

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Tuberculosis

journal homepage: <http://intl.elsevierhealth.com/journals/tube>

REVIEW

Informatics resources for tuberculosis – Towards drug discovery

Jagadish Chandrabose Sundaramurthi^a, S. Brindha^b, T.B.K. Reddy^c, Luke Elizabeth Hanna^{a,*}^a ICMR-Biomedical Informatics Centre, Tuberculosis Research Centre (ICMR), Chennai 600031, India^b Department of Biochemistry, University of Madras, Guindy Campus, Chennai 600025, India^c Department of Biochemistry, Stanford University, CA 94305, USA

ARTICLE INFO

Article history:

Received 25 April 2011

Received in revised form

3 August 2011

Accepted 22 August 2011

Keywords:

Tuberculosis

Bioinformatics

Cheminformatics

Database

Drug discovery

SUMMARY

Integration of biological data on gene sequence, genome annotation, gene expression, metabolic pathways, protein structure, drug target prioritization and selection, has resulted in several online bioinformatics databases and tools for *Mycobacterium tuberculosis*. Alongside there has been a growth in the list of cheminformatics databases for small molecules and tools to facilitate drug discovery. In spite of these efforts there is a noticeable lag in the drug discovery process which is an urgent need in the case of emerging and re-emerging infectious diseases. For example, more than 25 online databases are available freely for tuberculosis and yet these resources have not been exploited optimally. Informatics-centered drug discovery based on the integration and analysis of both bioinformatics and cheminformatics data could fill in the gap and help to accelerate the process of drug discovery. This article aims to review the current standing of developments in tuberculosis-bioinformatics and highlight areas where integration of existing resources could lead to acceleration of drug discovery against tuberculosis. Such an approach could be adapted for other diseases as well.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The availability of the full genome sequence of *Mycobacterium tuberculosis* H37Rv¹ brought hope and expectation that novel therapeutic agents would soon be developed against one of the oldest human diseases – tuberculosis. Free access to the complete genome sequences through public domain databases has enabled computational analysis of the sequence data and the development of specific online bioinformatics resources (Table 1), which could facilitate researchers to retrieve genomic sequences and perform computational analyses to identify functionally important genes and proteins, targets for drugs, vaccine candidates and diagnostic reagents. Cheminformatics has also grown, though not in parallel, resulting in significant development (Table 2). However, these developments did not bring any new drugs to the clinic, although a number of new drug candidates have entered clinical trials in recent years.² The apparent delay in translation of research to product points to the fact that there exist certain gaps between current developments and their optimal use. The integration, amalgamation, and convergence of knowledge and resources of bioinformatics (databases and tools) and cheminformatics could

facilitate the drug discovery process. This article presents a review of the available online resources for tuberculosis and our perspectives on the need to effectively integrate bioinformatics and cheminformatics resources to maximally benefit the process of drug discovery.

2. Resources for genomic data

The first complete genome sequence for *M. tuberculosis* was made available for H37Rv strain in 1998,¹ and now complete genome sequences of four more *M. tuberculosis* strains (CDC1551, H37Ra, F11 and KZN1435) are available in the GenBank and can be accessed via the accession numbers AE000516, CP000611, CP000717 and CP001658, respectively. TubercuList is a relational database that provides manually curated genomic information exclusively on the H37Rv strain. This resource is widely used with over 75,000 visits per month from around the world.³

3. Resources for gene regulation

Knowledge on gene regulatory networks in pathogens helps in the understanding of essential processes required for the survival of these organisms and their pathogenesis. Such knowledge is highly useful for identifying critical drug targets. MtbRegList is an exhaustive compilation of about 315 annotated DNA motifs that include 72 transcription start sites, 119 promoters, 121 transcription

* Corresponding author. Dept. of Clinical Research, Tuberculosis Research Centre, Mayor VR Ramanathan Road, Chetput, Chennai 600 031, India. Tel.: +91 44 28369597; fax: +91 44 28362528.

E-mail address: hannatrc@yahoo.com (L. E. Hanna).

Table 1
Online Bioinformatics Resources for Tuberculosis (non-exhaustive).

S. No.	Resource	Content	URL
<i>Genetic and Molecular Resources</i>			
1	TuberculList	Complete dataset of DNA and protein sequences of <i>M. tuberculosis</i> H37Rv with relevant annotation and functional insights for each gene	http://tuberculist.epfl.ch/
2	MtbRegList	Transcription start sites and DNA binding sites cross-referenced with respective transcription factors and predicted regulatory motifs	http://pages.usherbrooke.ca/gaudreau/MtbRegList/www/index.php
3	MycoperonDB	18053 genes organized as 8256 predicted operons and transcriptional units from five closely related species of mycobacteria viz. <i>M. tuberculosis</i> H37Rv, <i>M. tuberculosis</i> CDC1551, <i>M. bovis</i> , <i>M. avium</i> , <i>M. leprae</i>	http://cdfd.org.in/mycoperondb/home.html
4	MycoRegNet	460 gene regulatory interactions in <i>M. tuberculosis</i>	http://mycoregnet.cebitec.uni-bielefeld.de/v1/
5	OperonDB	Computationally predicted operon structures for over 800 genomes including that of <i>M. tuberculosis</i>	http://operondb.cbcb.umd.edu/cgi-bin/operondb/operons.cgi
6	ODB	2000 known operons in about 50 genomes, including that of <i>M. tuberculosis</i> , and about 13,000 putative operons in over 200 genomes	http://www.genome.sk.ritsumei.ac.jp/odb/
7	DOOR	932 predicted operons with 2755 genes from <i>M. tuberculosis</i> H37Rv	http://csbl1.bmb.uga.edu/OperonDB/index.php
<i>Gene Expression Related Resources</i>			
8	MTBreg	Transcriptomics and proteomics data of <i>M. tuberculosis</i> under different biological conditions	http://www.doe-mbi.ucla.edu/Services/MTBreg/
9	TBDB	Gene sequence and annotations for 30 different <i>M. tuberculosis</i> strains and over 3000 MTB microarrays, as well as a suite of comparative genomics and microarray analysis software	http://www.tbdb.org/
<i>Resources for Comparative Genomics</i>			
10	MGDD	SNPs, insertions, repeat expansions and regions showing sequence divergence in six organisms of the <i>M. tuberculosis</i> complex: <i>M. tuberculosis</i> strains H37Rv, CDC1551, H37Ra, F11, <i>M. bovis</i> AF2122/97 and <i>M. bovis</i> BCG str. Pasteur 1173P2	http://mirna.jnu.ac.in/mgdd/
11	GenoMycDB	Large-scale comparative analyses of completely sequenced six mycobacterial genomes (<i>M. tuberculosis</i> strains H37Rv and CDC1551, <i>M. bovis</i> AF2122/97, <i>M. avium</i> subsp. paratuberculosis K10, <i>M. leprae</i> TN, and <i>M. smegmatis</i> MC2 155)	http://157.86.176.108/~catanho/genomycdb/
12	MycoDB	Annotations from GenBank and Uniprot, facilitating genome comparison through a web-based graphical interface	http://xbase.bham.ac.uk/mycodb/
13	PATRIC	Genetic data of <i>M. tuberculosis</i> and a set of tools for comparative analysis	http://patricbrc.vbi.vt.edu/portal/portal/patric/Taxon?cType=taxon&cId=1763
14	MicrobesOnline	MicrobesOnline includes over 1000 complete genomes of bacteria, archaea and fungi and thousands of expression microarrays, and tools for comparative genome analysis, to identify co-regulated genes, operon predictions and a workbench for sequence analysis.	http://www.microbesonline.org/
15	COG	Orthologous genes for 66 microbial genomes; 1449 orthologous genes for <i>M. tuberculosis</i> H37Rv	http://www.ncbi.nlm.nih.gov/COG/
<i>Drug Resistance Related Resources</i>			
16	TBDReaMDB	Mutations associated with TB drug resistance with frequency of most common mutations associated with resistance to specific drugs	www.tbdreamdb.com/
17	<i>Mycobacterium tuberculosis</i> Comparative Database (Broad Institute)	Complete genome sequence of MDR and XDR <i>Mycobacterium tuberculosis</i> strains in addition to drug sensitive strains	http://broadinstitute.org/annotation/genome/mycobacterium_tuberculosis_spp/MultiHome.html
<i>Drug Target Related Resources</i>			
18	TBSCG	Gene sequence, protein sequence, structure and function, as well as homologs for each gene of <i>M. tuberculosis</i>	http://www.webtb.org
19	DEG	Essential genes of many organisms including <i>M. tuberculosis</i>	http://tubic.tju.edu.cn/deg/
20	VFDB	Genes involved in the virulence or pathogenesis of the several pathogens, including 143 virulence factors in <i>M. tuberculosis</i>	http://www.mgc.ac.cn/VFs/main.htm
21	TDR Targets	Prioritized list of entire genes of different pathogens including <i>M. tuberculosis</i> , based on several biological and pharmacological features, for identification of drug targets	http://tdrtargets.org/
22	DDTRP	Potential drug targets for several drug resistant pathogens including <i>M. tuberculosis</i> ; 85 proteins prioritized as alternative drug targets for drug resistant <i>M. tuberculosis</i>	http://bmi.icmr.org.in/DDTRP
<i>Metabolic Pathway Related Resources</i>			
23	MtbRvCyc	Part of BioCyc which is a collection of metabolic pathways for 1129 genomes; MtbRvCyc specifically contains list of metabolic pathways operating in <i>M. tuberculosis</i> H37Rv.	http://biocyc.org/MTBRV/organism-summary?object=MTBRV
24	KEGG	Manually drawn pathway maps representing molecular interaction and reaction networks for many organisms including 108 for <i>M. tuberculosis</i> H37Rv	http://www.genome.jp/kegg/
<i>Integrated Resource for Tuberculosis</i>			
25	TBrowse	More than half a million data-points of genomic data pertaining to <i>M. tuberculosis</i> H37Rv systematically collected from online resources and publications; 100 different bioinformatics resources are integrated for data retrieval and analysis	http://tbrowse.osdd.net/
<i>Consortium for Tuberculosis</i>			
26	OSDD	Consortium aimed at discovering novel therapies for tuberculosis through public participation	http://www.osdd.org/

Table 2
Resources for Virtual Screening (open source and non-exhaustive).

S. No.	Database	Content	URL
1	ChemBank	~900,000 small molecules and hundreds of bio-medically relevant assays	http://chembank.med.harvard.edu
2	ChemDB	~4.1 million commercially available compounds and 8.2 million counting isomers; user-friendly graphical interface, chemical reaction capabilities and unique search possibilities	http://cdb.ics.uci.edu/
3	ZINC	~727,842 molecules with 3D structure for docking applications	http://zinc.docking.org
4	PubChem	>25 million unique chemical structures and 90 million bioactivity outcomes associated with several thousand macromolecular targets; structures pre-clustered and cross-referenced by identity and similarity groups; calculated properties and descriptors for searching and filtering of chemical structures	http://pubchem.ncbi.nlm.nih.gov/
5	Open NCI Database	>250,000 small molecules for searching by numerous criteria; anti-cancer and anti-HIV screening results; calculated log <i>P</i> values and predicted biological activities for nearly all structures	http://129.43.27.140/ncidb2/
6	ArgusLab	An empirical scoring function for structure-based binding and affinity prediction	http://www.arguslab.com/
7	AutoDock	Uses Lamarckian genetic algorithm for molecular docking	http://autodock.scripps.edu/
8	DOCK Blaster	A free web-based service for molecular docking and virtual screening; employs DOCK as the docking program	http://blaster.docking.org/

factor binding sites and 3 terminators described in 56 research papers.⁴ MycoPeronDB contains comprehensive information on known and computationally identified transcriptional units and operons of five Mycobacterial genomes (*M. tuberculosis H37Rv*, *M. tuberculosis CDC1551*, *Mycobacterium bovis*, *Mycobacterium avium* and *Mycobacterium leprae*) along with related PubMed links.⁵ Operon structures play an important role in co-regulation of genes in prokaryotes. Operon information provides a basis and a reference for a comprehensive understanding on how the transcriptional controls are encoded in the genome. This information can be used by molecular biologists to identify targets with greater impact on the pathogen as compared to others. MycoRegNet contains information on 460 regulatory gene interactions occurring in *M. tuberculosis*.⁶ MycoRegNet has features required for data integration, analysis, visualization and reconstruction of mycobacterial transcriptional gene regulatory networks. Apart from the above specialized databases for *M. tuberculosis*, there are several databases where information on regulatory genes and operons of prokaryotes including *M. tuberculosis* are available. OperonDB, is one such database that contains results of a computational algorithm for locating operon structures in microbial genomes. It originally had information on operons for more than 550 bacterial and archeal genomes, comprising the complete collection of finished prokaryotic genomes available in Genbank in 2008.⁷ The updated version (2010) includes 1059 organisms.⁸ ODB⁹ and DOOR¹⁰ are other such resources which can be used by experimental biologists to predict functional linkages between operonic genes of mycobacteria, leading to their experimental characterization and validation.

4. Resources for gene expression

Branches of transcriptomics and proteomics provide wide scope for the identification of targets for drugs, vaccines and diagnostics. Such studies on a genomic scale have generated significant data for tuberculosis. MTBreg¹¹ houses data on differentially expressed genes and proteins which are research outcomes of different transcriptomics studies under experimental conditions like exposure to drugs, hypoxia, acidic pH, heat shock, nutrient starvation and source dependence, transcription factor alterations, regulatory mutations and in response to growth in macrophages or lungs and proteomics studies under conditions like interaction with

macrophages, oxygen depletion, nutrient starvation and aerobic versus anaerobic conditions. MTBreg is cross-linked to the Prolinks Database¹² and the *M. tuberculosis* Structural Genomics Consortium (TBSGC).¹³ TBDB (Tuberculosis Database), developed by a joint collaboration between Stanford and Broad Institute,^{14,15} houses data on nearly 3000 MTB microarrays. In addition to the gene expression data, TBDB also contains information on sequenced clinical strains, metabolic pathways, protein structure and epitope information. By bringing together *M. tuberculosis* genome annotation and gene expression data with a suite of analysis tools, TBDB provides a unique discovery platform for TB.

5. Resources for comparative genomics

The availability of many related genomes of the Mycobacterial complex has enabled us to perform comparative analysis to gain insight into the evolutionary and functional aspects of mycobacterial genes. Some such databases that can be used for comparative analysis of the *M. tuberculosis* genome are described here. MGDD (*M. tuberculosis* Genome Divergence Database) provides information on possible differences that exist between two strains or species of the *M. tuberculosis* complex in terms of SNPs, insertions, repeat expansion and divergent regions, helping us to understand the relationship between genes as well as their evolution.¹⁶ Organisms currently included in the database are *M. tuberculosis* strains H37Rv, CDC1551, H37Ra, F11, *M. bovis* AF2122/97 and *M. bovis* BCG strain Pasteur 1173P2. Pair-wise analyses of these genomes may be performed using MGDD. GenoMycDB, is an online resource for functional classification of mycobacterial proteins as well as large-scale comparative analyses of completely sequenced genome structures, their organization and evolution, based on their predicted protein content.¹⁷ The central structure of GenoMycDB comprises of results obtained after pair-wise sequence alignments of all predicted proteins coded by the genomes of six mycobacteria: *M. tuberculosis* (strains H37Rv and CDC1551), *M. bovis* AF2122/97, *M. avium* subsp. *paratuberculosis* K10, *M. leprae* TN and *Mycobacterium smegmatis* MC2 155. The database provides information on computed similarity parameters for every aligned pair, providing for each protein sequence the features of the corresponding gene, predicted sub-cellular localization of the protein, the assigned cluster of orthologous groups, and links to several important databases like GenBank, KEGG, PDB, Swiss-Prot/TrEMBL, COG,

PSORT, etc., making it a very versatile database. xBASE is a composite online database that can be used for comparative genomics studies of more than 800 bacteria.¹⁸ MycoDB is one of the component databases of xBASE that includes about 40 mycobacterial species and annotations from both GenBank and Uniprot. PATRIC is another comparative genomics database containing rich information on mycobacterial complex. Many software tools are incorporated in the database for analysis, making it very useful for TB researchers.¹⁹ MicrobesOnline is an integrated web resource for comparative and functional genomics with tools for analyzing genomes, genes, gene functions, gene expression data and metabolites.^{20,21} The database also provides information on operon structures by combining comparative genomics and genome-specific distance models.^{20,22} COGs (Clusters of Orthologous Groups of proteins) developed by Tatusov et al.^{23,24} is one of the most cited databases. Orthologs are genes in different species that evolved from a common ancestral gene by speciation and therefore typically have the same function. Identification of orthologs plays an important role in reliably predicting gene functions in newly sequenced genomes. Knowing the inventory of conserved genes responsible for housekeeping functions in an organism but missing in other organisms will pave the way for identification of the pool of potential targets for broad-spectrum antibiotics. At present, COGs houses genes from 66 genomes including that of *M. tuberculosis* H37Rv. COGs lists 1449 orthologs for *M. tuberculosis* H37Rv.

6. Resource for drug resistance

An estimated 390,000–510,000 cases of Multidrug-Resistant Tuberculosis (MDR-TB) emerged globally in 2008 and caused an estimated 150,000 deaths. Almost 50% of MDR-TB cases worldwide are estimated to occur in China and India.²⁵ From a microbiological perspective, drug resistance is caused by a genetic mutation that makes the drug ineffective against the mutant bacilli. A catalog of all drug resistance related mutations will help in the design of tools for detecting drug resistance. TBDRaMDB²⁶ has been developed by manual curation from primary literature. It provides a list of all

known drug resistant mutations, polymorphisms, position in the nucleotide and amino acid sequence, as well as information on the time of isolate collection, country of origin, susceptibility testing method, minimum inhibitory concentration and resistance pattern for *M. tuberculosis*. Further, the frequency data for mutations is also made available. High-confidence mutations may be useful in the design and validation of sequence-based diagnostic tools. Further, for the complete understanding of the molecular basis of drug resistance and the development of vaccines and newer drugs for MDR-TB and (Extensively Drug-Resistant Tuberculosis) XDR-TB, the complete genome sequence of drug resistant strains of *M. tuberculosis* is essential. MDR and XDR strains of *M. tuberculosis* have been completely sequenced and made available freely in the *M. tuberculosis* Comparative Database (Broad Institute).²⁷ Tools for comparative analysis are also available in this database for comparison of different drug resistant strains of the pathogen.

7. Resources for drug targets

Any protein that plays a critical role in the life of a pathogen, such as being essential for its survival or growth, vital to its maintenance in a dormant stage, important for its pathogenesis, involved in the reactivation of the disease, or beneficial to the pathogen in any other way could prove to be a good drug target. Several researchers have been working toward the goal of identifying ideal drug targets, and their findings are spread across literature. DEG (Database of Essential Genes) contains data on essential genes for various organisms including *M. tuberculosis*.²⁸ Six hundred and fourteen genes are listed in this database as essential to the TB bacilli based on a report by Sasseti et al. in 2003.²⁹ The database offers options to access data through *Blast*, *search* or *browse*. For each essential gene, the nucleotide and amino acid sequences are provided along with external hyperlinks to COG. Another important feature that can be exploited for identification of drug targets is the protein's role in pathogenesis. VFDB (Virulence Factors Database) provides a list of proteins involved in the virulence/pathogenesis of various medically significant pathogens,

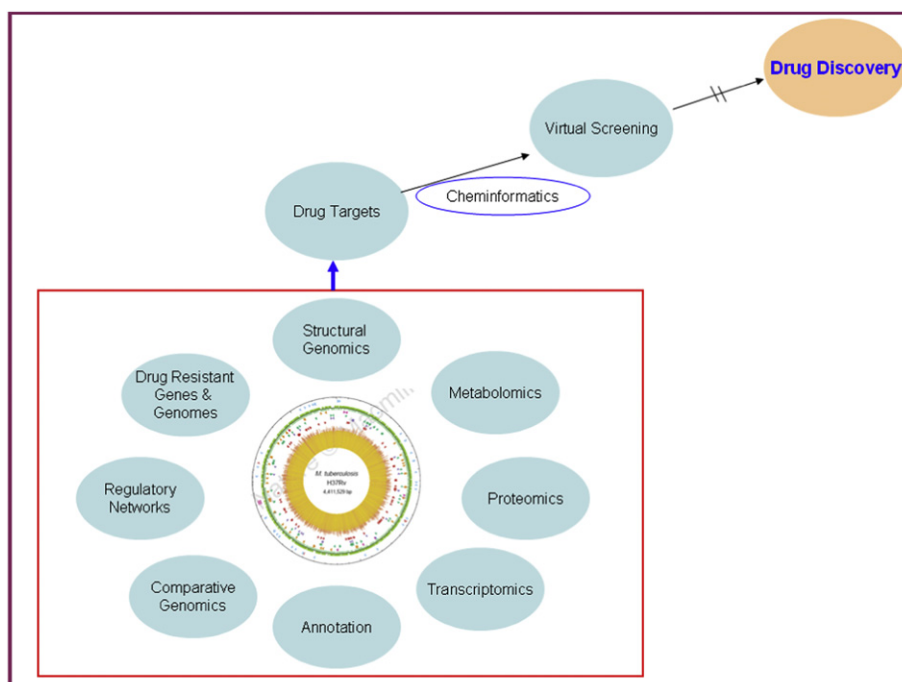


Figure 1. Need for integration of bioinformatics resources with cheminformatics resources toward drug discovery for tuberculosis

including 143 proteins of *M. tuberculosis*.³⁰ Compiling and making available such scattered but significant information would help researchers to move on with ease toward achieving related goals.

Drug targets are generally identified by experimental biology including transcriptomics, proteomics, gene knock-out studies, etc. The availability of three dimensional structures for the targets will go a long way in the process of drug discovery. TBSGC has an online resource for TB called webTB where genomic information, proteomic information, functional annotation, consortium information, homologous sequences, motifs and domains, as well as cross reference to various databases (UniProt, GenBank, Swiss-Prot, IBM, TuberculList, TB protein network, DIP, Sanger, PIR, Pedant) may be accessed for each ORF of *M. tuberculosis*.¹³ Among ~60,000 proteins listed in PDB,³¹ about 964 entries belong to *M. tuberculosis* (as on 29th March 2011), and structures have been resolved by TBSGC for 227 of these proteins (as on 29th March 2011). The status of each protein from the perspective of its potential as a drug target is also provided in the TBSGC portal. An effort from the TDR Consortium that focuses on research in tropical diseases, including tuberculosis, has resulted in a database called “The TDR Target Database”.³² This database facilitates researchers to prioritize drug targets for *M. tuberculosis* based on multiple biological and pharmacological features [viz. enzymatic activity, assayability, crystal structure, model structure, phylogeny (not present in mammals), essentiality, phenotype data, druggability, D index >0.6, compound desirability >0.6, etc]. A useful feature of this database is that each of the given targets can be easily linked to the associated compounds. In addition, the database gives antigenicity index for each of the proteins that could be used for the development of diagnostic tools and vaccines. Database of Drug Targets for Resistant Pathogens (DDTRP) is a database that provides a list of all current as well as alternative drug targets from metabolic pathways of current targets for drug resistant pathogens including *M. tuberculosis*. This database could be useful for novel drug discovery for TB.³³

8. Resources for metabolic pathways

Metabolic pathway analysis helps us to understand the relevance of proteins/enzymes to biochemical function and aid in the identification of potential drug targets. There are many computed metabolic pathway databases like KEGG,³⁴ BioCyc^{35,36} etc. One hundred and eight metabolic pathways are available in KEGG for *M. tuberculosis*, and about 20 other species of the mycobacterial complex. MtbRvCyc is a part of the composite database called BioCyc, which includes metabolic pathways of 1129 organisms including *M. tuberculosis* H37Rv as well as human beings.

9. Cheminformatics resources

Recent advancements in bioinformatics has resulted in many online resources for tuberculosis and improved our understanding about the pathogen and disease. A parallel development has happened in the field of cheminformatics. Due to the cost, time, and resources required for high-throughput screening of compound libraries, virtual screening is becoming very popular for the identification of experimentally active compounds.³⁷ Between 2007 and 2009, ligands for 20 proteins were identified by docking and subsequently confirmed experimentally.³⁸ The first FDA approved HIV-integrase inhibitor, *Isentress* (raltegravir) was initially screened by molecular docking using AutoDock.³⁹ AutoDock is one of the most widely used freely available software for molecular docking and virtual screening.⁴⁰ It accounts for more than 48% of citations among docking softwares.⁴¹ A non-exhaustive list of freely available resources for docking and databases of small molecules are

given in Table 2. Some of the popular small molecule databases from where small molecules can be accessed for virtual screening are ChemBank,^{42,43} ChemDB,⁴⁴ ZINC⁴⁵ and PubChem.⁴⁶ These open sources can be optimally used as an alternative for high-throughput screening for drug discovery.

10. Conclusion

Significant developments have taken place both in the fields of bioinformatics and cheminformatics. Effective amalgamation of the large number of existing resources for tuberculosis would be of great value to TB researchers, since extensive data can be obtained from a single portal with great ease. An online resource called TBrowse⁴⁷ is the outcome of integration of about half-million data-points from online databases and publications. Now is the right time to appropriately integrate bioinformatics and cheminformatics resources to exploit the available resources for discovery of novel drugs against tuberculosis (Figure 1). Efforts toward amalgamation of such resources would bridge the gap between knowledge and outcome.

Acknowledgments

The authors wish to acknowledge the Indian Council of Medical Research for providing the Biomedical Informatics facility.

Funding: None.

Conflict of interest: None of the authors have any conflict of interest to declare.

Ethical approval: Not required.

References

- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;**393**:537–44.
- Rivers EC, Mancera RL. New anti-tuberculosis drugs in clinical trials with novel mechanisms of action. *Drug Discov Today* 2008;**13**:1090–8.
- Lew JM, Kapopoulou A, Jones LM, Cole ST. TuberculList – 10 years after. *Tuberculosis (Edinb)* 2011;**91**:1–7.
- 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**:2563–5.
- Ranjan S, Gundu RK, Ranjan A. MycoPeronDB: a database of computationally identified operons and transcriptional units in Mycobacteria. *BMC Bioinform* 2006 **18**;7(Suppl 5):S9.
- Krawczyk J, Kohl TA, Goesmann A, Kalinowski J, Baumbach J. From *Corynebacterium glutamicum* to *Mycobacterium tuberculosis*—towards transfers of gene regulatory networks and integrated data analyses with MycoRegNet. *Nucleic Acids Res* 2009;**37**:97.
- Pertea M, Ayanbule K, Smedinghoff M, Salzberg SL. OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res* 2009;**37**:D479–82.
- Operon DB. Available from <http://operondb.cbcb.umd.edu/cgi-bin/operondb/operons.cgi> [accessed on 03.08.11].
- Okuda S, Katayama T, Kawashima S, Goto S, Kanehisa M. ODB: a database of operons accumulating known operons across multiple genomes. *Nucleic Acids Res* 2006;**34**:D358–462.
- Mao F, Dam P, Chou J, Olman V, Xu Y. DOOR: a database for prokaryotic operons. *Nucleic Acids Res* 2009;**37**:D459–63.
- The database of Conditionally Regulated proteins in *Mycobacterium tuberculosis*. Available from <http://www.doe-mbi.ucla.edu/Services/MTBReg/> [accessed on 4.04.11].
- Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, Eisenberg D. Pro-links: a database of protein functional linkages derived from coevolution. *Genome Biol* 2004;**5**:R35.
- Ioerger TR, Sacchetti JC. Structural genomics approach to drug discovery for *Mycobacterium tuberculosis*. *Curr Opin Microbiol* 2009;**12**:318–25.
- Reddy TB, Riley R, Wymore F, Montgomery P, DeCaprio D, Engels R, et al. TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* 2009;**37**:D499–508.
- Galagan JE, Sisk P, Stolte C, Weiner B, Koehrsen M, Wymore F, et al. TB database 2010: overview and update. *Tuberculosis (Edinb)* 2010;**90**:225–35.

16. Vishnoi A, Srivastava A, Roy R, Bhattacharya A. MGDD: Mycobacterium tuberculosis genome divergence database. *BMC Genomics* 2008;**9**:373.
17. Catanho M, Mascarenhas D, Degraive W, Miranda AB. GenoMycDB: a database for comparative analysis of mycobacterial genes and genomes. *Genet Mol Res* 2006;**5**:115–26.
18. Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, Pallen MJ. xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* 2008;**36**:D543–6.
19. PathoSystems Resource Integration Center (PATRIC). Available from <http://patricbr.vbi.vt.edu/portal/portal/patric/Taxon?ctype=taxon&cld=1763> [accessed 4.04.11].
20. Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, et al. The MicrobesOnline Web site for comparative genomics. *Genome Res* 2005;**15**:1015–22.
21. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 2010;**38**:D396–400.
22. Price MN, Huang KH, Alm EJ, Arkin AP. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res* 2005;**33**:880–92.
23. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science* 1997;**278**:631–7.
24. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinform* 2003;**4**:41.
25. World Health Organization: Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response. Available from http://www.who.int/tb/features_archive/world_tb_day_2010/en/index.html [accessed 4.04.11].
26. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. *PLoS Med* 2009;**6**:2.
27. Mycobacterium tuberculosis comparative database. Available from http://www.broadinstitute.org/annotation/genome/mycobacterium_tuberculosis_spp/MultiHome.html [accessed 4.04.11].
28. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res* 2009;**37**:D455–8.
29. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 2003;**48**:77–84.
30. Yang J, Chen L, Sun L, Yu J, Jin Q. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Res* 2008;**36**:D539–42.
31. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, et al. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* 2011;**39**:D392–401.
32. 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. *Nat Rev Drug Discov* 2008;**7**:900–7.
33. Database of Drug Targets for Resistant Pathogens (DDTRP). Available from <http://bmi.icmr.org.in/DDTRP> [accessed 4.04.11].
34. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 1999;**27**:29–34.
35. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 2004;**32**:D438–42.
36. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2010;**38**:D473–9.
37. Tang YT, Marshall GR. Virtual screening for lead discovery. *Methods Mol Biol* 2011;**716**:1–22.
38. Kolb P, Ferreira RS, Irwin JJ, Shoichet BK. Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol* 2009;**20**:429–36.
39. Schames JR, Henschman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. *J Med Chem* 2004;**47**:1879–81.
40. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998;**19**:1639–62.
41. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. *Proteins* 2006;**65**:15–26.
42. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, et al. ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* 2008;**36**:D351–9.
43. Strausberg RL, Schreiber SL. From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science* 2003;**300**:294–5.
44. Chen JH, Linstead E, Swamidass SJ, Wang D, Baldi P. ChemDB update—full-text search and virtual chemical space. *Bioinformatics* 2007;**23**:2348–51.
45. Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;**45**:177–82.
46. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. *Drug Discov Today* 2010;**15**:1052–7.
47. Bhardwaj A, Bhartiya D, Kumar N, Scaria V. TBrowse: an integrative genomics map of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* 2009;**89**:386–7.