

# Pathogenic-specific Databases

Peter J. Myler

Bioinformatics of Intracellular Pathogens

JNU, Feb 7-0, 2009

  
**researchtraining**  
on intracellular pathogens

 **SBRI**<sup>TM</sup>  
Seattle Biomedical  
Research Institute

  
Jawaharlal Nehru University

# What information do we need?

- **Sequence**
  - copy number
  - chromosomal location
  - variation (SNPs)
- **Expression pattern**
  - mRNA
  - protein
  - metabolites
- **Function**
  - predicted
  - experimental (essential?)
- **Structure**

# Where can we find the data?

- GenBank
- GeneDB
- ProtozoaDB
- EuPathDB
- BioHealthBase
- TDR Targets

NCBI National Center for Biotechnology Information  
National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for Go

Welcome to the GeneDB website Version 2.1

The Wellcome Trust Pathogen Sequencing Unit

Database Entry Point

Searches Sequence Searches Datasets

Go To: Fungi Choose

ProtozoaDB An Integrated Web2.0 Enabled Genome Database

TriTrypDB The Integrated Genome Resource

News

22 January 2009 Database release of TriTrypDB 3.1  
20 January 2009 International Molecular Cell Biology Meeting III  
20 January 2009 The 4th World Congress on Leishmaniasis  
1 November 2008 EuPathDB to take on the world

Mycobacterium tuberculosis BioHealthBase

Database Search

Organism Mycobacterium tuberculosis

What's New in BioHealthBase

New in BioHealthBase

- December 18, 2008 - Release 4.0
  - Perform a whole genome alignment using our integrated M.Miller tool
  - View information on M. tuberculosis drugs and their targets obtained from DrugBank
  - Our comparative genomics server has been significantly enhanced to visualize genomic prediction data
  - View interactive pI and molecular weight information on the gene details page
  - Full text publications are available on the full-text search capabilities on our website

Organism	Total Strains	Strains	Genome Publication
Mycobacterium tuberculosis	4	F11 H37Rv H37Hu CC-1091	PMID 1654239 PMID 1714672
Mycobacterium bovis	2	AF22267	PMID 1174672
Mycobacterium avium	2	H4 H4	PMID 1614677
Mycobacterium	1	BCD-165	

The TDR Targets Database  
Identification and ranking of targets against neglected tropical diseases

Home New Search History Selected lists of targets Drug targets survey Help/Manual

New in this release

Slideshow Tutorial

Drug Targets Survey

Example lists of prioritized targets for:

- M. tuberculosis
- P. falciparum
- T. brucei
- T. cruzi
- L. major
- B. malayi (Kumar et al.)

Disclaimer: these prioritized lists have been created and made public by users of the site and reflect the criteria and interests of the respective authors. They may contain complete or prioritized genomes or just subsets of genes (i.e. only keywords).

There are more updates in this release of TDR Targets. Check the release notes for a detailed description of all changes.

Welcome

This site is part of a WHO/TDR project seeking to exploit the availability of diverse datasets to facilitate the identification and prioritization of drug targets in pathogens causing neglected diseases.

The TDR Targets database functions both as a website where researchers can look for information on their targets of interest, and as a tool for prioritization of targets in whole genomes. Using the database as a tool, researchers can quickly prioritize a genome of interest by performing any number of individual queries on a species of interest, then assigning numerical weights to each query (in the history page) to finally obtain a ranked list of genes by combining the weighted queries.

Getting started

If you are a first time user and you want to leverage the full benefits of the database, you may want to quickly browse our slideshows tutorials.

To search for targets, click on search.

To examine and work with queries that you have previously run (for example to give scores to each query and then combine them to obtain a prioritized list of targets), click on history.

To examine queries that others have made publicly available, click on posted lists of targets.

To examine the entries in the targets Survey, click on drug targets survey.

News

New TDR Targets release, Feb 07, 2008

This is our second release since we started versioning the database. That is, check the release notes for a detailed description of changes in this release.

New genomes in TDR Targets, Feb 07, 2008

The genomes of *Brugia malayi* and its Wolbachia endosymbiont, *Allotholaxium* larvae, *Trypanosoma brucei* and *Plasmodium* views are now included in TDR Targets. For a summary of available data for these genomes, check the summary table at the bottom of this page.

TDR Targets site updated, Feb 07, 2007

The TDR Targets site has been updated. This update brings new functionalities that allow users to publish individual queries and/or query sets, and an update of the data from the ongoing curatorial effort that now includes *Leishmania* and *Plasmodium*.

Read all news posts here.

Supported by Special Programme for Research & Training in Tropical Diseases (SPRTD) sponsored by

# <http://www.genedb.org>



Welcome to the  
GeneDB website  
Version 2.1



## Database Entry Point

### Searches

Search for  
gene by  
ID/description in

All organisms

- Include description in search  
 Add wildcards to search term

Search Reset

### Sequence Searches

[omniBLAST](#)

(Multi-organism  
BLAST)

Go To single

organism BLAST:

Choose...

### Datasets

Fungi

Go To Choose...

Protozoa

Go To Choose...

Parasitic Helminths

Go To Choose...

Bacteria

Go To Choose...

Parasite Vectors

Go To Choose...

Viruses

Go To Choose...

Go to our [main search page](#), [complex querying page](#), [AmiGO](#) or [List Download](#)

### Information

[Guide to GeneDB](#)

What is GeneDB, and what's in it?

Navigating/Searching GeneDB

Contacting Us/Feedback

Privacy Policy

[Data Release Policy](#)

### Links

PSU Sequencing Projects

Software

[Prokaryotes](#)

[Eukaryotes \(Protozoa\)](#)

[Eukaryotes \(Fungi\)](#)

[ACT](#)

[Artemis](#)



## ProtozoaDB

An Integrated Web2.0 Enabled Genome Database



### Main Menu

#### Home

[ProtozoaDB](#)

[RoadMap](#)

#### Search Sequences

[All species](#)

[Trypanosoma cruzi](#)

[Trypanosoma brucei](#)

[Leishmania major](#)

[Plasmodium falciparum](#)

[Entamoeba histolytica](#)

#### Phylogenetics

[PhyloSearch](#)

#### Repetitions


[Search Repetition](#)

#### Web Services

[D2R Server](#)

## Welcome to Protozoa Database

Comparative approaches involving the integrative use of heterogeneous databases, analyses tools, distributed computing and (re) annotation systems as well as sensitive similarity detection algorithms have been catalyzed by a variety of sequenced genomes. In this purpose, the BiowebDB Consortium, partially funded by CNPq, present the **Protozoa Database**, a integrated, user-friendly and flexible platform that contains the several protozoa genomes: *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania major*, *Plasmodium falciparum* and *Entamoeba histolytica*.

Latest Version: 24 (08/15/2007)  [Database RSS \[?\]](#)

## Web2.0?

**ProtozoaDB** offers a modern Web interface, based on concepts inspired by what has been called the **Web2.0**. This new phase of the World Wide Web can be defined from the user perspective by its improved look-and-feel, more dynamic interfaces and facilitated collaboration and sharing between users. Some examples of Web2.0-inspired features provided by ProtozoaDB include:

- 'Auto Completion': to facilitate the use of ontologies, when the user starts typing a word, our system looks for ontology terms matching the typed text and suggests to the user through a pop-up menu;
- Tagging: you can add your own descriptions for our data, so that you can search for it later;
- Mashups: we fetch information for external sites and aggregate them in one page for you dynamically;
- ProtozoaDB API: we provide lightweight Web Services for third-party use of our data, so that you can also build your own mashups;
- RSS feeds to notify users of updates in our database: you can choose to track any changes to a certain gene, for example, and we will let you know when more annotations are available for that gene.
- SPARQL endpoint: Semantic Web enabled interface for complex queries to our data. On more information on how to use our SPARQL endpoint, please refer to the [D2R Server Technical Note](#) and our tutorial.

## Tag Cloud

We are pleased to announce our fourth annual EuPathDB Workshop, June 7-10, 2009. For more information and to apply for attendance, click here. The application deadline is February 6, 2009.

ApiDB/EuPathDB Bioinformatics Resource Center for Biodefense and Emerging/Re-emerging Infectious Diseases is a portal for accessing genomic-scale datasets associated with the eukaryotic pathogens (*Cryptosporidium*, *Giardia*, *Plasmodium*, *Toxoplasma* and *Trichomonas*).

Click on the logos below to access taxon-specific sites.



CryptoDB

Rel.3.7



GiardiaDB

Rel.1.1



PlasmoDB

Rel.5.5



ToxoDB

Rel.5.0



TrichDB

Rel.1.0

### Genomes in ApiDB/EuPathDB

(Mouse over organism for more information)

(M=Microarray, Pr=Proteomics, Pa=Pathway)

Organism/Strain	Last Updated	Genomic Sequence Size (Mb)	Gene Count	Multiple Strains	SNPs	ESTs	M	Pr	Pa
<b>Cryptosporidium</b>									
<i>C. hominis</i> TU502	02/2008	8.74	3956						• •
<i>C. muris</i>	02/2008	8.48				•			
<i>C. parvum</i> IOWA	02/2008	9.09	3888		•	•			• •
<b>Giardia</b>									
<i>G. lamblia</i> ATCC 50803	05/2008	11.19	4969*			•			
<b>Plasmodium</b>									
<i>P. berghei</i> ANKA	09/2008	18.00	12345				•	•	• •
<i>P. chabaudi</i> AS	09/2008	16.89	15095						• •
<i>P. falciparum</i> 3D7	09/2008	23.27	5695	•	•	•	•	•	• •
<i>P. gallinaceum</i>	09/2008	16.91							
<i>P. knowlesi</i> H	09/2008	25.44	5161						
<i>P. reichenowi</i>	09/2008	7.38			•				
<i>P. vivax</i> Salvador 1	09/2008	28.96	5507			•			•
<i>P. yoelii</i> 17XNL	09/2008	20.17	7971						•
<b>Theileria</b>									
<i>T. annulata</i> Ankara	03/2006	8.35				•			
<i>T. parva</i> Muguga	03/2006	8.35				•			
<b>Toxoplasma</b>									
<i>T. gondii</i>	11/2008	63.00	9239**	•	•	•	•	•	• •
<i>N. caninum</i>	11/2008	62.48	5761			•			
<b>Trichomonas</b>									
<i>T. vaginalis</i> G3	09/2007	176.41	60808			•			•

\* In addition, *G. lamblia* has 4778 deprecated genes that are not included in the official gene count.

\*\* *T. gondii* gene groups identified in ToxoDB across the three strains (ME49, GT1, VEG) and the Apicoplast.

### Tools

- ▶ SRT: Sequence Retrieval for all ApiDB organisms.
- ▶ ApiCyc: Metabolic pathways for all ApiDB organisms.
- ▶ BLAST: All ApiDB sequences.
- ▶ KEGG: Metabolic pathway maps with apicomplexan enzymes.

### Quick Search

ID  ▶go!

Keyword  ▶go!

### ApiDB/EuPathDB News

- ▶ EuPathDB to take on the kinetoplastida (CryptoDB, 1 December 2008)
- ▶ EuPathDB to take on the kinetoplastida (GiardiaDB, 1 December 2008)
- ▶ EuPathDB to take on the kinetoplastida (PlasmoDB, 1 December 2008)
- ▶ EuPathDB to take on the kinetoplastida (ToxoDB, 1 December 2008)

### My ApiDB/EuPathDB Account

Email:

Password:

### ApiDB/EuPathDB Outreach

- ▶ Events with ApiDB/EuPathDB Presence

- ▶ ApiDB/EuPathDB Workshops

Query Availability in Organism Specific Sites:  = CryptoDB  = GiardiaDB  = PlasmoDB  = ToxoDB  = TrichDB

### Identify Genes by:

Genomic Position	Gene Attributes	Other Attributes
<ul style="list-style-type: none"> <li>• Chromosomal Location <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Proximity to Centromeres <input type="checkbox"/></li> <li>• Proximity to Telomeres <input type="checkbox"/></li> <li>• Non-nuclear Genomes <input type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• Type (e.g. rRNA, tRNA) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Exon/Intron Structure <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• Keyword <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• List of IDs <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Species <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Available Reagents <input type="checkbox"/></li> </ul>
Transcript Expression	Protein Expression	Similarity/Pattern
<ul style="list-style-type: none"> <li>• Microarray Evidence <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• EST Evidence <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• SAGE Tag Evidence <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• ChIP chip Evidence <input type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• Mass Spec. Evidence <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• Protein Motif <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Interpro/Pfam Domain <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• BLAST similarity <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> </ul>
Predicted Proteins	Putative Function	Cellular Location
<ul style="list-style-type: none"> <li>• Molecular Weight <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Isoelectric Point <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Protein Structure <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Epitopes <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• GO Term <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• EC Number <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Metabolic Pathway <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Y2H Interaction <input type="checkbox"/></li> <li>• Predicted Interaction <input type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li>• Signal Peptide <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Transmembrane Domain <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/></li> <li>• Organellar Compartment <input type="checkbox"/></li> <li>• Exported to Host <input type="checkbox"/></li> </ul>
Evolution	Population Biology	

## News

**23 January 2009** Beta-release of TriTrypDB 1.0

**20 January 2009** Kinetoplastid Molecular Cell Biology Meeting III

**20 January 2009** The 4th World Congress on Leishmaniasis

**1 November 2008** EuPathDB to take on the kinetoplastida  
All TriTrypDB News

Community Links

Web Tutorials

Information and Help



This is a pre-release version of TriTrypDB that is under active development. There may be incomplete or inaccurate data and frequent site outages can be expected.



We are pleased to announce our fourth annual **EuPathDB Workshop**, June 7-10, 2009. For more information and to apply for attendance, click here. The application deadline is February 6, 2009.

### Identify Genes by:

Expand All | Collapse

- Text, IDs, Species
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- Protein Expression
- Cellular Location
- Putative Function
- Evolution

My Search History: 0

### Identify Other Data Types:

Expand All | Collapse

- Identify Genomic
- Identify EST
- Identify Assembly
- Identify ORF

My Search History: 0

### Tools:

#### BLAST

Identify Sequence Similarities

#### Sequence Retrieval

Retrieve Specific Sequences using IDs and coordinates

#### PubMed and Entrez

View the Latest *Trypanosoma* and *Leishmania* Pubmed and Entrez Results

#### GBrowse

View Sequences and Features in the GMOD Genome Browser



INFLUENZA  
VIRUS



MYCOBACTERIUM  
TUBERCULOSIS



FRANCISELLA  
TULARENSIS



MICROSPORIDIA



RICIN

The Biodefense and Public Health Database (BioHealthBase) Bioinformatics Resource Center (BRC) provides a comprehensive genomic and proteomic data repository for five pathogenic organism groups that pose a threat to public health. The BioHealthBase BRC also provides an analysis platform and appropriate tools to facilitate genomic and proteomic study of these pathogens. The goal of the BioHealthBase BRC is to provide a resource to the scientific research community to facilitate the development of vaccines, diagnostics and therapeutics for these pathogens. Click on your pathogen of interest above.

## What's New in BioHealthBase

### December 18 , 2008 - Release 4.6

- **Unique features for *influenza***
  - The influenza home page has been totally redesigned to highlight what interests our users the most.
  - See detailed information on **anti-viral drugs** and their targets.
  - **Keyword Search** now allows you to include or exclude genetically manipulated viruses.
  - Our novel **ModelCompare utility** suggests the best model of evolution for your set of sequences.
  - Our workbench now allows you to perform set operations on working sets and searches, combining them, finding the intersect, and subtracting one from another.
  - **Experimental database** – expanding available data
    - More than 3500 new avian surveillance records have been added.
    - Results from experiments can now be visualized using a multi-featured graphing tool.
  - We now provide a direct link to virus reagents available on the **BEI** resource
- **Unique features for *Francisella* and *Mycobacterium***
  - Perform a whole genome alignment using our integrated MUMmer tool.
  - View information on *M. tuberculosis* and *Francisella* drugs and their targets obtained from DrugBank.
  - Our comparative genomics viewer has been significantly enhanced to visualize operon prediction data.
  - View isoelectric pH and molecular weight information on the gene details page.
  - Our workbench now allows you to perform set operations on working sets and searches, combining them, finding the intersect, and subtracting one from another.
- **Unique features for *Microsporidia***
  - Added **Epitope Search** that allows user to search for curated IEDB epitopes in BioHealthBase.
  - View isoelectric pH and molecular weight information on the gene details page.
  - Our workbench now allows you to perform set operations on working sets and searches, combining them, finding the intersect, and subtracting one from another.

### Previous Releases

- [View previous releases](#)

### Data and Usage Summary

[View](#) a summary of the genome data in BioHealthBase and statistics about site usage.

### Data Source and Types

[View](#) a summary of the data types provided in BioHealthBase for each genome and how each was obtained

### Recent Updates

[View](#) a summary of when each data type was last updated.

### Related Links

[View](#) links to other web sites of interest to the BioHealthBase user community.

### Publications and Meetings

[View](#) a list of publications produced and meeting presentations made by the BioHealthBase Team.



# <http://tdrtargets.org>

## The TDR Targets Database

Identification and ranking of targets against neglected tropical diseases

[Login](#) | [Register](#) | [Documentation](#) | [Contact](#) | [FAQ](#)

[home](#) | [new search](#) | [history](#) | [posted lists of targets](#) | [drug targets survey](#) | [user manual](#)

### New in this release



[Slideshow Tutorials](#)

### Drug Targets Survey

Example lists of prioritized targets for:

[M. tuberculosis](#)  
[P. falciparum](#)  
[T. brucei](#)  
[T. cruzi](#)  
[L. major](#)  
[B. malayi \(Kumar et al.\)](#)

**Disclaimer:** these prioritized lists have been created and made public by users of the site and reflect the criteria and interests of the respective authors. They may contain complete prioritized genomes or just subsets of genes (i.e. only kinases).

There are more updates in this release of TDR Targets. Check the [release notes](#) for a detailed description of all changes.

## Welcome

This site is part of a WHO/TDR project seeking to exploit the availability of diverse datasets to facilitate the identification and prioritization of drug targets in pathogens causing neglected diseases.

The TDR Targets database functions both as a **website** where researchers can look for information on their targets of interest; and as a **tool** for prioritization of targets in whole genomes. Using the database as a tool, researchers can quickly prioritize a genome of interest by performing any number of individual queries on a species of interest, then assigning numerical weights to each query (in the history page) to finally obtain a ranked list of genes by combining the weighted queries.

### Getting started

If you are a first time user and you want to leverage the full benefits of the database, you may want to quickly browse our [slideshow tutorials](#).

To search for targets, click on [search](#).

To examine and work with queries that you have previously run (for example to give scores to each query and then combine them to obtain a prioritized list of targets), click on [history](#).

To examine queries that others have made publicly available, click on [posted lists of targets](#).

To examine the entries in the targets Survey, click on [drug targets survey](#).



Supported by

Special Programme for Research & Training  
in Tropical Diseases (TDR) sponsored by  
UNICEF/UNDP/World Bank/WHO

### News

**New TDR Targets release.**  
Posted: 30 Jun 2008

This is our second release (since we started versioning the database, that is). Check the [release notes](#) for a detailed description of changes in this release.

**New genomes in TDR Targets.**  
Posted: 7 Mar 2008

The genomes of *Brugia malayi* and its Wolbachia endosymbiont, *Mycobacterium leprae*, *Toxoplasma gondii* and *Plasmodium vivax* are now included in TDR Targets. For a summary of available data for these genomes, check the summary table at the bottom of this page.

**TDR Targets site updated**  
Posted: 7 June 2007

The TDR Targets site has been updated. This update brings new functionality that allow users to publish individual queries and/or query sets; and an update of the data from the ongoing curation effort that now includes *Leishmania* and *Plasmodium*.

Read old news posts [here](#).