruz.br](mailto:mcatanho@fiocruz.br)

1 Fiocruz, Instituto Oswaldo Cruz, Laboratório de Genômica Funcional e Bioinformática, Manguinhos, Rio de Janeiro, RJ, Brazil

2 Fiocruz, Instituto Oswaldo Cruz, Laboratório de Biologia Computacional e Sistemas, Manguinhos, Rio de Janeiro, RJ, Brazil

† These authors contributed equally.

## References

- [1] Denholm JT, Thevarajan I. Tuberculosis and the traveller: Evaluating and reducing risk through travel consultation. *Journal of Travel Medicine*. 2016;**23**(3):1-6. DOI: 10.1093/jtm/taw008
- [2] National Institute of Health-Bioinformatics Definition Committee. NIH working definition of bioinformatics and computational. *Biology*. 2000;**1** Available from: <http://www.binf.gmu.edu/jafri/math6390-bioinformatics/workingdef.pdf>
- [3] Hagen JB. The origins of bioinformatics. *Nature Reviews Genetics*. 2000;**1**(3):231-236. DOI: 10.1038/35042090
- [4] Britain G, Biology M, Road R. Margaret Oakley Dayhoff 1925–1983. *Bulletin of Mathematical Biology* [Internet]. 1984;**46**(4):467-472 Available from: <http://link.springer.com/10.1007/BF02459497>
- [5] Ouzounis CA, Valencia A. Early bioinformatics: The birth of a discipline—A personal view. *Bioinformatics* [Internet]. 2003;**19**(17):2176-2190. DOI: 10.1093/bioinformatics/btg309
- [6] Ouzounis C. Bioinformatics and the theoretical foundations of molecular biology. *Bioinformatics* [Internet]. 2002;**18**(3):377-378. DOI: 10.1093/bioinformatics/18.3.377
- [7] Joyce AR, Palsson BØ. The model organism as a system: Integrating “omics” data sets. *Nature Reviews Molecular Cell Biology*. 2006;**7**(3):198-210 Available from: <http://www.nature.com/doifinder/10.1038/nrm1857>
- [8] Mardis ER. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*. 2008;**9**(1):387-402 Available from: <http://www.annualreviews.org/doi/10.1146/annurev.genom.9.081307.164359>
- [9] Griffiths AJF, Wessler SR, Carroll SB, Doebley J. *Introduction to Genetic Analysis*. 11th ed. New York: W.H. Freeman and Company; 2015

- [10] Galagan JE. Genomic insights into tuberculosis. *Nature Reviews Genetics*. 2014;**15**(5):307-320 Available from: <http://www.nature.com/doifinder/10.1038/nrg3664>
- [11] Martin JA, Wang Z. Next-generation transcriptome assembly. *Nature Reviews Genetics*. 2011;**12**(10):671-682 Available from: <http://www.nature.com/doifinder/10.1038/nrg3068>
- [12] Sorek R, Cossart P. Prokaryotic transcriptomics: A new view on regulation, physiology and pathogenicity. *Nature Reviews Genetics*. 2010;**11**(1):9-16 Available from: <http://www.nature.com/doifinder/10.1038/nrg2695>
- [13] Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Research*. 2014;**42**(14):8845-8860. DOI: 10.1093/nar/gku555
- [14] Mukhopadhyay S, Nair S, Ghosh S. Pathogenesis in tuberculosis: Transcriptomic approaches to unraveling virulence mechanisms and finding new drug targets. *FEMS Microbiology Reviews*. 2012;**36**(2):463-485. DOI: 10.1111/j.1574-6976.2011.00302.x
- [15] Altelaar AFM, Munoz J, Heck AJR. Next-generation proteomics: Towards an integrative view of proteome dynamics. *Nature Reviews Genetics*. 2012;**14**(1):35-48 Available from: <http://www.nature.com/doifinder/10.1038/nrg3356>
- [16] Mehaffy MC, Kruh-Garcia NA, Dobos KM. Prospective on mycobacterium tuberculosis proteomics. *Journal of Proteome Research*. 2012;**11**(1):17-25. DOI: 10.1021/pr2008658
- [17] Bespyatykh JA, Shitikov EA, Ilina EN. Proteomics for the investigation of mycobacteria. *Acta Naturae*. 2017;**9**(1):15-25 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5406656/>
- [18] Zhang A, Sun H, Xu H, Qiu S, Wang X. Cell metabolomics. *OMICS: A Journal of Integrative Biology*. 2013;**17**(10):495-501. DOI: 10.1089/omi.2012.0090
- [19] Mirsaeidi M, Banoei MM, Winston BW, Schraufnagel DE. Metabolomics: Applications and promise in mycobacterial disease. *Annals of the American Thoracic Society*. 2015;**12**(9):1278-1287. DOI: 10.1513/AnnalsATS.201505-279PS
- [20] Swanepoel CC, Loots DT. The use of functional genomics in conjunction with metabolomics for mycobacterium tuberculosis research. *Disease Markers*. 2014;**2014**:1-12. DOI: 10.1155/2014/124218
- [21] Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assunção JA, Gao G, et al. Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. *Scientific Reports*. 2015;**5**(1):15443 Available from: <http://www.nature.com/articles/srep15443>
- [22] Zhu X, Chang S, Fang K, Cui S, Liu J, Wu Z, et al. MyBASE: A database for genome polymorphism and gene function studies of mycobacterium. *BMC Microbiology*. 2009;**9**:40. DOI: 10.1186/1471-2180-9-40
- [23] Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinburgh, Scotland)*. 2011;**91**(1):8-13. DOI: 10.1016/j.tube.2010.09.006

- [24] Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList—10 years after. *Tuberculosis* (Edinburgh, Scotland). 2011;**91**(1):1-7. DOI: 10.1016/j.tube.2010.09.008
- [25] Ding C, Yuan LF, Guo SH, Lin H, Chen W. Identification of mycobacterial membrane proteins and their types using over-represented tripeptide compositions. *Journal of Proteomics*. 2012;**77**:321-328. DOI: 10.1016/j.jprot.2012.09.006
- [26] Zhu P-P, Li W-C, Zhong Z-J, Deng E-Z, Ding H, Chen W, et al. Predicting the subcellular localization of mycobacterial proteins by incorporating the optimal tripeptides into the general form of pseudo amino acid composition. *Molecular BioSystems*. 2015;**11**(2):558-563 Available from: <http://pubs.rsc.org/en/content/articlehtml/2015/mb/c4mb00645c>
- [27] Lechat P, Hummel L, Rousseau S, Moszer I. GenoList: An integrated environment for comparative analysis of microbial genomes. *Nucleic Acids Research*. 2008;**36**(database issue):D469-D474. DOI: 10.1093/nar/gkm1042
- [28] Catanho M, Mascarenhas D, Degraeve W, Miranda AB. GenoMycDB: A database for comparative analysis of mycobacterial genes and genomes. *Genetics and Molecular Research*. 2006;**5**(1):115-126 Available: <https://www.geneticsmr.com/abstract/genomycdb-a-database-for-comparative-analysis-of-mycobacterial-genes-and-genomes-3115.html>
- [29] Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, Pallen MJ. xBASE2: A comprehensive resource for comparative bacterial genomics. *Nucleic Acids Research*. 2008;**36**(database issue):D543-D546. DOI: 10.1093/nar/gkm928
- [30] Mycobacterium tuberculosis Comparative. Available from: [http://www.broadinstitute.org/annotation/genome/mycobacterium\\_tuberculosis\\_spp/MultiHome.html](http://www.broadinstitute.org/annotation/genome/mycobacterium_tuberculosis_spp/MultiHome.html)
- [31] Iwai H, Kato-Miyazawa M, Kirikae T, Miyoshi-Akiyama T. CASTB (the comprehensive analysis server for the mycobacterium tuberculosis complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis*. 2015;**95**(6):843-844. DOI: 10.1016/j.tube.2015.09.002
- [32] Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. *Journal of Clinical Microbiology*. 2008;**46**(8):2692-2699. DOI: 10.1128/JCM.00540-08
- [33] Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: A web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Research*. 2010;**38**(Web Server issue):W326-W331. DOI: 10.1093/nar/gkq351
- [34] Benavente ED, Coll F, Furnham N, McNerney R, Glynn JR, Campino S, et al. PhyTB: Phylogenetic tree visualisation and sample positioning for *M. tuberculosis*. *BMC Bioinformatics*. 2015;**16**(1):155 Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0603-3>
- [35] Coll F, Preston M, Guerra-Assunção JA, Hill-Cawthorn G, Harris D, Perdigão J, et al. PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. *Tuberculosis*. 2014;**94**(3):346-354. DOI: 10.1016/j.tube.2014.02.005



- [36] Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, et al. SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infection, Genetics and Evolution*. 2012; **12**(4):755-766. DOI: 10.1016/j.meegid.2012.02.004
- [37] Couvin D, Zozio T, Rastogi N. SpolSimilaritySearch—A web tool to compare and search similarities between spoligotypes of *Mycobacterium tuberculosis* complex. *Tuberculosis*. 2017; **105**:49-52. DOI: 10.1016/j.tube.2017.04.007
- [38] Tang C, Reyes JF, Luciani F, Francis AR, Tanaka MM. spolTools: Online utilities for analyzing spoligotypes of the *Mycobacterium tuberculosis* complex. *Bioinformatics*. 2008; **24**(20):2414-2415. DOI: 10.1093/bioinformatics/btn434
- [39] Shabbeer A, Ozcaglar C, Yener B, Bennett KP. Web tools for molecular epidemiology of tuberculosis. *Infection, Genetics and Evolution*. 2012; **12**(4):767-781. DOI: 10.1016/j.meegid.2011.08.019
- [40] Abreu VAC, Almeida S, Tiwari S, Hassan SS, Mariano D, Silva A, et al. CMRegNet—An interspecies reference database for corynebacterial and mycobacterial regulatory networks. *BMC Genomics* [Internet]. 2015; **16**(1):452 Available from: <http://www.biomedcentral.com/1471-2164/16/452>
- [41] Desai AP, Razeghin M, Meruvia-Pastor O, Peña-Castillo L. GeNET: A web application to explore and share Gene Co-expression Network Analysis data. *PeerJ*. 2017; **5**(2012):e3678 Available from: <https://peerj.com/articles/3678>
- [42] Jacques PE, Gervais AL, Cantin M, Lucier JF, Dallaire G, Drouin G, et al. MtbRegList, a database dedicated to the analysis of transcriptional regulation in *Mycobacterium tuberculosis*. *Bioinformatics*. 2005; **21**(10):2563-2565. DOI: 10.1093/bioinformatics/bti321
- [43] Ranjan S, Gundu RK, Ranjan A. MycoperonDB: A database of computationally identified operons and transcriptional units in mycobacteria. *BMC Bioinformatics*. 2006; **7**(Suppl 5): S9. DOI: 10.1186/1471-2105-7-S5-S9
- [44] Ochoa-Montaña B, Mohan N, Blundell TL. CHOPIN: A web resource for the structural and functional proteome of mycobacterium tuberculosis. *Database (Oxford)*, 10. 2015; **2015**(September):1. DOI: 10.1093/database/bav026
- [45] Metri R, Hariharaputran S, Ramakrishnan G, Anand P, Raghavender US, Ochoa-Montaña B, et al. SInCRE—Structural interactome computational resource for *Mycobacterium tuberculosis*. *Database*. 2015; **2015**:1-10. DOI: 10.1093/database/bav060
- [46] Ragonnet R, Trauer JM, Denholm JT, Marais BJ, McBryde ES. A user-friendly mathematical modelling web interface to assist local decision making in the fight against drug-resistant tuberculosis. *BMC Infectious Diseases*. 2017; **17**(1):374 Available from: <http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-017-2478-6>
- [47] Dhanda SK, Vir P, Singla D, Gupta S, Kumar S, Raghava GPS. A web-based platform for designing vaccines against existing and emerging strains of *Mycobacterium tuberculosis*. *PLoS One*, Available from. 2016; **11**(4):1-11. DOI: 10.1371/journal.pone.0153771

- [48] Flandrois J-P, Lina G, Dumitrescu O. MUBII-TB-DB: A database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. BMC Bioinformatics. 2014;**15**(1): 107 Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-107>
- [49] Chaudhuri R, Kulshreshtha D, Raghunandan MV, Ramachandran S. Integrative immunoinformatics for mycobacterial diseases in R platform. Systems and Synthetic Biology. 2014;**8**(1):27-39 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3933634/>
- [50] Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: A web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. Journal of Clinical Microbiology. 2015;**53**(6):1908-1914. DOI: 10.1128/JCM.00025-15
- [51] Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS Medicine. 2009;**6**(2):e2. DOI: 10.1371/journal.pmed.1000002
- [52] Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Medicine. 2015;**7**(1):51 Available from: <http://genomemedicine.com/content/7/1/51>
- [53] Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, et al. Genomic-scale prioritization of drug targets: The TDR targets database. Nature Reviews. Drug Discovery. 2008;**7**(11):900-907 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3184002/>
- [54] Malhotra S, Mugumbate G, Blundell TL, Higuieruelo AP. TIBLE: A web-based, freely accessible resource for small-molecule binding data for mycobacterial species. Database [Internet]. 2017;**2017**(1):1-7 Available from: <https://academic.oup.com/database/article-lookup/doi/10.1093/database/bax041>
- [55] Radusky L, Defelipe LA, Lanzarotti E, Luque J, Barril X, Marti MA, et al. TuberQ: A *Mycobacterium tuberculosis* protein druggability database. Database. 2014;**2014**:1-10. DOI: 10.1093/database/bau035