CHAPTER 1

INTRODUCTION

1.1. Leishmaniasis, a complex parasitic disease

1.1.1 Overview

The leishmaniases are a group of vector-borne parasitic diseases that collectively affect around 12 million people around the globe, according to estimates by the World Health Organization. Approximately 310 million additional people worldwide are at risk of infection. The disease is spread by the bite of female sand flies of the genera *Lutzomyia* and *Phlebotomus*, and is caused by 20 different species of flagellated unicellular protozoa belonging to the genus *Leishmania*, although the taxonomic status of some of them is disputed. Clinical disease due to *Leishmania* presents in three major forms: cutaneous leishmaniasis, mucocutaneous leishmaniasis, and visceral leishmaniasis. Asymptomatic cases are known to exist in endemic areas, and may act as an important reservoir for re-infection (Singh, Hasker et al. 2014). The parasite may follow two different transmission cycles: a zoonotic cycle, with dogs being an especially important animal reservoir in addition to other mammals; and a strictly anthroponotic cycle, with humans as the only host, which is typically observed in densely inhabited urban areas.

Accurate epidemiological data from many endemic countries is lacking (Alvar, Velez et al. 2012), so the true burden of disease may be higher than official WHO estimates, especially if the mental health repercussions - in the form of social stigma and ostracization - associated with some clinical manifestations are appropriately factored in. In addition, HIV-*Leishmania* coinfection is a major complicating form of disease that has been systematically underreported in many endemic areas (WHO 2007), and for which poor clinical guidelines have been established (Diro, van Griensven et al. 2015). Opportunistic infection with *Leishmania* is an AIDS-defining illness in endemic settings, and the immunosuppressive effects of the parasitic infection are compounded by infection with HIV, often with irremediable consequences for the patient. *Leishmania* is widely considered to be the second biggest parasitic killer after malaria, and it is thought that global warming, anthropogenic environmental changes, and human migrations have led to an expansion of the geographic distribution of the leishmaniases (Desjeux 2004).

Visceral leishmaniasis (VL), also known as kala-azar, black fever, or Dumdum fever, is a systemic illness caused by infection of the liver, spleen, and bone marrow by the parasite. If left untreated, the disease is almost invariably lethal within two years of presentation of symptoms. According to the WHO Global Health Observatory, approximately 200 000-400 000 cases of VL occur per year, 90% of which are concentrated in 6 countries: Bangladesh, India, Ethiopia, Sudan, South Sudan, and Brazil. An estimated 20 000 to 30 000 deaths are due to VL each year, although the true number may be higher due to poor epidemiological surveillance in

many areas of active transmission. A serious complication that may occur in cases of VL following treatment is post-kala-azar dermal leishmaniasis (PKDL), a chronic syndrome characterized by the appearance, most notably on the face, of a multitude of papules, nodes, and patches, which require prolonged chemotherapy. In the Indian subcontinent, PKDL is rare, appears several years after successful therapy, and is particularly hard to treat. Conversely, in East Africa PKDL is more frequent, is noted a few months after initial therapy for VL, and often resolves spontaneously. Patients with PKDL may act as important reservoir hosts.

Mucocutaneous leishmaniasis (MCL), also known as espundia, is a severely disfiguring disease which typically occurs as a metastatic dissemination of the parasite from the site of initial infection on the skin to mucosal membranes of the body, primarily around the nose and mouth. Complete ulcerative destruction of the oro-pharyngeal tissues is often seen in the absence of timely treatment. Almost 90% of all cases of MCL occur in Bolivia, Brazil, and Peru. The disease is caused by New World species of *Leishmania* belonging to the *Viannia* subgenus, namely *L. [V.] braziliensis*, *L. [V.] panamensis*, *L. [V.] guyanensis*, and sometimes *L. amazonensis*. The first symptoms (e.g., persistent nosebleeds) may appear several months or even years after healing of the primary skin lesion. Disease pathogenesis is poorly understood, and is often associated with lack of appropriate treatment of cutaneous lesions.

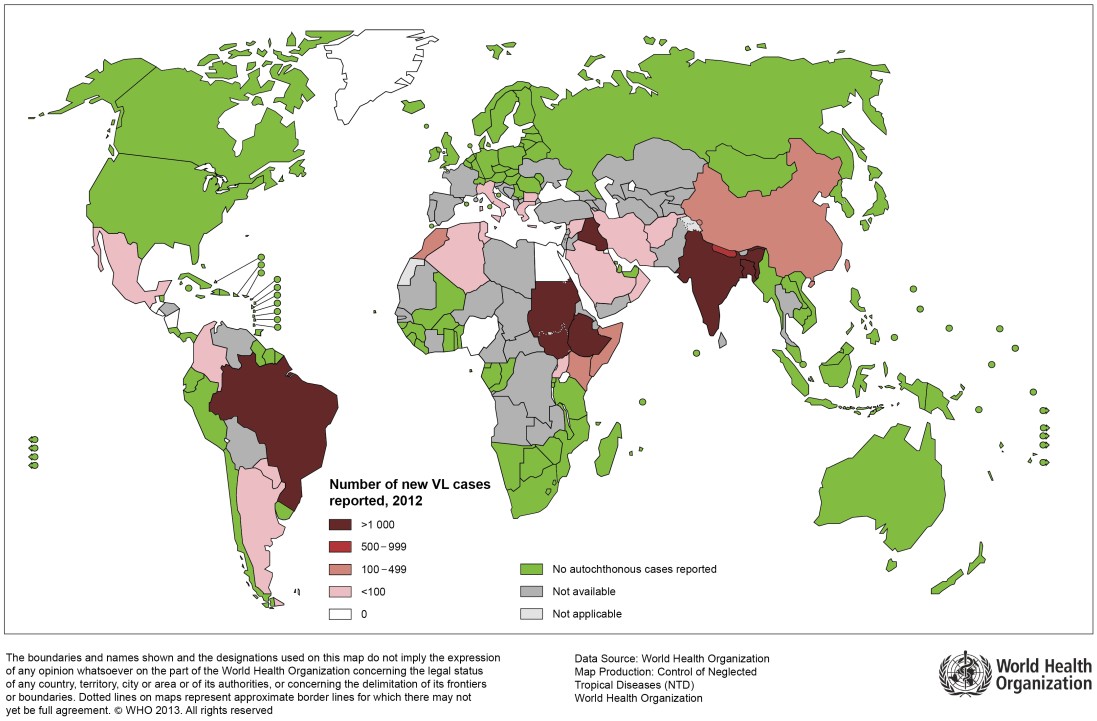

Figure 1.1: Distribution and endemicity of visceral leishmaniasis (VL) according to 2013 annual country reports. Countries in grey have no reliable epidemiological data or do not report disease incidence to the WHO Neglected Tropical Diseases (NTD) section. Countries in green had no autochthonous cases of VL reported in 2012 (source: WHO Global Health Observatory).

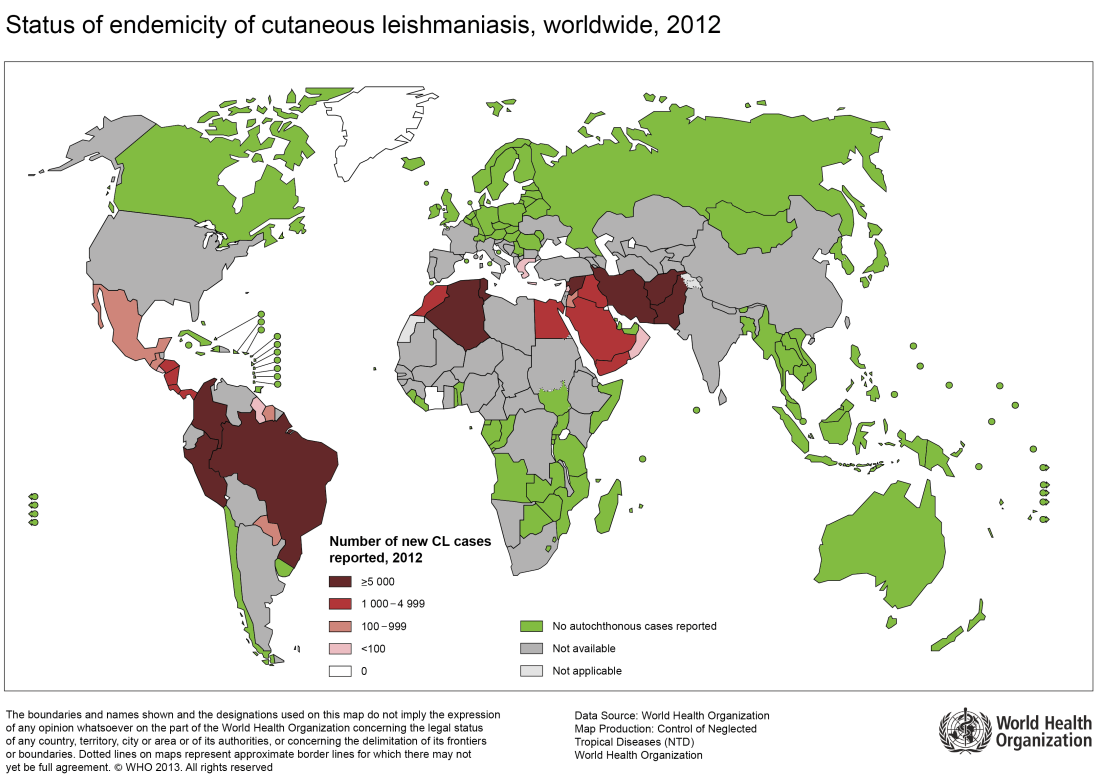

Figure 1.2: Distribution and endemicity of cutaneous leishmaniasis, (CL) according to 2013 country reports. Countries in grey have no reliable epidemiological data, or do not report disease incidence to the WHO Neglected Tropical Diseases (NTD) section. Countries in green had no autochthonous cases of CL reported in 2012 (source: WHO Global Health Observatory).

Cutaneous leishmaniasis (CL) is the most common form of disease, and is caused by both New World and Old World species of *Leishmania*. The disease manifests as ulcerative lesions that develop at the site of the sand fly bite, which can grow progressively larger and fail to heal naturally. These lesions often appear on exposed areas of the body such as the face, and are associated with significant social stigma and disability. If infected with bacteria, these lesions can be quite painful. Lymphadenopathy often precedes appearance of the lesion. Even with this localized form of disease, often multiple satellite lesions appear, and can persist for months or years. Successful treatment does not eliminate the noticeable scars that remain visible for life. Around 1.3 million new cases of CL occur annually worldwide according to the WHO Global Health Observatory.

This parasitic disease is one of the most neglected disease of the developing world, and is often entrenched in areas where poor sanitation, poor access to healthcare, and strained infrastructures due to war or social unrest lead to a combination of factors favorable to the establishment of recurrent epidemic cycles of transmission (Beyrer, Villar et al. 2007). A promising leishmaniasis elimination campaign has been championed by WHO since 2005 in Bangladesh, India, and Nepal, with the objective of reducing the incidence of VL to one case per 10 000 at the district or sub-district level by 2015. There were around 20 cases per 10 000 in the region in 2011, and the campaign has been making remarkable progress. The elimination of VL in this region is made achievable by the presence of a single sand fly vector species that is susceptible to insecticides; the distribution of cases in geographic clusters; and the fact that humans are the only reservoirs of infection.

The presence of asymptomatic carriers could complicate complete elimination in the region. Large-scale control and global elimination of all clinical forms of disease associated with *Leishmania* infection, on the other hand, is poised to be a significant challenge, given the remarkable differences observed in clinical presentation in different patient populations, the presence of zoonotic reservoirs, and the number of different parasite species causing significant disease. An improved understanding of disease pathogenesis and the transmissibility of the parasite in each clinical presentation can inform prioritization of different elimination strategies, as would the presence of an effective vaccine, the development of point-of-care diagnostics, and an affordable, easy-to-administer oral formulation for drug therapy (Matlashewski, Arana et al. 2014).

Currently, leishmaniasis is treated with a variety of remedies, ranging from first-line pentavalent antimonials, which are poorly tolerated in patients and to which many circulating parasite strains have developed resistance, to different regimens of amphotericin B, paromomycin, fluconazole, and the promising oral drug miltefosine, which recently received regulatory approval for use in India and the United States. No human vaccine is available, although there are several candidates in pre-clinical and clinical stages. The historical practice of "leishmanization", whereby live parasites are inoculated in the skin in a cosmetically acceptable part of the body, is the only fully effective way to gain immunity to CL. The fact that individuals who fully recover from VL are then resistant to re-infection suggests that immunity to symptomatic visceral disease can be achieved via vaccination (Working Group on Research Priorities for Development of Leishmaniasis, Costa et al. 2011).

1.1.2 The biology of the parasite

Leishmania belongs to a class of unicellular protists known as Kinetoplastida. The only kinetoplastid organisms known to cause disease in humans are: approximately 20 different *Leishmania* species; the two parasite species responsible for human African trypanosomiasis (HAT), or sleeping sickness, *Trypanosoma brucei brucei* and *T. brucei rhodesiense*; and *T. cruzi*, the parasite species responsible for Chagas disease in the Americas. All kinetoplastids, in addition to being flagellated for at least part of their life cycle, also share a unique DNA-containing organelle known as the kinetoplast, situated in a mitochondrian-like structure at the base of the flagellum. The kinetoplast contains multiple circular copies of kinetoplast DNA (kDNA), which serve the same function as the mitochondrial genome in more advanced eukaryotes.

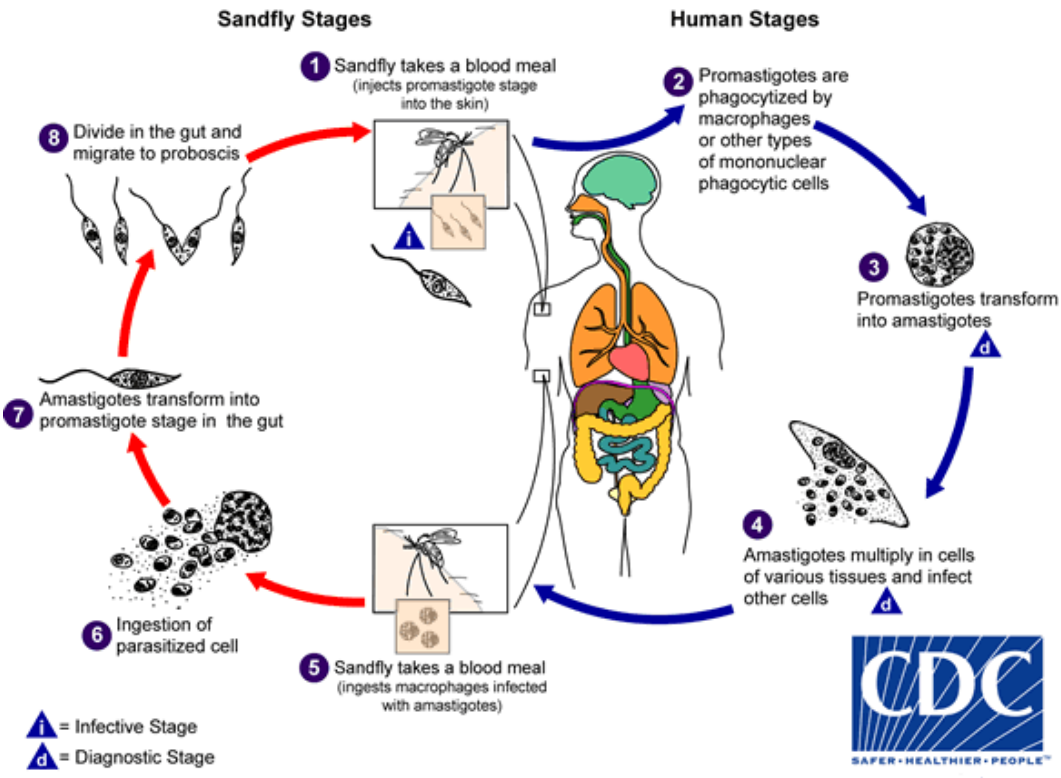

Figure 1.3. The *Leishmania* life cycle. (1) Upon bloodfeeding, the sandfly inoculates infectious metacyclic promastigotes into the host's skin; (2) once in the skin, promastigotes are ingested by phagocytic cells; (3) within the phagocytic cell the parasite differentiates into obligate intracellular amastigotes; (4) the parasite replicates intracellularly through multiple rounds of mitosis, invading neighbouring cells; amastigote-infected cells may localize to the skin lesion, or spread to other sites in the body; (5) circulating amastigote-infected macrophages are taken up in the blood meal of a sand fly; (6-7) amastigotes differentiate into extracellular promastigotes and attach to the midgut wall to survive excretion of the digested bloodmeal; (8) promastigotes migrate anteriorly and undergo a series of developmental

transitions to form infectious metacyclic promastigotes, encased within a PSG plug that blocks normal feeding of the sand fly (Source: US Center for Disease Control and Prevention, Division of Parasitic Diseases and Malaria).

Like *T. brucei* and *T. cruzi*, *Leishmania* has a digenetic life cycle, alternating between the sand fly vector and the mammalian host. When female sand flies blood feed on an appropriate host, the parasite is inoculated into the skin as metacyclic promastigotes, the infectious, extracellular, non-replicative stage (Bates 2007). These metacyclic promastigotes are lodged near the stomadeal valve in the anterior gut of the sand fly, and are encased in a gel-like "plug" created via secretion of PSG, or promastigote secretory gel. The sand fly is forced to regurgitate this plug into the skin as it takes its meal. Impaired uptake of blood leads to the sand fly attempting to feed with greater frequency, and thus increases the chance of parasite transmission (Rogers and Bates 2007).

The parasite is thus inoculated in the skin along with pro-inflammatory salivary components, where it then invades resident phagocytic cells within the skin tissues of the host. Once inside the host cell, the parasite differentiates into the obligate intracellular, non-flagellated form called the amastigote. The parasite continues to replicate by mitotic cell division, re-invading phagocytic cells as an intracellular amastigote, until it is taken up in the blood meal of the next sand fly. Both parasite and host factors are thought to be important in determining whether the infection is symptomatic and the type of pathology resulting from the infection. Tissue tropism of infecting parasites can vary, but VL is usually associated with

heavy infections of the liver, spleen, and bone marrow. Different symptomatology and distribution in the host tissues may determine differences in transmissibility of the parasite. Unusual tissue tropism has been observed in HIV-*Leishmania* coinfections, such as parasites in the gastroendothelial mucosa.

Once in the sand fly midgut, the amastigote forms differentiate into early procyclic promastigote stages, which are multiplicative and increase in numbers by cell division, while attaching to the interior wall of the sand fly midgut. Parasite attachment to the midgut wall is mediated by lipophospoglycan (LPG) covering the parasite cell surface. This molecule has a complex, branched structure composed by polysaccharides and polypeptides, and plays an important role in species-specific interactions between parasite and vector (Pimenta, Saraiva et al. 1994, Sacks 2001). By attaching to the midgut wall, the parasite survives expulsion of the digested blood meal as sand fly excrement. The parasite then differentiates into non-replicating nectomonad promastigotes, and migrates to the anterior part of the midgut where it resumes replication as leptomonad promastigotes. This stage is also responsible for production of promastigote secretory gel (PSG), and immediately precedes differentiation into mammalian-infective metacyclic promastigotes (Bates and Rogers 2004).

1.1.3 The burden of disease due to *L. tropica*

Among the approximately 20 species of *Leishmania* responsible for human disease, *L. tropica* appears to be related to *L. major*, and is considered by some to be

part of the same species complex as other CL-causing Old World species such as *L. aethiopica* and *L. killicki* (El Baidouri, Diancourt et al. 2013, Chaara, Ravel et al. 2015). *L. tropica* most often causes skin lesions similar to those caused by *L. major*, although lesions due to *L. tropica* tend to be larger, covered with scabs or crusts, and less responsive to treatment.

Skin lesions often recrudesce following apparent successful treatment, sometimes decades after the primary infection, in a syndrome unique to *L. tropica* known as leishmaniasis recidivans (LR). These encrusted or papular lesions are typically localized along the edge of the scarred tissue left by healing of the primary ulcer, and may be triggered by inflammation-activating events up to more than 40 years following healing of the primary lesion (Marovich, Lira et al. 2001).

In addition to these cutaneous manifestations, *L. tropica* has been associated with a variant form of VL known as viscerotropic leishmaniasis (VTL) that depending on both parasite and host factors may or may not resemble the classic symptomatology of VL disease, with weight loss, fever, splenohepatomegaly, general weakness and muscle pains, as well as dissemination of the parasite to internal organs and to the bone marrow, and subsequent immunosuppression and pancytopaenia in the more advanced stages of disease (Mebrahtu, Lawyer et al. 1989, Magill, Grogl et al. 1993, Sacks, Kenney et al. 1995, Alborzi, Rasouli et al. 2006, Weiss, Vogenthaler et al. 2009).

Cutaneous lesions due to *L. tropica* are treated with topical paromomycin, intralesional antimonials, and sometimes with topical miconazole. Systemic treatment is only recommended in the most serious cases or in LR, and it may

include the azoles ketoconazole and fluconazole, amphotericin B, parenteral antimonials, or a combination of several of these options (Monge-Maillo and Lopez-Velez 2013). The effectiveness of liposomal amphotericin B to treat CL due to *L. tropica* is only based on a few published case studies of patients where other treatment options had failed. No specific treatment guidelines exist for VTL due to *L. tropica*.

The *L. tropica* species complex has been found to be genetically extremely heterogeneous (Schwenkenbecher, Wirth et al. 2006), and *L. tropica* may follow different transmission cycles in different geographic areas. In densely populated urban or periurban areas of the Middle East, the parasite appears to follow an anthroponotic cycle, with no known zoonotic reservoir of infection and with human-to-human transmission contributing to the high incidence of disease. In other locations however, several animal reservoirs have been found, such as dogs in Morocco (Dereure, Rioux et al. 1991), jackals and foxes in Israel (Talmi-Frank, Kedem-Vaanunu et al. 2010), and rodents, such as rock hyraxes, belonging to the families Procaviidae and Ctenodactylidae in Tunisia, Israel, and East Africa (Sang, Njeru et al. 1994, Svobodova, Votypka et al. 2006, Talmi-Frank, Jaffe et al. 2010, Jaouadi, Haouas et al. 2011, Bousslimi, Ben-Ayed et al. 2012).

L. tropica is transmitted by several *Phlebotomus* species of sand flies. Most notably, it is transmitted by *P. sergenti* across the majority of its geographic distribution (Kamhawi, Modi et al. 2000), by *P. arabicus* in Israel (Svobodova, Volf et al. 2006), by *P. saevus* in Ethiopia (Gebre-Michael, Balkew et al. 2004), and by *P. guggisbergi* in Kenya (Lawyer, Mebrahtu et al. 1991). While *L. major* is known to be

primarily zoonotic, and the seasonal fluctuations in CL cases due to *L. major* follow seasonal changes in the population of the vector species *P. papatasi*, CL due to *L. tropica* most often occurs in urban areas, such as Aleppo in Syria and Kabul in Afghanistan, where it appears to be exclusively anthroponotic. In most areas in North Africa, however, the disease is hypoendemic, with only sporadic cases being reported in rural foci across Tunisia, Algeria, and Lybia, consistent with the presence of a zoonotic transmission cycle, although in Morocco a few large epidemics with hundreds of cases have also been reported in urban areas (Ajaoud, Es-sette et al. 2013).

Figure 1.4. Geographic distribution of Old World CL due to *L. tropica*, *L. aethiopica*, and related species (source: WHO Essential Leishmaniasis Maps).

Throughout its geographic range, cases of *L. tropica* seem to be increasing, possibly as a result of anthropogenic changes to the environment and subsequent expansion of the vector population to areas previously unaffected by the disease

(Reithinger, Dujardin et al. 2007). While *L. tropica* was rare in Kabul, Afghanistan before 1990, a marked increase in annual cases has been reported in recent years (Reithinger, Mohsen et al. 2003). In Aleppo, Syria, cases have also been steadily increasing in the last few decades, possibly reflecting changes in the vector population (Tayeh, Jalouk et al. 1997). Recent socio-political events may have precipitated these epidemics. According to the WHO, in 2013 Syria had the highest number of cases ever reported in the country, with 71,991 reported cases of CL. Afghanistan had 23,621 reported cases of CL in 2013. The majority of cases in both settings can be attributed to *L. tropica*.

New foci of infection have been reported in Northern Israel (Jacobson, Eisenberger et al. 2003), and large epidemics of *L. tropica* have been documented in refugee camps in Syria (Saroufim, Charafeddine et al. 2014) and in Pakistan (Rowland, Munir et al. 1999, Brooker, Mohammed et al. 2004). The disease is also highly endemic in the Arabian Peninsula and in Turkey, Iraq, Iran, and some areas of India. Displacement of populations due to armed conflict or civil unrest into areas with active transmission is known to be a major driver behind outbreaks of *L. tropica*. Although the majority of CL in East Africa is due to *L. major*, *L. tropica* is also present (Hotez, Savioli et al. 2012).

A reliable quantification of the burden of disease due to *L. tropica* that includes both CL and VTL manifestations has yet to be performed, although molecular typing of strains has confirmed earlier reports of *L. tropica* being a parasite species contributing to visceral disease in the region (Khanra, Datta et al. 2012, Krayter, Bumb et al. 2014).

1.1.4 Mechanisms of pathogenesis

The mechanisms by which non-healing, chronic lesions are generated upon inoculation in the skin are not fully understood. Previous exposure to uninfected sand fly bites are known to diminish the severity of disease (Kamhawi, Belkaid et al. 2000), so sand fly salivary proteins might play a role in the recruitment of phagocytic neutrophils to the site of the bite, which then act as an inflammatory "silent" route for the parasite to enter macrophages. Neutrophils are short-lived cells that undergo apoptosis, and are in turn phagocytosized by professional macrophages (van Zandbergen, Klinger et al. 2004, Peters, Egen et al. 2008). Components of the PSG plug are known to enhance infection (Rogers, Ilg et al. 2004). Although *Leishmania* can invade a variety of phagocytic and non-phagocytic mammalian cells, the parasite has a marked preference for macrophages. The metacyclic promastigote differentiates into the non-flagellated amastigote within the phagolysosome, where it carries out its entire replicative cycle. The only other microbe known to be capable of replicating within this cell compartment is the Gram-negative bacterium *Coxiella burnetti*. The LPG coat on the surface of the surface of the parasite is thought to play a role in initial uptake by neutrophils and macrophages by limiting the damage due to ROS generated during phagocytosis, although it then becomes strongly downregulated in the amastigote stages. Amastigote stages are highly opsonized by IgG antibodies, a characteristic that promotes uptake by macrophages via the Fc receptor and release of anti-inflammatory IL-10 (Kane and Mosser 2001).

Leishmania is auxotrophic for many aminoacids, which are required for protein biosynthesis, and must scavenge these from within the phagolysosome, along with carbon sources such as fatty acids and hexoses, and other essential nutrients such as heme, purines, and vitamins (McConville, de Souza et al. 2007). There is evidence that the parasite establishes a complex interplay with the host cell metabolism. Uptake of essential nutrients from the host cytosol is thought to occur via membrane proteins of the phagolysosome, and alternative activation of macrophages via Th2-associated cytokines such as IL-4 and IL-10 is associated with activation of the key host-encoded metabolic enzyme arginase-1, and subsequent increased production of amino acids essential to the parasite (Sacks and Anderson 2004).

Establishment of chronic infection has long been attributed to a Th2 cytokine profile in *Leishmania* infection (Heinzel, Sadick et al. 1991). Downregulation of effector T cells by regulatory T cells has been implicated in the establishment of chronic infection and immunity to reinfection at other sites of the body (Belkaid, Piccirillo et al. 2002). More recently, a role for NLRP3 activation and IL-18 has been found in the establishment of chronic infection (Gurung, Karki et al. 2015). In contrast to *L. major*, abrogation of IL-10 production is not sufficient to clear *L. tropica* from the site of infection (Anderson, Lira et al. 2008), suggesting important differences in pathogenesis between infection with *L. tropica* and with *L. major*.

Host genetics undoubtedly play a factor in pathogenesis. It should be noted that the majority of infections are completely asymptomatic in *L. infantum* and *L donovani* (Sakthianandeswaren, Foote et al. 2009). Recently, a large-scale case-

control study for susceptibility to VL in Brazilian and Indian populations found a strong association with the HLA locus, specifically, with the HLA class II regions HLA-DRB1 and HLA-DQA1 (LeishGEN Consortium, Wellcome Trust Case Control Consortium et al. 2013). CD4+ T cells are known to secrete interferon gamma to control infection in acute VL, and the HLA class II region may be involved in recognition of *Leishmania* antigen via classical interaction with TCRs (Kumar and Engwerda 2014). CD8+ cytotoxic T cells might also play a role (Mansueto, Vitale et al. 2007).

1.2. Alternative genetics in *Leishmania*

1.2.1 The unique genome of kinetoplastids

Leishmania parasites have a genome architecture and associated biology that are unlike those of almost any other eukaryote. Trypanosomatids as a family appear to be a very ancient clade of the eukaryotic tree of life (He, Fiz-Palacios et al. 2014), and the whole class Kinetoplastida appear to have remarkably unique features. Since publication of the *L. major* genome in 2005 (Ivens, Peacock et al. 2005), a series of discoveries have been made regarding *Leishmania* genome organization that are conserved across trypanosomatid species.

First, genes are organized head-to-tail into large clusters, often hundreds of kilobases in length, sharing the same direction of transcription. Chromosome 1, for instance, is transcribed in two large clusters, the first cluster of 29 genes on one

strand, and the remaining 50 genes on the other strand (Myler, Audleman et al. 1999). Transcription is thought to initiate and terminate in strand switch regions, with the transcribed polycistronic segments on opposite strands being either directionally divergent or convergent (Martinez-Calvillo, Yan et al. 2003, Martinez-Calvillo, Nguyen et al. 2004). Transcription is thought to primarily initiate in divergent strand switch regions, and terminate in convergent strand switch regions.

Second, in contrast to bacterial operons, polycistronic transcripts in *Leishmania* require further processing before translation. The mature mRNA transcript originates from coupled trans-splicing and polyadenylation of the initial polycistronic unit. A 39-nucleotide mini-exon sequence, called the spliced leader (SL), is trans-spliced to the 5' end of each gene (Sutton and Boothroyd 1986, Perry, Watkins et al. 1987). This process is mediated by the spliceosome complex, and is coupled with polyadenylation of the 3' end of the gene upstream of the splice site (LeBowitz, Smith et al. 1993, Matthews, Tschudi et al. 1994). The enzymatic machinery involved in mRNA maturation has not been fully characterized, although the complete set of small nuclear RNAs involved in trans-splicing has been found (Liang, Haritan et al. 2003).

Third, given the role played by polycistrons and their maturation by trans-splicing, individual *Leishmania* genes lack traditional eukaryotic promoters. Nuclear genome polycistrons appear to be constitutively transcribed by RNA polymerase II, and individual genes lack introns. The only gene known to be transcribed via eukaryotic promoter sequences in *Leishmania* is the SL RNA gene, which is well conserved in kinetoplastids (Gilinger and Bellofatto 2001). The primary transcript

of the SL RNA gene presents a distinctive three-loop secondary structure, as well as a terminal intronic sequence involved in binding with maturation factors. The intronic sequence contains two distinct domains, the -60 and -30 elements, so called based on their position with respect to the transcription start site (TSS) (Sturm, Fleischmann et al. 1998, Sturm and Campbell 1999, Sturm, Yu et al. 1999). The mature SL RNA transcript is capped at its 5' end with a modified nucleotide sequence called the ``4-cap'', due to the presence of 4 consecutive methylated bases. Trans-splicing of this capped SL RNA sequence to mature mRNAs requires an AG 3' acceptor site and U-rich polypyrimidine tract (PPT) in the target mRNA (Curotto de Lafaille, Laban et al. 1992, LeBowitz, Smith et al. 1993, Matthews, Tschudi et al. 1994). Recently, basal splicing factors were suggested to be involved in determining the exact splice site and thus the size of the 3' UTR. The size of the 3' UTR in turn may affect polyadenylation by inclusion or exclusion of a given regulatory sequence (Gupta, Carmi et al. 2013).

Fourth, kinetoplastid organisms also possess a kinetoplast-localized genome known as kDNA, which codes for mitochondrial metabolic enzymes involved in cellular respiration. The kDNA is organized into a network of interlocked minicircles and maxicircles. The minicircles, which are smaller (~1 kb) and more numerous (~100 copies), code for guide RNAs, while the maxicircles, which are larger (~25kb) and less abundant (~50 copies), code for pre-mRNAs. The pre-mRNAs encoded by the maxicircles mature by a unique process known as RNA editing, involving selective insertion and deletion of uridine residues. Ribosomal RNA and pre-mRNAs are transcribed by a phage-like polymerase in multicistronic units (Hajduk and

Ochsenreiter 2010). The guide RNAs are produced by 3' nucleolytic processing and uridylation of longer precursors encoded by the minicircles, and direct the editing reactions. Most proteins involved in the RNA editing process and all tRNAs are imported from the cytoplasm, although many of these molecules mature by additional modifications which occur in the kinetoplast (Aphasizhev and Aphasizheva 2011).

1.2.2 Karyotypic variation in *Leishmania*

The genomes of Old World species of *Leishmania* are organized into 36 heterologous chromosomes, while the New World species fall into two groups: species belonging to the *Viannia* subgenus have a 35-chromosome karyotype due to a fusion between chromosome 20 and chromosome 34, while species belonging to the *Leishmania* subgenus have a 34-chromosome karyotype, due to a fusion between chromosome 8 and 9 and between chromosome 20 and 36 (Britto, Ravel et al. 1998).

Overall, genome structure is remarkably consistent across species, with conserved synteny observed for approximately 99% of all genes. Overall, only ~200 genes are distributed in different regions of the genome in *L. major*, *L. infantum*, and *L. braziliensis*. Coding regions in *L. major* and *L. infantum* share up to 92% of the amino acid sequence, and 94% of the nucleotide sequence, while in the comparison with *L. braziliensis* these numbers drop to approximately 77% and 81%, respectively (Peacock, Seeger et al. 2007).

An important feature of *Leishmania* genomes is the extreme variation in chromosome copy number between strains of the same species, with individual chromosomes often having more than the two copies expected in diploid organisms (Rogers, Hilley et al. 2011, Sterkers, Lachaud et al. 2011, Lachaud, Bourgeois et al. 2014). Fluorescent *in situ* hybridization (FISH) experiments and whole genome sequencing (WGS) have shown that each heterologous chromosome may have a somy level different from that of the other chromosomes. Individual cells of any given lab-adapted strain have been shown to have variable patterns of aneuploidy, so that a given sample from a patient for instance, once grown in culture is rarely formed by a homogeneous population of cells with the same karyotype, a condition that has been named mosaic aneuploidy (Mannaert, Downing et al. 2012, Sterkers, Crobu et al. 2014). Monosomy of chromosomes has also been observed. For unexplained reason, chromosome 31 seems to be tetrasomic in many of the strains examined thus far.

In addition to the variation seen in chromosome number, *Leishmania* parasites are also susceptible to amplification or deletion of certain regions of the genome. Gene amplification has been documented as formation of linear or circular extrachromosomal amplicons, or as intrachromosomal gene duplication, which leads to the formation of tandem gene arrays. A conserved extrachromosomal circular amplicon was found in *L. donovani* field isolates (Downing, Imamura et al. 2011), and extrachromosomal amplicons are known to occur *in vitro* in response to drug pressure (Beverley, Coderre et al. 1984). The presence of short sequence repeats flanking individual genes was proposed as a mechanism facilitating

amplification and deletion via the molecular pathway involved in homologous recombination (Ubeda, Legare et al. 2008, Laffitte, Genois et al. 2014, Ubeda, Raymond et al. 2014).

1.2.3 Transcriptional regulation (or lack thereof)

Since *Leishmania* parasites lack eukaryotic promoters, much research has focused on how these organisms can modulate gene expression in response to environmental stimuli. The question of how parasites turn on and off genes involved in developmental transitions, such as in the differentiation from the extracellular promastigote to the intracellular amastigote stages, is an area of active inquiry. Constitutive changes in gene expression in response to selection, however, as described in Section 1.2.2, seem to arise over a limited number of generations via copy number variation at the level of individual genes, or at the level of whole chromosomes.

No control of transcriptional initiation has been so far found in *Leishmania*. Polycistrons appear to be constitutively transcribed into mRNA by RNA polymerase II. These transcriptional units resemble genomic elements lacking TATA boxes and Inr promoters found in lower eukaryotes as well as mammals (Carninci, Sandelin et al. 2006), suggesting that constitutive transcription may be the ancestral state common to all eukaryotes. In addition to protein-coding genes transcribed by RNA polymerase II, tRNA genes and rRNA genes in *Leishmania* are transcribed by RNA

polymerase III and RNA polymerase I, respectively, at defined initiator and terminator sequences (Das, Banday et al. 2008).

Transcriptional unit boundaries are enriched in histone acetylation marks: the H2A.Z and H2B.V marks have been functionally confirmed as essential, while H3.V is not required for normal transcriptional activity in both *Leishmania* (Thomas, Green et al. 2009, Anderson, Wong et al. 2013) and *Trypanosoma* (Siegel, Hekstra et al. 2009). While histone variants seem to have a conserved function in different kinetoplastids, other types of epigenetic modification appear to have a genus- or species-specific function. A hyper-modified base unique to kinetoplastid protozoa, glycosylated hydroxymethyluracil, also called base J, is enriched in telomeric regions, but is also present in strand switch regions in *Leishmania* (Genest, Ter Riet et al. 2007, van Luenen, Farris et al. 2012). This modified base is produced by a two-step hydroxylation and glycosylation reaction of thymine residues, and appears to serve a genome-wide function in *Leishmania* regulating RNA polymerase II transcription termination (van Luenen, Farris et al. 2012, Reynolds, Cliffe et al. 2014).

Gene expression studies have found a small number of genes differentially expressed between promastigote and amastigote stages. Global interspecies expression analyses have found that the majority of genes differentially regulated throughout development in one species, however, are not differentially regulated in others (Rochette, Raymond et al. 2008). In both *L. major* and *L. infantum*, only 7 to 9 percent of the genome is differentially expressed in these two life stages according to oligonucleotide microarray data (Rochette, Raymond et al. 2008). Interestingly,

up to 95 percent of the small set of genes found to be differentially expressed in *L. infantum* promastigotes are no longer differentially expressed if parasites are isolated from axenic culture as opposed to the sand fly midgut (Alcolea, Alonso et al. 2014), suggesting a significant bias introduced by *in vitro* culture. In the New World species *L. braziliensis*, only 9 percent of the genes are differentially expressed throughout the life cycle (Depledge, Evans et al. 2009). The small number of genes differentially expressed between promastigote and amastigote stages have been proposed to be a well-conserved pre-adaptation to intracellular survival.

Despite the small number of developmentally regulated genes, regulation of gene expression during development is thought to occur either post-transcriptionally, or by changes in epigenetic modifications affecting transcriptional activity. Changes in steady state transcript levels are thought to occur primarily via differences in the maturation and stability of individual mRNAs via interactions with RNA-binding proteins. Due to the small size of the sequence between the SL splice site and the start of the protein-coding sequence, *Leishmania* parasites have extremely short 5' UTRs. Regulatory sequences in the 3' UTR region of protein-coding genes have therefore been implicated in determining the stability of the trans-spliced mRNA transcript.

Translational and post-translational mechanisms determining steady state protein levels are also thought to be important in developmental regulation of *Leishmania* parasites (Bente, Harder et al. 2003). In particular, the increase in temperature associated with transmission from an insect vector to a mammalian host is thought to trigger a developmental programme via the activity of heat-shock

proteins such as HSP70, which act as chaperones and regulate production and maturation of proteins involved in the subsequent stress response.

1.3. A clonal, parasexual, or sexual organism?

1.3.1 The clonal theory

Since the 1990s, there has been ongoing debate as to the mating system and population structure of eukaryotic microbial pathogens such as *Leishmania*. These two biological aspects are closely linked to each other, and have important consequences for the epidemiology of transmissible diseases. The predominant view for many years has been that parasites such as *Leishmania* were primarily asexual, and that their population structure was essentially clonal (Tibayrenc, Kjellberg et al. 1990).

Support for the clonal theory has been based on the fact that several population genetic tests with a null hypothesis of panmixia, or random mating between individuals, have found significant deviations from Hardy-Weinberg expectations. A clonal population structure, defined by the predominance of a restricted mating system resulting in each generation being genetically identical to the generation that precedes it, is thus inferred as a result of statistical tests that reject the null hypothesis of unrestricted, random mating in the population. In this sense, asexuality and selfing via a sexual process may under most circumstances be virtually indistinguishable in population genetic data.

Several quantitative tests in *Leishmania* have yielded statistically significant results in this regard. These tests have focused on measures of linkage disequilibrium, or non-random associations between genotyped loci, and skewing of allele frequencies in the population due to prevailing homozygosity or heterozygosity (Tibayrenc and Ayala 2002). Qualitative observations have also been used in support of the clonal theory. These observations have focused on the effects one expects asexual reproduction and uniparental mating to have on the segregation and recombination of genetic markers in the population, and include considerations such as the widespread over-representation of identical genotypes, the absence of recombinant genotypes, the presence of fixed homozygosity or heterozygosity in a given population, and concordance in phylogenetic signal between independent sets of genetic markers (Ramirez and Llewellyn 2014).

In *L. tropica*, isoenzyme profiling of 27 isolates found significant heterogeneity and fixed heterozygosity (Le Blancq and Peters 1986). In addition, genotypes that should be segregating in natural panmictic populations, often cannot be found in *L. tropica* and several other Old World *Leishmania* species. The presence of ubiquitous genotypes such as the MON1 zymodeme in *L. infantum*, which is stable both across space and time (Rioux, Lanotte et al. 1990), has also been cited as evidence of predominant asexual reproduction. Species such as *L. infantum* that have a broad geographic range have been found to have extremely low genetic diversity and conserved linkage blocks by both MLEE and microsatellite MLST, with zymodemes such as MON1 being present across continents, but some sub-structuring detectable by MLST (Seridi, Amro et al. 2008, Kuhls, Alam et al. 2011). In

New World species of *Leishmania*, high concordance has been found between different typing methods, such as AFLP, MLEE, MLST and PFGE (Odiwuor, Veland et al. 2012). Both MLEE and RAPD data has been brought forward in support of linkage disequilibrium in the *Viannia* subgenus (Banuls, Jonquieres et al. 1999). Homogeneity between markers was also found in *L. donovani* in India (Alam, Kuhls et al. 2009).

1.3.2 Challenges to the clonal theory

Classic statistical tests employed in population genetics can only measure deviations from stated expectations. In the case of the clonal theory as originally proposed, quantitative and qualitative deviations from panmixia have been attributed to a predominance of self-mating (i.e. uniparental mating) and true asexual reproduction, with little differentiation between these two possible scenarios. However, in many instances rejecting the null hypothesis of panmixia doesn't in and of itself justify such broad conclusions. The original clonal theory has since been re-defined (Tibayrenc and Ayala 2012) to account for the distinction between self-mating and asexual reproduction, but some important issues remain.

Two main caveats need to be made explicit with respect to the body of evidence which has been used to argue in support of the clonal theory: these concern methodological limits due to sample size and choice of genetic markers, on one hand; and the theoretical models that may be invoked to explain trends in the population genetic data when panmixia is not observed, on the other.

Firstly, methodological issues may result in a statistically significant trend in the data, when in actuality there is none. It is a widely recognized fact in biology that inadequate sample sizes may yield artifacts due to insufficient power or to sampling bias, and it is beyond the scope of this introduction to analyse these aspects in detail. In addition, the choice of markers for MLST, AFLP, RAPD, and MLEE can significantly affect the results obtained. The fact that strains typed as belonging to the MON1 zymodeme break down into polyphyletic units when analysed with microsatellite MLST (Banuls, Hide et al. 1999) proves that high resolution markers can provide crucial additional information.

Secondly, a number of population processes may limit gene flow and recombination in sexually reproducing populations. Identifying the barriers to gene flow that may result in deviations from panmixia independently of biological features intrinsic to the parasite, however, is a challenging process. Such barriers may be posed by reproductive isolation due to geographic separation or to the presence of multiple, distinct transmission cycles, for example because of different vectors or reservoirs, even in the presence of an obligatory sexual life cycle. This has been called the Wahlund effect, and its defining feature is the detection of heterozygosity deficits as a result of substructuring of a population that is nonetheless following panmixia.

The first report of potential hybrids in natural populations concerned two inter-specific hybrids between *L. arabica* (now considered to be synonymous with *L. tropica*) and *L. major* that were isolated from zoonotic sources in Saudi Arabia, and were confirmed to be hybrids by a variety of analytical methods (Evans, Kennedy et

al. 1987, Kelly, Law et al. 1991). Putative hybrids between New World species of *Leishmania* were also found in Nicaragua (Belli, Miles et al. 1994), Ecuador (Banuls, Guerrini et al. 1997), and Venezuela (Delgado, Cupolillo et al. 1997). Natural inter-specific hybrids were later also found in Peru (Shani-Adir, Kamil et al. 2005, Nolder, Roncal et al. 2007). Hybrids between *L. major* and *L. infantum* were also reported in immunosuppressed patients in Portugal (Ravel, Cortes et al. 2006), and hybridization was found in *L. donovani* complex isolates from Turkey (Rogers, Downing et al. 2014).

The first experimental evidence of hybridization came from crosses with *L. major*, and clearly showed that *Leishmania* parasites have the capacity of undergoing sexual reproduction in the sand fly stages (Akopyants, Kimblin et al. 2009). Experimental hybrids have since also been generated in *L. donovani* (Sadlova, Yeo et al. 2011), and most recently between *L. infantum* and *L. major* (Romano, Inbar et al. 2014), indicating that cross-species hybrids may naturally occur and thus confirming previous reports of natural cross-species hybrids identified in the field. Indeed, a series of early microscope studies found evidence for cell fusion in *L. tropica* promastigotes (Lanotte and Rioux 1990). Later reports of nuclear fusion in amastigotes of several species of *Leishmania* as determined by DNA content (Kreutzer, Yemma et al. 1994) may have been detecting pre-existing aneuploids in the sampled cell population, rather than true 2*n* zygotes due to nuclear fusion events between gametes with *n* ploidy.

Several population genetics studies have since cast some doubt on the validity of the clonal theory as it was originally proposed. Using PFGE, recombinant

karyotypic variants were found in *L. infantum* (Blaineau, Bastien et al. 1992). Microsatellite MLST analyses of *L. donovani* complex populations found significant variation in inbreeding coefficients (Kuhls, Keilonat et al. 2007), and after expanding these analyses to New World *L. infantum* and *L. chagasi*, a range of inbreeding coefficients were found, reflecting possible Wahlund effects impacting the number of heterozygotes (Kuhls, Alam et al. 2011). In *L. braziliensis*, evidence for strong Wahlund effects was brought forward, with substantial heterozygote deficits and linkage disequilibrium due to sub-structuring of the population into several micro-foci of transmission (Rougeron, De Meeus et al. 2009), while in another New World species, *L. guyanensis*, very modest linkage disequilibrium was found in addition to an overrepresentation of homozygotes, suggesting substantial recombination (Rougeron, Banuls et al. 2011).

Focusing on the microgeographic scale, a mixed mode of reproduction was suggested for *L. donovani* in Sudan (Rougeron, De Meeus et al. 2011) and in Ethiopia (Gelanew, Kuhls et al. 2010). Recombinant genotypes were also found in *L. infantum* from Tunisia (Chargui, Amro et al. 2009). More evidence of a mixed mode of reproduction with an important role for recombination was offered by MLST analyses in the *Viannia* subgenus (Boite 2012, Kuhls 2013). When the local geographic scale of *Leishmania* transmission and the potential reproductive barriers present at this scale are disregarded, however, low linkage disequilibrium from MLEE or MLST markers can be misinterpreted as absence of recombination as a biological process, especially if the study is underpowered (El Baidouri, Diancourt et al. 2013).

In this dissertation, I follow the recommendations by Rougeron and colleagues (Rougeron, De Meeus et al. 2015), and advise against pooling of samples from different geographic areas and different time periods when making considerations regarding the mode of reproduction in *Leishmania*. I aimed to maintain a distinction between population genetics, which requires knowledge of the demographic units in this important human pathogen, the potential barriers to gene flow, and most importantly, access to a representative sample of individuals from each of these demographic units; and genomic analyses, which focus on genome-level processes at work in individual parasites, and allow accurate qualitative and quantitative comparative considerations to be made.

1.3.3 Models of asexuality, sexuality, and parasexuality

I chose to reject the misguiding terminology characteristic of the clonal theory, which focuses on patterns in the data rather than biological processes, and hereby describe three modes of reproduction that may be at work in *Leishmania*: asexual, sexual, and parasexual. Some overlap between these three modes is clearly possible, and as such they are not fully mutually exclusive, but for ease of reference I provide to the reader a codified description of each "model".

In the asexual model, parasites complete each transmission cycle, from human host to insect vector, without ever performing meiosis. Each cell reproduces mitotically, dividing by binary fission, and each daughter cell receives a full complement of *2n* heterologous chromosomes. Mosaic aneuploidy in this model

arises from unbalanced chromosome replication and/or unequal segregation of the chromosomes during mitotic division, whereby each daughter cell receives *Xn +/- m* chromosomes, where *X* is an integer, *n* is the full complement of heterologous chromosomes, and *m* is the absolute number of chromosomes in excess or in deficit with respect to the balanced ploidy *X*. Meiotic recombination is not observed, although homologous mitotic recombination may be possible under certain circumstances. The progeny can only inherit either the full set or a subset of the heterologous chromosomes present in the parental cell.

In the sexual model, the parasite performs meiosis as part of its transmission cycle, and generates haploid gametes that may or may not belong to separate mating types. Homologous meiotic recombination of the parental chromosomes occurs prior to segregation into the haploid gametes. Inexact cell division at any stage during meiosis may lead to gametes with chromosomes in excess or in deficit of the heterologous set *n*, thus resulting in aneuploidy of the progeny. Cell fusion thus occurs between compatible gametes, if mating types are indeed present, and gives rise to zygotes with *Xn* +/- *m* chromosomes that do not divide any further until mitosis is resumed.

In the parasexual model, there is no meiosis. Parasites perform cell fusion without prior generation of haploid gametes, and unequal segregation of the chromosomes in the daughter cells during subsequent cell division gives rise to progeny with a variable number of chromosomes *Xn* +/- *m*, that nevertheless successfully retain some combination of chromosomes from each parent. Homologous recombination may occur prior, during, or following cell fusion,

although it may affect only part of the genome. Distinct ameiotic mating types may still be present.

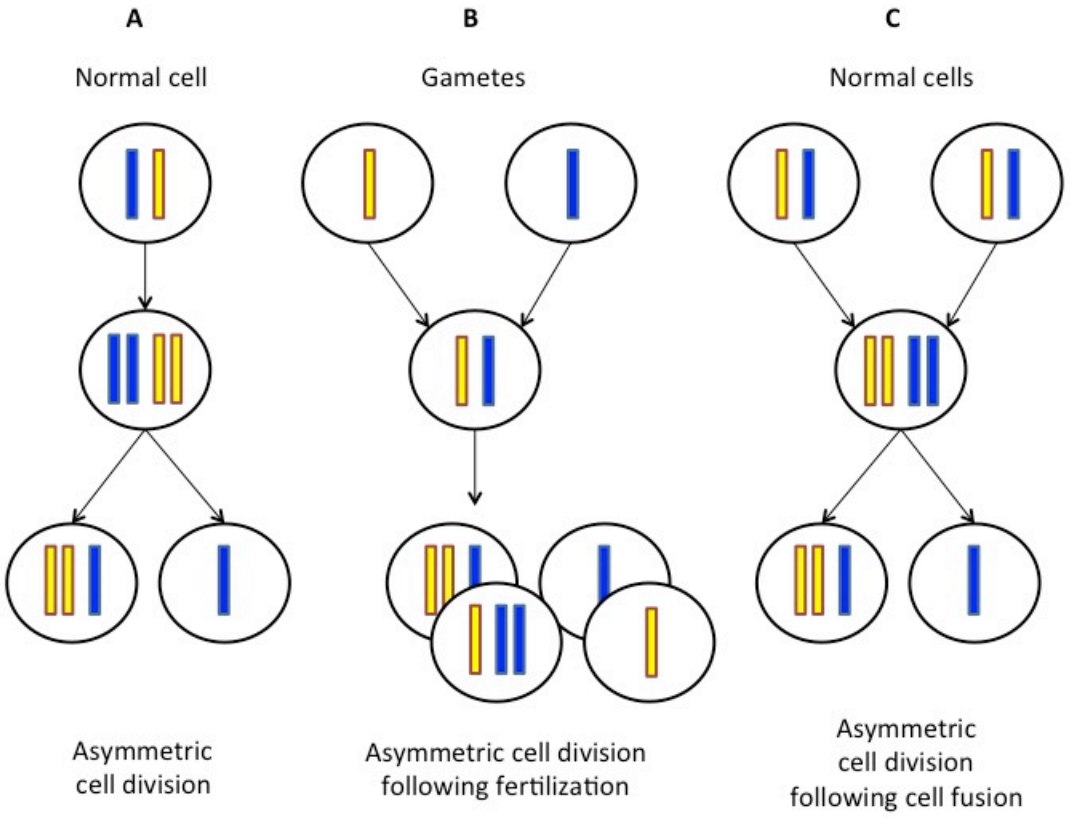

Figure 1.5. Schematic models of asexual, sexual and parasexual reproduction as referred to throughout this dissertation. (A) Asexual reproduction starts from normal somatic cells, which undergo chromosome replication and subsequent mitotic cell division. Asymmetric replication or asymmetric cell division may give rise to aneuploidy in the progeny. Homologous recombination is negligible. (B) Sexual reproduction starts with fertilization of gametes, possibly representing different mating types, which were

generated by a reductional meiotic process involving homologous recombination. Fertilization of the gametes gives rise to a zygotic form that may then generate aneuploidy daughter cells by asymmetric cell division in a mitotic process similar to that depicted in (A). (C) Parasexual reproduction starts with cell fusion of normal somatic cells, and then proceeds with reductional cell division that redistributes the chromosomes unevenly, thus generating aneuploid daughter cells. Homologous recombination may occur. (Source: adapted from Sterkers et al. 2014).

In both the sexual and parasexual models, cell divisions associated with a sexual or parasexual event may "reset" the ploidy each transmission cycle if there is balanced segregation of the chromosomes. The widespread mosaic aneuploidy reported in *Leishmania* may arise from subsequent inexact mitotic cell divisions, which may be called "somatic", rather than from the sexual or parasexual event itself.

Despite the many simplifications made in this section, I will refer back to this schematic representation of the sexual, parasexual, and asexual models throughout my dissertation as a useful way to make these discussions more accessible. In this simplified view, recombination plays a major role in shaping the *Leishmania* genome only in the parasexual or sexual models, although it may be present in all three models.

1.3.4 Aims and objectives

With this dissertation, I seek to address the open question of whether *Leishmania* parasites follow a sexual life cycle, and describe the changes that occur in their genome during hybridization. I describe how my findings fit with what we know about *Leishmania* genome biology, and with the models described in section 1.3.3. I then make observations on how genome plasticity and genetic exchange may affect parasite evolution in the field in response to selective pressures. Each chapter provides an overview of the experimental procedures and *in silico* analyses that I performed to address this central question in their respective Methods sections.

Chapter 2 describes the results from multi-locus sequence typing of 34 different isolates of *L. tropica* covering the entire geographic range of the species. This chapter then compares and contrasts the MLST data with whole-genome sequencing (WGS) data from some of these isolates, and makes some observations on how hybridization may have contributed to *L. tropica* population structure and genetic diversity.

Chapter 3 investigates in depth mechanisms of genome plasticity and their effect on gene expression via paired WGS and RNA-seq of a subset of 14 field isolates from the set of 34 discussed in Chapter 2. This set of clinical isolates was complemented with an additional 6 isogenic lines obtained by cloning of 4 of these isolates, with the aim of understanding the effects of mosaic aneuploidy on transcription.

Chapter 4 describes generation of transgenic single-drug resistant lines, sand fly feeding assays, and selection of double-drug resistant hybrids in sand fly lab-adapted colonies, and the implications of hybrid recovery rates from infected sand flies in laboratory crosses for estimating the frequency of hybridization in a natural setting.

Finally, Chapter 5 describes genetic exchange in experimental hybrids, and provides an exhaustive list of *de novo* mutations, recombination, structural rearrangements, and Mendelian violations in the inheritance of the genetic material of the two parental lines. I provide the first high-resolution, complete description of the effects of hybridization on genome sequence and structure in *L. tropica*.

In the Conclusions, I provide a summary of my findings and how they have shed new light on our current understanding of the processes generating and maintaining genetic variation in this important pathogen.

I hope that this thesis will contribute to the open debate regarding the predominant mode of reproduction in *Leishmania*. The work presented here provides for the first time experimental evidence in support of sexual or parasexual reproduction in the Old World species *L. tropica*, and on this basis I strongly reject the asexual model. Although a rigorous disqualification of either sexual or parasexual reproduction in favour of one or the other is not warranted by the present data, I provide to interested readers some indication of how future studies informed by our work can solve this debate once and for all.