

# Using Genomic Information to Understand *Leishmania* Biology

Juliano S. de Toledo, Elton J. R. Vasconcelos, Tiago R. Ferreira and Angela K. Cruz\*

*Department of Cell and Molecular Biology and Pathogens, School of Medicine, University of São Paulo at Ribeirão Preto, Av Bandeirantes, 3900, 14049-900, São Paulo, Brazil*

**Abstract:** The genomes of different species of *Leishmania* have been deciphered in recent years. We learned that the genome content and organization of *Leishmania major*, *Leishmania braziliensis* and *Leishmania infantum* are highly similar and annotation of these genomes revealed that there are few species-specific genes. Association of genome information with reverse and forward genetics approaches allows posing and answering relevant biological questions in a novel way. In this article we briefly present an overview of relevant aspects of genome organization of the *Leishmania* and how this information can be used to improve our understanding of the biology, pathogenesis, host-parasite interaction issues. We present some of the most useful bioinformatics tools/softwares, which are currently available and how each one of them can be used to explore the genome supporting a wide variety of queries. We included other computational tools which allow integrating the genome data with biochemical pathways revealing metabolic and regulatory networks to be investigated. Finally, we discuss reverse and forward genetic tools available and finalize with considerations on established and novel high-throughput approaches at the genome, transcriptome and proteome levels.

**Keywords:** *Leishmania*, genomic information, biology.

## 1. LEISHMANIA AND LEISHMANIASIS

Leishmaniasis is caused by the protozoa of the *Leishmania* genus and affects about 12 million people around the world. Leishmaniasis is considered the second most important parasitic disease in the world considering the number of people at risk in endemic areas, number of infected people and morbidity. It is estimated that more than 2 million new cases of leishmaniasis occur each year in 88 countries (<http://www.who.int/health-topics/leishmaniasis.htm>) affecting mainly the poorest population living in tropical and subtropical areas. Leishmaniasis is a spectral disease with multifaceted clinical manifestations varying from: *i*) cutaneous leishmaniasis (CL), that may produce localized and self-healing or disseminated lesions, caused mainly by *Leishmania (Leishmania) major*, *L. (Leishmania) tropica* and *L. (Leishmania) mexicana* species, *ii*) mucocutaneous leishmaniasis (MCL), a disfiguring form of the disease, typical of *L. (Viannia) braziliensis* infections, and *iii*) visceral leishmaniasis (VL) caused by *L. (Leishmania) donovani*, *L. (Leishmania) infantum* and *L. (Leishmania) chagasi* and it is fatal if untreated.

About 20 different species of the *Leishmania* genus are known to be pathogenic to humans. These protozoan parasites, kinetoplastids from the Trypanosomatidae family, complete their life cycle alternating between extreme forms, *i*) the free-living promastigotes occurring in the gut of the insect host, the female Phlebotominae and *ii*) the amastigotes, the intracellular form, which lives inside of phagolysosomes of phagocytic cell in a variety of vertebrate

organisms. The infective metacyclic promastigotes are transmitted by the Phlebotominae vector to the vertebrate host during its blood meal and to evade host's humoral defense, parasites invade the cells of macrophagic system [1]. Recently it was demonstrated by intravital two-photon microscopy that the parasites have more difficulty establishing an infection and surviving in mice lacking neutrophils, suggesting the relevance of this mechanism of silent entry into macrophages involving polymorphonuclear leukocytes as the first phagocytic cell to encounter the parasite in the host. The apoptotic infected cells are then ingested by macrophages functioning as a "Trojan horse" [2, 3].

*Leishmania* are mostly diploid organisms which divide mainly by binary partition, although sexual exchange has been recently evidenced in the invertebrate stage [4, 5]. *Leishmania* and the other eukaryotes of the order Kinetoplastida display some uncommon biochemical, genetic, and morphological features including the mitochondrial DNA organization, the kinetoplast, mitochondrial DNA editing [6, 7], glycosomes [8], polycistronic transcription [9], trans-splicing [10, 11], and GPI anchoring of membrane proteins [12].

*Leishmania* parasites have a very "plastic" genome and the ability to alter the copy number of individual gene, groups of genes, chromosomes or the entire genome as a means to fit adverse conditions. This singular characteristic is known as genetic plasticity and is manifested by chromosomal rearrangements, ploidy variation or occurrence of circular molecules that modify the expression levels of gene(s) involved in the process of "copy amplification". Sites of highly conserved repeat sequences appear to facilitate these rearrangements and the amplification of key genes during environmental changes. In the absence of sense or antisense repeats in the vicinity of a gene conferring a selective advan-

\*Address correspondence to this author at the Departamento de Biologia Celular e Molecular e Bioagentes, Patogênicos, Faculdade Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, no. 3900, CEP 14049-900, Ribeirão Preto, SP, Brazil; Tel: +55 16 3602 3318; Fax: +55 16 3633 6631; E-mail: akcruz@fmrp.usp.br

tage, *Leishmania* will make use of extra-copy chromosomes to increase the levels of a gene product [13-19].

Genes in trypanosomatids are transcribed as long polycistronic units. The primary transcripts are processed into individual mRNAs by a coupled mechanism of *trans-splicing* and polyadenylation [20, 21]. The *trans-splicing* machinery is responsible for transferring a capped small RNA, known as spliced-leader (SL), to the 5' end of the mRNAs [22]. The splice acceptor site (SAS) is an AG dinucleotide preceded by a polypyrimidine track located upstream from the open reading frame (ORF). The polypyrimidine track of a given gene is also responsible for directing the polyadenylation site of the upstream mRNA [20]. Therefore, control of gene expression is post-transcriptionally regulated by downstream events affecting mRNA stability and translation [23-25].

The success of the first genome projects and the urgent need for genetic tools and information to help improving knowledge about pathogenesis, virulence and host-parasite interaction issues were the driving forces to put forward a genome project for *Leishmania* in the early nineties. The enterprise should open new avenues for the rationale definition of molecular targets and compounds leading to low-toxicity and effective drugs, better diagnostic approaches and vaccine development.

## 2. THE LEISHMANIA GENOME PROJECT: AN OVERVIEW

The first Parasite Genome Network Planning Meeting was held in April 1994 in Rio de Janeiro, Brazil, and was sponsored by FIOCRUZ (Fundação Oswaldo Cruz) and World Bank/WHO (World Health Organization) Special Programme for Research and Training in Tropical Diseases (TDR). The Genome Projects of *Trypanosoma cruzi*, *Trypanosoma brucei*, and *Leishmania major* (CL) were planned at this meeting. Reference strains for the trypanosomes were chosen and *L. (L.) major* Friedlin (LmjF) was defined as the reference strain for the *Leishmania* Genome project a couple of years later. Several laboratories from all over the world joined the *Leishmania* Genome Network (LGN). The first tasks of the LGN were to determine the number of chromosomes of *Leishmania*, based on the linkage groups established previously by Bastien and co-workers [26], and to generate a physical map of a cosmid library representative of the parasite genome. Due to technology and financial restraints initially single pass sequencing of cDNA library clones was conducted to generate tags of the expressed genome (ESTs). A couple of years later the genomic sequencing was undertaken on a chromosome by chromosome approach based on the cosmid maps generated or by chromosome shotgun sequencing [27].

The chromosome 1 sequencing was initiated in 1996, as a pilot project of physical map-based sequencing, and was published by Myler and co-workers in 1999 [28]. The rest of the genome came afterwards, and the 32.8-Mb sequence representing the haploid genome was obtained by a combination of approaches ranging from hierarchical sequencing strategy, clone-by-clone approach, and Whole Chromosome Shotgun, which involves an initial fractionation step of individual or co-migrating chromosomes by PFGE. The entire *Leishmania* genome sequence was published along with the *T. brucei* and

*T. cruzi* genomes in 2005, and the genome annotation estimated the presence of 911 RNA genes, 39 pseudogenes, and 8272 protein-coding genes [29]. Since genome annotation is a continuous process, in 2007, Peacock *et al.* announced 8298 genes annotated as protein-coding and 97 pseudogenes for *L. (L.) major* [30]. As raised initially by El-Sayed *et al.*, those species-specific or trypanosomatid-specific genes or even those protein domains only found in trypanosomatids and not in their hosts are fertile grounds to be explored in the search for novel and safe anti-leishmanial or anti-trypanosomatid drugs.

The transcriptional analysis of *L. (L.) major* chromosome 1 by run-on assays showed that transcription initiates in both directions within a single region revealing a novel kind of gene organization, the Directional Gene Clusters (DGCs) [9]. This polarity of transcriptional units is noteworthy and it constitutes a singular characteristic of trypanosomatids. These DGCs are non homogeneous in size ranging from a few kb to more than 1Mb. These DGCs keep a strikingly conserved synteny among trypanosomatid genomes [31] but, differently from prokaryotic operons, genes in the same DGC are not functionally related. Sequences located between two divergent or convergent transcription units are called strand-switch regions (SSRs). The nucleotide sequences within these regions do not display a particular consensus sequence for any known RNA polymerase promoter but show a high AT content compared to the coding genome. These regions must contain the signaling sites for transcription initiation and termination.

Recent informatics and sequencing technology improvements allowed the expansion of the *Leishmania* genome project to *L. (L.) infantum* (VL) and *L. (V.) braziliensis* (MCL) species. These genomes were accomplished in less than a year, in a way that now genomic information is available for species representative of the three kinds of clinical manifestations of the disease.

## 3. COMPARATIVE GENOMICS OF LEISHMANIA

With the completion of the genome sequencing of three representative parasite species (two from the *Leishmania* subgenus: *L. (L.) major* and *L. (L.) infantum*; and one from the *Viannia* subgenus: *L. (V.) braziliensis*), a comparative analysis of the *Leishmania* spp. genomes was conducted and recently published [30]. This study is a landmark because now we can try to understand some basic aspects of *Leishmania* biology such as: *i)* the development of *L. (V.) braziliensis* in the hindgut of its sandfly vector, *ii)* the dogs as restricted reservoir hosts of infections with the *L. (L.) donovani*, *L. (L.) chagasi* or *L. (L.) infantum* or *iii)* the *Leishmania* genotype contribution to large spectrum of clinical manifestations in humans.

The identification of factors that allow three closely related organisms to cause a human disease with so many different manifestations is a great challenge for the *Leishmania* researchers. It was observed that gene families that determine the properties of the parasite cell surface, as the amastin gene repertoire, have some differentially expanded loci among the three species [32]. In addition, several functional studies of different genes involved in virulence or parasite survival within the host revealed that those genes are

not equally relevant in each species [33-36]. Overall, reverse genetics and genome information will be synergistic to improve understanding of host-parasite interaction similarities and differences among *Leishmania* species. Understanding *Leishmania* biology may help in the development of species-specific or common drugs for all pathogenic *Leishmania* species. Differential gene expression, species-specific genes and characterization of the specificity of the host response to the infection may determine the parasite tropism and virulence. A paradigm for a virulence related species-specific genes is the *L. donovani*-A2 multigenic family, only expressed in amastigotes of this species; transfection of A2 into *L. (L.) major* alters the strain virulence [37].

Despite the similar DNA content of around 33 Mb in sequenced *Leishmania* genomes and a remarkable conservation of gene content and synteny in orthologous chromosomes, karyotypic differences have been identified among the species. While *L. (L.) major* and *L. (L.) infantum* display 36 chromosomes, the *L. (V.) braziliensis* and *L. (L.) mexicana* present respectively 35 and 34 chromosomes. These differences were shown to be due to fusion of chromosomes 20 + 34 in *L. (V.) braziliensis*; 8 + 29 and 20 + 36 in *L. (L.) mexicana* [38].

Considering that 20-100 million years have gone since the divergence between the *Leishmania* species complexes and that major differences exist in the clinical manifestations of the disease, a large variation in genomic content would be expected. Nevertheless, only 2.5% of genes present in each *Leishmania* genome were found to be species-specific. Structure and sequence of the *Leishmania* genomes must have been under strong evolutionary pressure over time to maintain such high levels of genetic similarity between these species.

In the *Leishmania* comparative genome analysis it was noted that some genes are inactivated while others seem to be evolving faster. This analysis confirmed that the most divergent species among the three sequenced genomes is *L. (V.) braziliensis*, containing about 47 genes not shared by the other two. *L. (L.) infantum* has 26 species-specific genes while *L. (L.) major* has only 5 [30]. Compared to other trypanosomatids' genomes, such as *T. brucei* and/or *T. cruzi*, it was found that among the *L. (L.) major*, *L. (L.) infantum*, and *L. (V.) braziliensis* specific genes only one (1/5), ten (10/26) and seventeen (17/47), respectively, had orthologs in *T. brucei* and/or *T. cruzi* genomes [30].

Differently from other eukaryotes, where insertion/deletion events and sequence re-arrangements take place in gene diversification, degeneration of existing genes accounts for ~80% of the species differences in *Leishmania* [39].

The lack of intact mobile DNA elements in the *L. (L.) major* and *L. (L.) infantum* genomes in opposition to the presence of intact retrotransposon elements in *L. (V.) braziliensis* could explain the higher divergence of the *Viannia* subgenus [40]. Remarkably, sequencing of *L. (V.) braziliensis* genome revealed that only this species (and apparently other species from the *Viannia* subgenus, unpublished results) display genes encoding putative machinery for a functional RNAi pathway (S. Beverley, personal communica-

tion). The *L. (V.) braziliensis* Dicer and Argonaute-like genes were identified *in silico* based on its similarity to the functional genes from *T. brucei* [30, 31, 40-44].

The use of the RNAi machinery to knock-down different genes may turn into an invaluable tool to understand gene function and may bring novel insights into host-parasite interaction mechanisms. However, genetic plasticity of the *Leishmania* genome has been proposed as a mode by which *Leishmania* increases expression of a given gene or avoids loss of essential genes. The generation of extra copies of the wild-type alleles occurs during artificial attempts to inactivate essential genes in this protozoan [13-15, 19, 45-47]. This characteristic, which impedes generation of null mutants for essential genes, may also interfere with the outcome of RNAi experiments [13].

#### **4. EXPLORING THE GENOME: BIOINFORMATICS TOOLS**

The plethora of genome data available in recent years and the consequent comparative analyses are powerful resources to explore, at the molecular level, the genetic and biochemical similarities and dissimilarities in the *Leishmania* genus. At this point, the development of computational algorithms is crucial for analyzing the great amount of data from multiple organisms and it is a challenge to generate tools to allow the translation of genome data into relevant biological information and the postulation of novel routes of investigation. In the past twenty years, *Bioinformatics* has become a close partner from the *Genomics*, *Transcriptomics* and *Proteomics* fields providing essential tools frequently used by the geneticists, molecular biologists and molecular parasitologists. The algorithms are often open sourced and can be downloaded from the developers' webpages, so the expert users and bioinformaticians may alter them, run them locally on their own system and apply them in a specific pipeline if necessary. On the other hand, several developers also provide friendly web user interfaces of their algorithms, so the researchers may run them online in an easier way.

In this section we will highlight some important tools and databases available to work on the web (Table 1). It includes searching for genes of interest and getting all the information about them, searching for unique genes or by a group of them sharing some common structural and/or functional traits, and also the investigation of more complex data from these parasites such as regulatory networks and metabolic pathways. Possible approaches for rational identification of drug-targets will also be discussed here.

##### **4.1. GeneDB**

The first robust database created for accessing the raw data from *Leishmania* spp. genomes sequencing projects was GeneDB [48], a project developed by the Sanger Institute Pathogen Sequencing Unit (PSU), which provides access to 37 other genomes among viruses, bacteria, fungi, protozoa, parasitic vectors and helminths. The first *Leishmania* species to have a completely sequenced genome was *L. (L.) major* (<http://www.genedb.org/> Homepage/Lmajor - current annotation version 5.2). *L. (L.) infantum* (<http://www.genedb.org/> Homepage/Linfantum) and *L. (V.) braziliensis* (<http://www.genedb.org/> Homepage/Lbraziliensis) genomes came just after *L. (L.) major* and are in the annotation versions 3.0

and 2.0, respectively. These databases are kept under continual manual annotation and curation, so the versions might change with time and new features are frequently attributed to the genomes.

The most recent species being sequenced by the Sanger PSU is *L. (L.) mexicana* (<http://www.genedb.org/Homepage/Lmexicana>) which already have the assembled contigs and predicted peptide sequences available for searches on the genedb blast server ([http://www.genedb.org/blast/submitblast/GeneDB\\_Lmexicana](http://www.genedb.org/blast/submitblast/GeneDB_Lmexicana)). All the sequences data sets from these *Leishmania* species can be downloaded from the Sanger ftp server (<ftp://ftp.sanger.ac.uk/pub4/pathogens/>) where we can find FASTA files of the assembled contigs (whole genomes), predicted protein coding genes (CDSs), predicted proteins (translated CDSs), among others, and also the annotation files for each chromosome. All the annotation process of the trypanosomatids' genes/genomes is made by curators at the Sanger Institute and Artemis [49] bears it all in excellent graphical output. These files can also be downloaded on the Sanger ftp server mentioned above. Artemis is written in JAVA and can be executed on any OS that have JAVA installed.

On [genedb.org](http://genedb.org) one can search for genes by their accession number or by doing a full content text search and get all crucial information about them (DNA and protein sequences, genome location, peptide properties, domain information, gene ontology annotation, orthologues, among others). It is also possible to link all the genes directly to SWISS-PROT, Pfam, Interpro, and Gene Ontology (AmiGO) databases, and, clicking on the "view sequence" weblink, one can run local alignment search tools [50] inside a single genome (BLAST link) and/or using multiple selected genomes as databases (omniBLAST link). Other interesting tools, particularly for protein sequences of a chosen organism, are (I) *Motif Search* which finds proteins containing a specific pattern defined by the user, and (II) *EMOWSE* which identifies proteins in a genome that most likely correspond to a query peptide, not only by the peptide sequence but also by its mass, digestion type (e.g. trypsin), orientation (N-term to C-term or C-term to N-term) and other parameters. Both tools are available on <http://old.genedb.org//genedb/leish>.

## 4.2. TritrypDB

In January 2009, with financial support from the Bill & Melinda Gates Foundation, a beta-release version of a new component of the EuPathDB family of databases ([europathdb.org](http://europathdb.org)), the TriTrypDB ([tritrypdb.org](http://tritrypdb.org)), was announced. EuPathDB is considered a portal for accessing genomic-scale datasets associated with the eukaryotic pathogens (*Cryptosporidium*, *Giardia*, *Plasmodium*, *Toxoplasma*, *Trichomonas* and, more recently, *Leishmania* and *Trypanosoma*). The TritrypDB project represents a collaborative effort between the GeneDB team, researchers at the Seattle Biomedical Research Institute, and the EuPathDB team (an NIAID-supported Bioinformatics Resource Center) and, with less than one year of its creation, it has become a database with powerful bioinformatics tools able to integrate and to compare a large amount of data from the genomes of *T. brucei*, *T. cruzi*, *L. (L.) major*, *L. (L.) infantum* and *L. (V.) braziliensis*, and recently, *L. (L.) mexicana*, *L. (Sauvoleishmania) tarentolae*, *T. congolense* and *T. vivax*. Besides

the biological sequences and *in silico*-generated data, TriTrypDB also incorporates multiple experimental data sets and provides a refined Boolean search method that may be conducted by combining or subtracting a series of desired queries [51, 52]. Users' questions can be related to gene type (protein, snRNA, tRNA, snoRNA, rRNA coding sequences and misc RNA), gene putative functions (Gene Ontology terms and Enzyme Commission Number), cellular location (predicted transmembrane domains and signal peptides), transcript and protein expression (ESTs, microarray and mass spectrometry evidences), protein attributes (molecular weight, isoelectric point and 3D structure), protein features (epitope presence) and evolution (discovering orthologues and paralogs). In an example, one may search for all proteins from *L. (L.) major* which contain at least one predicted transmembrane domain AND a predicted signal peptide AND a high confidence epitope: the result for this search in *L. (L.) major* rescues 2 genes coding the surface antigen protein (GP63, metallo-peptidase).

It is also possible to do BLAST analysis and run Sequence Retrieval algorithm into the "Tools" link. FASTA files, including the newest *L. (S.) tarentolae* genome data, are available on the "Download" link.

It is important to highlight that a registered and logged user on TriTrypDB can save his/her search strategies and also add comments on his/her target-genes' page. The comments will be forwarded to the Annotation Center for review and possibly included in future versions of the Databank. As mentioned above, there is a partnership between TriTrypDB and GeneDB teams and both Databases should be maintained and updated evenly to keep a similar content (personal communication on the EuPathDB WorkShop 2009).

## 4.3. Regulatory Networks, Parasite Metabolism and Drug-target Discovery: KEGG, LeishCyc, TDR Targets, BRENDA and other web tools

The availability of the complete DNA sequence of the human and human pathogens genomes [31, 53] associated with the heavy expansion of information on chemical structures of known drugs and three-dimensional structures of potential new drug targets enables rational drug design [54-56]. Comparative genomic studies allow the identification of molecules or biochemical pathways that have already been targeted successfully in other pathogens [30, 31]. Concerning data integration into biochemical networks, the main databases to be consulted are KEGG and LeishCyc.

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a suite of databases and associated software for understanding and integrating the knowledge of the cell from its genomic information. It is maintained by Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo. This integrated resource contains databases for genomic, chemical and network information. It is designed to be a system for linking genomes to life at the cellular level, containing a complete set of genes and molecules (building blocks) linked in interaction networks (wiring diagrams) making possible to build a computational representation of the cell and the organism and enabling *in silico* analysis of biological systems [57, 58].

In 2009 an Australian group at the University of Melbourne developed the first biochemical pathways database devoted to *L. (L.) major*: LeishCyc ([www.leishcyc.org](http://www.leishcyc.org)) [59]. This database, a part of the BioCyc Project ([biocyc.org](http://biocyc.org)), was built based on the *L. (L.) major* genomic data (version 5.2) provided by the Sanger Institute and its main aim is to describe all chemical entities involved in *Leishmania* cellular processes (metabolites, proteins, enzymes, and parent genes), as well as their interactions. LeishCyc currently contains 8304 polypeptides, 1043 enzymatic reactions, 811 enzymes, and 785 compounds assembled into 214 metabolic pathways; all of them can be viewed at once by clicking on the Cellular Overview “Diagram” web link. Since a large amount of literature was consulted and experimental and bioinformatics studies were approached by the Australian research group, the LeishCyc development has refined the *L. (L.) major* genome data curation, finding errors in gene annotation, correcting them and increasing information about certain gene products [59].

Users can navigate into the target-gene pages where they may find a direct link to the GeneDB allowing rescue of genomic information. It is also possible to have access to specific metabolic pathways diagrams containing the target-proteins. *Leishmania* high-throughput data now has gained a practical and reduced-time-needed visualization of the results with the “Omics Viewer” tool of the LeishCyc database.

Chokepoints, reactions that consume unique substrates or synthesize unique products which are essential metabolites for parasite survival, can also be searched on the LeishCyc database, and enzymes that participate in unique chokepoints should be prioritized as potential drug targets [60].

The comparison between *Leishmania* and Human metabolic networks could be used to identify parasite-specific chokepoints. Besides LeishCyc, another useful tool to find chokepoints is The Pathway Hunter Tool (<http://pht.tu-bs.de/PHT/>). The search of druggable proteins in *Leishmania* genomes could be guided by rule of five (or ‘Lipinski’s rule of drug-likeness’). This rule sorts the drugs based on the physico-chemical properties that are necessary to increase the likelihood of oral bioavailability [61-63]. The Therapeutic Target Database (<http://xin.cz3.nus.edu.sg/group/cjtt/ttd.asp>) and DrugBank (<http://www.drugbank.ca>) are publicly available databases with information about virtually all known proteins, nucleic acid targets and drugs already described in the literature. They contain information about drugs and ligands directed to these targets [64, 65]. The use of docking softwares such as AUTODOCK allows prediction of the ability of small molecules to potentially fill protein pockets. However, current status of the *Leishmania* structural database restrains the utility of docking softwares (<http://www.sgpp.org>).

Several distinct metabolic pathways between parasite and host, such as purine salvage, peroxisome biogenesis, glycolysis and trypanothione redox-system, have been investigated and are target candidates to be tested in rational anti-leishmania drug discovery platforms. In addition to its importance in drug target discovery, the comparative analysis of the parasites’ genomes may reveal peerless parasitic en-

zymes able to convert a pro-drug in an active compound [66-68].

The accumulating genomic data for pathogens of neglected diseases is promising for the field of rational drug design. To facilitate the integration of data emerging from such studies and to help identify candidate drug targets in a user friendly platform the TDR Targets database (<http://tdrtargets.org>) was created [56]. In addition to assess the role of a gene in the pathogen, the user will rescue data on orthology relationships, a relevant issue to predict good drug targets and minimizing adverse effects. Importantly, when available, this database offers structural information on the target or related proteins, a helpful tool for drug designing. The physicochemical nature of small-molecule binding sites on the target and the availability of drug-like molecules directed to related proteins in other organisms may help prediction of druggability of a given target in a pathogen [63, 69].

Another useful database for rational drug discovery is BRENDA ([Braunschweig Enzyme Database](http://www.brenda-enzymes.org)), a comprehensive enzyme information system maintained and developed at the Institute of Biochemistry and Bioinformatics at the Technical University of Braunschweig, Germany (<http://www.brenda-enzymes.org>). BRENDA represents the largest freely available information system containing biochemical and molecular data on all classified enzymes as well as software tools for querying the database and calculating molecular properties. The data are manually curated and each entry is clearly linked to a literature reference, the origin organism and, where available, to the protein sequence [70].

The genomic information and curated databases make it possible to integrate data and to create metabolic networks *in silico*. Although *Leishmania spp.* functional data are far from complete, well characterized microorganisms such as *Escherichia coli* or *Saccharomyces cerevisiae* and their computational metabolic models can be used in the integration of high-throughput data to facilitate genome annotation process and drug target search. Chavali and co-workers reconstructed a metabolic network of *L. (L.) major*, the iAC560 [71], using a 560 genes repertoire (approximately 6.7% of the genome). The network encompasses 1112 reactions, most of which are gene related and the rest comprises intracellular reactions not associated to genes, or inter-compartment/extracellular transport. The authors proposed 25 new annotations based on iAC560 metabolic network. Seventeen were previously characterized as hypothetical proteins; the rest had an incorrect annotation, localization or EC (enzyme commission) classification. One example is the *LmjF23.1480* gene predicted to code for alanine racemase by the iAC560 metabolic network analysis, although no genes for alanine racemase had been predicted by KEGG or GeneDB. Using computational modeling the authors presented the theoretical network response to enzymatic inhibitors demonstrating the usefulness of network reconstruction for therapeutic targets research [71]. The comparison of metabolic network of *Leishmania* with that of humans can help identify peculiar features of *Leishmania* metabolism and aid the rational design of new drugs against leishmaniasis.

Another important database, mainly for immunologists, is ImmuneEpitope ([www.immuneepitope.org](http://www.immuneepitope.org)) which con-

tains ~1500 antibody or T cell response-related epitopes identified and described for trypanosomatids, being ~650 for the *Leishmania* genus. All information about the epitopes is presented there: reference, structure, source antigen, host, immunization and assay descriptions for B cell, T cell, MHC binding and MHC Ligand Elution. This is also an Analysis Resource containing a series of epitope prediction tools where the user can search for epitopes in protein sequences of interest using the suitable method for that particular case. It is important to remind that TriTrypDB uses ImmuneEpitope data on its “Protein Features – Epitope Presence” searching strategy. This kind of information is crucial for applied research on vaccine development.

## 5. APPROACHES FOR FUNCTIONAL ANALYSES OF THE PARASITE GENOME

The peculiarities of *Leishmania* gene expression among eukaryotes and the plasticity of its genome impaired rapid progress on parasite genetics studies. However, trypanosomatid researchers around the globe have developed a series of exceptionally useful approaches that enhanced the *Leishmania* molecular toolkit.

### 5.1. Forward and Reverse Genetics

The major difficulty in forward genetic approaches has always been the assumed asexual diploid life in *Leishmania* [72]. Recently, it has been shown that *Leishmania* clearly displays genetic exchange inside the sand fly vector [4]. Thus, infection of *Phlebotomus* could be a valuable experimental tool in investigation based on classic genetics.

Until the late 1980s there were no feasible direct means to conduct reverse genetics in *Leishmania* itself. It became evidently necessary to develop specific tools to facilitate trypanosomatid analyses. Tools such as transient and stable transfections, transposon-direct mutagenesis and gene replacement [73, 74] became major advantages in association with genome sequencing data. Circular shuttle vectors initially provided methods for functional genetic rescue of *Leishmania* mutants by complementation with DNA libraries [75]. Transfection of *Leishmania* with circular DNA molecules have allowed the identification of a variety of relevant genes related to pathogenesis and infectivity [75-77]. Transient and stable transfections of *Leishmania* are performed with vectors designed to be functional under the parasite odd regulatory genetic system [78, 79]. These vectors contain the genetic elements required for gene expression within the parasite. At the 5'UTR a dinucleotide AG associated with a polyuridine tract will work as the SL acceptor site and at the 3'UTR similar sequences will drive poly-A tail addition and cleavage. Any particular gene inserted in between these regions will be transcribed by the parasite RNA polymerase II in a promoterless background. On the other hand, RNA polymerase I drives strong initiation of transcription in trypanosomatids [80, 81].

Therefore, integration of a given gene in the rDNA locus is a convenient route for the overexpression under stable conditions. Targeted gene disruption is achievable in *Leishmania* by homologous recombination of linear DNA and it is a powerful and effective tool for genetic functional studies in the parasite [82]. Since gene replacement and double gene knockout strategies have been established in

*Leishmania* [74, 83], several essential and non-essential genes have been knocked-out and phenotypic changes have been analyzed [84-88]. Some of these studies confirmed the complexity of issues such as virulence, infectivity or pathogenesis [33-36]. Quite useful for gene function studies are the reporter genes, e.g. luciferase or GFP (green fluorescent protein), which are applicable to studying regulatory elements involved in mRNA stability or initiation of translation control [89-91] and to evaluate protein interactions or subcellular distribution.

In the toolkit of *Leishmania* reverse genetics a robust inducible kit was missing, such as the tetracycline repressor/inducer-dependent expression fully functional in *T. brucei* [92]. Recently, a reversible and tunable system has been adapted to *Leishmania*. It was originally described for mammalian cells and it is based on the targeting of FKBP (FK506-Binding Protein) to proteasomal degradation after rapamycin/FK506/Shld1 binding to its destabilizing domain [93]. This system was implemented in *L. (L.) major* and *L. (V.) braziliensis* to conditionally express destabilized UDP-galactopyranose mutase (UGM), a fundamental enzyme in lyphophosphoglycan (LPG) biosynthesis [94].

*Leishmania* reverse genetics is still lacking a functional and regulatable RNAi machinery to allow faster and global approaches of knocking down genes, despite the presence of orthologues of genes related to RNAi machinery in *L. (V.) braziliensis* [30, 95], as mentioned before. Therefore, reliable strategies such as the destabilization domain system represent promising benefits for molecular biology investigation [94].

### 5.2. *Leishmania* Biology Research Becomes High-Throughput

#### 5.2.1. Transcriptome and Proteome

Despite the fact that trypanosomatids are unicellular organisms, which should make genomic survey simpler than in more complex eukaryotes, they undergo drastic and diverse changes throughout infection of vertebrate and invertebrate hosts. Several efforts have been made to unveil the differential expression of genes during *Leishmania* growth and development in an attempt to explain modifications taking place during the parasite digenetic life-cycle. Understanding these modifications may help to comprehend genetic features supporting fitness of promastigotes and amastigotes for their distinct environments. Available genome data allow large-scale functional approaches to be undertaken. A diversity of methods or combination of these may be used such as RNA profiling (transcriptome), identification of the complete set of proteins (proteome) and the generation of genome-wide, single gene knockouts (or knockdowns via RNAi, a feasible goal). Besides expression profile comparison of different developmental stages, other comparative studies have been pursued to understand divergent behavior of *Leishmania* strains or species regarding drug resistance, clinical manifestations, tropism and host-pathogen interaction.

Transcriptome analyses based on microarrays have been extensively performed and proven to be an useful tool in *Leishmania* research [96-99]. So far, a variety of microarray analyses between diverse *Leishmania* samples revealed only small differences in expression profiling. Microarray ana-

lyses between procyclic and metacyclic promastigotes have revealed that only 1 to 2% of the genes (slides scoring ~7000 genes) are modulated by a 2-fold or larger factor in all studied conditions [100]. Preferential stage-specific genes constitute only ~0.2-5 % of the expressed genome (8160 genes surveyed) when analyzing promastigote and amastigote transcriptomes [98, 101]. Therefore, stage-specific genes are poorly represented in microarray analysis as shown for several *Leishmania* species (e.g. *L. (L.) donovani* axenic amastigotes: 5.5%) [100]. Altogether, most of the up-regulated genes are found in promastigotes.

An obstacle for amastigote versus promastigote comparative expression profile studies resides on the difficulty to obtain sufficient amounts of the intracellular form for the assays. With some of the *Leishmania* species it is feasible to obtain amastigotes in axenic conditions. These cells (axenic amastigotes) resemble amastigotes found in the vertebrate host and are widely used in life-cycle studies [102-104]. Nonetheless, recent global mRNA profiling analyses of axenic and intracellular amastigotes have shown important differences between their transcriptomes; they share only 12% of upregulated genes [105]. Therefore, it is necessary to go through validation steps before extrapolating data obtained with the axenic forms to intracellular amastigotes. A recent strategy to obtain pure amastigotes by combining macrophage/mice infection and fluorescence-activated cell sorting (FACS) has been described for *L. (L.) mexicana* expressing *DsRed* [106]. This approach provided a large parasite yield ( $> 2 \times 10^8$  cells/sort) mostly free of host cell material.

Microarray comparative analyses could not explain the differences observed between the life cycle stages of *Leishmania*. This probably happens because, considering genome architecture and demonstration of polycistronic nature of transcription in these parasites, it is believed that the whole genome is constitutively expressed [107] and that gene expression regulation happens at the levels of transcript stability and translational control, protein stability and post-

translational modifications. Therefore, transcriptomic analyses should be preferentially accompanied by proteomic approaches [97]. *Leishmania* post-translational modifications play important roles throughout the digenetic infectious life cycle and it has been emphasized that transcriptomic and proteomic analysis have a weak degree of correlation [97, 99, 107].

*Leishmania* proteome characterization based on two-dimensional gel electrophoresis (2DE) has been successfully conducted [108-110]. Even so, proteomics have not revealed evident differential expression between cell stages. Only 5% of proteins are differentially expressed between *L. (L.) infantum* promastigotes and amastigotes [111]. Around 90% of the *L. (L.) mexicana* proteome is qualitatively unchanged during the life cycle [108]. Apparently the most pronounced proteomic feature in promastigote-to-amastigote differentiation is the shifting from glucose to fatty acid catabolism which is explained by the lack of glucose availability inside the phagolysosome [112]. Amastigotes also synthesize more basic proteins [106], as shown for other microorganisms inhabiting acidic environments such as *Helicobacter pylori* [113] and *Coxiella burnetii* [114]. Still, proteomics is a more promising tool for estimating gene profiling than microarray analysis in *Leishmania*. Global protein profile analyses have contributed with noteworthy perception on energetic and metabolic requirements of amastigotes. Several hypothetical proteins have been identified by proteomic analysis [111, 115] confirming their expression for the first time. Studies have identified several protein isoforms differentially expressed during *Leishmania* life-cycle through successful enrichment of poorly represented proteins [99, 116]. Relevant results should come from characterization of organelle/structure-related proteomes and transcriptomes. The *T. brucei* flagellum proteome was completed and generated interesting further investigations [117].

New technologies based on isotope labeling (ICAT<sup>TM</sup> and iTRAQ<sup>TM</sup>) may solve some limitations of 2DE such as sub-representation of membrane bound and low-abundance

**Table 1. Important Web Resources on the Investigation of *Leishmania* Omics Data**

Web Tools	Capabilities	Address
1. GeneDB	1.1. Find genes (by Acc. numbers or text search) and run BLAST searches 1.2. Identify proteins containing a specific pattern (Motif Search) or by a query peptide mass/digestion type (Emowse)	1.1. <a href="http://www.genedb.org">http://www.genedb.org</a> 1.2. <a href="http://old.genedb.org//genedb/leish/">http://old.genedb.org//genedb/leish/</a>
2. TriTryp	Refined boolean search method, BLAST searches, sequence retrieval and more functions.	<a href="http://tritypdb.org">http://tritypdb.org</a>
3. KEGG	Find genome data and correlate genes with their specific(s) biochemical(s) pathway(s)	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>
4. LeishCyc	<i>L. (L.) major</i> metabolic pathways database: Cellular and Genome Overview and Omics Viewer for high-throughput data	<a href="http://leishcyc.bio21.unimelb.edu.au/">http://leishcyc.bio21.unimelb.edu.au/</a>
5. BRENDA	Comprehensive enzyme information system for rational drug discovery	<a href="http://www.brenda-enzymes.org/">http://www.brenda-enzymes.org/</a>
6. Pathway Hunter Tool (PHT)	Map reactions, metabolites and enzymes involved in chokepoints	<a href="http://pht.tu-bs.de/PHT/">http://pht.tu-bs.de/PHT/</a>
7. ImmuneEpitope	Database for antibody or T cell response-related epitopes	<a href="http://www.immuneepitope.org">www.immuneepitope.org</a>

proteins [107]. Proteome coverage is a major concern in studies of protein expression, because frequently only abundant proteins are detected. Around 2500 proteins have been identified so far in *Leishmania* genus covering approximately 30% of the predicted proteome [118]. Establishment of Multidimensional Protein Identification Technology (MudPIT) for *Leishmania* may be necessary to achieve a better unbiased proteome representation. In a recent study, Paape *et al.* used a gel free proteomic analysis to investigate *L. (L.) mexicana* proteins putatively secreted inside the host phagolysosome [118]. This approach allowed the identification of over 1000 novel proteins.

The prospect of a research filled with large amounts of information requires appropriate data handling. *In silico* tools based on the programming language Perl and the statistics package R are very useful for extracting and analyzing genomic data [119]. Both are free and computational open-source tools.

Taken together, data generated so far display an unexpected perspective related to *Leishmania* gene expression throughout parasitic development. Infectivity and contrasting adaptation may arise from the modulation of a small but relevant set of genes and/or the parasite life cycle is sustained by post-translational modifications yet to be investigated. Another possibility is the modulation of gene expression through chromatin-mediated epigenetic control between different forms.

### 5.2.2. Molecular Interactions on Chip

As described above, transcription initiation sites in *Leishmania* comprise divergent strand-switch regions at opposite polycistron junction sites [9]. It has been shown that such regions are enriched with acetylated H3 histone and transcription-related proteins, such as TATA-binding protein (TBP) and Small Nuclear Activating Protein complex (SNAP<sub>50</sub>), in *Leishmania* [120]. The group used a tiling array approach hybridized with DNA obtained from Chromatin ImmunoPrecipitation assays (ChIP-chip). Marked peaks of an acetylated form of the H3 histone were preferentially found in logarithmic instead of stationary phase promastigotes. Most of the acetylated H3 peaks were detected at divergent SSRs, although the presence of such peaks was shown at chromosome ends and within polycistronic gene clusters. These DGC-related peaks were associated with upstream TBP/SNAP50 peaks, evidence that they represent transcription start sites [120]. Therefore regulation of transcription initiation in *Leishmania* could be accomplished by controlling histone modification inside strand switch regions.

A similar work in *T. brucei* using sequencing-coupled ChIP (ChIP-seq) showed that transcription start sites (TSS) and termination sites (TTS) are enriched with specific distinct histone variants [121]. Noncanonical unstable nucleosomes are mainly associated to TSS in this organism, which results in an open conformation of chromatin in this region. A specific acetylated H4 histone is largely concentrated in TSS by a 300-factor over genomic background [121]. It would be also interesting to analyze the genomic distribution of histone variants in *Leishmania*. In *T. cruzi* sequencing of DNA from ChIP experiments revealed that

acetylated and methylated histones were found to be enriched in TSS but not in TTS [122].

Despite elevated costs, sequencing-coupled ChIP represents a large advantage over microarray analysis because it is not limited to the genome regions surveyed in the microarray slides and therefore potentially analyzes all types of enriched sequences throughout the genome [123]. Studies based on ChIP are promising and may still produce considerable data concerning transcription in *Leishmania*.

ChIP approaches rely mainly on the enrichment of interaction sites of a protein of interest. Protein microarrays, however, allow the identification of proteins bound to a known DNA sequence. There are three main types of protein microarrays used in biology research: analytical, functional and reverse phase microarrays [124]. Analytical microarrays provide insights at complex protein mixtures measuring protein profiling [125]. One example of analytical analysis is antibody microarray which may be used for important antigen screening in *Leishmania*. Functional microarrays are designed with full-length proteins or protein domains and are used for biochemical analysis of the entire proteome in a single run [125]. In this approach protein-protein, protein-DNA and protein-RNA interactions are targeted. In reverse phase microarrays, cellular proteins are bound to a nitrocellulose slide and screened for antibody interaction allowing the identification of labeled protein usually by a fluorescent/colorimetric signal. In contrast to DNA microarrays, there is a lag to the establishment of protein microarray in *Leishmania* research, due to the difficulty in the production of proteins compared with DNA. Different post-translational modifications may be surveyed by using protein microarrays, such as glycosylation and ubiquitination [126, 127].

### 5.2.3. Large-Scale Microscopy

Post-genomic era provided easier planning and development of mutations in specific genes to analyze their function and relevance in varied *Leishmania* biologic processes. Observation and analysis of modified pronounced phenotypes have been facilitated by the generation of feasible high-throughput large-scale microscopy. Systems based on confocal microplate readers such as the OPERA imaging platform (Evotec Technologies) are appropriate for the studies with parasitic microorganisms like *Leishmania*, mainly to analyze host-pathogen specific interactions and drug screening [128, 129]. These platforms recognize cells by fluorescence parameters, quantify the biologic process of interest and transform biological observations into numeric results.

## 6. PERSPECTIVES

The molecular comparisons provided by the experimental approaches discussed here may yet lead to a better appreciation of *Leishmania* biology and identification of novel targets for leishmaniasis control and treatment. The *Leishmania* toolkit available to research laboratories increases enormously in association with genome information. Several different approaches and combination of them may be chosen to answer biological questions under the light of the large data flow generated. In post-genomic era, trypanosomatid research is profiting from biological information made available by a combination of bioinformatics tools and databases; the standpoint for raising hypothesis has moved. The

technological advances in experimental large-scale approaches should catalyze *Leishmania*'s molecular genetics research and may facilitate drawing a "big picture" of the complex host-parasite interactions.

## ACKNOWLEDGEMENTS

The research laboratory is funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-2006/50323-7), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and WHO Special Programme for Research and Training in Tropical Diseases (TDR) (WHO/TDR). JST is supported by post-doctoral fellowship from CAPES/CNPq (Programa Nacional de Pós-Doutoramento (PNPD): 558966/2008-0; 151599/2008-4), EJRV is sponsored by PhD fellowship from FAPESP (2008/53929-9) and TRF is sponsored by MSc fellowship from FAPESP (2007/06443-0).

## REFERENCES

- [1] Sacks D, Kamhawi S. Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. *Annu Rev Microbiol* 2001; 55: 453-83.
- [2] Peters NC, Egen JG, Secundino N, et al. *In vivo* imaging reveals an essential role for neutrophils in leishmaniasis transmitted by sand flies. *Science* 2008; 321: 970-4.
- [3] Ritter U, Frischknecht F, van Zandbergen G. Are neutrophils important host cells for *Leishmania* parasites? *Trends Parasitol* 2009; 25: 505-10.
- [4] Akopyants NS, Kimblin N, Secundino N, et al. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science* 2009; 324: 265-8.
- [5] Rougeron V, De Meeus T, Hide M, et al. Extreme inbreeding in *Leishmania braziliensis*. *Proc Natl Acad Sci USA* 2009; 106: 10224-9.
- [6] Simpson L, Shaw J. RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* 1989; 57: 355-66.
- [7] Ibrahim ME, Mahdi MA, Bereir RE, Giha RS, Wasunna C. Evolutionary conservation of RNA editing in the genus *Leishmania*. *Infect Genet Evol* 2008; 8: 378-80.
- [8] Michels PA, Bringaud F, Herman M, Hannaert V. Metabolic functions of glycosomes in trypanosomatids. *Biochim Biophys Acta* 2006; 1763: 1463-77.
- [9] Martinez-Calvillo S, Yan S, Nguyen D, et al. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* 2003; 11: 1291-9.
- [10] Boothroyd JC, Cross GA. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5' end. *Gene* 1982; 20: 281-9.
- [11] Liang XH, Haritan A, Uliel S, Michaeli S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell* 2003; 2: 830-40.
- [12] Mensa-Wilmot K, Garg N, McGwire BS, et al. Roles of free GPIs in amastigotes of *Leishmania*. *Mol Biochem Parasitol* 1999; 99: 103-16.
- [13] Cruz AK, Titus R, Beverley SM. Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. *Proc Natl Acad Sci USA* 1993; 90: 1599-603.
- [14] Gueiros-Filho FJ, Beverley SM. Selection against the dihydrofolate reductase-thymidylate synthase (DHFR-TS) locus as a probe of genetic alterations in *Leishmania major*. *Mol Cell Biol* 1996; 16: 5655-63.
- [15] Genest PA, ter Riet B, Dumas C, et al. Formation of linear inverted repeat amplicons following targeting of an essential gene in *Leishmania*. *Nucleic Acids Res* 2005; 33: 1699-709.
- [16] Squina FM, Pedrosa AL, Nunes VS, Cruz AK, Tosi LR. Shuttle mutagenesis and targeted disruption of a telomere-located essential gene of *Leishmania*. *Parasitology* 2007; 134: 511-22.
- [17] Ubeda JM, Legare D, Raymond F, et al. Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biol* 2008; 9: R115.
- [18] Leprohon P, Legare D, Raymond F, et al. Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Res* 2009; 37: 1387-99.
- [19] Martinez-Calvillo S, Stuart K, Myler PJ. Ploidy changes associated with disruption of two adjacent genes on *Leishmania major* chromosome 1. *Int J Parasitol* 2005; 35: 419-29.
- [20] LeBowitz JH, Smith HQ, Rusche L, Beverley SM. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 1993; 7: 996-1007.
- [21] Ullu E, Matthews KR, Tschudi C. Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol* 1993; 13: 720-5.
- [22] Campbell DA, Thomas S, Sturm NR. Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect* 2003; 5: 1231-40.
- [23] Wallach M, Fong D, Chang KP. Post-transcriptional control of tubulin biosynthesis during leishmanial differentiation. *Nature* 1982; 299: 650-2.
- [24] Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* 2007; 156: 93-101.
- [25] Haile S, Papadopoulou B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Curr Opin Microbiol* 2007; 10: 569-77.
- [26] Wincker P, Ravel C, Blaineau C, et al. The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res* 1996; 24: 1688-94.
- [27] Uliana SR, Ruiz JC, Cruz AK. *Leishmania* genomics: where do we stand? chapter B02. In: bioinformatics in tropical disease research: a practical and case-study approach. Gruber A, Durham AM, Huynh C, del Portillo H, Eds. Bethesda, MD: National Library of Medicine, National Center for Biotechnology Information, 2008.
- [28] Myler PJ, Audleman L, DeVos T, et al. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci USA* 1999; 96: 2902-6.
- [29] Ivens AC, Peacock CS, Worley EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005; 309: 436-42.
- [30] Peacock CS, Seeger K, Harris D, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* 2007; 39: 839-47.
- [31] El-Sayed NM, Myler PJ, Blandin G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 2005; 309: 404-9.
- [32] Jackson AP. The evolution of amastigote surface glycoproteins in trypanosomatid parasites. *Mol Biol Evol* 2010; 27: 33-45.
- [33] Zhang WW, Matlashewski G. Loss of virulence in *Leishmania donovani* deficient in an amastigote-specific protein, A2. *Proc Natl Acad Sci USA* 1997; 94: 8807-11.
- [34] Hubel A, Krobitsch S, Horauf A, Clos J. *Leishmania major* Hsp100 is required chiefly in the mammalian stage of the parasite. *Mol Cell Biol* 1997; 17: 5987-95.
- [35] Joshi PB, Kelly BL, Kamhawi S, Sacks DL, McMaster WR. Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. *Mol Biochem Parasitol* 2002; 120: 33-40.
- [36] Hille JD, Zawadzki JL, McConville MJ, Coombs GH, Mottram JC. *Leishmania mexicana* mutants lacking glycosylphosphatidylinositol (GPI)-protein transamidase provide insights into the biosynthesis and functions of GPI-anchored proteins. *Mol Biol Cell* 2000; 11: 1183-95.
- [37] Zhang WW, Mendez S, Ghosh A, et al. Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J Biol Chem* 2003; 278: 35508-15.
- [38] Britto C, Ravel C, Bastien P, et al. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene* 1998; 222: 107-17.
- [39] Eschenlauer SC, Coombs GH, Mottram JC. PFPI-like genes are expressed in *Leishmania major* but are pseudogenes in other *Leishmania* species. *FEMS Microbiol Lett* 2006; 260: 47-54.
- [40] Smith DF, Peacock CS, Cruz AK. Comparative genomics: from genotype to disease phenotype in the leishmanias. *Int J Parasitol* 2007; 37: 1173-86.

- [41] Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc Natl Acad Sci USA 1998; 95: 14687-92.
- [42] Shi H, Tschudi C, Ullu E. An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. RNA 2006; 12: 2063-72.
- [43] Shi H, Ullu E, Tschudi C. Function of the Trypanosome Argonaute 1 protein in RNA interference requires the N-terminal RGG domain and arginine 735 in the Piwi domain. J Biol Chem 2004; 279: 49889-93.
- [44] Shi H, Tschudi C, Ullu E. Depletion of newly synthesized Argonaute1 impairs the RNAi response in *Trypanosoma brucei*. RNA 2007; 13: 1132-9.
- [45] Dumas C, Ouellette M, Tovar J, et al. Disruption of the trypanothione reductase gene of *Leishmania* decreases its ability to survive oxidative stress in macrophages. EMBO J 1997; 16: 2590-8.
- [46] Mottram JC, McCready BP, Brown KG, Grant KM. Gene disruptions indicate an essential function for the LmmCRK1 cdc2-related kinase of *Leishmania mexicana*. Mol Microbiol 1996; 22: 573-83.
- [47] Victoir K, Dujardin JC, de Doncker S, et al. Plasticity of gp63 gene organization in *Leishmania* (*Viannia*) *braziliensis* and *Leishmania* (*Viannia*) *peruviana*. Parasitology 1995; 111 ( Pt 3): 265-73.
- [48] Hertz-Fowler C, Peacock CS, Wood V, et al. GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Res 2004; 32: D339-43.
- [49] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000; 16: 944-5.
- [50] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-10.
- [51] Aurrecoechea C, Brestelli J, Brunk BP, et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 2009; 37: D539-43.
- [52] Aslett M, Aurrecoechea C, Berriman M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 2010; 38: D457-62.
- [53] El-Sayed NM, Myler PJ, Bartholomew DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005; 309: 409-15.
- [54] Wishart DS. *In silico* drug exploration and discovery using DrugBank. Curr Protoc Bioinformatics 2007; Chap. 14: Unit 14.4.
- [55] Davis AJ, Murray HW, Handman E. Drugs against leishmaniasis: a synergy of technology and partnerships. Trends Parasitol 2004; 20: 73-6.
- [56] Aguero F, Al-Lazikani B, Aslett M, et al. Genomic-scale prioritization of drug targets: the TDR targets database. Nat Rev Drug Discov 2008; 7: 900-7.
- [57] Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res 2002; 30: 42-6.
- [58] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004; 32: D277-80.
- [59] Doyle MA, MacRae JI, De Souza DP, et al. LeishCyc: a biochemical pathways database for *Leishmania major*. BMC Syst Biol 2009; 3: 57.
- [60] Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB. Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. Genome Res 2004; 14: 917-24.
- [61] Zhang MQ, Wilkinson B. Drug discovery beyond the 'rule-of-five'. Curr Opin Biotechnol 2007; 18: 478-88.
- [62] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001; 46: 3-26.
- [63] Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov 2002; 1: 727-30.
- [64] Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36: D901-6.
- [65] Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006; 34: D668-72.
- [66] Moyersoen J, Choe J, Fan E, Hol WG, Michels PA. Biogenesis of peroxisomes and glycosomes: trypanosomatid glycosome assembly is a promising new drug target. FEMS Microbiol Rev 2004; 28: 603-43.
- [67] Jaeger T, Flohé L. The thiol-based redox networks of pathogens: unexploited targets in the search for new drugs. Biofactors 2006; 27: 109-20.
- [68] Datta AK, Datta R, Sen B. Antiparasitic chemotherapy: tinkering with the purine salvage pathway. Adv Exp Med Biol 2008; 625: 116-32.
- [69] Myler PJ. Searching the Tritryp genomes for drug targets. Adv Exp Med Biol 2008; 625: 133-40.
- [70] Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. BRENDa, AMENDA and FRENDa the enzyme information system: new content and tools in 2009. Nucleic Acids Res 2009; 37: D588-92.
- [71] Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. Mol Syst Biol 2008; 4: 177.
- [72] Panton LJ, Tesh RB, Nadeau KC, Beverley SM. A test for genetic exchange in mixed infections of *Leishmania major* in the sand fly *Phlebotomus papatasii*. J Protozool 1991; 38: 224-8.
- [73] Clayton CE. Genetic manipulation of kinetoplastida. Parasitol Today 1999; 15: 372-8.
- [74] Cruz A, Coburn CM, Beverley SM. Double targeted gene replacement for creating null mutants. Proc Natl Acad Sci USA 1991; 88: 7170-4.
- [75] Beverley SM, Turco SJ. Identification of genes mediating lipophosphoglycan biosynthesis by functional complementation of *Leishmania donovani* mutants. Ann Trop Med Parasitol 1995; 89 Suppl 1: 11-7.
- [76] Turco S, Descoteaux A, Ryan K, Garraway L, Beverley S. Isolation of virulence genes directing GPI synthesis by functional complementation of *Leishmania*. Braz J Med Biol Res 1994; 27: 133-8.
- [77] Beverley SM, Turco SJ. Lipophosphoglycan (LPG) and the identification of virulence genes in the protozoan parasite *Leishmania*. Trends Microbiol 1998; 6: 35-40.
- [78] LeBowitz JH, Coburn CM, McMahon-Pratt D, Beverley SM. Development of a stable *Leishmania* expression vector and application to the study of parasite surface antigen genes. Proc Natl Acad Sci USA 1990; 87: 9736-40.
- [79] Curotto de Lafaille MA, Laban A, Wirth DF. Gene expression in *Leishmania*: analysis of essential 5' DNA sequences. Proc Natl Acad Sci USA 1992; 89: 2703-7.
- [80] Gay LS, Wilson ME, Donelson JE. The promoter for the ribosomal RNA genes of *Leishmania chagasi*. Mol Biochem Parasitol 1996; 77: 193-200.
- [81] Uliana SR, Fischer W, Stempliuk VA, Floeter-Winter LM. Structural and functional characterization of the *Leishmania amazonensis* ribosomal RNA promoter. Mol Biochem Parasitol 1996; 76: 245-55.
- [82] Beverley SM, Akopyants NS, Goyard S, et al. Putting the *Leishmania* genome to work: functional genomics by transposon trapping and expression profiling. Philos Trans R Soc Lond B Biol Sci 2002; 357: 47-53.
- [83] Cruz A, Beverley SM. Gene replacement in parasitic protozoa. Nature 1990; 348: 171-3.
- [84] Stewart J, Curtis J, Spurck TP, et al. Characterisation of a *Leishmania mexicana* knockout lacking guanosine diphosphate-mannose pyrophosphorylase. Int J Parasitol 2005; 35: 861-73.
- [85] Reguera RM, Balana-Fouce R, Showalter M, Hickerson S, Beverley SM. *Leishmania major* lacking arginase (ARG) are auxotrophic for polyamines but retain infectivity to susceptible BALB/c mice. Mol Biochem Parasitol 2009; 165: 48-56.
- [86] Jain M, Madhubala R. Characterization and localization of ORFF gene from the LD1 locus of *Leishmania donovani*. Gene 2008; 416: 1-10.
- [87] Mukherjee A, Roy G, Guimond C, Ouellette M. The gamma-glutamylcysteine synthetase gene of *Leishmania* is essential and involved in response to oxidants. Mol Microbiol 2009; 74(4): 914-27.
- [88] Spath GF, Schlesinger P, Schreiber R, Beverley SM. A novel role for Stat1 in phagosome acidification and natural host resistance to intracellular infection by *Leishmania major*. PLoS Pathog 2009; 5: e1000381.
- [89] Haile S, Dupe A, Papadopoulou B. Deadenylation-independent stage-specific mRNA degradation in *Leishmania*. Nucleic Acids Res 2008; 36: 1634-44.

- [90] Bringaud F, Muller M, Cerqueira GC, et al. Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog* 2007; 3: 1291-307.
- [91] Wu Y, El Fakhry Y, Sereno D, Tamar S, Papadopoulou B. A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins. *Mol Biochem Parasitol* 2000; 110: 345-57.
- [92] Wirtz E, Clayton C. Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science* 1995; 268: 1179-83.
- [93] Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 2006; 126: 995-1004.
- [94] Madeira da Silva L, Owens KL, Murta SM, Beverley SM. Regulated expression of the *Leishmania major* surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins. *Proc Natl Acad Sci USA* 2009; 106: 7583-8.
- [95] Beverley S. Retention and loss of RNA interference pathways in Trypanosomatid protozoans. Proceedings of the XXIV Meeting of the Brazilian Society of Protozoology 2008; Aguas de Lindoia, SP, Brazil; 2008.
- [96] Saxena A, Worthey EA, Yan S, et al. Evaluation of differential gene expression in *Leishmania major* Friedlin procyclics and metacyclics using DNA microarray analysis. *Mol Biochem Parasitol* 2003; 129: 103-14.
- [97] Duncan R. DNA microarray analysis of protozoan parasite gene expression: outcomes correlate with mechanisms of regulation. *Trends Parasitol* 2004; 20: 211-5.
- [98] Holzer TR, McMaster WR, Forney JD. Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *Leishmania mexicana*. *Mol Biochem Parasitol* 2006; 146: 198-218.
- [99] McNicoll F, Drummondsmith J, Muller M, et al. A combined proteomic and transcriptomic approach to the study of stage differentiation in *Leishmania infantum*. *Proteomics* 2006; 6: 3567-81.
- [100] Cohen-Freue G, Holzer TR, Forney JD, McMaster WR. Global gene expression in *Leishmania*. *Int J Parasitol* 2007; 37: 1077-86.
- [101] Lynn MA, McMaster WR. *Leishmania*: conserved evolution-diverse diseases. *Trends Parasitol* 2008; 24: 103-5.
- [102] Pan AA, Pan SC. *Leishmania mexicana*: comparative fine structure of amastigotes and promastigotes *in vitro* and *in vivo*. *Exp Parasitol* 1986; 62: 254-65.
- [103] Bates PA, Robertson CD, Tetley L, Coombs GH. Axenic cultivation and characterization of *Leishmania mexicana* amastigote-like forms. *Parasitology* 1992; 105 ( Pt 2): 193-202.
- [104] Doyle PS, Engel JC, Pimenta PF, da Silva PP, Dwyer DM. *Leishmania donovani*: long-term culture of axenic amastigotes at 37 degrees C. *Exp Parasitol* 1991; 73: 326-34.
- [105] Rochette A, Raymond F, Corbeil J, Ouellette M, Papadopoulou B. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *Leishmania infantum*. *Mol Biochem Parasitol* 2009; 165: 32-47.
- [106] Paape D, Lippuner C, Schmid M, et al. Transgenic, fluorescent *Leishmania mexicana* allow direct analysis of the proteome of intracellular amastigotes. *Mol Cell Proteomics* 2008; 7: 1688-701.
- [107] Leifso K, Cohen-Freue G, Dogra N, Murray A, McMaster WR. Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Mol Biochem Parasitol* 2007; 152: 35-46.
- [108] Nugent PG, Karsani SA, Wait R, Temporo J, Smith DF. Proteomic analysis of *Leishmania mexicana* differentiation. *Mol Biochem Parasitol* 2004; 136: 51-62.
- [109] Gongora R, Acester N, Quadroni M, et al. Mapping the proteome of *Leishmania Viannia* parasites using two-dimensional polyacrylamide gel electrophoresis and associated technologies. *Bio medica* 2003; 23: 153-60.
- [110] Walker J, Vasquez JJ, Gomez MA, et al. Identification of developmentally-regulated proteins in *Leishmania panamensis* by proteome profiling of promastigotes and axenic amastigotes. *Mol Biochem Parasitol* 2006; 147: 64-73.
- [111] El Fakhry Y, Ouellette M, Papadopoulou B. A proteomic approach to identify developmentally regulated proteins in *Leishmania infantum*. *Proteomics* 2002; 2: 1007-17.
- [112] Rosenzweig D, Smith D, Opperdoes F, et al. Retooling *Leishmania* metabolism: from sand fly gut to human macrophage. *FASEB J* 2008; 22: 590-602.
- [113] Jungblut PR, Bumann D, Haas G, et al. Comparative proteome analysis of *Helicobacter pylori*. *Mol Microbiol* 2000; 36: 710-25.
- [114] Seshadri R, Paulsen IT, Eisen JA, et al. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci USA* 2003; 100: 5455-60.
- [115] Cuervo P, De Jesus JB, Saboia-Vahia L, et al. Proteomic characterization of the released/secreted proteins of *Leishmania (Viannia) braziliensis* promastigotes. *J Proteomics* 2009; 73: 79-92.
- [116] Morales MA, Watanabe R, Laurent C, et al. Phosphoproteomic analysis of *Leishmania donovani* pro- and amastigote stages. *Proteomics* 2008; 8: 350-63.
- [117] Broadhead R, Dawe HR, Farr H, et al. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 2006; 440: 224-7.
- [118] Paape D, Barrios-Llerena ME, Le Bihan T, Mackay L, Aebsicher T. Gel free analysis of the Proteome of intracellular *Leishmania mexicana*. *Mol Biochem Parasitol* 2010; 169(2): 108-14.
- [119] Sturm NR, Martinez LL, Thomas S. Kinetoplastid genomics: the thin end of the wedge. *Infect Genet Evol* 2008; 8: 901-6.
- [120] Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ. Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* 2009; 10: 152.
- [121] Siegel TN, Hekstra DR, Kemp LE, et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* 2009; 23: 1063-76.
- [122] Respuela P, Ferella M, Rada-Iglesias A, Aslund L. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *J Biol Chem* 2008; 283: 15884-92.
- [123] Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009; 10: 669-80.
- [124] Tao SC, Chen CS, Zhu H. Applications of protein microarray technology. *Comb Chem High Throughput Screen* 2007; 10: 706-18.
- [125] Hall DA, Ptacek J, Snyder M. Protein microarray technology. *Mech Ageing Dev* 2007; 128: 161-7.
- [126] Gupta R, Kus B, Fladd C, et al. Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. *Mol Syst Biol* 2007; 3: 116.
- [127] Patwa TH, Zhao J, Anderson MA, Simeone DM, Lubman DM. Screening of glycosylation patterns in serum using natural glycoprotein microarrays and multi-lectin fluorescence detection. *Anal Chem* 2006; 78: 6411-21.
- [128] Osorio y Fortea J, Prina E, de La Llave E, et al. Unveiling pathways used by *Leishmania amazonensis* amastigotes to subvert macrophage function. *Immunol Rev* 2007; 219: 66-74.
- [129] Lang T, Lecoeur H, Prina E. Imaging *Leishmania* development in their host cells. *Trends Parasitol* 2009; 25: 464-73.

Received: November 15, 2009

Revised: August 10, 2010

Accepted: August 11, 2010

© de Toledo et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.