

Published in final edited form as:

*Int J Parasitol.* 2007 September ; 37(11): 1173–1186. doi:10.1016/j.ijpara.2007.05.015.

## Comparative genomics: From genotype to disease phenotype in the leishmaniases

Deborah F. Smith<sup>a,\*</sup>, Christopher S. Peacock<sup>b</sup>, and Angela K. Cruz<sup>c</sup>

<sup>a</sup> Immunology and Infection Unit, Department of Biology/Hull York Medical School, University of York, Heslington, York YO10 5YW, UK

<sup>b</sup> Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

<sup>c</sup> Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Brazil

### Abstract

Recent progress in sequencing the genomes of several *Leishmania* species, causative agents of cutaneous, mucocutaneous and visceral leishmaniasis, is revealing unusual features of potential relevance to parasite virulence and pathogenesis in the host. While the genomes of *Leishmania major*, *Leishmania braziliensis* and *Leishmania infantum* are highly similar in content and organisation, species-specific genes and mechanisms distinguish one from another. In particular, the presence of retrotransposons and the components of a putative RNA interference machinery in *L. braziliensis* suggest the potential for both greater diversity and more tractable experimentation in this *Leishmania Viannia* species.

### Keywords

*Leishmania*; Comparative genomics; Disease phenotype; Retrotransposons; RNAi

## 1. Introduction

Species of the genus *Leishmania* cause a diverse spectrum of infectious diseases, the leishmaniases, in tropical and subtropical regions of the world (reviewed in Murray et al., 2005). Belonging to the order Kinetoplastida and the family Trypanosomatidae, mammalian-infective species of the genus can be divided into two subgenera, *Leishmania (Leishmania)* and *Leishmania (Viannia)*, which differ principally in the site of parasite development within the female sandfly vector (reviewed in Sacks and Kamhawi, 2001). Once flagellated metacyclic promastigote stages are transmitted from vector to mammalian host, parasites enter resident tissue macrophages and transform into aflagellated replicative

© 2007 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

\*Corresponding author. Tel.: +44 01904 328843; fax: +44 01904 328844. E-mail address: dfs501@york.ac.uk (D.F. Smith).

**Publisher's Disclaimer:** This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit: <http://www.elsevier.com/copyright>

amastigotes, often inducing immuno-inflammatory responses and persistent tissue infection. The fate of these intracellular parasites determines disease type, which can range from cutaneous or mucocutaneous infection to diffuse cutaneous or visceral disease. The principal factors contributing to disease outcome are the infecting parasite species and immunological status of the host, including a well-established role for host genetic variation (reviewed in Lipoldova and Demant, 2006). Attributing the relative importance of these factors to disease outcome in humans is a complex area, however, particularly given the overlap in pathology associated with infection with different *Leishmania* species (Guerbouj et al., 2001; McMahon-Pratt and Alexander, 2004; Wilson et al., 2005; BenSaid et al., 2006). While multiple species cause cutaneous disease (see Table 1) and only three are usually considered as visceralising types (*Leishmania donovani*, *Leishmania infantum* and *Leishmania chagasi*), cutaneous lesions attributed to these latter species have also been reported (Gramiccia et al., 1995; Noyes et al., 1997; Gumy et al., 2004; BenSaid et al., 2006), while isolated cases of visceral disease due to cutaneous species have been recorded as well (Barral et al., 1991). Variable disease phenotype has also been observed in naturally and experimentally infected animals. Teasing out the relative importance of all contributing factors to clinical disease is a challenging objective.

A number of parasite factors affecting pathogenesis in the host have already been well-characterised. These include the surface zinc metalloprotease GP63 (or leishmanolysin) which protects promastigotes against host lysis by cleavage of complement, enhances phagocytosis and has recently been shown to protect against antimicrobial peptide-induced apoptotic killing (reviewed in Yao et al., 2005; Kulkarni et al., 2006); several parasite-derived cysteine proteases which can function as virulence factors (Mottram et al., 2004) and the complex glycoconjugates, lipophosphoglycan and the proteophosphoglycans, which promote parasite survival in the vector and host by a range of mechanisms (reviewed by Beverley and Turco, 1998; Ilg, 2000; Sacks and Kamhawi, 2001; Turco et al., 2001; Descoteaux and Turco, 2002; Lodge and Descoteaux, 2005).

Sequencing the genomes of representative parasite species provides a global framework within which to investigate the contribution of these and other parasite factors to the diverse forms of leishmaniasis and address the question: what is the contribution of parasite genotype to disease phenotype?

## 2. Sequencing the *Leishmania* genomes: which species?

The decision to initiate sequencing programmes for three kinetoplastid organisms, *Leishmania*, *Trypanosoma brucei* and *Trypanosoma cruzi* (the Tritryp genome projects), was made in principal at a Oswaldo Cruz Foundation (FIOCRUZ) – World Health Organization Special Programme for Research and Training in Tropical Diseases (WHO/TDR) Parasite Genome Network Planning Meeting in Rio de Janeiro in 1994. Although the available technology for completion of such large genomes was lacking at that time, these projects were considered feasible given the organisation of a committed consortium of participating laboratories, sufficient funding and anticipated advances in sequencing technology. Thirteen years later, three completed *Leishmania* genomes (*Leishmania major*, *L. infantum*, *Leishmania braziliensis*) are in the public domain (Ivens et al., 2005; Peacock et al., 2007), with sequencing of a fourth (*Leishmania mexicana*) in progress (<http://www.genedb.org/>). These first four species were chosen to represent the main species complexes of the *L. Leishmania* subgenus together with the best-characterised species of the *L. Viannia* subgenus. The strains selected were all infective, allowing study of host responses in suitable rodent models in vivo, and adapted for maintenance and manipulation in the laboratory (Laurentino et al., 2004; Ivens et al., 2005; Denise et al., 2006).

Importantly, the three completed genomes represent species that usually give rise to distinct disease types (Table 1).

*Leishmania major* was the first species to be sequenced and has provided the model for subsequent genomic analyses. *L. major* is a causative agent of cutaneous leishmaniasis (CL), the most common disease form that is usually self-healing in humans but leaves unsightly scars in those affected (Table 1). Widespread in Africa and Asia, CL caused by *L. major* is predominantly a zoonotic infection, with parasites maintained in rodent reservoir hosts following transmission by phlebotomine sandflies. In the laboratory, *L. major* has been one of the principal experimental species used to dissect the roles of T<sub>H</sub>1/T<sub>H</sub>2 cells in response to infection in susceptible and resistant hosts (reviewed in Sacks and Noben-Trauth, 2002).

The case for sequencing a visceralising species of *Leishmania* as the second genome to be analysed was over-whelming: visceral leishmaniasis (VL) is the most serious disease form and frequently fatal if left untreated. Closely related species of the *L. donovani* complex are found in different geographical regions: *L. donovani* is the primary cause of VL in the Indian subcontinent and East Africa, *L. infantum* in the Mediterranean region and *L. chagasi* in the New World (Table 1). Humans are the only known reservoir of *L. donovani* while canines, especially domestic and stray dogs, provide the reservoir hosts for *L. infantum* and *L. chagasi*. While the last two species are generally considered to be genetically identical (Mauricio et al., 2000), all three species are very similar as demonstrated in recent comprehensive phylogenetic studies using both non-coding and protein-coding sequences (Kuhls et al., 2005, 2007; Mauricio et al., 2006; Lukes et al., 2007). The choice of a prevalent *L. infantum* strain for genome sequencing was made on the basis of its virulence in animals, transmissibility in sandflies and adaptability to laboratory experimentation (Denise et al., 2006).

The New World species *L. braziliensis*, within the subgenus *L. (Viannia)*, is the third and most divergent species sequenced to date – and the principal cause of CL and mucocutaneous leishmaniasis (MCL) in tropical America. Originally associated with close contact with the forest environment, it is now accepted that *L. braziliensis* transmission may occur in peri-domestic and domestic habitats. More recently, studies on the reservoir hosts strongly suggest that a number of rodent species are involved, including domestic rats (Grimaldi and Tesh, 1993; Brandao-Filho et al., 1994; Oliveira et al., 2005).

The most severe pathology associated with this species is the disease affecting mucous membranes, mainly the anterior nasal region and occasionally the pharynx and larynx, which can result in destruction of the cartilagenous tissue. Most frequently, mucosal disease occurs after the appearance of cutaneous lesions and its diagnosis may happen months or years after treatment of the primary lesion (Zajtchuk et al., 1989). However, recent clinical investigation indicates that mucosal infection may be present, but not diagnosed, concomitantly with the primary cutaneous lesion (Boaventura et al., 2006). This is a relevant issue because the patient usually does not recognise MCL until partial destruction of the nose mucosa and/or cartilage occurs. Treatment of such cases is often difficult and death may occur as a result of uncontrollable secondary infections.

### 3. The *L. major* genome

The *L. major* genome is composed of both nuclear and independently replicating mitochondrial or kinetoplast DNA. The structure and function of kinetoplastid DNA have been reviewed recently (Lukes et al., 2005) and will not be discussed further here.

Prior to publication of the Tritryp kinetoplastid genomes in 2005, relatively little was known about the overall physical architecture of kinetoplastid chromosomes. However, earlier

studies on antigenic variation in *T. brucei* had defined telomeric expression sites for variant surface glycoprotein (VSG) as polycistronic transcription units (reviewed in Pays et al., 1994), while work on RNA processing mechanisms in *T. brucei* and *Leishmania* had demonstrated linkage between 3'-polyadenylation sites and 5'-*trans*-splicing of adjacent genes, suggesting a coupled mechanism that could be processive along a chromosome (LeBowitz et al., 1993; Ullu et al., 1993; Matthews et al., 1994). The first whole chromosome sequence of *L. major* (Myler et al., 1999) revealed an unusual pattern of gene distribution, consistent with the polycistronic transcription model. Of the 79 identified genes, 29 were present as a contiguous unit on one DNA strand with the remaining 50 as a second unit on the opposite strand. Further chromosome sequencing revealed that the genes on other *L. major* chromosomes are similarly aligned in long arrays that appear to be transcribed as a single unit (Martinez-Calvillo et al., 2003; see Fig. 1) prior to *trans*-splicing and polyadenylation. These directional gene clusters (DGCs) range in size from a few to hundreds of genes stretching over 1 Mb of DNA. DGCs are separated by AT-rich strand switch regions (SSRs) believed to contain sites for transcription initiation and termination (Martinez-Calvillo et al., 2004). Whole genome microarray experiments have recently confirmed that most genes are constitutively transcribed in *Leishmania* (Holzer et al., 2006; Leifso et al., 2007). Differential gene expression between different life cycle stages is subsequently mediated by downstream processing events affecting mRNA stability and translation (Boucher et al., 2002; Myung et al., 2002).

The DGC type of chromosomal organization is remarkably well-conserved in all the kinetoplastids sequenced to date, with most of the *L. major* chromosomes being completely syntenic with corresponding regions in the larger *T. brucei* chromosomes (El-Sayed et al., 2005a), despite separation of these two species by at least 200 million years of evolution (Stevens et al., 2001). DGCs do not contain clusters of genes of related function similar to the organisation of prokaryotic operons, however.

In contrast to the high conservation of DGC gene order shared by *L. major* and the other Trypanosoma, organization of the *L. major* chromosome ends is distinctive, with short and highly conserved subtelomeric regions bordering distinctive repeat sequences that precede heterogeneous telomeres (Ivens et al., 2005). Interestingly, there are no large subtelomeric multigene families encoding proteins involved in host immune evasion – these are features not only of the *T. brucei* and *T. cruzi* genomes but also of other protozoan pathogen genomes such as those of *Plasmodium* spp. (Hall et al., 2002). Associated with this lack of telomeric repeated genes is a lack of transposable elements in the *L. major* genome (Bringaud et al., 2006) – these mobile sequence units are associated with gene rearrangements in the subtelomeric regions of *Trypanosoma* spp. (Berriman et al., 2005).

#### 4. The *L. major* genome – gene content

The 33 Mb genome of *L. major* is predicted to contain ~8,300 genes, a figure likely to be accurate given the distinct codon bias in the coding regions and the rarity of *cis*-splicing in this species, features that make gene prediction reliable (Ivens et al., 2005; Peacock et al., 2007). Within this gene complement are a number of tandemly arrayed gene families, a kinetoplastid feature that may overcome the limitations of polycistronic transcription by increasing gene dosage (see below). Properties of the *L. major* genome, in comparison with the *L. infantum*, *L. braziliensis* and *T. brucei* genomes, are listed in Table 2.

The *L. major* genes can be divided into three broad categories. Around 85% encode the core proteome common to all kinetoplastids. Many of these function in metabolism and can be described as house-keeping genes, similar to those found in higher eukaryotes. A second category contains genes restricted to kinetoplastids, with no orthologues in higher

eukaryotes. These include sequences encoding cytoskeletal proteins, RNA editing components and other metabolic enzymes. A third category appears to be restricted to *Leishmania*: these number ~1,000, ~12% of the genome (El-Sayed et al., 2005a). Many of these genus-specific genes have no conserved or recognisable functional domains and it is reasonable to predict that some of them function in host/pathogen interactions.

Annotation of the *L. major* genome benefited considerably from the comparative genomics approach developed in the Tritryp projects: based on chromosomal position and sequence similarity, orthologues of genes experimentally characterised in the trypanosomes could be identified in *Leishmania* and vice versa. This approach was used to great effect in the identification of *L. major* metabolic pathways, which are more extensive than in either *T. brucei* or *T. cruzi* (Berriman et al., 2005). *Leishmania* metabolism reflects parasite lifestyle in the vector and the host and is influenced by the availability of nutrients in both locations (reviewed in Zilberstein, 1993; Tielens and Van Hellemond, 1998). As with the other Tritryps, *Leishmania* lacks the biosynthetic pathways for essential amino acids which must therefore be host-derived, explaining the presence of the large amino acid transporter gene families found in *L. major* (Ivens et al., 2005). Similarly, while all the Tritryps encode the full machinery for uptake and degradation of glucose (Berriman et al., 2005), *L. major* alone appears to have the capability to hydrolyse disaccharide sugars, probably as a consequence of sandfly nectar feeding that provides an alternative source of energy for the promastigote stage of the parasite (Muller and Schlein, 2004). *L. major* also encodes enzymes required for haem synthesis, which are lacking from the other Tritryps (Ivens et al., 2005). Haem is an important *Leishmania* growth factor that is required in both extracellular and intracellular stages of the life cycle (Chang and Chang, 1985). The last three enzymes in the biosynthetic pathway, encoded solely by *Leishmania*, share a high similarity to genes in prokaryotic pathways, like many of the apparent kinetoplastid gene acquisitions that lead to new metabolic activities – in the Tritryp analysis, as many as 50 genes were identified as likely candidates for lateral gene transfer from prokaryotic sources (Berriman et al., 2005). It should be noted, however, that synthesis of haem has not been demonstrated biochemically in *Leishmania* and genes coding for enzymes required for the rest of the pathway have yet to be identified.

While *L. major* lacks the highly expanded and divergent large subtelomeric immune evasion gene families common to other protozoan pathogens, the sub-telomeric regions are not totally devoid of genes, with some sequence families found in these locations (see Fig. 1). The largest and best studied gene family codes for a family of phosphoglycan sc $\beta$ -galactosyltransferases (SCGs). These enzymes are involved in side chain modifications of the phosphoglycan repeats of the major promastigote glycoconjugate, LPG (Dobson et al., 2003). There are 14 members spread across three subfamilies (*SCG*, *SCGR* and *SCGL*) but seven of the “core” *SCG* genes are found exclusively at distinct subtelomeric sites (Dobson et al., 2006). Interestingly, *SCG* expression does not appear to be restricted to a single location that might support an antigenic variation model in *Leishmania*, similar to that operating with the *T. brucei* VSGs (reviewed in Barry and McCulloch, 2001). Further analysis is required to rigorously test this “single-cell exclusive-expression” model (Dobson et al., 2006). It is clear, however, that these sub-telomeric locations can facilitate gene conversion events that may ultimately support the strain-specific diversity in LPG galactosylation known to be critical for successful sandfly transmission of different *Leishmania* strains and species (reviewed in Sacks and Kamhawi, 2001; Kamhawi, 2006).

In contrast to the sub-telomeric *SCG* genes, the tandemly arrayed gene families coding for *Leishmania* surface proteins are found at internal chromosome sites. A good example is the *GP63* gene family which encodes glycoproteins critical for host invasion and virulence (Joshi et al., 2002; Yao et al., 2003). These zinc metalloproteases are not confined to

*Leishmania* species however: a single tandem *GP63* gene array is present in all three Trityp species (El-Sayed et al., 2005a), while recent studies suggest that GP63-type proteins are conserved and functional in insect and plant trypanosomatid species (Masini Masini d'Avila-Levy et al., 2006) as well as the more distant *Trichomonas vaginalis* (Carlton et al., 2007).

In *Leishmania*, the *GP63* locus is limited to four to six copies in *L. major* and *L. infantum* but expanded approximately eightfold in the *L. braziliensis* genome (Peacock et al., 2007). Studies with multiple isolates of *L. braziliensis* and *Leishmania peruviana* (both *L. Viannia* spp., see Table 1) have shown extensive size variation and evidence of gene re-arrangement at this locus during long-term maintenance in the laboratory, suggesting that the *GP63* locus is susceptible to dynamic change in this sub-genus (Victoir et al., 1995). This does not seem to be the case in a more limited study in *L. major* (Sadlova et al., 2006). Recent analysis of sequence diversity within a single *GP63* gene copy in multiple *L. donovani* complex strains has confirmed earlier speculation that evolution of this family is influenced by mosaic gene conversion and provides strong genetic evidence for GP63 as an important marker for host preference in infection (Mauricio et al., 2006).

Another family of leucine-rich membrane surface proteins, the PSA-2 or GP46 family, is restricted to *Leishmania* spp.; analysis at both the DNA and protein levels failed to detect GP46 in the *L. braziliensis* complex (McMahon-Pratt et al., 1992). Confirming this earlier study, only one *PSA-2*-like gene has been annotated in the *L. braziliensis* genome. This is in contrast to the genome strains of *L. major* (with 25 *PSA-2* genes) and *L. infantum* (also with a large gene complement, although the precise copy number has been difficult to determine accurately). The *PSA-2*s are known to bind to host cell macrophages and protect the parasite from complement-mediated lysis during infection with *L. chagasi* and *L. infantum* (Kedzierski et al., 2004; Lincoln et al., 2004). In *L. infantum*, different strains that cause either visceral or cutaneous disease show considerable correlative variation in the size of the repeat regions containing the *PSA-2* genes (Guerbouj et al., 2001).

The amastin genes, the largest family in the *Leishmania* genome, encode small surface-expressed glycoproteins of as yet unknown function that are conserved in *T. cruzi*. Importantly, a majority of these genes are expressed specifically in intracellular amastigotes in *L. major* and therefore presumably contribute to survival in the host (Rochette et al., 2005). Unlike most other large *Leishmania* gene families, the amastin genes are not present in a single array but found on seven different chromosomes in the *L. major* genome (Ivens et al., 2005). Most of these sites contain small groups of genes (one to four copies) but two large arrays are also present, with the largest containing amastin genes alternating with the highly conserved tuzin genes; a similar arrangement is found in *T. cruzi* but is absent from *T. brucei* (El-Sayed et al., 2005b). The second large gene array, not present in *T. cruzi*, contains amastin genes separated by degenerate tuzin genes. A likely explanation for this gene complexity is that a common kinetoplastid ancestor possessed a single amastin/tuzin array, which underwent duplication, re-arrangement and diversification after separation of the *Leishmania* and *Trypanosoma* lineages. With the exception of a lone copy on *L. major* chromosome 36, *L. infantum* and *L. braziliensis* have amastin gene arrays at the same locations, although sequence read depth suggests that the large amastin array with intact tuzin genes is shorter in *L. braziliensis*.

Other *Leishmania*-specific proteins that are not necessarily surface-exposed may also be important regulators of host-parasite interactions. One interesting example is an apparent homologue of the human cytokine macrophage inhibitory factor (MIF). The first cytokine to be identified, MIF is also one of the most pleiotropic, implicated in a wide range of cellular activities that could influence parasite survival (Calandra and Roger, 2003). Addition of purified recombinant human MIF to *L. major*-infected macrophages activates parasite killing

by inducible nitric oxide synthase and TNF- $\alpha$  (Juttner et al., 1998). In malaria, variations in the expression of MIF have been associated with susceptibility to high density parasitemia, while reduced MIF production may lead to increased severity of disease (Awandare et al., 2006). MIF-deficient mice also show increased susceptibility to *L. major* (Satoskar et al., 2001), highlighting the important role of this cytokine in mounting an effective immune response to infection.

MIF sequences are not confined to the host, however. All three *Leishmania* spp. sequenced to date have two similar *MIF* gene copies located at a strand switch region (see <http://www.genedb.org/>). Functional analysis of these sequences may be of considerable interest, given that MIF expression in the filarial worm *Brugia malayi* influences the progression of the CD4+ T cell response to infection towards a T<sub>H</sub>2 pathway (Falcone et al., 2001). The *Leishmania MIF* genes show greatest similarity to bacterial orthologues, suggesting that they too may have been acquired by lateral gene transfer following diversification from the trypanosome lineage.

## 5. Comparison of the *L. major*, *L. infantum* and *L. braziliensis* genomes

From an evolutionary perspective, phylogenetic analyses have suggested a neotropical origin for the *Leishmania* genus (Stevens et al., 2001) and, while there has been some controversy in this designation (see Kerr, 2000), this has been largely resolved in a recent multifactorial genetic study (Lukes et al., 2007). Irrespective of this debate, *L. braziliensis* is the most genetically and biologically divergent of the three sequenced species. Divergence between the *Leishmania* species complexes is estimated to have occurred 15–50 million years ago (Lukes et al., 2007), within the same range as two potential host species, mouse and human. Given this period of isolation, it was expected that there would be significant differences in both genome architecture and gene repertoire between *L. braziliensis*, *L. infantum* and *L. major*. Indeed, while the genomes have a similar DNA content of around 33 Mb, karyotypic differences had already been identified by linkage group analysis (Britto et al., 1998): *L. major* and *L. infantum*, in common with other Old World species, have a haploid content of 36 chromosomes, while the New World species have either 35 (*L. braziliensis* complex) or 34 (*L. mexicana* complex). These differences were shown to be due to fusion of pairs of chromosomes (chromosomes 20 + 34 in *L. braziliensis*; chromosomes 8 + 29 and 20 + 36 in *L. mexicana*), with the former observation now confirmed in the sequencing project.

Surprisingly, comparison of the respective orthologous chromosomes has revealed remarkable conservation of both gene content and gene order in all three genome species. Despite the differences in gene copy number within some of the major protein-coding families described above, not a single chromosomal re-arrangement has been identified between *L. major* and *L. infantum* across the whole genome, while *L. braziliensis* has only a few possible sequence re-arrangements (Peacock et al., 2007). Equally surprising, from the total content of ~8,300 genes in each species, only ~200 can be identified as differentially distributed between the three genomes. The most divergent, *L. braziliensis*, possesses 47 genes that are absent from the other two species. In comparison, *L. infantum* has 27 species-specific genes while *L. major* has only five (see Fig. 2). A number of the other differentially distributed genes are found in two out of the three species. Some of these species-specific sequences have already been analysed at the molecular level. Examples include the *L. infantum A2* gene that encodes an amastigote-specific repeat-containing protein previously characterised in *L. donovani*, the only *Leishmania* sequence to date that confers a change in virulence phenotype when introduced into *L. major* by genetic transfection (Zhang et al., 2003); and the HASP and SHERP genes, expressed from a single locus (absent in *L. braziliensis*) in infective stages of *L. major* and *L. infantum*, with their protein products

localizing to the plasma membrane and intracellular membranes, respectively, in these species (Denny et al., 2000; Knuepfer et al., 2001).

In the Tritryp genome analyses, most genes specific to each of the representative species were found either at the ends of the DGCs or in the subtelomeric regions of the chromosomes, regions that appear to be more tolerant to genome re-arrangement. However, comparison of the three *Leishmania* genomes has revealed that gene variation is not predominantly restricted to the subtelomeric regions or even the SSRs but is evenly distributed across the genome (Peacock et al., 2007).

*Leishmania* is also distinctive from other eukaryotes in the apparent mechanism by which species-specific gene variation occurs. Whereas insertions/deletions and sequence re-arrangements play major roles in gene diversification in most other eukaryotes characterised to date, degeneration of existing genes (leading to probable loss of function) accounts for ~80% of the species differences in *Leishmania*. These degenerate sequences have in-frame stop codons and frame shifts, generating truncated open reading frames that are presumably not translated. One example is the gene encoding cysteine peptidase *Pfp1*, which is present as an intact gene and translated in *L. major* (Eschenlauer et al., 2006). However, there are five in-frame stop codons and a frame shift in the *L. infantum* orthologue, while the syntenic region in *L. braziliensis* is even more degenerate.

*Pfp1* like some of the other species-specific genes, appears to be another candidate for lateral gene transfer from bacteria. Of the remaining species-specific sequences not caused by loss of function, many also fall into this category. One example is the cyclopropane fatty acyl synthase (*CFAS*) gene, present in *L. infantum* and *L. braziliensis* but absent from *L. major*. Acquisition of novel genes in this way may be a mechanism for environmental adaptation to promote survival; similar adaptations to stress or other stimuli may lead to the redundancy of other sequences clearly identified as pseudogenes in the *Leishmania* genomes (Peacock et al., 2007). In the case of *CFAS*, acquisition of this gene may have an impact on parasite survival in the host, since the *CFAS* orthologue in *Mycobacterium tuberculosis* is associated with increased virulence and persistence, functions that apparently require cyclopropanation of a mycolic acid substrate in the bacterial cell wall (Rao et al., 2005).

Despite its chromosomal plasticity (Martinez-Calvillo et al., 2005), the incredible conservation of synteny revealed by comparative genomic analyses of these three species suggests that the *Leishmania* genome is highly stable and has not undergone major genomic re-arrangements during speciation. One contributing factor to this stability could be a lack of mobile DNA elements, as originally demonstrated in the *L. major* genome and verified in *L. infantum*. The comparative sequencing project has revealed some surprising observations, however, one of the most striking being the presence of transposable elements in *L. braziliensis*.

## 6. Retrotransposons in *L. braziliensis*

Repeat sequences are a common feature of both prokaryotic and eukaryotic genomes, with different categories contributing 50% or more to some eukaryotic genomes. Among these groups are the mobile genetic elements which, by virtue of their ability to move to different chromosomal locations, are considered important in shaping the course of genome evolution. Mobile elements are widespread in the biological kingdom. When actively proliferating (Wickstead et al., 2003), they can create and destroy or modify existing genes and regulatory elements, thus reshaping the structure and function of their host genomes. When no longer able to proliferate (“dead repeats”), they are still important markers for evolutionary events, a palaeonto-logical record.



Although some protozoan parasites carry no mobile elements or any vestige of them (Katinka et al., 2001; Carlton et al., 2002; Gardner et al., 2002; Xu et al., 2004), different types of transposable elements have been found in kinetoplastids. The genomes of *T. brucei* and *T. cruzi* contain both highly reiterated interspersed elements (Bhattacharya et al., 2002; Wickstead et al., 2003; El-Sayed et al., 2005b) and retroelements, transposons that require reverse transcription from an RNA intermediate. Depending on their mechanism of transposition, these retroelements are classified either as long terminal repeat (LTR) retroelements (or retrotransposons) or non-LTR retroelements (the retroposons), which include the short and long interspersed nuclear elements (Singer, 1982).

In some kinetoplastids, retrotransposons (VIPER), site-specific retroposons (SLACS, CZAR, CRE) and non-site-specific retroposons (Ingi and L1Tc) have been described (Kimmel et al., 1987; Murphy et al., 1987; Martin et al., 1995; Teng et al., 1995; Wickstead et al., 2003). *Trypanosoma cruzi* bears the highest number of repetitive sequences, which occupy at least 50% of the parasite genome. Long terminal repeat (LTR) and non-LTR retroelements account for ~5% and 2% of the haploid *T. cruzi* and *T. brucei* genomes, respectively (El-Sayed et al., 2005b). These mobile elements probably played an important role in shaping these parasites' genome organization and evolution.

In contrast, analysis of several *Leishmania* species originally indicated a lack of active retroelements (Bhattacharya et al., 2002). In the *L. major* and *L. infantum* genomes, it was possible to detect remnants of retroposons, named DIREs (degenerate ingi/L1Tc-related elements), which had been previously described in trypanosomes (Ghedini et al., 2004; Bringaud et al., 2006). Other relics of retroelements, distinct from *ingi/L1Tc*, were also detected in *L. major*. These are conserved in some Old and New World *Leishmania* species, but absent from the *T. cruzi* and *T. brucei* genomes (Pedrosa et al., 2006).

Against this background, in addition to these degenerate retroelements, it was surprising to detect potentially active retroposons in the genome of *L. braziliensis* (Peacock et al., 2007). Similar to the organization described for the SLACS and CZAR elements in *T. brucei* and *T. cruzi*, these non-LTR retrotransposons are associated with the spliced leader array in *L. braziliensis* (Aksoy et al., 1987; Villanueva et al., 1991).

In addition, the organization of the *L. braziliensis* telomeres is remarkably distinct from the Old World species. They accommodate the so called TATEs (Telomeric Associated Transposable Elements), a family of 20–30 novel DNA transposable elements, each including putative reverse transcriptase, phage integrase (site-specific recombinase) and DNA/RNA polymerase domains. These are highly conserved retroelements with a precise site of insertion within the telomeric hexamer repeats and a typical TT duplication flanking the transposon. The clusters of tandemly arrayed TATEs at the ends of chromosomes resemble ribosomal mobile element (RIME)/Ingi retroelements that are also organised in clusters within the subtelomeric regions of *T. brucei* chromosomes (Bringaud et al., 2002; Peacock et al., 2007). Based on current annotation of the *L. braziliensis* genome, TATEs are present on at least 12 chromosomes, precisely inserted in the middle of the GGGTTA repeats (Peacock et al., 2007). The TATE-containing *L. braziliensis* chromosome ends share longer stretches of sequence with a higher number of heterologous chromosomes, compared with the corresponding chromosomes in *L. major* (Brito, L.O. et al., unpublished data). On the other hand, only remnants of TATE elements are detected at *L. braziliensis* internal chromosomal sites. Besides the bias of abundance of these elements at the chromosome ends, selection against disruption of coding regions at chromosome central domains could explain TATE distribution patterns. To understand the role of the TATEs in shaping the *L. braziliensis* genome, it will be necessary to improve genome assembly and further analyse TATE distribution in different strains and species of the *Viannia* subgenus.

## 7. RNAi in *L. braziliensis*

Among different functional classes of non-coding RNAs, there is a large class of post-transcriptional regulators implicated in gene expression control in various organisms. Two classes of small RNAs, the short interfering RNAs (siRNAs) and the microRNAs (miRNAs) have been extensively studied in gene silencing. Initially described in *Caenorhabditis elegans* (Lee et al., 1993; Wightman et al., 1993), the term RNA interference (RNAi) was used to describe these events (Fire et al., 1998) and link them to similar post-transcriptional silencing phenomena in plants and fungi (Jorgensen, 1990; Romano and Macino, 1992).

Currently, RNAi is the best-known mechanism of gene silencing, an evolutionary conserved process in which double-stranded RNA induces the sequence-specific degradation of homologous mRNA (Fire et al., 1998). Besides its role in the control of endogenous mRNAs, the process seems to have an important defensive role against viruses and mobile element activity.

An RNAi mechanism was rapidly identified in *T. brucei* (Ngo et al., 1998) but has not been demonstrated in other kinetoplastids including *L. major* and *T. cruzi* (Robinson and Beverley, 2003; El-Sayed et al., 2005b). Therefore, one of the most unexpected differences between the sequenced *Leishmania* genomes is the identification of genes implicated in the RNAi pathway in *L. braziliensis*. One of the hallmarks of this pathway is *Dicer* activity, which converts double-stranded RNA (dsRNA) into small interfering RNA (siRNA). Annotation of the genomes of three *Leishmania* species and the trypanosomes did not reveal a *Dicer* orthologue (Berriman et al., 2005; El-Sayed et al., 2005b; Ivens et al., 2005; Peacock et al., 2007). Instead, it was speculated that a combination of independent proteins carrying the relevant domains DSRBD, DEAD/H box RNA helicase and RNase III could be responsible for *Dicer* activity in *T. brucei*. Orthologues of genes carrying these domains have also been annotated in the *L. braziliensis* genome (see Fig. 3 for an outline of the putative RNAi machinery in *L. braziliensis*, compared with *T. brucei*).

Curiously, it has recently been shown that a *T. brucei* *Dicer*-like protein (TbDcl1), quite divergent from the typical *Dicer* proteins described for other eukaryotes, is required for the generation of siRNA-size molecules and for RNAi (Shi et al., 2006a). Interestingly, a TbDcl1 gene is also present in *L. braziliensis* (Peacock et al., 2007), although there are no orthologues in *T. cruzi*, *L. major* or *L. infantum*.

Another crucial component of the RNAi pathway is Argonaute, an endonuclease involved in the dsRNA-triggered cleavage of mRNA. Unlike *L. major* and *L. infantum*, *L. braziliensis* contains one orthologue for *Ago1*, the functional argonaute gene of *T. brucei*. *TbAGO1* can be functionally replaced by the human gene *Argonaute 2*, which suggests that *TbAGO1* has the endonuclease activity involved in mRNA target degradation in the trypanosome RNAi pathway (Shi et al., 2006b). The *L. braziliensis* gene contains the typical argonaute domains PAZ and PIWI, the latter containing conserved amino acid residues that are essential for functional TbAGO1 (Shi et al., 2004). In addition, an N-terminal RGG domain, present in TbAGO1 and shown to be essential for its association with polyribosomes, is also present in the *L. braziliensis* *Ago1* gene (Shi et al., 2004). Thus, the *L. braziliensis* gene product may also participate in the control of translation.

It is an intriguing observation that RNAi machinery is present in only one of the three, highly syntenic *Leishmania* species sequenced to date. While this could indicate a lack of biological relevance for this pathway in the parasite, a more interesting possibility is that the availability of RNAi in *L. braziliensis* plays a role in the biological differences observed between the two subgenera. The demonstration of RNAi and its biological consequences in *L. braziliensis* awaits further experimentation.

Whether RNAi machinery has been lost or acquired across the kinetoplastid species and genera is an open question. Examination of the syntenic regions in *L. major* and *L. infantum* identified remnants of *Argonaute* in both, an observation that favours the proposal that RNAi was lost in the subgenus *Leishmania*. This speculation is reinforced by the conservation of key RNAi members between organisms as distant as *T. brucei* and humans (Shi et al., 2006a,b). It is relevant to stress that RNA viruses have been detected in species of the subgenus *L. Viannia* (Guilbride et al., 1992; Widmer and Dooley, 1995), suggesting that *L. braziliensis* could have retained RNAi as an antiviral defense mechanism (although there is also a single report of an RNAi virus in *L. major*; Scheffter et al., 1995). In addition, RNAi could have a role in regulating the function of the *L. braziliensis*-specific transposable elements described above. A role for RNAi as an endogenous mechanism to hamper expression of specific genes and defend cells from viruses and mobile elements has been demonstrated in a range of organisms, including *T. brucei* (Wickstead et al., 2003; Ullu et al., 2004).

The discovery of a *Leishmania* RNAi mechanism could also be of practical importance, given the wide and effective use of inducible RNAi methods to study gene function in other eukaryotes. An RNAi system for reducing gene expression would be especially relevant in *Leishmania*, a diploid pathogen that lacks a sexual cycle, and provide an interesting alternative to the classical two-step gene knockout strategy (Cruz et al., 1991). RNAi has been rigorously tested for in *L. major* and *L. donovani* with no success (Robinson and Beverley, 2003). Detection of an RNAi machinery so similar to that of *T. brucei* in *L. braziliensis*, however, suggests that this mechanism may be shown to be a useful tool in *L. braziliensis*. A word of caution however: the genetic plasticity of the *Leishmania* genome, which impedes the generation of null mutants for essential genes, may also interfere with the outcome of RNAi experiments. In fact, downregulation of *T. brucei* essential genes by RNAi led to selection of cell lines that escaped from the deleterious effect of RNA interference (Ullu et al., 2004). Notwithstanding this concern, the combination of RNAi and classical reverse genetic approaches promises to bring novel insights into the functional biology of *Leishmania* parasites.

## 8. How do the genome sequences impact on our understanding of parasite tropism?

It is premature to speculate on how knowledge of the genome sequences of different *Leishmania* species will impact on our global understanding of disease phenotype. However the primary comparative analysis (Peacock et al., 2007) supports some preliminary observations. Firstly, the genes that differ between the three species are good candidates for functional analysis to determine their roles in establishment of infection. With the *A2* gene as a precedent (Zhang et al., 2003), it can be speculated that one or more of these sequences, expressed separately or together, may impact upon the course of disease. Testing this hypothesis will require the generation and functional testing of specific mutations in suitable models of infection. Murine models of CL are well-established, with experimental *L. major* infection reproducing distinct aspects of the clinical spectrum of human disease (reviewed in Gummy et al., 2004). The establishment of a dermal infection model, that introduces small numbers of infective *L. major* metacyclics into the host dermis via sandfly feeding, more closely mimics the natural course of parasite transmission (Belkaid et al., 1998). Murine models for *L. infantum* and *L. braziliensis* are not as well advanced, partly because these species are more difficult to grow and differentiate in vitro and partly because of the variable sensitivity of model mouse strains to infection with these parasites. Progress has recently been made, however, in the development of dermal models for both *L. infantum* and *L. braziliensis* infection (Ahmed et al., 2003; de Moura et al., 2005). For all species, more work

is required to relate the immune correlates of protection in mice to those that are important in human and, as appropriate, canine infection.

Second, as most of the species-specific genes encode proteins sharing little identity with other eukaryotic sequences, the use of a combination of genetic, biochemical and immunological analyses will be essential to help delineate relevant biological functions. It will also be necessary to unravel the importance of those tandemly repeated genes that vary in copy number between species, given that this mechanism can support increased gene expression and potentially lead to detectable phenotypic effects. In this context, the recently published whole genome microarray analyses suggest that RNA expression is largely constitutive throughout the *Leishmania* life cycle with few genes showing increased expression at the RNA level in different parasite stages (Holzer et al., 2006; Leifso et al., 2007). Further studies are required to verify these observations in different parasite species and correlate RNA and protein expression levels, using high resolution proteomic methods (Foucher et al., 2006).

A third observation from the comparative genome analysis is that, despite the well-conserved synteny between the three model species, *L. braziliensis* may have the capacity for more extensive genomic reorganisation, given the putative retrotransposons and RNAi machinery found in this species. Clearly, whether RNAi actually operates as a functional mechanism in *L. braziliensis* awaits definitive experimental validation, while the functional relevance of the retroelements requires more extensive multiple strain analysis to determine possible levels of sequence variation. However, it cannot be discounted that retrotransposons play a role in shaping the *L. braziliensis* genome, impacting on the different disease outcomes associated with this species.

## 9. Future perspectives

Sequencing the genomes of three *Leishmania* species, with a fourth in progress, has provided a wealth of new information that is reshaping research initiatives for leishmaniasis. Extending this type of analysis to further species and virulent strains may reveal further surprises. Of most immediate practical application perhaps are the unlimited number of potential vaccine candidates that are now available and require robust evaluation, to identify the most favourable molecules to incorporate into clinical trials (Kedzierski et al., 2006). Similarly, the *Leishmania*-specific metabolic pathways revealed by comparative analysis are underpinning focused biochemical analyses to identify valid drug targets for development. New, more sensitive and specific diagnostic reagents for the leishmaniasis should also emerge from exploring the different genome sequences.

More fundamentally, genomic knowledge is revealing fascinating insights into *Leishmania* biology. In particular, the current comparative analyses are generating testable hypotheses relevant to understanding the importance of parasite genotype in pathogenesis. It is possible that just a few species-specific parasite genes are crucial to disease outcome and/or alternatively, that expression levels of key conserved parasite genes differ considerably both between and within species (perhaps as a consequence of variation in gene copy number). It is also possible that the parasite genome may play only a minor role in determining disease phenotype. In-depth functional analyses may ultimately unravel the relative contribution of parasite genes to disease progression in this spectrum of crippling infections.

## Acknowledgments

Research in the laboratory of D.F.S. is supported by the Wellcome Trust and the BBSRC and in the laboratory of A.K.C., by FAPESP, CNPq and TDR/WHO. C.P. is also supported by the Wellcome Trust through core funding to

the Pathogen Sequencing Unit, Sanger Institute. Space restraints require us to restrict the number of references for this review; we apologise to those colleagues whose work has not been cited here.

## References

- Ahmed S, Colmenares M, Soong L, Goldsmith-Pestana K, Munstermann L, Molina R, McMahon-Pratt D. Intradermal infection model for pathogenesis and vaccine studies of murine visceral leishmaniasis. *Infect. Immun.* 2003; 71:401–410. [PubMed: 12496190]
- Aksoy S, Lalor TM, Martin J, Van der Ploeg LH, Richards FF. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. *EMBO J.* 1987; 6:3819–3826. [PubMed: 2828038]
- Awandare GA, Ouma C, Keller CC, Were T, Otieno R, Ouma Y, Davenport GC, Hittner JB, Ong'echa JM, Ferrell R, Perkins DJ. A macrophage migration inhibitory factor promoter polymorphism is associated with high-density parasitemia in children with malaria. *Genes Immun.* 2006; 7:568–575. [PubMed: 16929348]
- Barral A, Pedral-Sampaio D, Grimaldi G Junior, Momen H, McMahon-Pratt D, Ribeiro de Jesus A, Almeida R, Badaro R, Barral-Netto M, Carvalho EM, et al. Leishmaniasis in Bahia, Brazil: evidence that *Leishmania amazonensis* produces a wide spectrum of clinical disease. *Am. J. Trop. Med. Hyg.* 1991; 44:536–546. [PubMed: 2063957]
- Barry JD, McCulloch R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv. Parasitol.* 2001; 49:1–70. [PubMed: 11461029]
- Belkaid Y, Kamhawi S, Modi G, Valenzuela J, Noben-Trauth N, Rowton E, Ribeiro J, Sacks DL. Development of a natural model of cutaneous leishmaniasis: powerful effects of vector saliva and saliva preexposure on the long-term outcome of *Leishmania major* infection in the mouse ear dermis. *J. Exp. Med.* 1998; 188:1941–1953. [PubMed: 9815271]
- BenSaid M, Guerbouj S, Saghrouni F, Fathallah-Mili A, Guizani I. Occurrence of *Leishmania infantum* cutaneous leishmaniasis in central Tunisia. *Trans. R. Soc. Trop. Med. Hyg.* 2006; 100:521–526. [PubMed: 16356518]
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivans A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM. The genome of the African trypanosome *Trypanosoma brucei*. *Science.* 2005; 309:416–422. [PubMed: 16020726]
- Beverley SM, Turco SJ. Lipophosphoglycan (LPG) and the identification of virulence genes in the protozoan parasite *Leishmania*. *Trends Microbiol.* 1998; 6:35–40. [PubMed: 9481823]
- Bhattacharya S, Bakre A, Bhattacharya A. Mobile genetic elements in protozoan parasites. *J. Genet.* 2002; 81:73–86. [PubMed: 12532039]
- Boaventura VS, Cafe V, Costa J, Oliveira F, Bafica A, Rosato A, de Freitas LA, Brodskyn C, Barral-Netto M, Barral A. Concomitant early mucosal and cutaneous leishmaniasis in Brazil. *Am. J. Trop. Med. Hyg.* 2006; 75:267–269. [PubMed: 16896130]
- Boucher N, Wu Y, Dumas C, Dube M, Sereno D, Breton M, Papadopoulou B. A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element. *J. Biol. Chem.* 2002; 277:19511–19520. [PubMed: 11912202]
- Brandao-Filho SP, de Carvalho FG, de Brito ME, Almeida Fde A, Nascimento LA. American cutaneous leishmaniasis in Pernambuco, Brazil: eco-epidemiological aspects in 'Zona da Mata' region. *Mem. Inst. Oswaldo Cruz.* 1994; 89:445–449. [PubMed: 7476230]

- Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell*. 2002; 1:137–151. [PubMed: 12455980]
- Bringaud F, Ghedin E, Blandin G, Bartholomeu DC, Caler E, Levin MJ, Baltz T, El-Sayed NM. Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol. Biochem. Parasitol*. 2006; 145:158–170. [PubMed: 16257065]
- Britto C, Ravel C, Bastien P, Blaineau C, Pages M, Dedet JP, Wincker P. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene*. 1998; 222:107–117. [PubMed: 9813266]
- Calandra T, Roger T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat. Rev. Immunol*. 2003; 3:791–800. [PubMed: 14502271]
- Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK, Quackenbush J, Sedegah M, Shoaihi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*. 2002; 419:512–519. [PubMed: 12368865]
- Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UC, Besteiro S, Sicheritz-Ponten T, Noel CJ, Dacks JB, Foster PG, Simillion C, Van de Peer Y, Miranda-Saavedra D, Barton GJ, Westrop GD, Muller S, Dessi D, Fiori PL, Ren Q, Paulsen I, Zhang H, Bastida-Corcuera FD, Simoes-Barbosa A, Brown MT, Hayes RD, Mukherjee M, Okumura CY, Schneider R, Smith AJ, Vanacova S, Villalvazo M, Haas BJ, Pertea M, Feldblyum TV, Utterback TR, Shu CL, Osoegawa K, de Jong PJ, Hrdy I, Horvathova L, Zubacova Z, Dolezal P, Malik SB, Logsdon JM Jr, Henze K, Gupta A, Wang CC, Dunne RL, Upcroft JA, Upcroft P, White O, Salzberg SL, Tang P, Chiu CH, Lee YS, Embley TM, Coombs GH, Mottram JC, Tachezy J, Fraser-Liggett CM, Johnson PJ. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*. 2007; 315:207–212. [PubMed: 17218520]
- Chang CS, Chang KP. Heme requirement and acquisition by extracellular and intracellular stages of *Leishmania mexicana amazonensis*. *Mol. Biochem. Parasitol*. 1985; 16:267–276. [PubMed: 4058483]
- Cruz A, Coburn CM, Beverley SM. Double targeted gene replacement for creating null mutants. *Proc. Natl. Acad. Sci. USA*. 1991; 88:7170–7174. [PubMed: 1651496]
- Denise H, Poot J, Jimenez M, Ambit A, Herrmann DC, Vermeulen AN, Coombs GH, Mottram JC. Studies on the CPA cysteine peptidase in the *Leishmania infantum* genome strain JPCM5. *BMC Mol. Biol*. 2006; 7:42. [PubMed: 17101050]
- de Moura TR, Novais FO, Oliveira F, Clarencio J, Noronha A, Barral A, Brodskyn C, de Oliveira CI. Toward a novel experimental model of infection to study American cutaneous leishmaniasis caused by *Leishmania braziliensis*. *Infect. Immun*. 2005; 73:5827–5834. [PubMed: 16113301]
- Denny PW, Gokool S, Russell DG, Field MC, Smith DF. Acylation-dependent protein export in *Leishmania*. *J. Biol. Chem*. 2000; 275:11017–11025. [PubMed: 10753904]
- Descoteaux A, Turco SJ. Functional aspects of the *Leishmania donovani* lipophosphoglycan during macrophage infection. *Microbes Infect*. 2002; 4:975–981. [PubMed: 12106791]
- Dobson DE, Scholtes LD, Valdez KE, Sullivan DR, Mengeling BJ, Cilmi S, Turco SJ, Beverley SM. Functional identification of galactosyltransferases (SCGs) required for species-specific modifications of the lipophosphoglycan adhesin controlling *Leishmania major*-sand fly interactions. *J. Biol. Chem*. 2003; 278:15523–15531. [PubMed: 12604613]
- Dobson DE, Scholtes LD, Myler PJ, Turco SJ, Beverley SM. Genomic organization and expression of the expanded SCG/L/R gene family of *Leishmania major*: internal clusters and telomeric localization of SCGs mediating species-specific LPG modifications. *Mol. Biochem. Parasitol*. 2006; 146:231–241. [PubMed: 16464509]
- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN,

Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 2005a; 309:404–409. [PubMed: 16020724]

- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasc AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 2005b; 309:409–415. [PubMed: 16020725]
- 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. [PubMed: 16790017]
- Falcone FH, Loke P, Zang X, MacDonald AS, Maizels RM, Allen JE. A *Brugia malayi* homolog of macrophage migration inhibitory factor reveals an important link between macrophages and eosinophil recruitment during nematode infection. *J. Immunol.* 2001; 167:5348–5354. [PubMed: 11673551]
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998; 391:806–811. [PubMed: 9486653]
- Foucher AL, Papadopoulou B, Ouellette M. Prefractionation by digitonin extraction increases representation of the cytosolic and intracellular proteome of *Leishmania infantum*. *J. Proteome Res.* 2006; 5:1741–1750. [PubMed: 16823982]
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 2002; 419:498–511. [PubMed: 12368864]
- Ghedin E, Bringaud F, Peterson J, Myler P, Berriman M, Ivens A, Andersson B, Bontempi E, Eisen J, Angiuoli S, Wanless D, Von Arx A, Murphy L, Lennard N, Salzberg S, Adams MD, White O, Hall N, Stuart K, Fraser CM, El-Sayed NM. Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* 2004; 134:183–191. [PubMed: 15003838]
- Gramiccia M, Gradoni L, Troiani M. Heterogeneity among zymodemes of *Leishmania infantum* from HIV-positive patients with visceral leishmaniasis in south Italy. *FEMS Microbiol. Lett.* 1995; 128:33–38. [PubMed: 7744236]
- Grimaldi G Jr, Tesh RB. Leishmaniases of the New World: current concepts and implications for future research. *Clin. Microbiol. Rev.* 1993; 6:230–250. [PubMed: 8358705]
- Guerbouj S, Guizani I, Speybroeck N, Le Ray D, Dujardin JC. Genomic polymorphism of *Leishmania infantum*: a relationship with clinical pleomorphism? *Infect. Genet. Evol.* 2001; 1:49–59. [PubMed: 12798050]
- Guilbride L, Myler PJ, Stuart K. Distribution and sequence divergence of LRV1 viruses among different *Leishmania* species. *Mol. Biochem. Parasitol.* 1992; 54:101–104. [PubMed: 1518522]
- Gumy A, Louis JA, Launois P. The murine model of infection with *Leishmania major* and its importance for the deciphering of mechanisms underlying differences in Th cell differentiation in

- mice from different genetic backgrounds. *Int. J. Parasitol.* 2004; 34:433–444. [PubMed: 15013733]
- Hall N, Pain A, Berriman M, Churcher C, Harris B, Harris D, Mungall K, Bowman S, Atkin R, Baker S, Barron A, Brooks K, Buckee CO, Burrows C, Cherevach I, Chillingworth C, Chillingworth T, Christodoulou Z, Clark L, Clark R, Corton C, Cronin A, Davies R, Davis P, Dear P, Dearden F, Doggett J, Feltwell T, Goble A, Goodhead I, Gwilliam R, Hamlin N, Hance Z, Harper D, Hauser H, Hornsby T, Holroyd S, Horrocks P, Humphray S, Jagels K, James KD, Johnson D, Kerhornou A, Knights A, Konfortov B, Kyes S, Larke N, Lawson D, Lennard N, Line A, Maddison M, McLean J, Mooney P, Moule S, Murphy L, Oliver K, Ormond D, Price C, Quail MA, Rabbinowitsch E, Rajandream MA, Rutter S, Rutherford KM, Sanders M, Simmonds M, Seeger K, Sharp S, Smith R, Squares R, Squares S, Stevens K, Taylor K, Tivey A, Unwin L, Whitehead S, Woodward J, Sulston JE, Craig A, Newbold C, Barrell BG. Sequence of *Plasmodium falciparum* chromosomes 1, 3–9 and 13. *Nature.* 2002; 419:527–531. [PubMed: 12368867]
- 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. [PubMed: 16430978]
- Ilg T. Proteophosphoglycans of *Leishmania*. *Parasitol. Today.* 2000; 16:489–497. [PubMed: 11063860]
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchetti G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'Neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ. The genome of the kinetoplastid parasite, *Leishmania major*. *Science.* 2005; 309:436–442. [PubMed: 16020728]
- Jorgensen R. Altered gene expression in plants due to trans interactions between homologous genes. *Trends Biotechnol.* 1990; 8:340–344. [PubMed: 1366894]
- 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. [PubMed: 11849703]
- Juttner S, Bernhagen J, Metz CN, Rollinghoff M, Bucala R, Gessner A. Migration inhibitory factor induces killing of *Leishmania major* by macrophages: dependence on reactive nitrogen intermediates and endogenous TNF- $\alpha$ . *J. Immunol.* 1998; 161:2383–2390. [PubMed: 9725234]
- Kamhawi S. Phlebotomine sand flies and *Leishmania* parasites: friends or foes? *Trends Parasitol.* 2006; 22:439–445. [PubMed: 16843727]
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivares CP. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature.* 2001; 414:450–453. [PubMed: 11719806]
- Kedzierski L, Montgomery J, Bullen D, Curtis J, Gardiner E, Jimenez-Ruiz A, Handman E. A leucine-rich repeat motif of *Leishmania* parasite surface antigen 2 binds to macrophages through the complement receptor 3. *J. Immunol.* 2004; 172:4902–4906. [PubMed: 15067069]
- Kedzierski L, Zhu Y, Handman E. *Leishmania* vaccines: progress and problems. *Parasitology.* 2006; 133:S87–S112. [PubMed: 17274851]
- Kerr SF. Palaeartic origin of *Leishmania*. *Mem. Inst. Oswaldo Cruz.* 2000; 95:75–80. [PubMed: 10656708]

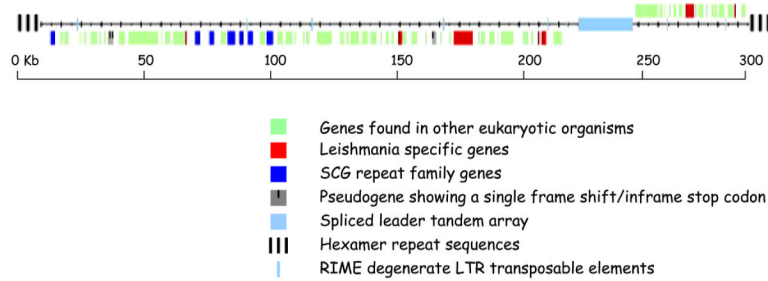


- Kimmel BE, ole-MoiYoi OK, Young JR. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Mol. Cell Biol.* 1987; 7:1465–1475. [PubMed: 3037321]
- Knuepfer E, Stierhof YD, McKean PG, Smith DF. Characterization of a differentially expressed protein that shows an unusual localization to intracellular membranes in *Leishmania major*. *Biochem. J.* 2001; 356:335–344. [PubMed: 11368759]
- Kuhls K, Mauricio IL, Pratlong F, Presber W, Schonian G. Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex. *Microbes Infect.* 2005; 7:1224–1234. [PubMed: 16002315]
- Kuhls K, Keilonat L, Ochsenreither S, Schaar M, Schweynoch C, Presber W, Schonian G. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. *Microbes Infect.* 2007; 9:334–343. [PubMed: 17307010]
- Kulkarni MM, McMaster WR, Kamysz E, Kamysz W, Engman DM, McGwire BS. The major surface-metalloprotease of the parasitic protozoan, *Leishmania*, protects against antimicrobial peptide-induced apoptotic killing. *Mol. Microbiol.* 2006; 62:1484–1497. [PubMed: 17074074]
- Laurentino EC, Ruiz JC, Fazelinia G, Myler PJ, Degrave W, Alves-Ferreira M, Ribeiro JM, Cruz AK. A survey of *Leishmania braziliensis* genome by shotgun sequencing. *Mol. Biochem. Parasitol.* 2004; 137:81–86. [PubMed: 15279954]
- 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. [PubMed: 8504937]
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell.* 1993; 75:843–854. [PubMed: 8252621]
- 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. [PubMed: 17188763]
- Lincoln LM, Ozaki M, Donelson JE, Beetham JK. Genetic complementation of *Leishmania* deficient in PSA (GP46) restores their resistance to lysis by complement. *Mol. Biochem. Parasitol.* 2004; 137:185–189. [PubMed: 15279966]
- Lipoldova M, Demant P. Genetic susceptibility to infectious disease: lessons from mouse models of leishmaniasis. *Nat. Rev. Genet.* 2006; 7:294–305. [PubMed: 16543933]
- Lodge R, Descoteaux A. Modulation of phagolysosome biogenesis by the lipophosphoglycan of *Leishmania*. *Clin. Immunol.* 2005; 114:256–265. [PubMed: 15721836]
- Lukes J, Hashimi H, Zikova A. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr. Genet.* 2005; 48:277–299. [PubMed: 16215758]
- Lukes J, Mauricio IL, Schonian G, Dujardin JC, Soteriadou K, Dedet JP, Kuhls K, Tintaya KW, Jirku M, Chocholova E, Haralambous C, Pratlong F, Obornik M, Horak A, Ayala FJ, Miles MA. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc. Natl. Acad. Sci. USA.* 2007; 104:9375–9380. [PubMed: 17517634]
- Martin F, Maranon C, Olivares M, Alonso C, Lopez MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzymes. *J. Mol. Biol.* 1995; 247:49–59. [PubMed: 7534829]
- Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell.* 2003; 11:1291–1299. [PubMed: 12769852]
- Martinez-Calvillo S, Nguyen D, Stuart K, Myler PJ. Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot. Cell.* 2004; 3:506–517. [PubMed: 15075279]
- 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; 35:419–429. [PubMed: 15777918]
- Masini d'Avila-Levy C, de Almeida Dias F, Nogueira de Melo AC, Martins JL, De Carvalho Santos Lopes AH, Souza Dos Santos AL, Vermelho AB, Branquinha MH. Insights into the role of gp63-like proteins in lower trypanosomatids. *FEMS Microbiol. Lett.* 2006; 254:149–156. [PubMed: 16451193]

- Matthews KR, Tschudi C, Ullu E. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* 1994; 8:491–501. [PubMed: 7907303]
- Mauricio IL, Stothard JR, Miles MA. The strange case of *Leishmania chagasi*. *Parasitol. Today.* 2000; 16:188–189. [PubMed: 10782075]
- Mauricio IL, Gaunt MW, Stothard JR, Miles MA. Glycoprotein 63 (gp63) genes show gene conversion and reveal the evolution of Old World *Leishmania*. *Int. J. Parasitol.* 2006; 27:27.
- McMahon-Pratt D, Traub-Cseko Y, Lohman KL, Rogers DD, Beverley SM. Loss of the GP46/M-2 surface membrane glycoprotein gene family in the *Leishmania braziliensis* complex. *Mol. Biochem. Parasitol.* 1992; 50:151–160. [PubMed: 1542309]
- McMahon-Pratt D, Alexander J. Does the *Leishmania major* paradigm of pathogenesis and protection hold for New World cutaneous leishmaniases or the visceral disease? *Immunol. Rev.* 2004; 201:206–224. [PubMed: 15361243]
- Mottram JC, Coombs GH, Alexander J. Cysteine peptidases as virulence factors of *Leishmania*. *Curr. Opin. Microbiol.* 2004; 7:375–381. [PubMed: 15358255]
- Muller G, Schlein Y. Nectar and honeydew feeding of *Phlebotomus papatasi* in a focus of *Leishmania major* in Neot Hakikar oasis. *J. Vector Ecol.* 2004; 29:154–158. [PubMed: 15266752]
- Murphy NB, Pays A, Tebabi P, Coquelet H, Guyaux M, Steinert M, Pays E. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *J. Mol. Biol.* 1987; 195:855–871. [PubMed: 3656436]
- Murray HW, Berman JD, Davies CR, Saravia NG. Advances in leishmaniasis. *Lancet.* 2005; 366:1561–1577. [PubMed: 16257344]
- Myler PJ, Audleman L, deVos T, Hixson G, Kiser P, Lemley C, Magness C, Rickel E, Sisk E, Sunkin S, Swartzell S, Westlake T, Bastien P, Fu G, Ivens A, Stuart K. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl. Acad. Sci. USA.* 1999; 96:2902–2906. [PubMed: 10077609]
- Myung KS, Beetham JK, Wilson ME, Donelson JE. Comparison of the post-transcriptional regulation of the mRNAs for the surface proteins PSA (GP46) and MSP (GP63) of *Leishmania chagasi*. *J. Biol. Chem.* 2002; 277:16489–16497. [PubMed: 11856749]
- Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. USA.* 1998; 95:14687–14692. [PubMed: 9843950]
- Noyes H, Chance M, Ponce C, Ponce E, Maingon R. *Leishmania chagasi*: genotypically similar parasites from Honduras cause both visceral and cutaneous leishmaniasis in humans. *Exp. Parasitol.* 1997; 85:264–273. [PubMed: 9085923]
- Oliveira FS, Pirmez C, Pires MQ, Brazil RP, Pacheco RS. PCR-based diagnosis for detection of *Leishmania* in skin and blood of rodents from an endemic area of cutaneous and visceral leishmaniasis in Brazil. *Vet. Parasitol.* 2005; 129:219–227. [PubMed: 15845276]
- Pays E, Vanhamme L, Berberof M. Genetic controls for the expression of surface antigens in African trypanosomes. *Annu. Rev. Microbiol.* 1994; 48:25–52. [PubMed: 7826007]
- Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M, Kerhornou A, Ivens A, Fraser A, Rajandream MA, Carver T, Norbertczak H, Chillingworth T, Hance Z, Jagels K, Moule S, Ormond D, Rutter S, Squares R, Whitehead S, Rabbinowitsch E, Arrowsmith C, White B, Thurston S, Bringaud F, Baldauf SL, Faulconbridge A, Jeffares D, Depledge DP, Oyola SO, Hilley JD, Brito LO, Tosi LR, Barrell B, Cruz AK, Mottram JC, Smith DF, Berriman M. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet.* 2007; 39:839–847. [PubMed: 17572675]
- Pedrosa AL, Silva AM, Ruiz JC, Cruz AK. Characterization of LST-R533: uncovering a novel repetitive element in *Leishmania*. *Int. J. Parasitol.* 2006; 36:211–217. [PubMed: 16368097]
- Rao V, Fujiwara N, Porcelli SA, Glickman MS. Mycobacterium tuberculosis controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule. *J. Exp. Med.* 2005; 201:535–543. [PubMed: 15710652]
- Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol. Biochem. Parasitol.* 2003; 128:217–228. [PubMed: 12742588]

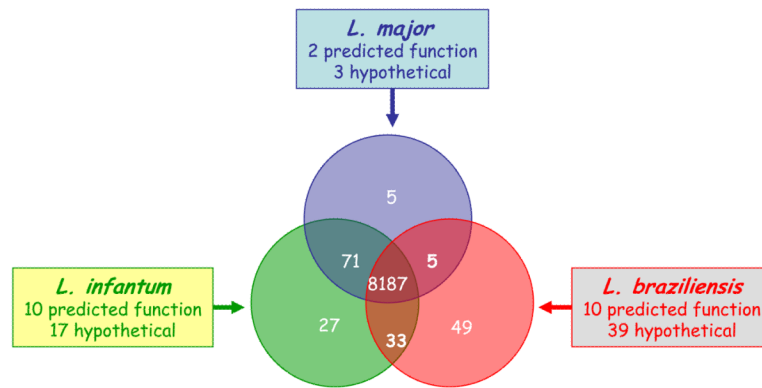
- Rochette A, McNicoll F, Girard J, Breton M, Leblanc E, Bergeron MG, Papadopoulou B. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. *Mol. Biochem. Parasitol.* 2005; 140:205–220. [PubMed: 15760660]
- Romano N, Macino G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* 1992; 6:3343–3353. [PubMed: 1484489]
- Sacks D, Kamhawi S. Molecular aspects of parasite–vector and vector–host interactions in leishmaniasis. *Annu. Rev. Microbiol.* 2001; 55:453–483. [PubMed: 11544364]
- Sacks D, Noben-Trauth N. The immunology of susceptibility and resistance to *Leishmania major* in mice. *Nat. Rev. Immunol.* 2002; 2:845–858. [PubMed: 12415308]
- Sadlova J, Volf P, Victoir K, Dujardin JC, Votypka J. Virulent and attenuated lines of *Leishmania major*: DNA karyotypes and differences in metalloproteinase GP63. *Folia Parasitol. (Praha)*. 2006; 53:81–90. [PubMed: 16898121]
- Satoskar AR, Bozza M, Rodriguez Sosa M, Lin G, David JR. Migration-inhibitory factor gene-deficient mice are susceptible to cutaneous *Leishmania major* infection. *Infect. Immun.* 2001; 69:906–911. [PubMed: 11159984]
- Scheffter SM, Ro YT, Chung IK, Patterson JL. The complete sequence of *Leishmania* RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology.* 1995; 212:84–90. [PubMed: 7676652]
- 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; 279:49889–49893. [PubMed: 15383544]
- Shi H, Tschudi C, Ullu E. An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA.* 2006a; 12:2063–2072. [PubMed: 17053086]
- Shi H, Tschudi C, Ullu E. Functional replacement of *Trypanosoma brucei* Argonaute by the human slicer Argonaute2. *RNA.* 2006b; 12:943–947. [PubMed: 16611939]
- Singer MF. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell.* 1982; 28:433–434. [PubMed: 6280868]
- Stevens JR, Noyes HA, Schofield CJ, Gibson W. The molecular evolution of Trypanosomatidae. *Adv. Parasitol.* 2001; 48:1–56. [PubMed: 11013754]
- Teng SC, Wang SX, Gabriel A. A new non-LTR retrotransposon provides evidence for multiple distinct site-specific elements in *Crithidia fasciculata* miniexon arrays. *Nucleic Acids Res.* 1995; 23:2929–2936. [PubMed: 7659515]
- Tielens AG, Van Hellemond JJ. Differences in energy metabolism between trypanosomatidae. *Parasitol. Today.* 1998; 14:265–272. [PubMed: 17040781]
- Turco SJ, Spath GF, Beverley SM. Is lipophosphoglycan a virulence factor? A surprising diversity between *Leishmania* species. *Trends Parasitol.* 2001; 17:223–226. [PubMed: 11323305]
- 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; 13:720–725. [PubMed: 8417363]
- Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. *Cell Microbiol.* 2004; 6:509–519. [PubMed: 15104593]
- Victoir K, Dujardin JC, de Doncker S, Barker DC, Arevalo J, Hamers R, Le Ray D. Plasticity of gp63 gene organization in *Leishmania* (*Viannia*) *braziliensis* and *Leishmania* (*Viannia*) *peruviana*. *Parasitology.* 1995; 111:265–273. [PubMed: 7567095]
- Villanueva MS, Williams SP, Beard CB, Richards FF, Aksoy S. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol. Cell Biol.* 1991; 11:6139–6148. [PubMed: 1719380]
- Wickstead B, Ersfeld K, Gull K. Repetitive elements in genomes of parasitic protozoa. *Microbiol. Mol. Biol. Rev.* 2003; 67:360–375. [PubMed: 12966140]
- Widmer G, Dooley S. Phylogenetic analysis of *Leishmania* RNA virus and *Leishmania* suggests ancient virus–parasite association. *Nucleic Acids Res.* 1995; 23:2300–2304. [PubMed: 7610059]
- Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell.* 1993; 75:855–862. [PubMed: 8252622]

- Wilson ME, Jeronimo SM, Pearson RD. Immunopathogenesis of infection with the visceralizing *Leishmania* species. *Microb. Pathog.* 2005; 38:147–160. [PubMed: 15797810]
- Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA. The genome of *Cryptosporidium hominis*. *Nature.* 2004; 431:1107–1112. [PubMed: 15510150]
- Yao C, Donelson JE, Wilson ME. The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Mol. Biochem. Parasitol.* 2003; 132:1–16. [PubMed: 14563532]
- Yao C, Luo J, Hsiao C, Donelson JE, Wilson ME. Internal and surface subpopulations of the major surface protease (MSP) of *Leishmania chagasi*. *Mol. Biochem. Parasitol.* 2005; 139:173–183. [PubMed: 15664652]
- Zajtcuk JT, Casler JD, Netto EM, Grogl M, Neafie RC, Hessel CR, de Magalhaes AV, Marsden PD. Mucosal leishmaniasis in Brazil. *Laryngoscope.* 1989; 99:925–939. [PubMed: 2671555]
- Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL, Matlashewski G. Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J. Biol. Chem.* 2003; 278:35508–35515. [PubMed: 12829719]
- Zilberstein D. Transport of nutrients and ions across membranes of trypanosomatid parasites. *Adv. Parasitol.* 1993; 32:261–291. [PubMed: 8237616]

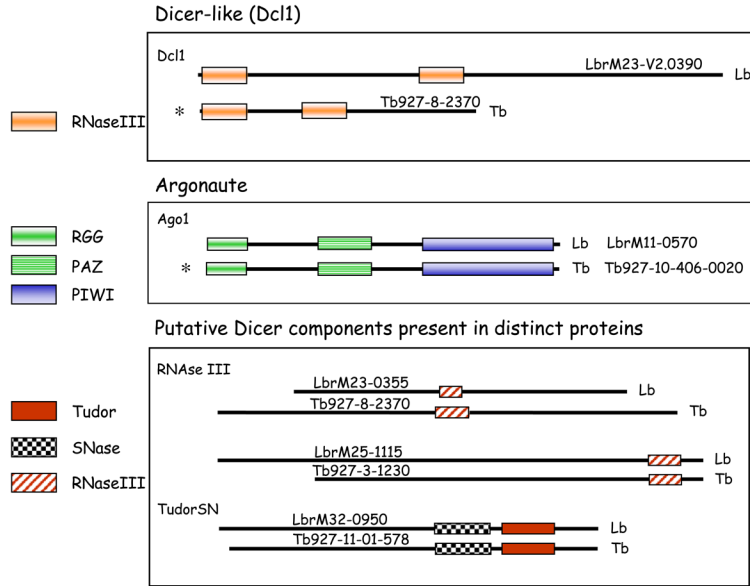


**Fig. 1.**

The structure of a typical *Leishmania* chromosome: chromosome 2 of *Leishmania major*. The location and coding strands of the 74 protein-coding genes are shown as coloured boxes, coded according to the categories indicated below. The majority are genes conserved with other eukaryotes; distribution of the *Leishmania*-specific and sc $\beta$ -galactosyltransferase (SCG) repeat family genes are also shown, together with three pseudogenes. Other chromosomal features include the telomeric hexamer repeats, the spliced leader tandem array and the ribosomal mobile element (RIME) degenerate LTR transposons.



**Fig. 2.** *Leishmania* species-specific genes. The Venn diagram shows how many of the 8,187 protein-coding genes are species-specific or shared between two of the three sequenced *Leishmania* species. These genes are subdivided into those that have a predicted function (mostly through sequence identity) and those that are of unknown function at the time of publication (the “hypotheticals”).



**Fig. 3.** RNAi machinery: *Leishmania braziliensis* versus *Trypanosoma brucei*. Schematic representation of the main components of the RNAi pathways identified by sequence similarity and/or functional evaluation in *T. brucei* (Tb) and *L. braziliensis* (Lb). Comparisons of the gene products and their domains in each species are accompanied by their GeneDB gene IDs; Tb927, *T. brucei* genes; LbrM, *L. braziliensis* genes. Domain IDs are depicted on the left. RGG indicates a domain containing arginine-glycine-glycine repeats; SNase indicates the Staphylococcal nuclease domain present in TudorSN proteins. \*Indicates genes confirmed to be functionally active in the *T. brucei* RNAi pathway.

**Table 1**Human-infective species of the *Leishmania* genus

Old World species	New World species	Disease type
<b><i>L. major</i> complex</b>	<b><i>L. mexicana</i> complex</b>	Cutaneous
<u><i>L. (L.) major</i></u>	<u><i>L. (L.) mexicana</i></u>	
<i>L. (L.) tropica</i>	<i>L. (L.) amazonensis</i>	
<i>L. (L.) aethiopica</i>	<i>L. (L.) pifanoi</i>	
	<i>L. (L.) venezuelensis</i>	
	<b><i>L. (Viannia) subgenus</i></b>	
	<u><i>L. (V.) braziliensis</i></u>	
	<i>L. (V.) panamensis</i>	
	<i>L. (V.) guyanensis</i>	
	<i>L. (V.) peruviana</i>	
	<i>L. (V.) lansonii</i>	
	<u><i>L. (V.) braziliensis</i></u>	Mucocutaneous
<i>L. (L.) aethiopica</i>	<i>L. (L.) amazonensis</i>	Diffuse cutaneous
	<i>L. (L.) pifanoi</i>	
<b><i>L. donovani</i> complex</b>		Visceral
<i>L. (L.) donovani</i>		
<u><i>L. (L.) infantum</i></u> *	<i>L. (L.) chagasi</i> *	

The main species complexes and subgenus are shown in bold; the species with complete genome sequences either available or being generated are underlined.

\* Species that can also be associated with cutaneous leishmaniasis.



Table 2

Genome facts: the *Leishmanias* versus *Trypanosoma brucei*

	<i>Leishmania major</i>	<i>Leishmania infantum</i>	<i>Leishmania braziliensis</i>	<i>Trypanosoma brucei</i>
Total size (Mb)	32.8	32.1	32.0	26.1
Contigs	36	562	1,041	30
No. of chromosomes	36	36	35	11*
Chromosome size range (Mb)	0.3–2.8	0.3–2.8	0.3–2.8	1–5.2
Overall G + C content %	59.7	59.3	60.4	46.4
No. of genes	8,298	8,154	8,153	9,068
No. of pseudogenes	97	41	161	904
Average gene size (bp)	1,894	1,868	1,873	1,592
Gene density (per Mb)	252	235	258	317
Coding percentage	48.0	44.0	48.5	50.5
Coding G + C content %	62.5	62.4	60.4	50.9
No. of DGCs	133	133	n/a	127
Average DGC length (kb/genes)	240/61	n/a	n/a	204/71
No. of tRNAs	83	62	66	65
No. of snoRNAs	693	n/a	n/a	353
No. of snRNAs	6	n/a	n/a	5
No. of rRNAs	63	n/a	n/a	56
Average intergenic size	1,939	2,049	1,976	1,279
Active mobile elements	None (degenerate RIME/ DIRE)	None (degenerate RIME/ DIRE)	TATEs, SLACS	ingi, RIME, DIRE, SLACs, SIRE, VIPER

Data included in this Table are correct as of February 2007 (for *Leishmania species*) and July 2005 (for *T. brucei*).

\* only the megabase chromosome are included, not the intermediate or mini-chromosomes. DGC, directional gene cluster; RIME, ribosomal mobile element; DIRE, degenerate *ingi*/L1Tc-related element; TATE, telomere-associated transposable element; SLACS, Spliced Leader Associated Conserved Sequence; SIRE, short interspersed repetitive element; VIPER, LTR retroelement related to SIRE.