

RESEARCH ARTICLE

TriTrypDB: An integrated functional genomics resource for kinetoplastida

Achchuthan Shanmugasundram¹, David Starns¹, Ulrike Böhme¹, Beatrice Amos¹, Paul A. Wilkinson¹, Omar S. Harb², Susanne Warrenfeltz³, Jessica C. Kissinger³, Mary Ann McDowell⁴, David S. Roos², Kathryn Crouch^{5*}, Andrew R. Jones^{1*}

1 Department of Biochemistry and Systems Biology, Institute of Integrative, Systems and Molecular Biology, University of Liverpool, Liverpool, United Kingdom, **2** Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **3** Center for Tropical & Emerging Global Diseases, Department of Genetics, Institute of Bioinformatics, University of Georgia, Athens, Georgia, United States of America, **4** Department of Biological Sciences, Eck Institute for Global Health, University of Notre Dame, Notre Dame, Indiana, United States of America, **5** School of Infection and Immunity, University of Glasgow, Glasgow, United Kingdom

* Kathryn.Crouch@glasgow.ac.uk (KC); andrew.jones@liverpool.ac.uk (ARJ)



OPEN ACCESS

Citation: Shanmugasundram A, Starns D, Böhme U, Amos B, Wilkinson PA, Harb OS, et al. (2023) TriTrypDB: An integrated functional genomics resource for kinetoplastida. *PLoS Negl Trop Dis* 17(1): e0011058. <https://doi.org/10.1371/journal.pntd.0011058>

Editor: João L. Reis-Cunha, Universidade Federal de Minas Gerais, BRAZIL

Received: October 12, 2022

Accepted: December 23, 2022

Published: January 19, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pntd.0011058>

Copyright: © 2023 Shanmugasundram et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and the supporting database, <https://tritrypdb.org/>. All our bioinformatics tools

Abstract

Parasitic diseases caused by kinetoplastid parasites are a burden to public health throughout tropical and subtropical regions of the world. TriTrypDB (<https://tritrypdb.org>) is a free online resource for data mining of genomic and functional data from these kinetoplastid parasites and is part of the VEuPathDB Bioinformatics Resource Center (<https://veupathdb.org>). As of release 59, TriTrypDB hosts 83 kinetoplastid genomes, nine of which, including *Trypanosoma brucei brucei* TREU927, *Trypanosoma cruzi* CL Brener and *Leishmania major* Friedlin, undergo manual curation by integrating information from scientific publications, high-throughput assays and user submitted comments. TriTrypDB also integrates transcriptomic, proteomic, epigenomic, population-level and isolate data, functional information from genome-wide RNAi knock-down and fluorescent tagging, and results from automated bioinformatics analysis pipelines. TriTrypDB offers a user-friendly web interface embedded with a genome browser, search strategy system and bioinformatics tools to support custom *in silico* experiments that leverage integrated data. A Galaxy workspace enables users to analyze their private data (e.g., RNA-sequencing, variant calling, etc.) and explore their results privately in the context of publicly available information in the database. The recent addition of an annotation platform based on Apollo enables users to provide both functional and structural changes that will appear as ‘community annotations’ immediately and, pending curatorial review, will be integrated into the official genome annotation.

Author summary

Kinetoplastid parasites cause severe infections in humans including African sleeping sickness, Chagas disease and leishmaniasis, which are classified as neglected tropical diseases. With the advancement of sequencing technologies, more and more genome sequences are being generated and deposited in archival repositories. TriTrypDB (<https://tritrypdb.org>),

and resources are maintained in the VEuPathDB GitHub repository: <https://github.com/VEuPathDB>.

Funding: This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID, <https://www.niaid.nih.gov/>) — National Institutes of Health (NIH) Contract #75N93019C00077 to DSR, MAM & JCK and Wellcome Trust (<https://wellcome.org>) grants 218288/Z/19/Z and 212929/Z/18/Z to ARJ, KC, DSR & JCK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

a component database of the VEuPathDB resource (<https://veupathdb.org>) is a free data mining resource, which currently hosts a subset of these parasite genomes that are of clinical and scientific importance. TriTrypDB also integrates functional genome-scale datasets (e.g. transcript expression, protein expression, genetic variation data) and information predicted from automated bioinformatics pipelines and from manual curation. TriTrypDB provides a user-friendly web interface and a number of tools and functions for users to conduct *in silico* experiments to ask questions and generate hypotheses. Researchers can also contribute their expertise via the User Comments form and Apollo annotation platform, and utilize our cloud-based workspace to analyze their own data. TriTrypDB has been extensively used by the research community over the last decade and serves as a primary resource for communities working with these organisms.

Introduction

TriTrypDB (<https://tritrypdb.org>) is a component database of the Eukaryotic Pathogen, Vector and Host Informatics Resource (VEuPathDB, <https://veupathdb.org>) Bioinformatics Resource Center [1] and is supported by the US National Institutes of Allergy and Infectious Diseases (NIAID) [2] and the Wellcome Trust UK. TriTrypDB is a free online resource for data mining and multi-omic analysis of Kinetoplastid parasites. Beyond TriTrypDB, VEuPathDB also provides resources for other eukaryotic protist parasites (Apicomplexa, Amoeba, Giardia, etc.), fungi (both pathogens and non pathogens, <https://fungidb.org>) [3], vectors (arthropods and molluscs, <https://vectorbase.org>) [4], selected mammalian host data (<https://hostdb.org>), orthology determination and phylogenetic inference (<https://orthomcl.org>) [5], and clinical epidemiological (<https://clinepidb.org>) [6] and microbiome data (<https://microbiome.org>) [7].

As of release 59 (30th August 2022), TriTrypDB hosts 83 genomes from 36 kinetoplastid species. Although TriTrypDB predominantly hosts genomes of *Leishmania* and *Trypanosoma* species, it also includes genomes of species belonging to other clades in the Kinetoplastea class, such as *Angomonas*, *Crithidia* and *Leptomonas*. TriTrypDB integrates a wide range of other data types, including transcriptomic, protein expression, epigenomic and genetic variation data, phenotypes and experimental cellular localization data from fluorescent tagging. These data are obtained either from repositories, such as International Nucleotide Sequence Database Collaboration (INSDC) for genome assemblies or Sequence Read Archive (SRA) for functional sequencing data, or directly from providers. Several data types are analyzed by TriTrypDB and made available for computational interrogation. Processing is carried out using standard workflows and integrated using an ontology-driven framework to ensure data comparability across studies.

TriTrypDB provides an easy-to-use web interface with record pages that compile all data for entities such as genes, genomic sequences, single nucleotide polymorphisms (SNPs) and metabolic pathways, a genome browser for visualization of sequence data, publicly available bioinformatics tools for data analysis, a search strategy system for interrogation of pre-analyzed data, and a private Galaxy workspace [8] for analyzing primary data and examining it in the context of public data already loaded into TriTrypDB. Expert knowledge from the user community is captured in the form of User Comments and community annotations via Apollo [9] and this information is reviewed and incorporated to improve gene models and functions. Here we present a general overview of the current state of TriTrypDB, highlighting the major developments in the last decade and since the initial publication of this resource in 2010 [10].

Methods

Data integration

Integrated datasets. TriTrypDB release 59 hosts 83 genomes and 176 other functional datasets relating to *Trypanosoma*, *Leishmania* and a few other kinetoplastid species. New datasets and functionality are incorporated into TriTrypDB via bimonthly releases. Integrated datasets can be found on the datasets page under the ‘Data’ menu and those integrated in a recent release can be found on the news section (<https://tritrypdb.org/tritrypdb/app/static-content/TriTrypDB/news.html>). An overall trend of datasets (both genomes and other functional datasets) integrated between release 1.0 and release 59 is illustrated in Fig A in [S1 File](#).

Of these 83 genomes, 31 are from *Leishmania* and 43 are from *Trypanosoma* genera. TriTrypDB also has genomes of other parasitic kinetoplastids including, *Angomonas deanei* [11], *Blechnomonas ayalai*, *Crithidia fasciculata* [12], *Endotrypanum monterogei* [12], *Leptomonas pyrrocoris* [13], *Leptomonas seymouri* [14], *Paratrypanosoma confusum* [15], and *Porcisia her-tigi* [16,17] and free-living nonparasitic kinetoplastid *Bodo saltans*. Gene annotations are available for 66 of these sequences, while the remaining 17 are genome assemblies without annotations (Table 1). Of these 66 genome sequences with annotations, 36 are classified as ‘reference’ (or ‘representative’) sequences representing distinct species, while the remaining 30 are either additional strains of the already existing reference species or resequencing of already available strains.

The other omics data types available in TriTrypDB are listed in Table 1, which includes single-cell RNA-Seq datasets from *T. brucei* [19] that were integrated for the first time in release

Table 1. Summary of integrated datasets.

	Dataset Categories	Number of datasets
Genomics	Genome sequences and annotations	66
	Genome sequences only	17
Transcriptomics	Expressed sequence tags (ESTs)	1 ^a
	RNA-Seq	39
	Single-cell RNA-Seq ^b	2
	Microarrays	10
Proteomics (MS-based)	Identification	12
	Quantitative	10
	Organellar	16
	PTMs	13
Epigenomics	ChIP-ChIP	2
	ChIP-Seq	7
Variation	Sequence variation	28
	Array probes	1
Phenomics	Quantitative (high throughput RNAi-target sequencing)	1
	Curated	1
Sequence sites, features and motifs	Origins of replication	5
	BAC, PAC and Cosmid-end sequences	6
	Maxicircle annotations	2
	Additional gene predictions	1
	Transcript alignments from previous annotations	3

^aESTs are obtained from the dbEST database [18].

^bThis is a new data type integrated for the first time in release 59.

<https://doi.org/10.1371/journal.pntd.0011058.t001>

59. Other key data types integrated into TriTrypDB include cellular localization images, orthology profiles assigned with the OrthoMCL algorithm [5] and metabolic pathways. TriTrypDB hosts microscopy images and annotations from the TrypTag project, which aims to tag and determine the cellular localization of every protein encoded in *T. brucei* TREU927 genome [20]. Metabolic pathways are integrated from KEGG [21], MetaCyc [22], TrypanoCyc [23] and LeishCyc [24] and are represented using Cytoscape JS [25], an open source graph library.

Since release 59, TriTrypDB integrates protein structure predictions from AlphaFold [26,27], an artificial intelligence system created by DeepMind (<https://www.deepmind.com/>). The integrated AlphaFold predictions are the EMBL-EBI predictions, currently covering sequences from the UniProt reference proteome and these predictions will be automatically updated with new releases from AlphaFold in the future. These predictions are integrated in TriTrypDB by mapping TriTrypDB gene IDs to UniProt IDs when there is a corresponding entry in UniProt and by sequence similarity when TriTrypDB genes do not have an exact match in UniProt. These protein structures can be visualized via the 3D viewer in the gene pages. Gene or protein features are cross-referenced with external databases including Chemical Entities of Biological Interest (ChEBI) [28], Enzyme Nomenclature [29], Gene Ontology [30,31], PDB [32–34], IEDB [35] and NCBI Taxonomy [36] and information from these external resources are also integrated into TriTrypDB. These datasets can be downloaded or used with the site tools available under the ‘Tools’ menu.

Updated data processing. Genome sequences loaded into TriTrypDB are first obtained from an INSDC repository (<https://www.insdc.org>) and processed by European Bioinformatics Institute (EBI) workflows. During EBI processing, assembly core statistics are generated and DNA features are predicted using RepeatMasker (<https://www.repeatmasker.org>), DustMasker [37] and Tandem repeat finder [38]. Protein features are predicted with InterProScan [39], SignalP [40], TMHMM [41], Panther [42], and PSIPRED [43]. RNA features are predicted with Rfam [44], tRNAscan [45] and miRBase [46]. Genomic data from the EBI workflow are supplemented with data generated from in-house pipelines (<https://github.com/VEuPathDB>) including open reading frames (ORFs), EST alignments and synteny plots generated using Mercator and MAVID [47]. A small number of legacy genomes that were not submitted to INSDC repository by the genome providers are processed by in-house pipelines from VEuPathDB, although we no longer accept genome assemblies that are not submitted to an INSDC repository.

All DNA, RNA and protein datasets are aligned to their respective reference genome sequences. Alignments are used for downstream processing including variant calling, copy number variation analysis, and differential gene expression analysis. The EBI RNA-seq alignment pipeline uses HISAT2 [48] for alignment to the reference genome sequence and HTSeq-count [49] for counting reads aligned to genome features. This is followed by in-house processing to generate TPM-normalized data for plotting and fold-change queries, normalized bigwig files for visualization in JBrowse, and to run DESeq2 [50] analysis on all pairwise conditions for differential expression queries. SNP and copy number variation (CNV) analyses are conducted using VEuPathDB pipelines where Bowtie2 [51], Samtools [52] and VarScan [53] are utilized for calling SNPs and normalized coverage data is used to predict chromosome and gene-scale copy number variations. Functional data such as gene names and product descriptions are assigned using the Ensembl Xref pipeline that links functional annotation for proteins analyzed by Uniprot. Large scale parallel compute is conducted by EBI and the generated data are loaded into a relational database at VEuPathDB. More details on the architecture of the database and data processing pipelines can be found in Amos *et al.* [1].

Results

A user-friendly web interface

The VEuPathDB portal and all component sites including the TriTrypDB database share the same backend infrastructure and web interface. The VEuPathDB user interface is continually improved to provide convenient and consistent access to data, searches, tools and help information.

Homepage. The homepage (Fig 1) features a header that is present on all pages, a main panel, an expandable ‘News & Tweets’ section on the right, and a footer with clickable icons to access other VEuPathDB resources (Fig 1G). The center of the header includes a site search



Fig 1. Home page of TriTrypDB. (A) Header on all site pages that includes site search, menu bar providing access to all searches, data and tools, and links for social media, registration, login and user profile. (B) The recently implemented ‘My Organism Preferences’ filter. (C) The left hand search panel contains searches of different data in the database, organized into categories. (D) The overview of resources and tools section provides vignettes to help users get started on specific tools and resources of interest. (E) The expandable news and tweets section (collapsed by default) offers quick access to news releases and recent tweets. (F) Links to more detailed step-by-step instructional material. (G) Footer consisting of hyperlinked logos to other VEuPathDB resources in addition to the Gitter Community Chat button.

<https://doi.org/10.1371/journal.pntd.0011058.g001>

box (see below) and a ‘menu’ bar that appears below the site search that provides quick access to ‘My Strategies’, ‘Searches’, ‘Tools’, ‘My Workspace’, ‘Data’, ‘About’, ‘Help’ and ‘Contact Us’ sections (Fig 1A). Social media, login, registration and user profile links are displayed on the right corner of the header.

A newly implemented feature named ‘My Organism Preferences’ is found below the right corner of the header on all pages (Fig 1B). This function offers the users the option to filter menus to a subset of organisms of their interest. TriTrypDB will therefore function as a personalized database containing only their chosen organisms. This feature can easily be enabled or disabled at any time from the button located next to the ‘My Organism Preferences’ on any page.

The left sidebar of the main panel of the home page provides access to all available searches from the database (Fig 1C). Searches are organized into expandable categories and the users can also refine the list of searches based on key words using the filter above the menu (Fig 1C, red box). The top of the central portion of the main panel displays a list of scrollable vignette buttons, which when clicked provide useful information on various tools and resources available in TriTrypDB (Fig 1D). A section on ‘Tutorials and Exercises’ is displayed below the vignettes, which provides access to step-by-step tutorials downloadable in PDF format (Fig 1E). An expandable section on ‘News and Tweets’ to the right of the vignettes provides a quick exploration of the website news and recent tweets (Fig 1E).

Record pages. Gene records compile all the available information about a particular gene into a single web page. The information about a gene and its function that are available in the gene page are generated from integrated datasets and automated pipelines (See ‘Data integration’ above). These record pages are updated on a regular basis both to provide better user experience and to ensure the relevance of content displayed in the record pages.

TriTrypDB gene record pages can be navigated via the thumbnail ‘Shortcuts’ at the top and the collapsible ‘Contents’ menu on the left. The data are categorized into 19 different categories as displayed in the ‘Contents’ menu. Help icons (blue question mark icons) are available at multiple locations in the gene pages to help users find additional information where it is appropriate. Gene pages host a multitude of data including gene model information, functional annotation (see ‘Functional annotation’ section below), ortholog and paralog predictions from the OrthoMCL pipeline [5], experimental data including transcript expression, protein expression and phenotyping data (see ‘Integrated datasets’ section above), immune epitopes from IEDB database [35], link outs to external databases such as UniProtKB [54], TrypsNetDB [55] and PDB [32–34], link outs to relevant literature from PubMed, protein structure predictions from AlphaFold [26,27], and predictions of protein features and properties such as molecular weight, InterPro motifs [56], signal peptides and transmembrane domains.

TriTrypDB provides different types of data representations (e.g. tables, bar charts and plots) to facilitate better exploration and visualization of omics data on the gene pages, including a summary graph (using Plotly, <https://plotly.com>) of a gene’s expression values across all integrated RNA-Seq datasets. This representation is useful for observing overall trends in expression of a gene across experiments and to identify outliers that may require further exploration. Another example is CELLxGENE (<https://cellxgene.cziscience.com>), a single-cell visualization platform developed by the Chan-Zuckerberg initiative, which was implemented in release 59 to facilitate exploration of single-cell RNA-Seq data. TriTrypDB gene pages provide static images and links to explore the data dynamically in the CELLxGENE platform, enabling the user to select or paint groups of cells based on gene expression, experimentally derived metadata such as clusters or metadata derived from the experimental design such as condition or replicate. An example dataset can be viewed at https://tritypdb.org/cellxgene/view/tbruTREU927_briggs_wt_cellxgene_RSRC.h5ad/.

Similar to gene record pages, TriTrypDB also provides pages for other record types including popset isolate sequences, genomic sequences (scaffolds), genomic segments, SNPs, ESTs, metabolic pathways and chemical compounds. These record pages can be accessed by conducting searches (see Search strategy system section below) for the respective entity types other than genes (e.g. popset isolate sequences) and accessing the record IDs for the retrieved entities (Fig B in [S1 File](#)). Several of these record types (e.g. metabolic pathways) can also be accessed from the gene record pages via clicking links for the record IDs under the respective sections in gene pages. The record pages of these additional entity types are also organized in a similar fashion to gene pages with the data organized into multiple categories and the displayed data can be navigated via the collapsible content menu on the left.

Tools

Site search. The search bar present in the header of all TriTrypDB web pages ([Fig 1A](#)) allows users to perform a site-wide search with gene identifiers and free text. This search returns a categorized list of results with filters available for users to define categories or organisms of interest. The site search results include pages of datasets, news items and tutorials in addition to feature record pages (genes, pathways etc.). The site search results corresponding to records can be exported as a step in the search strategy system allowing further data exploration and download.

Search strategy system. The search strategy system available in TriTrypDB provides a unique and powerful mechanism to mine the vast amount of omics datasets and to integrate results in a multi-step '*in silico*' experiment ([Fig 2](#)). Multi-step search strategies are built one step at a time, choosing the first search from either the 'Search for. . .' on the home page ([Fig 1C](#)) or the 'Searches' menu on the header ([Fig 1A](#)). Searches are also available for other feature types (listed above in the 'Record pages' section) such as genomic segments, SNPs and pathways.

Search results are displayed in a newly designed 'My Search Strategies' page. The top graphic panel on this page displays growing search strategies ([Fig 2A](#)), and these search strategies can be extended by clicking 'Add a step' ([Fig 2A](#), red box). The options for adding steps include 'Combine' for use with similar records using union, intersection and minus operations, 'Transform' to find records cross-referenced to the current results (orthologs, metabolic pathways and compounds) and genomic colocation searches to find features that are related by their location in the genome (genes, genomic segments and SNPs) ([Fig 2C](#)). Search strategies can be made, copied, edited and deleted by any user. Users with a free account can additionally save strategies and share them with others using a private link. The users can add relevant details to the description section when saving a strategy, which will help them remember key details when they access it at a later time point. Saving a strategy retains the order of steps and parameter values, but not the actual results as subsequent database versions containing new data may alter the results.

The results table appears below the graphic panel ([Fig 2B](#)) and it includes a list of resulting feature IDs and associated data. Columns of associated data can be added to the results table via 'Add Columns' and results can be downloaded locally using the 'Download' option ([Fig 2B](#), red box). The recently implemented 'Send To' dropdown menu ([Fig 2B](#), red box) allows users to save their search results as an ID list in the 'My Data Sets' page, as a downloadable text file for future analyses, or to transfer the list from TriTrypDB to VEuPathDB and analyze in the context of all VEuPathDB organisms. By saving both search strategies and search results from those strategies, users can compare results from different versions of the database in the future. The collapsible organism filter that appears on the left of the results table ([Fig 2B](#), red arrow)

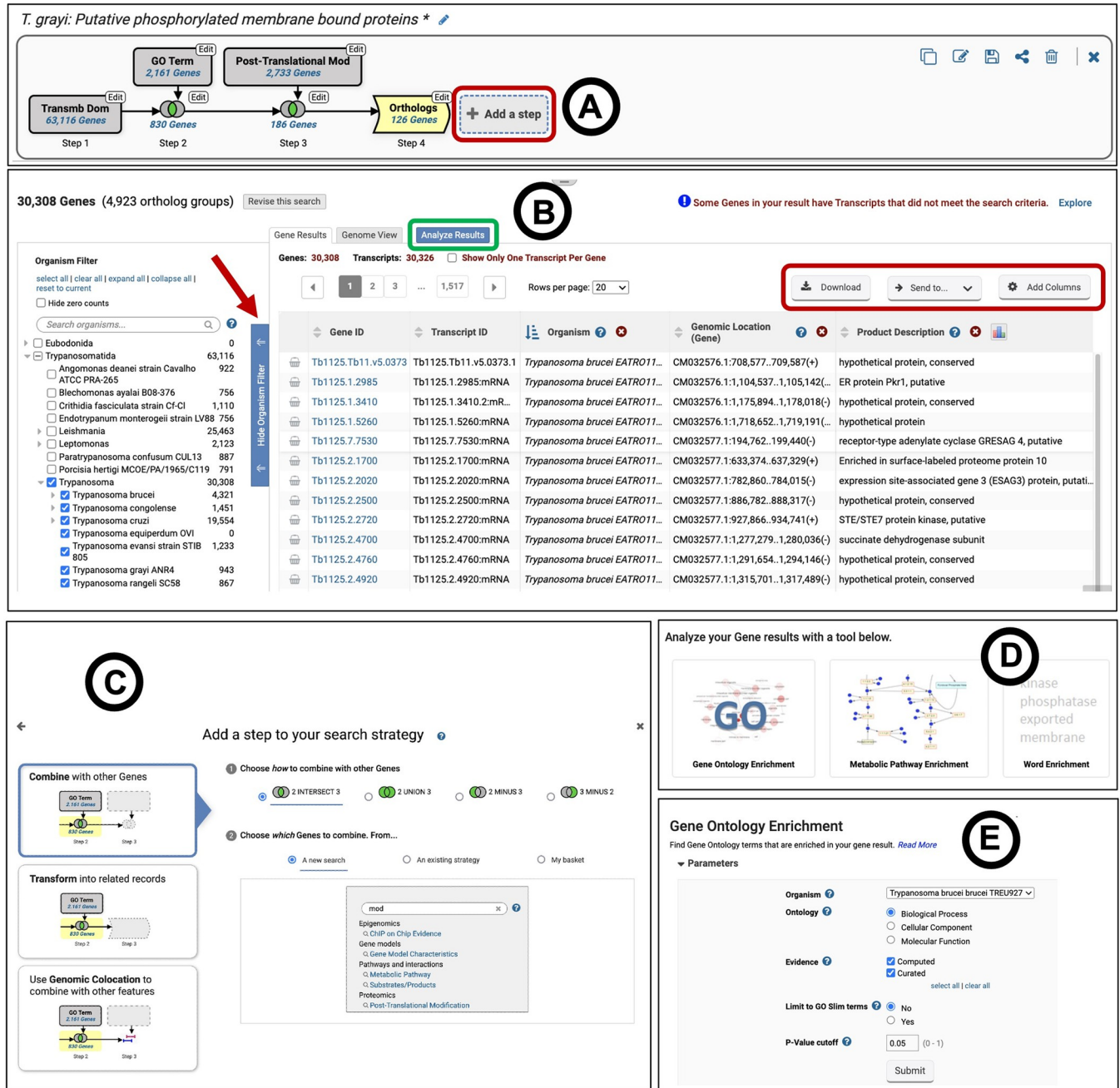


Fig 2. Search strategies as *in silico* experiments and functional enrichment analysis. (A) The graphic panel shows an example multi-step search strategy (<https://tritrypdb.org/tritrypdb/app/workspace/strategies/import/9c17640460e66cd5>). (B) Results of the first step of the search strategy, with the redesigned vertical organism filter on the left (red arrow) and the results table on right. (C) Redesigned ‘Add Step’ popup showing the three options to add steps, ‘Combine’, ‘Transform’ and ‘Genomic Collocation’, and the details of the chosen option. (D) The three options available for the analysis of gene results: GO enrichment, metabolic pathway enrichment and word cloud. (E) The form for selecting appropriate parameters for GO enrichment analysis.

<https://doi.org/10.1371/journal.pntd.0011058.g002>

shows the distribution of the results across the organisms searched and the results can be limited to organisms of interest using the filter.

Enrichment analysis. TriTrypDB provides tools for users to perform functional enrichment of gene results arising from search strategies or user-supplied gene lists (Fig 2B, highlighted in green box). Available tools include word, gene ontology (GO) [30,31] and

metabolic pathway enrichment analyses (Fig 2D). These tools use Fisher's exact test in combination with multiple test corrections implemented using both Bonferroni and Benjamini-Hochberg methodologies. Simple word enrichment is useful in detecting keywords that are enriched in annotations such as product descriptions, user comments and annotator notes when more formal GO and metabolic pathway annotations are not available for a gene. An option to limit the GO terms to the slim subset is available with the GO enrichment analysis to reduce the redundancy of enriched terms (Fig 2E). GO enrichment results can also be exported to REVIGO [57], which facilitates data visualization via a variety of interactive tools. For metabolic pathway enrichment analysis, TriTrypDB provides an option to choose pathways from KEGG [21] and MetaCyc [22].

Genome and protein browsers. TriTrypDB currently embeds the JBrowse genome browser [58] to facilitate the dynamic visualization of annotations and functional data on genome sequences. JBrowse is an open source and configurable platform that offers improved browsing and zooming speed and the ability to save and share personalized views. JBrowse enables users to select and display tracks with aligned transcriptomic, proteomic, epigenomic and variation data on the genome browser. Transmembrane domains (TMHMM predictions [41]), protein domains from InterPro [56] and synteny views across multiple genomes can be accessed via the protein browser.

Sequence analysis tools. The protein features and properties section of gene record pages currently provides direct access to six external bioinformatics tools for analysis of protein sequences. These are BLAST-P [59,60], InterProScan [39], big-PI predictor [61], MitoProt [62], STRING [63] and WoLF PSORT [64]. The users can submit a protein sequence for analysis by clicking a tool's 'Submit' button from the record page of the relevant gene, which opens a new tab and initiates the respective query on the external web interface of the chosen tool. MitoProt, WoLF PSORT and big-PI predictor are used for the prediction of mitochondrial targeting signal, subcellular localization site and glycosylphosphatidylinositol anchor respectively. STRING accesses visualizations of both known and predicted protein-protein interactions.

The Clustal Omega [65] tool is embedded in the orthologs and paralogs section of record pages. Users can align a gene sequence with the sequences of its homologs by selecting the genes from 'Orthologs and Paralogs within TriTrypDB' table, choosing the sequence type and the output format, and clicking the 'Run Clustal Omega for selected genes' button.

Galaxy interface. The VEuPathDB Galaxy [8] interface offers an environment for users to privately analyze their own data as well as data available at INSDC repositories. Users can either upload their own data or use the sample accession numbers from European Nucleotide Archive (ENA) or Sequence Read Archive (SRA) for data transfer into Galaxy. Individual samples can be placed into dataset collections for efficient organization and downstream processing. Currently, TriTrypDB provides preconfigured workflows for RNA-seq data (for identification of transcript expression from single and paired end stranded and non-stranded Illumina data and for differential expression analysis), variant calling and for mapping proteins to ortholog groups using the OrthoMCL algorithm [5]. There are a host of tools that are available to use individually or for users to create their own workflows using the built-in user interface. Analysis results can either be downloaded or exported into TriTrypDB for private exploration through custom searches, the search strategy system or genome browser visualizations.

Curation and annotation

Nine selected kinetoplastid genomes (four *Leishmania* and five *Trypanosoma*) that were sequenced by the Parasite Genomics group at Wellcome Sanger Institute and previously

hosted in GeneDB [66] were manually curated by expert curators from the VEuPathDB project by utilizing the curation infrastructure from the Wellcome Sanger Institute. This joint effort between GeneDB and TriTrypDB in improving these parasite genomes was made possible by continuous funding from the Wellcome Trust for the TriTrypDB database since 2012. The updated annotations are regularly integrated into TriTrypDB through bimonthly releases. As GeneDB was taken offline last year, TriTrypDB serves as the sole authoritative resource for the annotation of these genomes. TriTrypDB continues to curate these genomes and functional annotations were updated in releases from the last year.

Although structural and functional annotations are updated by VEuPathDB, improving genome assemblies with new data is beyond the scope of this project. However, VEuPathDB sources new and improved assemblies of already integrated genomes from INSDC repositories as detailed above. These new genomes can either replace already existing genomes or be integrated as additional genomes of already existing species/ strains in TriTrypDB, depending on the data quality or perceived community importance. VEuPathDB has also obtained permissions from the majority of the genome data providers to update annotations in INSDC repositories. Genomes with significant annotation updates (either from VEuPathDB curation or from community curation via Apollo) are planned to be updated in INSDC repositories in the future.

Structural annotation. Re-annotation efforts have resulted in sequence changes and gene model updates for the nine curated kinetoplastid genomes. These annotation changes include creation of new genes and obsolescence of existing genes including pseudogenes, addition of alternative transcripts, changes to existing gene models, transcripts and coding sequences, and changes to gene IDs (Table 2). The gene IDs remain stable despite changes to gene models in the majority of cases and the gene IDs are not versioned with structural annotation updates. However, all changes are tracked in the curation database and the changes are documented in the annotation logs associated with the news of our bimonthly releases. TriTrypDB also stores previous identifiers of genes with new gene IDs in the database and users can access the record pages of these genes with new IDs by searching with their previous IDs. In addition, putative gene models removed from official annotations due to lack of evidence for them to be real are made obsolete rather than being deleted and they can be reverted back and added to official annotations when new evidence becomes available. In addition all sequences and annotations from previous releases are made available in the downloads section and users can compare different versions of annotations and conduct analyses with these previous genome versions available in downloads, if required.

VEuPathDB annotation efforts are focussed on improving the structural annotations of these curated genomes in order to provide high quality data for these key parasite species. We

Table 2. Summary of structural annotation changes for curated organisms since the initial integration of genomes in TriTrypDB.

Organism	Coding genes			Pseudogenes		Changed Gene IDs	Total genes with updates
	New	Deleted	Changed	New	Deleted		
<i>Trypanosoma brucei</i> TREU927	405	568	613	165	5	121	1877
<i>Trypanosoma brucei</i> gambiense	0	1894	88	74	0	2	2058
<i>Trypanosoma congolense</i>	0	1556	68	85	0	3	1712
<i>Trypanosoma cruzi</i>	2	0	156	7	8	9	182
<i>Trypanosoma vivax</i>	1	462	324	340	0	2	1129
<i>Leishmania braziliensis</i>	49	1	25	16	18	4	113
<i>Leishmania donovani</i>	5	3	58	4	0	57	127
<i>Leishmania major</i>	120	1	1474	4	6	1	1606
<i>Leishmania mexicana</i>	14	2	12	4	0	2	34

<https://doi.org/10.1371/journal.pntd.0011058.t002>

rely on our data providers to supply genome assemblies and annotations. The improvement of genome assemblies and generation of first-pass genome annotations for genome assemblies without annotations are beyond the remit of the VEuPathDB project. However, VEuPathDB curators have also generated first-pass structural annotations for genomes relevant to our user community, when requested by them. An example is the *Leishmania amazonensis* genome [67], which was annotated by the VEuPathDB project using the Companion genome annotation pipeline (<https://companion.ac.uk/>) [68], an external tool that can be accessed from the ‘Tools’ section of TriTrypDB.

Functional annotation. Functional annotation dominates the curatorial efforts undertaken by the TriTrypDB project. Functional annotation attributes include gene names/ symbols, synonyms, product descriptions, GO annotations, EC numbers, annotator notes, literature citations, previous database identifiers and external database references. Between releases 1.0 and 59, a total of 79,242 genes had functional annotation updates and these include 8,334 gene names and synonyms, 37,995 product descriptions and 92,041 GO annotations. A summary of complete functional annotation changes made over the last decade can be found in Table 3.

GO annotations are curated using the standards developed by the Gene Ontology (GO) Consortium and the curated GO annotations are accompanied by relevant metadata such as evidence codes (<http://geneontology.org/docs/guide-go-evidence-codes/>), references (PubMed IDs) and additional evidence in support of annotations (with/ from). In addition to these internally curated GO annotations, TriTrypDB also hosts electronically annotated GO annotations from InterPro2GO and UniProt and curated GO annotations from the TrypTag project. These metadata and source of annotations (e.g. GeneDB, InterPro, TrypTag) are displayed in the GO annotations section of gene pages (Fig C in S1 File) and in the GO annotation (GAF) files in the downloads section to help users to understand the methods used to assign different GO terms and to let them decide on whether to trust an annotation.

TriTrypDB collaborates with the Gene Ontology Consortium to create new terms to represent kinetoplastid biology, particularly cellular components and biological processes. A few examples of these newly obtained biological processes include acidocalcisome organization (GO:0106117), ciliary basal body segregation (GO:0120312), procyclogenesis (GO:0120324) and kinetoplast DNA replication (GO:0140909), and cellular components include flagellum attachment zone (GO:0120119), reservosome (GO:0106123), ciliary microtubule quartet (GO:0120260) and ciliary centriole arm (GO:0120269).

Table 3. Summary of functional annotation updates for curated organisms since the initial integration of genomes in TriTrypDB.

Organism	Gene products	Names/ synonyms	GO annotations	EC numbers	PubMed IDs	Annotator notes	Total genes with annotations
<i>Trypanosoma brucei</i> TREU927	4525	2140	14565	344	6687	15242	9547
<i>Trypanosoma brucei</i> gambiense	2175	719	8326	76	109	241	7346
<i>Trypanosoma congolense</i>	8724	577	7041	81	71	260	11322
<i>Trypanosoma cruzi</i>	6733	1599	19629	333	3822	3939	20304
<i>Trypanosoma vivax</i>	1935	574	7424	67	49	208	5454
<i>Leishmania braziliensis</i>	2777	556	7498	154	620	530	4990
<i>Leishmania donovani</i>	2632	580	6732	70	67	208	4885
<i>Leishmania major</i>	3059	767	8215	167	926	1119	6069
<i>Leishmania mexicana</i>	2679	729	7165	75	334	297	5098

<https://doi.org/10.1371/journal.pntd.0011058.t003>

Harnessing community expertise for genome annotation

The generation of genome sequences has been expanding at a scale larger than ever before and the number of kinetoplastid genomes in TriTrypDB has increased from five in release 1.0 to 83 in release 59. As only a limited number of genomes are curated by staff curators at VEuPathDB and the level of curation is much lower than that of model organism databases (MODs) even for the curated species, we offer User Comments and the Apollo web-based genome annotation platform [9] for research communities to contribute their expertise to improve annotations.

User Comments. User Comments offer the fastest way to add information to gene records and to alert the curation team in the case of curated genomes. TriTrypDB strongly encourages the user community to offer their expertise by submitting comments about new findings or publications, or even negative results. A new comment can be added by clicking on the ‘Add a comment’ link available on all gene record pages (Fig 3A). The submission of these comments requires users to create an account with TriTrypDB. Users can enter descriptive information about gene structure or function, upload a reference and files (e.g. images of protein localization), and add other gene identifiers (Fig 3B). All comments become immediately visible on the gene pages, searchable via either the site search or the text search from the menu, and can be modified or deleted at any time by the same user.

TriTrypDB currently (as of 30th August 2022) has 5,079 user submitted comments covering a total of 10,038 genes from 51 annotated genomes. Of these, 4,793 comments are associated

(A) User comment
 Tb927.8.6390 lysophospholipase, putative
 Name: TbLysoPLA
 Gene Type: protein coding gene
 Biotype Classification: protein_coding
 Chromosome: 08
 Location: Tb927_08_v5.1:1,855,721..1,857,075(+)
 Species: *Trypanosoma brucei*
 Strain: brucei TREU927
 Status: Curated Reference Strain
 View 1 user comment, or add a comment
 View and update community annotations in Apollo

(B) Add a comment to gene Tb927.8.6390
 Please add only scientific comments to be displayed on the gene page for Tb927.8.6390. If you want to report a problem, use the support page. Your comments are approved. They will be forwarded to the Annotation Center for review and possibly included in future releases of the genome. gene, please also add a comment in the corresponding Genome Sequence

(C) Apollo
 Leishmania infantum JP
 User-created Annotations
 Promastigote transcriptome - Linf.JPC45 Promastigote (unique) Coverage
 Promastigote transcriptome Density - Unique Only
 Promastigote transcriptome XYPlot - Unique And Non-Unique

(D) Apollo
 Details
 Gene Product: RNA-binding protein RBP23
 Evidence: ECO:0001
 Reference: PMID:34705820

(E) Apollo
 This gene is available in Apollo for community annotation. To find out more about Apollo, please visit this help page.
 View in JBrowse genome browser
 Annotate in Apollo

(F) Apollo
 Gene Details
 Gene: LINF_170011800-T1
 Position: LinJ.17:289034..290663 (- strand)
 Links: Apollo
 Apollo Gene Product: RNA-binding protein RBP23
 Apollo Evidence Code: ECO:0000314
 Apollo DBxref: PMID:34705820
 Transcript Details

(G) Annotation, curation and identifiers
 Community annotations from Apollo
 Annotations are derived from gene models in Apollo where an annotator has added additional functional information. To view or edit the annotations, press the 'Annotate in Apollo' button above on this page.

Apollo Product Description	Apollo Transcript ID	Publication	Gene symbol	GO annotation	Most recent change by	Date
RNA-binding protein RBP23	LINF_170011800-T1	PMID:34705820	RBP23	N/A	Uli Boehme, VEuPathDB	2022-06-22

Fig 3. Community curation via User Comments and Apollo. (A) On the top of the gene record page are links to add User Comments and to access Apollo. (B) User Comment interface. (C) To annotate a gene in Apollo, the current gene model needs to be added to the User-created Annotations Area. (D) Interface in Apollo to add gene symbols, product descriptions, GO terms, database references and comments. Once the annotation is finalized, the status needs to be set as finished. (E) Finished annotations are shown on the gene page in the Gene models graphic in the track ‘Community annotations from Apollo’. (F) A popup shows Apollo product descriptions, evidence code and the publication that is associated with the gene. (G) Apollo product description, gene symbol and publication are shown in the Annotation, curation and identifiers section on the gene record page.

<https://doi.org/10.1371/journal.pntd.0011058.g003>

with 9,320 genes from the nine currently curated genomes. Approximately 94% (4,506) of these comments from curated genomes have now been integrated into official annotations, while another 97 (~2%) comments have been reviewed, but not integrated into TriTrypDB.

Community annotation via Apollo. TriTrypDB facilitates manual curation by community experts via Apollo, a collaborative genome annotation and curation platform [9]. This tool allows the editing of existing structural annotations and the creation of new annotations. Users can also update product descriptions and add other functional annotation attributes such as gene names/ symbols, GO terms, and publications associated with the gene. All TriTrypDB reference genomes and selected non-reference genomes (39 in release 59) are available in Apollo and can be accessed via the tools menu and links on gene record pages (Fig 3A). The only prerequisite to access Apollo is to create a TriTrypDB account and to log into TriTrypDB, as is the case with User Comments.

To initiate an annotation in Apollo, the gene model needs to be added into the User-created Annotations area. This can either be done by dragging and dropping or alternatively by using the right-click Apollo menu. Once the gene model is in the Annotation area, it can be modified based on evidence (Fig 3C) and functional annotations such as product descriptions can be added (Fig 3D). Users can also add comments, database references and literature references (PubMed IDs) in the annotation editor window in order to provide supporting evidence for structural or functional annotations made by them. By setting the status to finished, the curator indicates that the annotation is complete. Finished genes are represented in the community annotations track on the genome browser embedded under 'Gene Models' section of gene pages on the following day (Fig 3E and 3F). Similarly, added product descriptions can be found under 'Community annotations from Apollo' in the 'Annotation, curation and identifiers' section of gene pages (Fig 3G). These community annotations entered in Apollo are now indexed and are available for search via the site search. After review by VEuPathDB curators, community annotations are periodically integrated into the official gene set.

Discussion

Recent science enabled by TriTrypDB

The availability of the TriTrypDB resource has supported the advancement of kinetoplastid science, both in basic discovery and translational research over the last decade. Below are a few recent examples of how TriTrypDB data and tools have been utilized by the research community to conduct their own research. TriTrypDB genomes and annotations have been used to characterize individual genes [69–73] or gene families [74,75], identify orthologs across species [55,70,76], conduct genome-wide analyses to study genetic variations such as SNPs, CNVs and hybridization events [77–79], perform comparative genomic [80–82] and phylogenetic/ phylogenomic [82,83] analyses, and as reference genomes for the assembly and/or annotation of newly generated genomes [16,84]. Similarly, the genome assemblies and annotations have also been utilized for the analysis of differential gene expression [19,72,85], protein expression [85–87], the identification of post-translational modifications [72] and potential new genes missing in the official gene sets [88]. Genome data in TriTrypDB has also been instrumental in the development of other kinetoplastid-specific database resources such as TrypsNetDB [55] and TrypTag.org [20]. Moreover, TriTrypDB gene pages are hyperlinked from the record pages of external databases including UniProtKB [54], TrypanoCyc [23] and TDR Targets [89].

Some of the TriTrypDB tools that are popularly used by researchers include the genome browser [83,90], GO enrichment [19,86] and BLAST [71,79]. In addition, researchers have utilized integrated datasets including MS-based proteomics [91,92], RNAi [87] and SNPs [93], implemented searches ranging from gene IDs [94], text/ keywords [76,94], EC numbers [95],

protein features and properties such as InterPro domains, signal peptides and transmembrane domains [86,96,97], and the search strategy system [86] to ask their research questions or test their hypotheses. TriTrypDB serves as an invaluable tool for the selection of drug targets as reviewed in Osorio-Méndez *et al.* [98] and some of the studies discussed above have illustrated the role of TriTrypDB in the identification of potential targets for the design of drugs [73,80] and vaccines [96,97].

Future perspectives

TriTrypDB will continue to add and improve tools, functional datasets and genome annotations in the future including functionality for integrated analysis and visualization of host-parasite interactions. Two host-parasite RNA-Seq datasets investigating the responses of *T. cruzi* infections on human cell lines [99,100] have already been integrated in VEuPathDB, with human data in HostDB (<https://hostdb.org>) and the corresponding parasite data in TriTrypDB. Currently, the search strategy system can be used for the separate interrogation of human and parasite data on respective databases, future development will include functionality to explore both host and parasite data in the context of one another. Maxi-circle sequences from *Leishmania* have also been loaded onto TriTrypDB and can currently be visualized via the genome browser; complete integration of these organellar genome sequences and development of appropriate gene record pages and tools for exploration of these data via search strategies is planned for the future.

TriTrypDB's future development plans include infrastructure for the functional integration of MS-based metabolomics data with sequence-based information, and tools for visual representation of different types of phenotypic information, loss-of-heterozygosity and haplotype data. A significant forthcoming challenge will be the integration of phased genomes, including tools for exploration of structural and sequence differences between haplotypes. With at least one kinetoplastid example already published [101] and others in progress, we anticipate that this will be a priority for the TriTrypDB community. VEuPathDB's efforts are also focused on synchronization of genome sequences and annotations with the INSDC data repositories, rolling out dedicated outreach activities for Apollo-based community annotations, development of a gateway resource for integrated exploration of VEuPathDB and bacterial/ viral BRC (<https://www.bv-brc.org>) and additional workflows for users to analyze their data with the Galaxy workspace.

Supporting information

S1 File. Additional supplementary figures. Fig A Overall trend of genomes and other functional datasets available in TriTrypDB between release 1.0 (October 2009) and release 59 (October 2022). **Fig B** Accessing record pages of popset isolate sequences by conducting a dedicated search from the home page. **Fig C The Gene Ontology terms table from the gene pages.** An example from gene Tb927.8.4470 (chaperone protein DnaJ, putative, J40) showcasing annotations from multiple sources such as GeneDB, UniProt and TrypTag databases. The descriptions of data available in the different columns of this GO terms table are also provided here. (DOCX)

Acknowledgments

We would like to acknowledge the support of all the VEuPathDB team, including past and present team members (<https://tritrypdb.org/tritrypdb/app/static-content/personnel.html>)

and all members of the GeneDB team. We also wish to thank members of TriTrypDB research communities for their willingness to share their data, often prior to publication and for their continuous feedback to improve the resource.

Author Contributions

Conceptualization: Achchuthan Shanmugasundram, Jessica C. Kissinger, Mary Ann McDowell, David S. Roos, Kathryn Crouch, Andrew R. Jones.

Funding acquisition: Jessica C. Kissinger, Mary Ann McDowell, David S. Roos, Kathryn Crouch, Andrew R. Jones.

Project administration: Mary Ann McDowell, David S. Roos, Kathryn Crouch, Andrew R. Jones.

Writing – original draft: Achchuthan Shanmugasundram, David Starns, Ulrike Böhme.

Writing – review & editing: Achchuthan Shanmugasundram, David Starns, Ulrike Böhme, Beatrice Amos, Paul A. Wilkinson, Omar S. Harb, Susanne Warrenfeltz, Jessica C. Kissinger, Mary Ann McDowell, David S. Roos, Kathryn Crouch, Andrew R. Jones.

References

1. Amos B, Aurrecochea C, Barba M, Barreto A, Basenko EY, Bazant W, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.* 2022; 50: D898–D911. <https://doi.org/10.1093/nar/gkab929> PMID: 34718728
2. Greene JM, Collins F, Lefkowitz EJ, Roos D, Scheuermann RH, Sobral B, et al. National Institute of Allergy and Infectious Diseases Bioinformatics Resource Centers: New Assets for Pathogen Informatics. *Infection and Immunity.* 2007; 75: 3212–3219. <https://doi.org/10.1128/IAI.00105-07> PMID: 17420237
3. Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *J Fungi (Basel).* 2018; 4: 39. <https://doi.org/10.3390/jof4010039> PMID: 30152809
4. 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.* 2021; 50: 100860. <https://doi.org/10.1016/j.cois.2021.11.008> PMID: 34864248
5. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics.* 2011; 35: 6.12.1–6.12.19. <https://doi.org/10.1002/0471250953.bi0612s35> PMID: 21901743
6. Ruhamyankaka E, Brunk BP, Dorsey G, Harb OS, Helb DA, Judkins J, et al. ClinEpiDB: an open-access clinical epidemiology database resource encouraging online exploration of complex studies. *Gates Open Res.* 2019; 3: 1661. <https://doi.org/10.12688/gatesopenres.13087.2> PMID: 32047873
7. Oliveira FS, Brestelli J, Cade S, Zheng J, Iodice J, Fischer S, et al. MicrobiomeDB: a systems biology platform for integrating, mining and analyzing microbiome experiments. *Nucleic Acids Res.* 2018; 46: D684–D691. <https://doi.org/10.1093/nar/gkx1027> PMID: 29106667
8. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018; 46: W537–W544. <https://doi.org/10.1093/nar/gky379> PMID: 29790989
9. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, et al. Apollo: Democratizing genome annotation. *PLoS Comput Biol.* 2019; 15: e1006790. <https://doi.org/10.1371/journal.pcbi.1006790> PMID: 30726205
10. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 2010; 38: D457–62. <https://doi.org/10.1093/nar/gkp851> PMID: 19843604
11. Davey JW, Catta-Preta CMC, James S, Forrester S, Motta MCM, Ashton PD, et al. Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*. *G3.* 2021; 11(1): jkaa018. <https://doi.org/10.1093/g3journal/jkaa018> PMID: 33561222

12. Warren WC, Akopyants NS, Dobson DE, Hertz-Fowler C, Lye L-F, Myler PJ, et al. Genome assemblies across the diverse evolutionary spectrum of *Leishmania* protozoan parasites. *Microbiol Resour Announc.* 2021; 10: e00545–21. <https://doi.org/10.1128/MRA.00545-21> PMID: 34472979
13. Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, et al. Genome of *Leptomonas pyrrocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep.* 2016; 6: 23704. <https://doi.org/10.1038/srep23704> PMID: 27021793
14. Kraeva N, Butenko A, Hlaváčová J, Kostygov A, Myšková J, Grybchuk D, et al. *Leptomonas seymouri*: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with *Leishmania donovani*. *PLOS Pathogens.* 2015; 11(8): e1005127. <https://doi.org/10.1371/journal.ppat.1005127> PMID: 26317207
15. Skalický T, Dobáková E, Wheeler RJ, Tesařová M, Flegontov P, Jirsová D, et al. Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-branching trypanosomatid. *Proc Natl Acad Sci U S A.* 2017; 114: 11757–11762. <https://doi.org/10.1073/pnas.1712311114> PMID: 29078369
16. Almutairi H, Urbaniak MD, Bates MD, Jariyapan N, Kwakye-Nuako G, Thomaz Soccol V, et al. Chromosome-scale genome sequencing, assembly and annotation of six genomes from subfamily *Leishmaniinae*. *Sci Data.* 2021; 8: 234. <https://doi.org/10.1038/s41597-021-01017-3> PMID: 34489462
17. Almutairi H, Urbaniak MD, Bates MD, Al-Salem WS, Dillon RJ, Bates PA, et al. Chromosome-Scale Assembly of the Complete Genome Sequence of *Porcisia heritigi*, Isolate C119, Strain LV43. *Microbiol Resour Announc.* 2021; 10: e00651–21. <https://doi.org/10.1128/MRA.00651-21> PMID: 34647802
18. Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for “expressed sequence tags”. *Nat Genet.* 1993; 4: 332–333. <https://doi.org/10.1038/ng0893-332> PMID: 8401577
19. Briggs EM, Rojas F, McCulloch R, Matthews KR, Otto TD. Single-cell transcriptomic analysis of bloodstream *Trypanosoma brucei* reconstructs cell cycle progression and developmental quorum sensing. *Nat Commun.* 2021; 12: 5268. <https://doi.org/10.1038/s41467-021-25607-2> PMID: 34489460
20. Dean S, Sunter JD, Wheeler RJ. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol.* 2017; 33: 80–82. <https://doi.org/10.1016/j.pt.2016.10.009> PMID: 27863903
21. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021; 49: D545–D551. <https://doi.org/10.1093/nar/gkaa970> PMID: 33125081
22. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.* 2020; 48: D445–D453. <https://doi.org/10.1093/nar/gkz862> PMID: 31586394
23. Shameer S, Logan-Klumpler FJ, Vinson F, Cottret L, Merlet B, Achcar F, et al. TrypanoCyc: a community-led biochemical pathways database for *Trypanosoma brucei*. *Nucleic Acids Res.* 2015; 43: D637–44. <https://doi.org/10.1093/nar/gku944> PMID: 25300491
24. Doyle MA, MacRae JI, De Souza DP, Saunders EC, McConville MJ, Likić VA. LeishCyc: a biochemical pathways database for *Leishmania major*. *BMC Syst Biol.* 2009; 3: 57. <https://doi.org/10.1186/1752-0509-3-57> PMID: 19497128
25. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics.* 2016; 32: 309–311. <https://doi.org/10.1093/bioinformatics/btv557> PMID: 26415722
26. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021; 596: 583–589. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
27. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022; 50: D439–D444. <https://doi.org/10.1093/nar/gkab1061> PMID: 34791371
28. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 2016; 44: D1214–9. <https://doi.org/10.1093/nar/gkv1031> PMID: 26467479
29. Bairoch A. The ENZYME database in 2000. *Nucleic Acids Res.* 2000; 28(1): 304–305. <https://doi.org/10.1093/nar/28.1.304> PMID: 10592255
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25: 25–29. <https://doi.org/10.1038/75556> PMID: 10802651

31. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 2021; 49: D325–D334. <https://doi.org/10.1093/nar/gkaa1113> PMID: 33290552
32. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* 2019; 47: D520–D528. <https://doi.org/10.1093/nar/gky949> PMID: 30357364
33. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28: 235–242. <https://doi.org/10.1093/nar/28.1.235> PMID: 10592235
34. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol.* 2003; 10: 980. <https://doi.org/10.1038/nsb1203-980> PMID: 14634627
35. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2019; 47: D339–D343. <https://doi.org/10.1093/nar/gky1006> PMID: 30357391
36. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database.* 2020; 2020: 1–21. <https://doi.org/10.1093/database/baaa062> PMID: 32761142
37. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol.* 2006; 13: 1028–1040. <https://doi.org/10.1089/cmb.2006.13.1028> PMID: 16796549
38. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999; 27(2): 573–580. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982
39. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014; 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
40. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* 2022; 40: 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3> PMID: 34980915
41. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001; 305: 567–580. <https://doi.org/10.1006/jmbi.2000.4315> PMID: 11152613
42. Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 2021; 49: D394–D403. <https://doi.org/10.1093/nar/gkaa1106> PMID: 33290554
43. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* 2019; 47: W402–W407. <https://doi.org/10.1093/nar/gkz297> PMID: 31251384
44. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* 2021; 49: D192–D200. <https://doi.org/10.1093/nar/gkaa1047> PMID: 33211869
45. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol.* 2019; 1962: 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1 PMID: 31020551
46. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019; 47(D1): D155–D162. <https://doi.org/10.1093/nar/gky1141> PMID: 30423142
47. Dewey CN. Aligning multiple whole genomes with Mercator and MAVID. *Methods Mol Biol.* 2007; 395: 221–236. https://doi.org/10.1007/978-1-59745-514-5_14 PMID: 17993677
48. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019; 37: 907–915. <https://doi.org/10.1038/s41587-019-0201-4> PMID: 31375807
49. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics.* 2022; 38: 2943–2945. <https://doi.org/10.1093/bioinformatics/btac166> PMID: 35561197
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
51. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics.* 2019; 35(3): 421–432. <https://doi.org/10.1093/bioinformatics/bty648> PMID: 30020410
52. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCftools. *Gigascience.* 2021; 10(2): giab008. <https://doi.org/10.1093/gigascience/giab008> PMID: 33590861

53. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012; 22: 568–576. <https://doi.org/10.1101/gr.129684.111> PMID: 22300766
54. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021; 49: D480–D489. <https://doi.org/10.1093/nar/gkaa1100> PMID: 33237286
55. Gazestani VH, Yip CW, Nikpour N, Berghuis N, Salavati R. TrypsNetDB: An integrated framework for the functional characterization of trypanosomatid proteins. *PLoS Negl Trop Dis.* 2017; 11: e0005368. <https://doi.org/10.1371/journal.pntd.0005368> PMID: 28158179
56. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 2021; 49: D344–D354. <https://doi.org/10.1093/nar/gkaa977> PMID: 33156333
57. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE.* 2011; 6(7): e21800. <https://doi.org/10.1371/journal.pone.0021800> PMID: 21789182
58. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 2016; 17: 66. <https://doi.org/10.1186/s13059-016-0924-1> PMID: 27072794
59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
60. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
61. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol.* 1999; 292: 741–758. <https://doi.org/10.1006/jmbi.1999.3069> PMID: 10497036
62. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem.* 1996; 241: 779–786. <https://doi.org/10.1111/j.1432-1033.1996.00779.x> PMID: 8944766
63. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021; 49: D605–D612. <https://doi.org/10.1093/nar/gkaa1074> PMID: 33237311
64. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007; 35: W585–7. <https://doi.org/10.1093/nar/gkm259> PMID: 17517783
65. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018; 27: 135–145. <https://doi.org/10.1002/pro.3290> PMID: 28884485
66. Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, et al. GeneDB—an annotation database for pathogens. *Nucleic Acids Res.* 2012; 40: D98–108. <https://doi.org/10.1093/nar/gkr1032> PMID: 22116062
67. Real F, Vidal RO, Carazzolle MF, Mondego JMC, Costa GGL, Herai RH, et al. The genome sequence of *Leishmania (Leishmania) amazonensis*: functional annotation and extended analysis of gene models. *DNA Res.* 2013; 20: 567–581. <https://doi.org/10.1093/dnares/dst031> PMID: 23857904
68. Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, et al. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* 2016; 44: W29–34. <https://doi.org/10.1093/nar/gkw292> PMID: 27105845
69. Bakari-Soale M, Ikenga NJ, Scheibe M, Butter F, Jones NG, Kramer S, et al. The nucleolar DExD/H protein Hel66 is involved in ribosome biogenesis in *Trypanosoma brucei*. *Sci Rep.* 2021; 11: 18325. <https://doi.org/10.1038/s41598-021-97020-0> PMID: 34526538
70. Paris Z, Svobodová M, Kachale A, Horáková E, Nenarokova A, Lukeš J. A mitochondrial cytidine deaminase is responsible for C to U editing of tRNA to decode the UGA codon in *Trypanosoma brucei*. *RNA Biol.* 2021; 18: 278–286. <https://doi.org/10.1080/15476286.2021.1940445> PMID: 34224320
71. Perdomo D, Berdance E, Lallinger-Kube G, Sahin A, Dacheux D, Landrein N, et al. TbKINX1B: a novel BILBO1 partner and an essential protein in bloodstream form *Trypanosoma brucei*. *Parasite.* 2022; 29: 14. <https://doi.org/10.1051/parasite/2022015> PMID: 35262485
72. Picchi-Constante GFA, Guerra-Slompo EP, Tahira AC, Alcántara MV, Amaral MS, Ferreira AS, et al. Metacyclogenesis defects and gene expression hallmarks of histone deacetylase 4-deficient *Trypanosoma cruzi* cells. *Sci Rep.* 2021; 11: 21671. <https://doi.org/10.1038/s41598-021-01080-1> PMID: 34737385

73. Pezza A, Tavernelli LE, Alonso VL, Perdomo V, Gabarro R, Prinjha R, et al. Essential Bromodomain BDF2 as a Drug Target against Chagas Disease. *ACS Infect Dis*. 2022; 8: 1062–1074. <https://doi.org/10.1021/acinfecdis.2c00057> PMID: 35482332
74. Hickson J, Athayde LFA, Miranda TG, Junior PAS, Dos Santos AC, da Cunha Galvão LM, et al. *Trypanosoma cruzi* iron superoxide dismutases: insights from phylogenetics to chemotherapeutic target assessment. *Parasit Vectors*. 2022; 15: 194. <https://doi.org/10.1186/s13071-022-05319-2> PMID: 35668508
75. Gúiza J, García A, Arriagada J, Gutiérrez C, González J, Márquez-Miranda V, et al. Unnexins: Homologs of innexin proteins in Trypanosomatidae parasites. *J Cell Physiol*. 2022; 237: 1547–1560. <https://doi.org/10.1002/jcp.30626> PMID: 34779505
76. Yao C, Wilson ME. Dynamics of sterol synthesis during development of *Leishmania* spp. parasites to their virulent form. *Parasit Vectors*. 2016; 9: 200. <https://doi.org/10.1186/s13071-016-1470-0> PMID: 27071464
77. Rosa-Teijeiro C, Wagner V, Corbeil A, d'Annessa I, Leprohon P, do Monte-Neto RL, et al. Three different mutations in the DNA topoisomerase 1B in *Leishmania infantum* contribute to resistance to antitumor drug topotecan. *Parasit Vectors*. 2021; 14: 438. <https://doi.org/10.1186/s13071-021-04947-4> PMID: 34454601
78. Lypaczewski P, Matlashewski G. *Leishmania donovani* hybridisation and introgression in nature: a comparative genomic investigation. *Lancet Microbe*. 2021; 2: e250–e258. [https://doi.org/10.1016/S2666-5247\(21\)00028-8](https://doi.org/10.1016/S2666-5247(21)00028-8) PMID: 35544170
79. Antonia AL, Barnes AB, Martin AT, Wang L, Ko DC. Variation in *Leishmania* chemokine suppression driven by diversification of the GP63 virulence factor. *PLoS Negl Trop Dis*. 2021; 15: e0009224. <https://doi.org/10.1371/journal.pntd.0009224> PMID: 34710089
80. Prava J, Pan A. *In silico* analysis of *Leishmania* proteomes and protein-protein interaction network: Prioritizing therapeutic targets and drugs for repurposing to treat leishmaniasis. *Acta Trop*. 2022; 229: 106337. <https://doi.org/10.1016/j.actatropica.2022.106337> PMID: 35134348
81. Oldrieve GR, Malacart B, López-Vidal J, Matthews KR. The genomic basis of host and vector specificity in non-pathogenic trypanosomatids. *Biol Open*. 2022; 11. <https://doi.org/10.1242/bio.059237> PMID: 35373253
82. Viana de Almeida L, Luís Reis-Cunha J, Coqueiro-Dos-Santos A, Flávia Rodrigues-Luís G, de Paula Baptista R, de Oliveira Silva S, et al. Comparative genomics of *Leishmania* isolates from Brazil confirms the presence of *Leishmania major* in the Americas. *Int J Parasitol*. 2021; 51: 1047–1057. <https://doi.org/10.1016/j.ijpara.2021.05.009> PMID: 34329650
83. Tesan FC, Lorenzo R, Alleva K, Fox AR. AQPX-cluster aquaporins and aquaglyceroporins are asymmetrically distributed in trypanosomes. *Commun Biol*. 2021; 4: 953. <https://doi.org/10.1038/s42003-021-02472-9> PMID: 34376792
84. Zakharova A, Albanaz ATS, Oppendoes FR, Škodová-Sveráková I, Zagirova D, Saura A, et al. *Leishmania guyanensis* M4147 as a new LRV1-bearing model parasite: Phosphatidate phosphatase 2-like protein controls cell cycle progression and intracellular lipid content. *PLoS Negl Trop Dis*. 2022; 16: e0010510. <https://doi.org/10.1371/journal.pntd.0010510> PMID: 35749562
85. Kalesh K, Wei W, Mantilla BS, Roumeliotis TI, Choudhary J, Denny PW. Transcriptome-Wide Identification of Coding and Noncoding RNA-Binding Proteins Defines the Comprehensive RNA Interactome of *Leishmania mexicana*. *Microbiol Spectr*. 2022; 10: e0242221. <https://doi.org/10.1128/spectrum.02422-21> PMID: 35138191
86. de Pablos LM, Ferreira TR, Dowe AA, Forrester S, Parry E, Newling K, et al. The mRNA-bound Proteome of *Leishmania mexicana*: Novel Genetic Insight into an Ancient Parasite. *Mol Cell Proteomics*. 2019; 18: 1271–1284. <https://doi.org/10.1074/mcp.RA118.001307> PMID: 30948621
87. Turra GL, Liedgens L, Sommer F, Schneider L, Zimmer D, Vilurbina Perez J, et al. Structure-Function Analysis and Redox Interactomes of *Leishmania tarentolae* Erv. *Microbiol Spectr*. 2021; 9: e0080921. <https://doi.org/10.1128/Spectrum.00809-21> PMID: 34585988
88. Tinti M, Kelner-Mirôn A, Marriott LJ, Ferguson MAJ. Polysomal mRNA Association and Gene Expression in *Trypanosoma brucei*. *Wellcome Open Res*. 2021; 6: 36. <https://doi.org/10.12688/wellcomeopenres.16430.3> PMID: 34250262
89. Urán Landaburu L, Berenstein AJ, Videla S, Maru P, Shanmugam D, Chernomoretz A, et al. TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Res*. 2020; 48: D992–D1005. <https://doi.org/10.1093/nar/gkz999> PMID: 31680154
90. Yagoubat A, Croub L, Berry L, Kuk N, Lefebvre M, Sarrazin A, et al. Universal highly efficient conditional knockout system in *Leishmania*, with a focus on untranscribed region preservation. *Cell Microbiol*. 2020; 22: e13159. <https://doi.org/10.1111/cmi.13159> PMID: 31909863

91. Tinti M, Ferguson MAJ. Visualisation of proteome-wide ordered protein abundances in *Trypanosoma brucei*. Wellcome Open Res. 2022; 7: 34. <https://doi.org/10.12688/wellcomeopenres.17607.1> PMID: 35284642
92. Irwin NAT, Pittis AA, Richards TA, Keeling PJ. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. Nat Microbiol. 2022; 7: 327–336. <https://doi.org/10.1038/s41564-021-01026-3> PMID: 34972821
93. Contreras Garcia M, Walshe E, Stekete PC, Paxton E, Lopez-Vidal J, Pearce MC, et al. Comparative Sensitivity and Specificity of the 7SL sRNA Diagnostic Test for Animal Trypanosomiasis. Front Vet Sci. 2022; 9: 868912. <https://doi.org/10.3389/fvets.2022.868912> PMID: 35450136
94. Silva LA, Vinaud MC, Castro AM, Cravo PVL, Bezerra JCB. *In silico* search of energy metabolism inhibitors for alternative leishmaniasis treatments. Biomed Res Int. 2015; 2015: 965725. <https://doi.org/10.1155/2015/965725> PMID: 25918726
95. de Azevedo-Martins AC, Ocaña K, de Souza W, Vasconcelos ATR de, Teixeira MMG, Camargo EP, et al. The Importance of Glycerophospholipid Production to the Mutualist Symbiosis of Trypanosomatids. Pathogens. 2022; 11(1): 41. <https://doi.org/10.3390/pathogens11010041> PMID: 35055989
96. Autheman D, Crosnier C, Clare S, Goulding DA, Brandt C, Harcourt K, et al. An invariant *Trypanosoma vivax* vaccine antigen induces protective immunity. Nature. 2021; 595: 96–100. <https://doi.org/10.1038/s41586-021-03597-x> PMID: 34040257
97. Michel-Todó L, Reche PA, Bigey P, Pinazo M-J, Gascón J, Alonso-Padilla J. *In silico* Design of an Epitope-Based Vaccine Ensemble for Chagas Disease. Front Immunol. 2019; 10: 2698. <https://doi.org/10.3389/fimmu.2019.02698> PMID: 31824493
98. Osorio-Méndez JF, Cevallos AM. Discovery and Genetic Validation of Chemotherapeutic Targets for Chagas' Disease. Front Cell Infect Microbiol. 2018; 8: 439. <https://doi.org/10.3389/fcimb.2018.00439> PMID: 30666299
99. Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, et al. Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. PLoS Pathog. 2016; 12: e1005511. <https://doi.org/10.1371/journal.ppat.1005511> PMID: 27046031
100. Belew AT, Junqueira C, Rodrigues-Luiz GF, Valente BM, Oliveira AER, Polidoro RB, et al. Comparative transcriptome profiling of virulent and non-virulent *Trypanosoma cruzi* underlines the role of surface proteins during infection. PLoS Pathog. 2017; 13: e1006767. <https://doi.org/10.1371/journal.ppat.1006767> PMID: 29240831
101. Cosentino RO, Brink BG, Siegel TN. Allele-specific assembly of a eukaryotic genome corrects apparent frameshifts and reveals a lack of nonsense-mediated mRNA decay. NAR Genom Bioinform. 2021; 3(3): lqab082. <https://doi.org/10.1093/nargab/lqab082> PMID: 34541528