



# Databases and web resources for neglected tropical disease research

Praffulla Kumar Arya, Krishnendu Barik, Ajay Kumar Singh, Anil Kumar\* 

Department of Bioinformatics, Central University of South Bihar, Fatehpur, India.

---

## ARTICLE INFO

Received on: 09/01/2023

Accepted on: 08/07/2023

Available Online: 04/08/2023

---

### Key words:

Neglected tropical diseases, infectious diseases, therapeutic target, drug discovery.

---

## ABSTRACT

The term “neglected tropical diseases” (NTDs) refers to a group of infectious diseases that are common in developing tropical and subtropical nations. There are 20 such diseases that are prevalent in 149 countries according to the World Health Organization. These diseases have a devastating impact on the economies of affected countries and result in enormous losses in terms of morbidity and mortality. All NTDs need to receive more attention to prevent widespread epidemics and outbreaks that could result in a high death toll globally. The few drugs that are already on the market are no longer effective against many infectious diseases, and various other problems, such as toxicity and pharmacokinetics concerns, also prohibit the eradication of these diseases. One of the crucial steps in the drug discovery pipeline is the identification and validation of a therapeutic target based on the biology of the organism. This review provides an overview of more than 35 web-accessible NTD databases categorized under general, specialized, and other NTD databases and highlights recent advances. These databases can be very useful for research, especially drug design and discovery against NTDs. The majority of these databases are simple to use and freely accessible online.

---

## INTRODUCTION

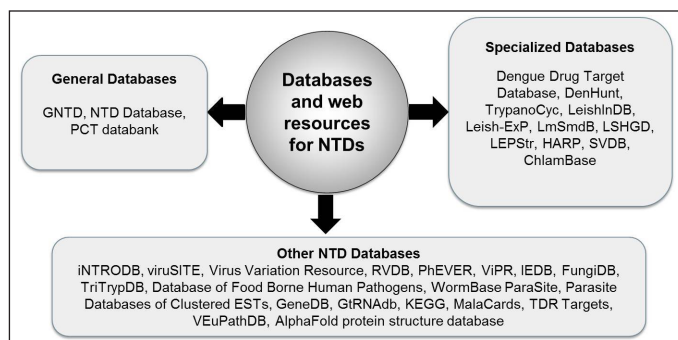
Neglected tropical diseases (NTDs) are a collection of infections caused by various types of pathogens, such as bacteria, viruses, fungi, and parasites (Ferreira *et al.*, 2023). These diseases are common in many tropical and subtropical regions of developing countries where poverty is prevalent (Mukherjee, 2023). It is a worldwide problem that affects a significant portion of the world’s population, approximately one billion people (Mitra and Mawson, 2017). Since many NTDs are asymptomatic and have long incubation periods, their significance has often been underestimated. Sometimes it is difficult to determine whether a death is related to a long-dormant tropical disease (Getaz *et al.*, 2016). Moreover, geographic isolation often complicates treatment and prevention in areas of high endemicity. People affected by NTDs are mostly from impoverished and marginalized communities, with little visibility and political clout (Hotez *et al.*, 2007; Silva *et al.*, 2021). Pharmaceutical companies do

not prioritize such diseases because they are not considered a health risk in the world’s wealthier regions (Biswas and Mandal, 2023; Hotez, 2013). There is no doubt that NTDs have significant impacts on children’s cognitive and physical development, pregnancy complications, and the productivity of the worker class, particularly in impoverished rural communities where people rely on manual labor for their livelihoods (Lenk *et al.*, 2016). The severity, duration, and type of the disease can all contribute to decreased productivity, potentially leading to reduced efficiency while at work, absenteeism or job loss (Conteh *et al.*, 2010; Hotez *et al.*, 2007).

NTDs defined by the World Health Organization (WHO) are African trypanosomiasis, buruli ulcer, Chagas disease, chromoblastomycosis and other deep mycoses, dengue, dracunculiasis, lymphatic filariasis, echinococcosis, foodborne trematode, leprosy, mycetoma, leishmaniasis, schistosomiasis, rabies, onchocerciasis, neurocysticercosis, and taeniasis, soil-transmitted helminthiases, scabies, and other ectoparasites, trachoma, yaws, and snakebite envenoming (Casulli, 2021; Mukherjee, 2023). This review summarizes the currently available databases on NTDs. These databases are classified as general, specialized, or other NTD databases (Fig. 1). The general database is comprehensive in two aspects. In addition to covering almost all NTDs, it provides general information such

---

\*Corresponding Author  
Anil Kumar, Department of Bioinformatics, Central  
University of South Bihar, Fatehpur, India.  
E-mail: [kumaranil@cub.ac.in](mailto:kumaranil@cub.ac.in)



**Figure 1.** Databases and web resources for NTDs research.

as risk models, targeted control interventions, disease monitoring, and surveillance data, and historical and current survey data. The second is disease-specific, providing details on a specific NTD or disease-causing organism. These databases contain genomics, transcriptomics, proteomics, and metabolomics data. The third category includes databases that store information about various diseases and disease-causing organisms, including NTDs. All of the databases are freely accessible and may be useful to clinicians as well as researchers.

## GENERAL NTD DATABASES

Databases in this category provide general information on NTDs, such as historical and recent survey data, model-based disease risk estimations, survey population data, survey period, diagnostic tests used, geographical information of survey locations, disease risk modeling, control intervention targeting, disease monitoring, surveillance data, number of infected individuals, distribution of diseases, disease-specific epidemiological situations, and disease-specific epidemiological situations. [Table 1](#) lists general NTD databases.

### Global neglected tropical diseases database (GNTD)

GNTD is a free, global, public online resource that includes historical and recent NTD survey data, as well as the first model-based disease risk estimations. It was initiated by a project funded by the European Union to achieve optimal control and transmission surveillance of schistosomiasis in Sub-Saharan Africa. It was created to promote epidemiological research and provide up-to-date disease risk estimates ([Brodish and Singh, 2016](#)). In the mid-1980s, [Doumenge \*et al.\* \(1987\)](#) performed the first globally comprehensive survey of schistosomiasis prevalence. Later, [Brooker \*et al.\* \(2010\)](#) collected schistosomiasis and soil-transmitted helminthiasis data, but accessed data were limited. Other NTDs added to the GNTD database include lymphatic filariasis, leishmaniasis, Chagas disease, leprosy, soil-transmitted helminthiasis (infections caused by *Trichuris trichiura*, *Ascaris lumbricoides*, and hookworm), and Buruli ulcer ([Hürlimann \*et al.\*, 2011](#)). The database contains basic information such as the authors, nature of the publication and publication year as well as study-oriented information such as demographic survey, research period, diagnostic test applied and number of infected individuals categorized by age and gender among those examined ([Ezeh \*et al.\*, 2019](#); [Kokaliaris \*et al.\*, 2022](#)). The database did not incorporate hospital research, information on certain specific vulnerable populations (infected with viruses) or post intervention studies.

Geographical information in the GNTD database facilitates the consideration of potential spatial correlations and therefore makes it possible to develop more realistic models ([Schur \*et al.\*, 2011](#)).

### NTD database

The NTD Database is a database that offers special opportunities for modeling disease risk, focusing control interventions, tracking disease progression, and providing surveillance services from projects backed by the United States Agency for International Development (USAID) aimed at managing and eradicating NTDs ([Rust \*et al.\*, 2022](#)). It can also be used to evaluate the spatial and temporal distribution of diseases ([Schur \*et al.\*, 2011](#)). In the NTD Database, data are gathered from a variety of USAID NTD projects, including the NTD Control Program (2006–2012), End Neglected Tropical Diseases (END) in Africa (2010–2014), END in Asia (2010–2014), and the ENVISION Project (2011–2016). The NTD Database contains data from various USAID NTD projects, including the NTD Control Program (2006–2012), the ENVISION Project (2011–2016), END in Asia (2010–2014) and END in Africa (2010–2014). The database was created as part of USAID's ENVISION project. The database is only accessible to authorized users. The program collects data on all people treated and the drugs they use on a regular basis, as recorded by community drug distributors during mass drug administration rounds and reported to the national level via the health system ([Lemoine \*et al.\*, 2016](#)).

### Preventive chemotherapy and transmission control (PCT) databank

The WHO established the PCT databank resource to provide access and dissemination of information relevant to national programs with stakeholders indulged in NTD control ([Kumapley \*et al.\*, 2015](#)). The resources support users in becoming familiar with Excel and the Global Health Observatory so that they can construct summaries as quickly as possible ([Yajima \*et al.\*, 2012](#)). This platform provides comprehensive data and analyses on key preventative chemotherapy health themes, as well as current trends ([Turner \*et al.\*, 2022](#)). There are also links to additional resources. The database contains data on disease-oriented epidemiological conditions, the geographically overlapped NTD regions and updates on control intervention in all NTD-endemic countries. Onchocerciasis, lymphatic filariasis, schistosomiasis, soil-transmitted helminthiasis, and trachoma diseases are all covered in the database ([Yajima \*et al.\*, 2012](#)). The PCT database only incorporates WHO-reported data and leaves out survey information collected from other sources, such as (nongovernmental organizations), academic researchers and other partners ([Brooker \*et al.\*, 2010](#)). The administrator of the PCT database compares reported treatment through the inventory to the previously reported treatment submitted in the PCT Databank by WHO regional offices and Ministries of Health. The assessment is restricted to reported therapies that have actually been used; it lists the quantity of medications that were donated to other organizations rather than given directly to communities as treatments ([Gallo \*et al.\*, 2013](#)).

### SPECIALIZED NTD DATABASES

The Specialized NTD database category includes databases devoted to a specific NTD or disease-causing

**Table 1.** General NTD databases.

| S. No. | Database name | URL   |
|--------|---------------|---|
| 1.     | GNTD          | <a href="http://www.gntd.org">www.gntd.org</a>  |
| 2.     | NTD database  | <a href="https://www.ntddatabase.org/">https://www.ntddatabase.org/</a>   |
| 3.     | PCT databank  | <a href="https://www.who.int/teams/control-of-neglected-tropical-diseases/data-platforms/pct-databank">https://www.who.int/teams/control-of-neglected-tropical-diseases/data-platforms/pct-databank</a> |

organism. These databases contain information about genomics, transcriptomics, proteomics, and metabolomics. Table 2 lists specialized NTD databases.

### Dengue drug target database (DengueDT-DB)

DengueDT-DB is a web-based, freely accessible database for dengue disease drug targets that provides the most beneficial data related to genomic studies, gene information, and therapeutic target information for researchers (to develop better dengue vaccines) and doctors for better treatment of dengue patients. Genome sequences were obtained from GenBank, which includes strand information in a graphical format, drug targets from DrugBank, and analog targets from a Basic Local Alignment Search Tools (BLAST) search query. The platform also provides links to global sequence databases such as National Center for Biotechnology Information (NCBI), DNA Data Bank of Japan, European Molecular Biology Laboratory (EMBL), Protein Data Bank (PDB), ExPasy, and medical literature such as PubMed. The database's web interface is user-friendly and well-organized, allowing anyone to easily access the database. This database is accessible at URL: <http://www.bioinformatics.org/dengueDTDB/Pages/main.htm>.

### Dengue human interaction database (DenHunt)

DenHunt is a comprehensive and integrated network resource for dengue virus interaction with humans (Zhang *et al.*, 2023). The database includes information on 4,120 human genes that are differentially expressed in dengue-infected cell lines and patients, as well as experimentally verified data on approximately 682 direct interactions of human proteins with dengue viral components, 382 indirect interactions and 4,120 direct interactions. The database portrays the dynamic network connections of viral pathogenesis by mapping dengue-human interactions onto the host interactome. The database is categorized into parts that contain pathways, expression variants, direct and indirect interactions and (dengue virus host-dependent factors). The database supports the scientific community in comprehending dengue virus interactions with its human host by offering a variety of information, including disease stage, cell type, gene expression patterns, gene silencing studies, direct and indirect connections, virus serotype, etc. (Karyala *et al.*, 2016).

### TrypanoCyc

TrypanoCyc is a community annotated pathway/genome repository of *T. brucei* (*Trypanosoma brucei*), the agent responsible for African trypanosomiasis (in humans) or nagana disease (in domestic animals) (Basher *et al.*, 2020; Prakash *et al.*, 2022). The French National Institute of Agricultural Research or Institut Nationale de la Recherche Agronomique maintains TrypanoCyc to gather and make public data on *T. brucei* metabolism. TrypanoCyc

was developed with the use of a team-based web platform called TrypAnnot, which allowed the annotation work to be distributed across teams of specialists working on different pathways of interest. TrypanoCyc enhances automatic metabolic network reconstruction by incorporating compartment information as well as information on enzyme activity specific to the developmental stage. These network metadata will help with the development of customized metabolic networks and a better comprehension of parasite metabolism. The database is continuously updated as novel pathways and reactions are found in the literature and choke point analyses are improved to find prospective therapeutic targets (Shameer *et al.*, 2015). TrypanoCyc reported 94% of clinically proven therapeutic targets for African sleeping sickness and 100% of clinically validated drug targets for Nagana disease. Drug targets found in the *T. brucei* network can be prioritized or modeled based on druggability, and similar drugs targeting homologues in other species can be designed (Chukualim *et al.*, 2008).

### Leishmania inhibitor database (LeishInDB)

The LeishInDB is a web-based resource that contains manually curated compounds from the literature that have been shown to inhibit *Leishmania* species. The database currently contains 8,273 records culled from over 600 literature reports. The database is designed so that users can search for chemicals based on a number of parameters, including experiment type, target organism, target enzyme, predicted oral toxicity, predicted carcinogenicity, IC50 range and so on. In addition, each compound's toxicity can be predicted. The database can be used to search for compounds with specific activity against *Leishmania* and use the results to repurpose the compounds or select compounds for synergistic studies. In addition, chemical characteristics such as the number of rotatable bonds, molecular weight, logP value, 2D/3D structural data, number of hydrogen bond donors and acceptors, etc., were incorporated. AdmetSAR was used to predict the toxicity of each molecule, and the resulting data were made available to perform the filtered search. Additionally, a substructure search feature, a dynamic search across a number of fields, and the ability to download all of the molecule structures that match the search criteria were included. Active links exist between each target enzyme and TriTrypDB (Vijayakumar *et al.*, 2020). The user can perform simple text searches to complicated searches with numerous range values, fields and dynamic chemical substructures using the simple graphical user interface of the LeishInDB portal. Bootstrap, Hypertext Markup Language, and Hypertext Preprocessor were used to develop the LeishInDB user interface. Javascript was used for efficient navigation and data retrieval from the database. My Structured Query Language was used to manage data. A chemical substructure search is performed using an integrated Practical Extraction and Reporting Language script. The (Java Molecular Editor) plugin, a tool for creating

**Table 2.** Specialized NTD databases.

| S. No. | Database name                                | URL   |
|--------|--|---|
| 1.     | Dengue drug target database                  | <a href="http://www.bioinformatics.org/dengueDTDB/Pages/main.htm">http://www.bioinformatics.org/dengueDTDB/Pages/main.htm</a> |
| 2.     | DenHunt                                      | <a href="http://proline.biochem.iisc.ernet.in/DenHunt/">http://proline.biochem.iisc.ernet.in/DenHunt/</a>                     |
| 3.     | TrypanoCyc                                   | <a href="http://vm-trypanocyc.toulouse.inra.fr/">http://vm-trypanocyc.toulouse.inra.fr/</a>                                   |
| 4.     | LeishInDB                                    | <a href="http://leishindb.biomedinformri.com/">http://leishindb.biomedinformri.com/</a>                                       |
| 5.     | Leish-Exp                                    | <a href="http://www.hpppi.iicb.res.in/Leish-ex/index.html">http://www.hpppi.iicb.res.in/Leish-ex/index.html</a>               |
| 6.     | LmSmdB                                       | <a href="http://www.nccs.res.in/LmSmdb/">http://www.nccs.res.in/LmSmdb/</a>   |
| 7.     | LSHGD  | <a href="http://www.vit.ac.in/leprosy/leprosy.htm">http://www.vit.ac.in/leprosy/leprosy.htm</a>                               |
| 8.     | LEPStr- <i>M. leprae</i> : tandem repeats DB | <a href="https://www.jalma-icmr.org.in/LEPStr/">https://www.jalma-icmr.org.in/LEPStr/</a>                                     |
| 9.     | HARP   | <a href="https://harp-leprosy.org/">https://harp-leprosy.org/</a>   |
| 10.    | SVDB   | <a href="https://www.snakevenomdb.org/">https://www.snakevenomdb.org/</a>   |
| 11.    | ChlamBase                                    | <a href="https://chlambase.org/">https://chlambase.org/</a>   |

complex structures, is used by the LeishInDB interface to perform the substructure search. Js modules were used to create graphs and plots within the database (Vijayakumar *et al.*, 2019).

#### Leishmania exclusivprotein database (Leish-Exp)

The Leish-Exp database contains a number of distinctive proteins that are exclusively found in *Leishmania* (Prakash *et al.*, 2022). It contains the protein sequences of five different *Leishmania* species (*Leishmania major*, *Leishmania infantum*, *Leishmania braziliensis*, *Leishmania mexicana*, and *Leishmania donovani*) (Fall *et al.*, 2022). The database could be useful in the future for target-based leishmaniasis pharmaceutical therapy. The database is divided into two parts, each of which can be queried separately: 1) *Leishmania* species-specific protein information and 2) *Leishmania* genus-specific protein information. Users can conduct searches by entering the *Leishmania* protein ID or the UniProt ID. Furthermore, the Leish-Exp database enables the BLAST server to compare query data to the Leish-Exp database (Das *et al.*, 2020). This database is accessible at URL: <http://www.hpppi.iicb.res.in/Leish-ex/index.html>.

#### LmSmdB

LmSmdB is an integrated database that incorporates the genomic sequences of *Schistosoma mansoni* and *Leishmania major* (Das *et al.*, 2020; Prakash *et al.*, 2022). It is listed on the (National Center for Cell Science) online page. It takes into account biological networks and regulatory pathways that have been established by computational methods. It is the first database to show both the product's simulation pattern and network design. The database aims to create a broad canopy for the control of the lipid metabolic response in parasites, enabling metabolism to be controlled at the genetic level by integrating regulatory genes, transcription factors, and protein products. The integration approaches offered by this web-based database provide a platform for the generation of hypotheses for major human disorder systems. It will be easy to identify a therapeutic target and model a disease network if it is established that the conservation of two species is appropriate. The conservation of species allows us to search for homologous and orthologous genes in a model organism. One can search LmSmdB using the gene name and Kyoto Encyclopedia of Genes and Genomes (KEGG)/NCBI cluster identifier. LmSmdB

is continuously updated via manual screening of new articles and a review of the literature. Additionally, LmSmdB links users to a KEGG pathway for a specific gene that is directly linked with the network or pathway (Patel *et al.*, 2016). LmSmdB contains no information about the inhibitors.

#### Leprosy susceptible human gene database (LSHGD)

LSHGD is a freely accessible web-based database that integrates leprosy and human-associated genes through a thorough literature search. The database offers a user-friendly and fascinating environment for exploring the function of Single Nucleotide Polymorphisms (human polymorphisms) in leprosy as well as the independent genetic control of both *Mycobacterium leprae* multidrug resistance and susceptibility to leprosy. This is the first genetic database on human leprosy, with the aim of delivering details on leprosy polymorphisms, related genes, corresponding protein sequences, and accessible three-dimensional structures. Users can access relevant information on the web, which serves as a useful information platform and versatile tool. LSHGD provides comprehensive information for researchers looking to investigate the linkage and involvement of the human genome with leprosy (Doss *et al.*, 2012).

#### LEPStr

LEPStr is an easy-to-use interactive web-based database designed to find *M. leprae* minisatellites and microsatellites, as well as primers to amplify the repeat areas. It can be used to identify *M. leprae* strains and track the dynamics of leprosy transmission. The database contains 166 detected short tandem repeats screened from the entire genome of *M. leprae* (Mohanty *et al.*, 2020).

#### Hansen's disease antimicrobial resistance profiles (HARP)

HARP is a web-based platform that predicts the structural consequences of *M. leprae* therapeutic target mis-sense mutations. This web library contains sequence- and structure-based predictions of mutation-induced changes in stability and vibrational entropy that have been computed. It is a website devoted to learning about how known and emerging mutations influence protein–ligand, protein–protein, and protein–nucleic acid affinities as well as the antibiotic resistance that leads to leprosy (Vedithi *et al.*, 2020). The mycobacterial scientific community



benefits from HARP's knowledge of the structural consequences of mutations that cause antibiotic resistance in leprosy (Maladan *et al.*, 2021). HARP was developed in Blundell's laboratory at the Department of Biochemistry at the University of Cambridge (Vedithi *et al.*, 2020).

### Snake venom database (SVDB)

The SVDB is a web-based, automated, subject-specific, nonredundant generic database for storing, disseminating, and analyzing data on poisonous snakes, venom compositions, and functions (Rahman *et al.*, 2022; Soodeh *et al.*, 2021). To aid in data integration, SVDB offers autonomous linkages that may retrieve significant information both asynchronously and on-demand. SVDB includes precise, nonredundant, and up-to-date scientific data on a wide range of subjects, including small molecules, sequences, taxonomy, structures, literature, and more. The SVDB platform also offers external links to BLAST, CLUSTALW-1, phylogeny, SWISS-MODEL and other toxin-related resources. The SVDB information architecture is unique and may be included in any generic data gathering process for a given topic via NCBI. On Earth, there are approximately 375 different species of poisonous snakes. The composition of snake venom is complex and comprises a variety of poisons, including cardiotoxins, neurotoxins, myotoxins, and hemotoxins, as well as different combinations of these poisons. Snake venom serves a number of therapeutic and pharmacological purposes in addition to treating human snake bites. Experimental results reported that data on snake venom may be quickly crawled and displayed using SVDB (Hossain *et al.*, 2018).

### ChlamBase

ChlamBase is a freely accessible web-based database that contains collaboratively organized genomic and proteomic data, as

well as functional annotations for four model *Chlamydia* species. It is a community-curated open source database based on the Wikidata and WikiGenomes application architecture. ChlamBase provides a single platform to access genomic and proteomic data for the *Chlamydia* research community. ChlamBase combines crucial data collected from the literature as well as from a number of external databases. ChlamBase allows users to search, edit and add gene annotations based on evidence, developmental gene expression, orthologous gene comparisons, engineered mutant strains, protein interactions, functions, processes, and more. In addition, ChlamBase provides a widget for sequence alignment of either DNA or amino acids of *Chlamydia* orthologs, as well as a display of the organism genome using JBrowse at the top of the main page (Putman *et al.*, 2019).

### OTHER IMPORTANT NTD DATABASES

Other important NTD database categories include databases that deal with more than one NTD or a variety of disease-causing organisms. These databases provide much more comprehensive genomics-, transcriptomics-, proteomics- and metabolomics-related information than specialized databases. Other important databases are listed in Table 3.

### INTRODB

The iNTRODB was created through a partnership between Japanese institutes and the (Drugs for Neglected Diseases Initiative), a nonprofit research and development group for drugs. iNTRODB is the world's first integrated database for NTD drug discovery research. The four NTDs targeted in this database were Chagas disease, African trypanosomiasis, leishmaniasis and dengue fever (Namatame, 2016). It provides a variety of information, such as protein structure, potential inhibitors, and experimental assay results, to assist researchers in identifying and selecting promising

**Table 3.** Other important NTD databases.

| S.No. | Database name  | URL   |
|-------|--|---|
| 1.    | iNTRODB  | <a href="http://www.cb.cs.titech.ac.jp/introdb/">http://www.cb.cs.titech.ac.jp/introdb/</a>   |
| 2.    | virusITE—integrated database for viral genomics                | <a href="http://www.virusite.org/">http://www.virusite.org/</a>   |
| 3.    | VVR  | <a href="http://www.ncbi.nlm.nih.gov/genome/viruses/variation/">http://www.ncbi.nlm.nih.gov/genome/viruses/variation/</a>                         |
| 4.    | RVDB   | <a href="https://rvdb.dbi.udel.edu/">https://rvdb.dbi.udel.edu/</a>   |
| 5.    | PhEVER   | <a href="http://pbil.univ-lyon1.fr/databases/phever/help.html">http://pbil.univ-lyon1.fr/databases/phever/help.html</a>                           |
| 6.    | ViPR   | <a href="https://www.viprbrc.org/">https://www.viprbrc.org/</a>   |
| 7.    | IEDB   | <a href="https://www.iedb.org/">https://www.iedb.org/</a>   |
| 8.    | FungiDB  | <a href="https://fungidb.org/">https://fungidb.org/</a>   |
| 9.    | TriTrypDB  | <a href="https://tritrypdb.org">https://tritrypdb.org</a>   |
| 10.   | Database of food borne human pathogens                         | <a href="https://bioinfo.bisr.res.in/cgi-bin/project/foodpath/bacteria.cgi">https://bioinfo.bisr.res.in/cgi-bin/project/foodpath/bacteria.cgi</a> |
| 11.   | WormBase ParaSite  | <a href="https://parasite.wormbase.org/index.html">https://parasite.wormbase.org/index.html</a>   |
| 12.   | Parasite databases of clustered expressed sequence tags (ESTs) | <a href="http://www.cbil.upenn.edu/ParaDBs/">http://www.cbil.upenn.edu/ParaDBs/</a>   |
| 13.   | GeneDB   | <a href="https://www.genedb.org/">https://www.genedb.org/</a>   |
| 14.   | GtRNADB  | <a href="http://gtRNADB.ucsc.edu/">http://gtRNADB.ucsc.edu/</a>   |
| 15.   | KEGG database  | <a href="https://www.kegg.jp/">https://www.kegg.jp/</a> or <a href="https://www.genome.jp/kegg/">https://www.genome.jp/kegg/</a>                  |
| 16.   | MalaCards: the human disease database                          | <a href="http://www.malacards.org/">www.malacards.org/</a>  |
| 17.   | TDR targets  | <a href="http://tdrtargets.org/">http://tdrtargets.org/</a>   |
| 18.   | VEuPathDB  | <a href="https://veupathdb.org/">https://veupathdb.org/</a>   |
| 19.   | AlphaFold protein structure database                           | <a href="https://alphafold.ebi.ac.uk/">https://alphafold.ebi.ac.uk/</a>   |

therapeutic target proteins (Yasuo *et al.*, 2021). The iNTRODB can help advance drug development research for NTDs worldwide because it is open to all researchers. Each participating institute contributed to drug discovery by various approaches (Namatame, 2016).

#### **VirusSITE—integrated database for viral genomics**

VirusSITE is an integrated repository of viral genes and genomes that includes all genomes from viroids, viruses and satellites that have been published in the Reference Sequence Database (RefSeq) of NCBI (Je *et al.*, 2019; Stano *et al.*, 2016). It is a comprehensive database of viral genomics knowledge (Aris-Brosou *et al.*, 2019). The data were mined from Universal Protein Resource Knowledgebase, NCBI RefSeq, ViralZone, Gene Ontology and PubMed and integrated under human supervision. The database has details on the virus gene, genome characteristics, taxonomy, host range, sequential relatedness, protein characteristics and functions. The database entries are connected to several sources of information. The database also includes viruses that cause dengue, chikungunya, and rabies. The database is simple to use and has been built to offer a straightforward, user-friendly environment. It offers a taxonomy-based browsing system along with a key text searching mechanism. A sequence search-based alternate strategy is also provided by virusSITE. The genome browser visualizes viral genomes graphically. Users can retrieve and visualize data as well as execute comparative genomics analyses using a range of tools (Stano *et al.*, 2016).

#### **Virus variation resource (VVR)**

The VVR is an online comprehensive resource supported by the NCBI. It comprises modules for seven different viral groups: dengue, west Nile, influenza, Middle East respiratory syndrome coronavirus, ebolavirus, Zika, and rotavirus (Hatcher *et al.*, 2017). The resource includes a suite of sequence data visualizations as well as specialized search tools (Ibrahim *et al.*, 2018). Sequence annotation and database loading algorithms produce reliable protein and gene annotations, collect sequence attributes from sequence records, and map these attributes to a predetermined vocabulary. Sequences can be identified using a range of clinical and biological parameters using the web-based search tool of the database. Once the sequences have been identified, they may either be downloaded in a variety of formats or analyzed with a number of programs (Brister *et al.*, 2014). VVR initially focuses on influenza and dengue viruses. Its tool is versatile and can be used for other viruses in the future. Its data are managed via the relational database system Microsoft Structured Query Language Server 2005, which makes use of a straightforward schema for storing protein and nucleic acid sequences (Resch *et al.*, 2009).

#### **Reference viral database (RVDB)**

The RVDB is a publicly available database that enables users to analyze high-throughput sequencing data for the diagnosis of new emerging and known viruses, with the exception of bacteriophages (Bigot *et al.*, 2020). The database also includes viruses that cause dengue, chikungunya, and rabies. RVDB was created at the Food and Drug Administration's Center for Biologics Evaluation and Research. The RVDB contains partial and complete sequences as well as endogenous retroviruses, endogenous

nonretroviral components, and retrotransposons from a variety of viral families. The reduction of cellular sequences is an advantage of RVDB; it improves the efficiency with which transcriptomic and genomic data are processed, increasing detection specificity. RVDB provides data in the form of nucleotide sequence files that are clustered (C-) or unclustered (U-). The database is regularly updated to include newly uploaded viral sequences from GenBank (Goodacre *et al.*, 2018). An RVDB-prot (proteic version of the nucleic acid reference virus database) is available to support analysis. Marc Eloit's group at Institut Pasteur created this version. RVDB-prot seeks to provide precise, reliable, and distinctly annotated entries in addition to a repository of (Hidden Markov Model) protein profiles for distant protein searches (Bigot *et al.*, 2020).

#### **Phylogenetic exploration of virus evolutionary relationship (PhEVER)**

PhEVER is a one-of-a-kind web-based database of homologous families that provides precise evolutionary and phylogenetic information to help researchers better understand virus-host and virus-virus lateral gene transfers. It combines data from nonredundant viral genomes (2,426), nonredundant prokaryotic genomes (1,007), and eukaryotic genomes (43) to generate protein clusters, alignments, and phylogenies containing at least one viral sequence (Palmeira *et al.*, 2011). The database also includes viruses that cause dengue fever, chikungunya, and rabies.

#### **Virus pathogen database and analysis resource (ViPR)**

ViPR is a freely accessible comprehensive repository with integrated analysis resources of various virus families. The (National Institute of Allergy and Infectious Diseases) maintains the ViPR (Bukhari *et al.*, 2022). ViPR offers tools for searching, visualizing, analyzing, saving, and sharing data for key viruses of biodefense importance as well as other viruses that cause infectious diseases. ViPR is one of five (Bioinformatics Resource Centers) maintained by the NIH (National Institutes of Health). The ViPR database contains genomes from the families Flaviviridae, Coronaviridae, Arenaviridae, Herpesviridae, Poxviridae, Bunyaviridae, Reoviridae, Caliciviridae, Filoviridae, Hepeviridae, Paramyxoviridae, Togaviridae, and Picornaviridae (Pickett *et al.*, 2012). ViPR facilitates bioinformatics processes for a variety of human viruses and associated viruses, including the whole Coronaviridae family. ViPR serves as a platform for accessing sequence information, protein 3D structures, protein and associated gene annotations, immunological epitopes, host factors, and other data types through an integrative web-based search resource. The results of the searches can then be put through web-based studies such as phylogenetic inference, (multiple sequence alignment), BLAST comparison, determination of sequence variation, and statistical comparative genomics analysis. ViPR tools are freely available and assist in the development of diagnostics, prevention measures, vaccinations and therapies for human diseases (Vita *et al.*, 2019).

#### **Immune epitope database (IEDB)**

IEDB is a web-based database that groups experimental data on epitopes, antibodies, and Major Histocompatibility

Complex (MHC) binding studies in nonhuman primates, humans, and other animal species responsible for allergies, communicable disease, autoimmune, and related diseases. It is openly accessible and simple to search. Additionally, the IEDB includes resources for the prediction and analysis of epitopes (Dhanda *et al.*, 2019). Users can also utilize a query Application Programming Interface to perform most of the inquiries provided on the IEDB home page and work with the data in their own context. The database is regularly updated with improved query and reporting features to meet the needs of users by summarizing data that are becoming more complex and large in volume (Zaib *et al.*, 2023). To date, this database contains 1,481,848 peptidic epitopes, 3,133 nonpeptidic epitopes, 439,003 T-cell assays, 1,323,282 B-cell assays, 4,391,152 MHC ligand assays, 4,186 epitope source organisms and 963 restricting MHC alleles (Vita *et al.*, 2019). The database also incorporates viruses that are responsible for dengue, chikungunya and rabies.

### FungiDB

FungiDB is an online web-based integrated database for functional genomics analysis and data mining of oomycete and fungal species (Basenko *et al.*, 2018). The database also contains fungi that are responsible for mycetoma, chromoblastomycosis, and other deep mycoses NTDs. It is freely available at [fungidb.org](http://fungidb.org). One of the largest parts of the Eukaryotic Pathogen Database (EuPathDB) platform, FungiDB, combines different forms of data with phenotypic, genomic, proteomic, and transcriptome datasets for free-living, parasitic, pathogenic, and nonpathogenic species. The genomes in FungiDB were collected from sources such as Aspergillus Genome Database, GenBank, Joint Genome Institute, Broad Institute, Ensembl, and others (Bertuzzi *et al.*, 2021). Early in 2011, EuPathDB and Stajich's *et al.* (2012) collaborated to establish FungiDB; nevertheless, by the end of 2015, FungiDB was incorporated into the EuPathDB bioinformatic resource center. FungiDB uses the same user interface and infrastructure as EuPathDB, enabling key and integrated searches through a simple graphical interface. FungiDB features a straightforward user interface and integrated bioinformatics tools for customized in silico research. Users can also create custom pipelines for large-scale data analysis using a Galaxy-based workspace. FungiDB also integrates cell cycle hyphal growth RNA-sequence data, microarray data and two hybrid interaction data. FungiDB Data provides a powerful resource for in silico experimentation by integrating sequence and functional annotation data with additional data from the FungiDB standard analysis pipeline and assigning weight to the orthology (Stajich *et al.*, 2012).

### TriTrypDB

TriTrypDB is a free web-based repository of unified functional genomic and genomics database of Trypanosomatidae pathogens. The database accommodates organisms from both the *Leishmania* and *Trypanosoma* genera. TriTrypDB was developed as a result of a collaborative project between the Seattle Biomedical Research Institute, EuPathDB database, and GeneDB database (Shanmugasundram *et al.*, 2023). The objective of this database is to employ computational infrastructure to incorporate genome annotation and analysis from GeneDB and other sources with a comprehensive range of functional genomics information

provided by collaborators of the international research community, frequently before publication. Later, TriTrypDB incorporated *L. infantum*, *L. braziliensis*, *Leishmania tarentolae*, *L. major*, *Trypanosoma cruzi*, and *T. brucei* datasets. Users can explore chromosomal regions or specific genes in their genomic context (Aslett *et al.*, 2010). This database is accessible from the URL: <https://tritrypdb.org>.

### Database of food borne human pathogens

This is a Web-based database of unique probes and primers for detecting food-borne human pathogens. Data are collected using molecular technology such as immunological detection, adenosine triphosphate bioluminescence, in situ hybridization, biosensors, polymerase chain reaction -based genotyping, and pyrosequencing, which is faster and more specific. The database includes 76 pathogens that are food-borne and have been connected to a number of human diseases. The database contains information on bacteria (26), fungi (13), protozoa (5), viruses (9), and helminths (23). This was developed to solve an issue with conventional techniques for identifying food-borne infections, such as nutrient media culture followed by various biochemical testing. Conventional techniques are less specific, labor-intensive, and more time consuming. The Bioinformatics Center at the Birla Institute of Scientific Research created the database. This database is accessible from the URL <https://bioinfo.bisr.res.in/cgi-bin/project/foodpath/bacteria.cgi>.

### WormBase parasite

WormBase ParaSite is an integrated repository of flatworm (helminths) genomes and parasitic nematodes. It contains gene expression analysis, gene comparison analysis, and uniform functional annotation for more than 150 different helminth species (Doyle, 2022). The database offers a variety of ways to access the data, including bulk downloads, a query wizard, programmatic interfaces, gene and genome summary pages, a choice of genome browsers, text and sequence searching.

The database includes helminths responsible for NTDs such as dracunculiasis, onchocerciasis, echinococcosis, lymphatic filariasis, schistosomiasis, taeniasis/cysticercosis, food-borne trematodiasis, food-borne helminthiasis, and soil-transmitted helminthiasis. WormBase ParaSite is structured in partnership with WormBase and Ensembl Genomes (Bolt *et al.*, 2018). It is free to use and funded by the UK Biotechnology and Biological Sciences Research Council.

### Parasite databases of clustered ESTs

Parasite Databases of Clustered ESTs is an umbrella of integrated web-based databases that store consensus sequences that were generated by screening organism-oriented ESTs from dbEST, stripping out polyA sequences from the ends and splicing of 5' and 3' regions with more than 25% Ns in a window size of 20 base pairs. The resultant quality sequences were then aligned using the cap2 tool. The database delivers useful analyses and information to the research community without curating the outcomes. To date, seven databases exist under the parasite database. Out of seven databases, four (*Brugia malayi*, *S. mansoni*, *T. cruzi* and *T. brucei*) correspond to organisms responsible for NTDs. This database is accessible from the URL: <http://www.cbil.upenn.edu/ParaDBs/>.

## GeneDB

GeneDB is a generic web-based database resource built to curate and annotate sequencing data for short genomes along with a wide range of genomic and proteomic data (Hertz-Fowler and Peacock, 2002). The database contains the genomes of pathogens that are prokaryotic, eukaryotic and organisms that are closely related. The resource offers a platform for genome annotation and sequencing, which was initially generated by the Pathogen Genomics group at the Wellcome Trust Sanger Institute and other cooperating sequencing labs (Logan-Klumpler *et al.*, 2012; Manske *et al.*, 2019). It combines information from ESTs and genomic studies with curated annotations that can be found using a single web resource and combined with search, sort, and download functionality (Hertz-Fowler *et al.*, 2004). The database also incorporates fungi, trypanosomes and worm parasite organisms that are responsible for NTDs such as Chagas disease, giardiasis, mycetoma, chromoblastomycosis, mycoses, human African trypanosomiasis, and leishmaniasis.

## Genomic tRNA database (GtRNAdb)

The GtRNAdb serves as a platform for predictions of tRNA genes using the program tRNAscan-SE on complete or partial genomes (Lowe and Eddy, 1997; Marygold *et al.*, 2022). GtRNAdb accommodates more than 367,000 tRNAs from the genomes of 4,032 bacteria, 155 eukaryotes and 184 archaea. The database can be queried by gene and characteristics of a sequence. GtRNAdb provides a thorough overview of all tRNAs, their primary sequences and predicted secondary structures, identification information of tRNAscan-SE display, multigene alignments and direct links to the Archaeal/Microbial Genome Browser and University of California Santa Cruz genome browsers and reveals genomic context-relevant information about tRNA genes. Additionally, the GtRNAdb BLAST server allowed the study of tRNA gene similarity across all sequences present in the GtRNAdb (Chan and Lowe, 2016).

## KEGG DISEASE DATABASE

KEGG is a web-based resource/database that contains data on chemicals, genomics and a category of health-related data that is specific to humans (Kanehisa *et al.*, 2021, 2023). It is used to analyze high-throughput biological data, including genomic sequences (Kanehisa *et al.*, 2019). The database covers all NTDs except snakebite envenoming. The genome information is stored in the GENES database, which includes gene libraries of partially and fully sequenced genomes with updated gene function annotations (Kanehisa *et al.*, 2010). The primary objective of the KEGG database project is to attribute functional values to genes and genomes at both molecular and higher levels. The KEGG Orthology (KO) database stores molecular-level functions, with each KO being a functional ortholog of a gene or protein. Networks of KEGG modules, BRITE hierarchies, and KEGG pathway maps serve as representations of higher-level functions (Kanehisa *et al.*, 2017; Kibbe *et al.*, 2015). The majority of diseases affecting humans are complicated multifactorial diseases caused by the interaction of many hereditary and environmental variables. In the KEGG database, drugs are seen as perturbants of the molecular system, and diseases are seen as perturbed states. Molecule/gene lists and pathway maps are two computational forms for disease information (Kanehisa *et al.*, 2010). The DISEASE database

contains an assortment of disease entries, each of which includes a list of the associated genes, pathogens for infections, carcinogens and other environmental factors. The DRUG database contains a thorough compilation of data on drug targets, approved drugs and drug metabolism, the latter of which is subsequently classified into metabolizing transporters and enzymes as well as by substrates, inhibitors and inducers. To comprehend disrupted molecular networks, the DRUG and DISEASE databases are combined with other molecular network and PATHWAY databases (Kanehisa *et al.*, 2010). Each entry is designated by an H number and includes a list of known environmental factors, infections, environmental factors, genetic factors (disease genes), and therapeutic drugs.

The KEGG DISEASE entry merely provides membership data, but it may also represent the underlying molecular network. Rather than drawing altered route maps for single-gene disorders, the disease entries link the causal genes to the normal pathway maps (Kanehisa *et al.*, 2019).

## MalaCards: the human disease database

MalaCards is a web-based database resource of human diseases and associated annotations (Wu *et al.*, 2022). Its design and development is based on the architecture and technique of the GeneCards database. It covers all NTDs. MalaCards categorizes diseases into disease cards, with each card containing a list of numerous known aliases for each disease and prioritized information as well as a wide range of connections and annotations between diseases. There are annotations for related diseases, genes, medications, publications, clinical trials, symptoms, and more. The MalaCards disease database includes rare diseases, inherited diseases, complex disorders, and other ailments in addition to general and specialist disease categories (Rappaport *et al.*, 2014, 2017). MalaCards divides diseases into two separate categories: global and anatomical. With over 18,000 distinct diseases, the global disease division is divided into six categories: infectious, cancer, fetal, metabolic, genetic, and rare diseases. More than 26,000 disorders encompassing all bodily regions, including bone, blood, muscle, immunological, and reproductive diseases, are organized into 18 broad categories under the division of anatomical diseases. The names of the diseases were gathered from primary and secondary disease databases. Overall, the database contains more than 19,000 disease entries that map to over 13,000 genes from over 70 well-known databases and websites. Moreover, there are approximately 13,000 unique entries in the database. MalaCards makes use of a variety of language systems, medical classification methods, and related ontologies. Genetic variations, genetic information and disease-related associated information are also included along with mapped tables, matrices and annotations (Espe, 2018).

## Tropical disease research (TDR) targets

The TDR Targets Database was created as an online resource that can be used by researchers to access information on selected targets as well as a tool for prioritizing targets in whole genomes with an emphasis on NTDs causing pathogens (Magariños *et al.*, 2012). Researchers can quickly prioritize genes of interest using the TDR targets database by executing simple queries such as searching for small proteins and enzymes with reliable quality 3D models by allocating numerical weights to each query and integrating the outcomes to produce a prioritized list of potential



targets (Landaburu *et al.*, 2020). The TDR Targets database is a resource that compiles all necessary data for drug target prioritization for multiple diseases in one place (Agüero *et al.*, 2008; Pollastri and Campbell, 2011). The TDR Target database provides tools for searching therapeutic targets in *Mycobacterium tuberculosis* (tuberculosis pathogen), *M. leprae* (leprosy pathogen), *Plasmodium falciparum* and *Plasmodium vivax* (malaria parasite). Additionally, the database contains information on the genomes of *L. major* (leishmaniasis pathogen), *Toxoplasma gondii* (toxoplasmosis pathogen), *T. cruzi* (American trypanosomiasis or Chagas disease pathogen), *T. brucei* (sleeping sickness pathogen or African trypanosomiasis), *B. malayi*, and *Wolbachia bancrofti* (filariasis pathogens). The database's name incorporates the acronym 'TDR,' which stands for Tropical Disease Research and was a WHO special program. The database incorporates specific genomic data of the pathogen along with functional data such as phylogeny, essentiality and expression data obtained from a variety of sources, including literature mining (Agüero *et al.*, 2008).

### VEuPathDB

VEuPathDB is a eukaryotic pathogen, vector, and host informatics resource created by combining the VectorBase and EuPathDB databases (Giraldo-Calderón *et al.*, 2022). The database also contains information on organisms that cause NTDs, such as *Leishmania* species and *T. cruzi*, which cause leishmaniasis and Chagas disease, respectively. The database is funded by the Wellcome Trust and the NIH. VEuPathDB integrates over 1,700 preanalyzed datasets and accompanying metadata with over 500 different organisms and offers cutting-edge search functionality, visualization, and analytical tools. The database's creation aims to provide researchers with access to omics data and bioinformatics studies. A proprietary OrthoMCL algorithm for predicting orthology is used in the analysis of various data sources using standardized procedures. This innovative data mining tool simplifies the comparison of datasets, data types, and species (Amos *et al.*, 2022). This database is freely accessible from the URL: <https://veupathdb.org/>.

### AlphaFold protein structure database

The AlphaFold database is a freely available database that contains approximately 992,316 predicted protein structures of the human proteome and other important proteins. This database currently contains predicted structures for 48 different organisms. Users can narrow their searches by protein name, gene name, UniProt accession number, or organism name. According to the January 2022 release, the database gains 27 additional organisms and almost 200,000 new protein structural predictions specific to pathogens responsible for NTDs and antibiotic resistance (Varadi *et al.*, 2022). Seventeen of the 27 new organisms are on the WHO's NTD list. The database was developed by DeepMind with EMBL (European Bioinformatics Institute of EMBL) to provide open access to protein structure models of interest to the research community. It is an artificial intelligence-based system that predicts a 3D structure of a protein based on its amino acid sequence. The database is regularly updated and maintains accuracy comparable to experiments (Borkakoti and Thornton, 2023).

### CONCLUSIONS AND FUTURE PROSPECTIVE

NTDs are a class of infectious diseases that are strongly linked to poverty. Many other poverty-related diseases, such as

Ebola virus, *Zika virus*, and SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), are not on the WHO list but must be monitored by governments and international organizations (Carabelli *et al.*, 2023; Giraldo *et al.*, 2023; Khubchandani *et al.*, 2020). Unfortunately, such diseases receive attention only when there is an outbreak, such as with SARS-CoV-2, which has received little attention since its discovery. Studies have revealed that symptoms from other coronaviruses are worse than those from SARS-CoV (Cheng *et al.*, 2007). This virus, which caused the Coronavirus disease 2019 pandemic, prompted a rush to develop new treatments and vaccines using various biological/chemical databases and computational methods. All NTDs and emerging diseases require greater attention to prevent widespread outbreaks and epidemics that could result in a high death toll on a worldwide scale.

### AUTHOR CONTRIBUTIONS

PKA: Data collection, compilation, drafting and art work of the manuscript. KB: Manuscript editing. AKS: Manuscript moderation. AK: Primary investigator of the work presented.

### FUNDING

There is no funding to report.

### CONFLICT OF INTEREST

The authors declare that there are no competing conflicts of interest.

### ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

### DATA AVAILABILITY

All data generated and analyzed are included in this research article.

### PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

### REFERENCES

- Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AE, Chen F, Crowther GJ. Genomic-scale prioritization of drug targets: the TDR targets database. *Nat Rev Drug Discov*, 2008; 7(11):900–7.
- Amos B, Aurrecochea C, Barba M, Barreto A, Basenko EY, Belnap R, Blevins AS, Böhme U, Brestelli J, Brunk BP, Caddick M. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res*, 2022; 50(D1):D898–911.
- Aris-Brosou S, Parent L, Ibeh N. Viral long-term evolutionary strategies favor stability over proliferation. *Viruses*, 2019; 11(8):677.
- Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*, 2010; 38(suppl\_1):D457–62.
- Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, Warrenfeltz S, Aurrecochea C, Stoeckert Jr CJ, Kissinger JC, Roos DS. FungiDB: an integrated bioinformatic resource for fungi and oomycetes. *J Fungi*, 2018; 4(1):39.
- Basher ARMA, McLaughlin RJ, Hallam SJ. Metabolic pathway inference using multi-label classification with rich pathway features. *PLOS Comput Biol*, 2020; 16(10):e1008174.

- Bertuzzi M, Van Rhijn N, Krappmann S, Bowyer P, Bromley MJ, Bignell EM. On the lineage of *Aspergillus fumigatus* isolates in common laboratory use. *Med Mycol*, 2021; 59(1):7–13.
- Bigot T, Temmam S, Pérot P, Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Research*, 2020; 8:530.
- Biswas N, Mandal L. Treatment strategies against selected common tropical parasitic diseases. In: Bagchi D, Das A, Downs BW (Eds.). *Viral, parasitic, bacterial, and fungal infections*, Academic Press, Cambridge, MA, pp 383–8, 2023.
- Bolt BJ, Rodgers FH, Shafie M, Kersey PJ, Berriman M, Howe KL. Using WormBase ParaSite: an integrated platform for exploring helminth genomic data. *Methods Mol Biol*, 2018; 1757:471–91.
- Borkakoti N, Thornton JM. AlphaFold2 protein structure prediction: implications for drug discovery. *Curr Opin Struct Biol*, 2023; 78:102526.
- Brisler JR, Bao Y, Zhdanov SA, Ostapchuck Y, Chetvernin V, Kiryutin B, Zaslavsky L, Kimelman M, Tatusova TA. Virus variation resource—recent updates and future directions. *Nucleic Acids Res*, 2014; 42(D1):D660–5.
- Brodish PH, Singh K. Association between *Schistosoma haematobium* exposure and human immunodeficiency virus infection among females in Mozambique. *Am J Trop Med Hyg*, 2016; 94(5):1040.
- Brooker S, Hotez PJ, Bundy DAP. The global atlas of helminth infection: mapping the way forward in neglected tropical disease control. *PLoS Negl Trop Dis*, 2010; 4(7):e779.
- Bukhari SN, Webber J, Mehbodniya A. Decision tree based ensemble machine learning model for the prediction of *Zika virus* T-cell epitopes as potential vaccine candidates. *Sci Rep*, 2022; 12(1):7810.
- Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, COVID-19 Genomics UK Consortium, Peacock SJ, Barclay WS, de Silva TI, Towers GJ, Robertson DL. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol*, 2023; 18:1–6.
- Casulli A. New global targets for NTDs in the WHO roadmap 2021–2030. *PLoS Negl Trop Dis*, 2021; 15(5):e0009373.
- Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res*, 2016; 44(D1):D184–9.
- Cheng VC, Lau SK, Woo PC, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin Microbiol Rev*, 2007; 20(4):660–94.
- Chukualim B, Peters N, Fowler C, Berriman M. TrypanoCyc—a metabolic pathway database for *Trypanosoma brucei*. *BMC Bioinform*, 2008; 9:P5.
- Conteh L, Engels T, Molyneux DH. Socioeconomic aspects of neglected tropical diseases. *Lancet*, 2010; 375(9710):239–47.
- Das A, Biswas N, Chakrabarti S. Leish-Exp: a database of exclusive proteins from Leishmania parasite. *bioRxiv*, 2020:2020–05.
- Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, Sette A. IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Res*, 2019; 47(W1):W502–6.
- Doumenge JP, Mott KE, Cheung C. Atlas of the global distribution of schistosomiasis continued. Presses Universitaires de Bordeaux, Bordeaux, France, 1987.
- Doss CGP, Nagasundaram N, Srajan J, Chiranjib C. LSHGD: a database for human leprosy susceptible genes. *Genomics*, 2012; 100(3):162–6.
- Doyle SR. Improving helminth genome resources in the post-genomic era. *Trends Parasitol*, 2022; 38(10):831–40.
- Espe S. MalaCards: the human disease database. *J Med Libr Assoc*, 2018; 106(1):140.
- Ezeh CO, Onyekwelu KC, Akinwale OP, Shan L, Wei H. Urinary schistosomiasis in Nigeria: a 50 year review of prevalence, distribution and disease burden. *Parasite*, 2019; 26:19.
- Fall F, Mamede L, Schioppa L, Ledoux A, De Tullio P, Michels P, Frédéric M, Quetin-Leclercq J. *Trypanosoma brucei*: metabolomics for analysis of cellular metabolism and drug discovery. *Metabolomics*, 2022; 18(4):20.
- Ferreira AF, Heukelbach J, Costa CH, Souza EA, Maciel AM, Correia D, Ramos Jr AN. Scientometric review of research on neglected tropical diseases: a 31-year perspective from the journal of the Brazilian society of tropical medicine. *Rev Soc Bras Med Trop*, 2023; 56:e0403–2022.
- Gallo K, Mikhailov A, Hailemeskal MB, Koporc K, Mbabazi PS, Addiss D. Contributions of non-governmental organizations to WHO targets for control of soil-transmitted helminthiasis. *Am J Trop Med Hyg*, 2013; 89(6):1186.
- Getaz L, Silva-Santos D, Wolff H, Vitoria M, Serre-Delcor N, Lozano-Becerra JC, Chappuis F, Albajar-Viñas P. Persistent infectious and tropical diseases in immigrant correctional populations. *Rev Espanola Sanid Penit*, 2016; 18(2):57–66.
- Giraldo MI, Gonzalez-Orozco M, Rajsbaum R. Pathogenesis of *Zika Virus* infection. *Annu Rev Pathol Mech Dis*, 2023; 18:181–203.
- Giraldo-Calderón GI, Harb OS, Kelly SA, Rund SS, Roos DS, McDowell MA. VectorBase.org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms. *Curr Opin Insect Sci*, 2022; 50:100860.
- Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS. A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *MSphere*, 2018; 3(2):e00069–18.
- Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brisler JR. Virus variation resource—improved response to emergent viral outbreaks. *Nucleic Acids Res*, 2017; 45(D1):D482–90.
- Hertz-Fowler C, Peacock CS. Introducing geneDB: a generic database. *Trends Parasitol*, 2002; 18(10):465–7.
- Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, Parkhill J. GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res*, 2004; 32:D339–43.
- Hossain MM, Haque A, Mazid SZ, Khan A, Ullah TR, Rumeen ST. Snake venom database (SVDB) a potential resource for complementary & alternative medicine and drug designing. *Proceedings of the 2018 2nd International Conference on Computational Biology and Bioinformatics—ICCB 2018*, ACM Press, Bari, Italy, 2018, pp 32–6.
- Hotez PJ. NTDs V. 2.0: “blue marble health”—neglected tropical disease control and elimination in a shifting health policy landscape. *PLoS Negl Trop Dis*, 2013; 7(11):e2570.
- Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, Sachs JD, Savioli L. Control of neglected tropical diseases. *N Engl J Med*, 2007; 357(10):1018–27.
- Hürlimann E, Schur N, Boutsika K, Stensgaard AS, Laserna de Himpel M, Ziegelbauer K, Laizer N, Camenzind L, Di Pasquale A, Ekpo UF, Simoonga C. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *PLoS Negl Trop Dis*, 2011; 5(12):e1404.
- Ibrahim B, McMahon DP, Hufsky F, Beer M, Deng L, Le Mercier P, Palmarini M, Thiel V, Marz M. A new era of virus bioinformatics. *Virus Res*, 2018; 251:8–90.
- Je M, Son HS, Kim H. Data-processing pipeline and database design for integrated analysis of mycoviruses. *Int J Adv Smart Converg*, 2019; 8(3):115–22.
- Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, 2021; 49(D1):D545–51.
- Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*, 2023; 51(D1):D587–92.
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, 2017; 45(D1):D353–61.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*, 2010; 38:D355–60.
- Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*, 2019; 47(D1):D590–5.
- Karyala P, Metri R, Bathula C, Yelamanchi SK, Sahoo L, Arjunan S, Sastri NP, Chandra N. DenHunt—a comprehensive database of the intricate network of dengue-human interactions. *PLoS Negl Trop Dis*, 2016; 10(9):e0004965.

- Khubchandani J, Jordan TR, Yang YT. Ebola, Zika, Corona... what is next for our world? *Int J Environ Res Public Health*, 2020; 17(9):3171.
- Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res*, 2015; 43(D1):D1071–8.
- Kokaliaris C, Garba A, Matuska M, Bronzan RN, Colley DG, Dorkenoo AM, Ekpo UF, Fleming FM, French MD, Kabore A, Mbonigaba JB. Effect of preventive chemotherapy with praziquantel on schistosomiasis among school-aged children in sub-Saharan Africa: a spatiotemporal modelling study. *Lancet Infect Dis*, 2022; 22(1):136–49.
- Kumapley RS, Kupka R, Dalmiya N. The role of child health days in the attainment of global deworming coverage targets among preschool-age children. *PLoS Negl Trop Dis*, 2015; 9(11):e0004206.
- Landaburu LU, Berenstein AJ, Videla S, Maru P, Shanmugam D, Chernomoretz A, Agüero F. TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Res*, 2020; 48(D1):D992–1005.
- Lemoine JF, Desormeaux AM, Monestime F, Fayette CR, Desir L, Direny AN, Carciunoiu S, Miller L, Knipes A, Lammie P, Smith P. Controlling neglected tropical diseases (NTDs) in Haiti: implementation strategies and evidence of their success. *PLoS Negl Trop Dis*, 2016; 10(10):e0004954.
- Lenk EJ, Redekop WK, Luyendijk M, Rijnsburger AJ, Severens JL. Productivity loss related to neglected tropical diseases eligible for preventive chemotherapy: a systematic literature review. *PLoS Negl Trop Dis*, 2016; 10(2):e0004397.
- Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I. GeneDB—an annotation database for pathogens. *Nucleic Acids Res*, 2012; 40(D1):D98–108.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 1997; 25(5):955–64.
- Magariños MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, Van Voorhis WC, Agüero F. TDR targets: a chemogenomics resource for neglected diseases. *Nucleic Acids Res*, 2012; 40(D1):D1118–27.
- Maladan Y, Krismawati H, Agus R, Hutapea HM, Tanjung R, Cahyani VD, Rokhmad MF, Parikesit AA. Mutation of gyra gene found in *Mycobacterium Leprae* from leprosy patient in West Papua and Papua, Indonesia. *Indones J Pharm*, 2021; 32:35–42.
- Manske M, Böhme U, Püthe C, Berriman M. GeneDB and wikidata. *Wellcome Open Res*, 2019; 14:4.
- Marygold SJ, Chan PP, Lowe TM. Systematic identification of tRNA genes in *Drosophila melanogaster*. *MicroPubl Biol*, 2022; 2022:10.17912/micropub.biology.000560.
- Mitra AK, Mawson AR. Neglected tropical diseases: epidemiology and global burden. *Trop Med Infect Dis*, 2017; 2(3):36.
- Mohanty PS, Saikia D, Kalra S, Naaz F, Bansal AK, Pawar HS, Mohanty KK, Sharma S, Singh M, Patil SA. LEPStr: a database for *Mycobacterium leprae* short tandem repeats. *Inform Med Unlocked*, 2020; 19:100322.
- Mukherjee S. The United States food and drug administration (FDA) regulatory response to combat neglected tropical diseases (NTDs): a review. *PLoS Negl Trop Dis*, 2023; 17(1):e0011010.
- Namatame I. Activity of NTDs drug-discovery research consortium. *YAKUGAKU ZASSHI*, 2016; 136(2):231–6.
- Palmeira L, Penel S, Lotteau V, Rabourdin-Combe C, Gautier C. PhEVER: a database for the global exploration of virus–host evolutionary relationships. *Nucleic Acids Res*, 2011; 39:D569–75.
- Patel P, Mandlik V, Singh S. LmSmdB: an integrated database for metabolic and gene regulatory network in *Leishmania major* and *Schistosoma mansoni*. *Genomics Data*, 2016; 7:115–8.
- Pickett BE, Greer DS, Zhang Y, Stewart L, Zhou L, Sun G, Gu Z, Kumar S, Zaremba S, Larsen CN, Jen W. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses*, 2012; 4(11):3209–26.
- Pollastri MP, Campbell RK. Target repurposing for neglected diseases. *Future Med Chem*, 2011; 3(10):1307–15.
- Prakash DN, Routaray CB, Bhor R, Pai K. Recent advances in microbial databases with special reference to kinetoplastids. In: Gunjal A, Shinde S (Eds.). *Microbial diversity in hotspots*, Academic Press, Cambridge, MA, pp 291–316, 2022.
- Putman T, Hybiske K, Jow D, Afrasiabi C, Lelong S, Cano MA, Stupp GS, Waagmeester A, Good BM, Wu C, Su AI. ChlamBase: a curated model organism database for the *Chlamydia* research community. *Database*, 2019; 2019:baz041.
- Rahman S, Anim N, Rume ST. EzPubMed: a user-friendly snapshot of NCBI PubMed repository. *Int J Inf Technol Decis Mak*, 2022; 4(2):21–6.
- Rappaport N, Twik M, Nativ N, Stelzer G, Bahir I, Stein TI, Safran M, Lancet D. MalaCards: a comprehensive automatically-mined database of human diseases. *Curr Protoc Bioinforma*, 2014; 47(1):1–24.
- Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, Gershoni M, Morrey CP, Safran M, Lancet D. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*, 2017; 45(D1):D877–87.
- Resch W, Zaslavsky L, Kiryutin B, Rozanov M, Bao Y, Tatusova TA. Virus variation resources at the national center for biotechnology information: dengue virus. *BMC Microbiol*, 2009; 9(1):1–7.
- Rust J, Clark A, Woodgate M, Koch C, Mohammed T, Steinmann P, Krentel A, Torres-Vitolas CA, Carlin A, Pavluck A. Innovate to eliminate: a prerequisite in NTD programmes. *Int Health*, 2022; 14:ii20–4.
- Schur N, Hürlimann E, Garba A, Traore MS, Ndir O, Ratard RC, Tchuem Tchuenté LA, Kristensen TK, Utzinger J, Vounatsou P. Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged ≤ 20 years in West Africa. *PLoS Negl Trop Dis*, 2011; 5(6):e1194.
- Shameer S, Logan-Klumpler FJ, Vinson F, Cottret L, Merlet B, Achcar F, Boshart M, Berriman M, Breitling R, Bringaud F, Bütikofer P. TrypanoCyc: a community-led biochemical pathways database for *Trypanosoma brucei*. *Nucleic Acids Res*, 2015; 43(D1):D637–44.
- Shanmugasundram A, Starns D, Böhme U, Amos B, Wilkinson PA, Harb OS, Warrenfeltz S, Kissinger JC, McDowell MA, Roos DS, Crouch K. TriTrypDB: an integrated functional genomics resource for kinetoplastida. *PLoS Negl Trop Dis*, 2023; 17(1):e0011058.
- Silva JV, Santos SD, Machini MT, Giarolla J. Neglected tropical diseases and infectious illnesses: potential targeted peptides employed as hits compounds in drug design. *J Drug Target*, 2021; 29(3):269–83.
- Soodeh O, Masoumeh M, Morteza O, Abbas A, Fatemeh KL, Najmeh Y. Venomics and antivenomics data: current and future perspective. *Arch Biotechnol Biomed*, 2021; 5:026–31.
- Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, Stoeckert Jr CJ. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res*, 2012; 40(D1):D675–81.
- Stano M, Beke G, Klucar L. viruSITE—integrated database for viral genomics. *Database*, 2016; 2016:baw162.
- Turner HC, Ottesen EA, Bradley MH. A refined and updated health impact assessment of the global programme to eliminate lymphatic filariasis (2000–2020). *Parasit Vectors*, 2022; 15(1):1–9.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*, 2022; 50(D1):D439–44.
- Vedithi SC, Malhotra S, Skwark MJ, Munir A, Acebrón-García-De-Eulate M, Waman VP, Alsulami A, Ascher DB, Blundell TL. HARP: a database of structural impacts of systematic missense mutations in drug targets of *Mycobacterium leprae*. *Comput Struct Biotechnol J*, 2020; 18:3692–704.
- Vijayakumar S, Kant V, Das P. LeishInDB: a web-accessible resource for small molecule inhibitors against *Leishmania* sp. *Acta Trop*, 2019; 190:375–9.
- Vijayakumar S, Ranjan R, Kumar R, Das P. TrypInDB: a searchable online resource of small molecule inhibitors against *Trypanosoma* sp. *Parasitol Int*, 2020; 78:102131.

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*, 2019; 47(D1):D339–43.

Wu Y, Zhu Y, Xie N, Wang H, Wang F, Zhou J, Qu F. A network pharmacology approach to explore active compounds and pharmacological mechanisms of a patented Chinese herbal medicine in the treatment of endometriosis. *PLOS ONE*, 2022; 17(2):e0263614.

Yajima A, Mikhailov A, Mbabazi PS, Gabrielli AF, Minchiotti S, Montresor A, Engels D. Preventive chemotherapy and transmission control (PCT) databank: a tool for planning, implementation and monitoring of integrated preventive chemotherapy for control of neglected tropical diseases. *Trans R Soc Trop Med Hyg*, 2012; 106(4):215–22.

Yasuo N, Ishida T, Sekijima M. Computer aided drug discovery review for infectious diseases with case study of anti-Chagas project. *Parasitol Int*, 2021; 83:102366.

Zaib S, Rana N, Hussain N, Alrbyawi H, Dera AA, Khan I, Khalid M, Khan A, Al-Harrasi A. Designing multi-epitope monkeypox virus-specific vaccine using immunoinformatics approach. *J Infect Public Health*, 2023; 16(1):107–16.

Zhang Y, Guo J, Gao Y, Li S, Pan T, Xu G. Dynamic transcriptome analyses reveal m6A regulated immune non-coding RNAs during dengue disease progression. *Heliyon*, 2023; 9(1):e12690.

**How to cite this article:**

Arya PK, Barik K, Singh AK, Kumar A. Databases and web resources for neglected tropical disease research. *J Appl Pharm Sci*, 2023; 13(08):043–054.