Virulence and adaptation of *Leishmania donovani* parasites

Patrick Lypaczewski

Department of Microbiology and Immunology

McGill University, Montreal

December 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dla mojej rodziny

TABLE OF CONTENTS

Note: the electronic copy of this thesis contains navigation bookmarks

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADR	Arginine deprivation response
BLAST	Basic local alignment search tool
CDC	Center for Disease Control
CL	Cutaneous leishmaniasis
DNA	Deoxyribonucleic acid
ER	Endoplasmic reticulum
GFP	Green fluorescent protein
GIPL	Glycoinositolphospholipids
GPI	Glycosylphosphatidylinositol
HIV	Human immunodeficiency virus
IGV	Integrative Genomics Viewer
ISC	Indian subcontinent
IV	Increased virulence
LAMP	Loop-mediated isothermal amplification
LPG	Lipophosphoglycans
ML	Mucocutaneous leishmaniasis
mRNA	messenger RNA
mTORC1	mammalian Target Of Rapamycin Complex 1
mTORC2	mammalian Target Of Rapamycin Complex 2
MVT	Multivesicular tubule lysosome
MYA	Million years ago
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Bioinformatic Information
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PI	Phosphatidylinositol

PI3K	Phosphatidylinositol 3-kinase
PKDL	Post kala-azar dermal leishmaniasis
PNA	Peanut agglutinin
PPG	Proteophosphoglycan
RDT	Rapid diagnostic test
RNA	Ribonucleic acid
ROS	Radical oxygen species
rRNA	ribosomal RNA
RT	Room temperature
SNARE	SNAP Receptor
SNP	Single nucleotide polymorphism
SL	Spliced leader
SLC38A9	Solute carrier family 38 member 9
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SRA	Sequencing Read Archive
tER	transitional endoplasmic reticulum
TOR	Target Of Rapamycin
tRNA	transfer-RNA
UTR	Untranslated region
UV	Ultraviolet
VL	Visceral leishmaniasis
WHO	World Health Organization

ABSTRACT

Leishmaniasis is a neglected tropical disease arising from *Leishmania* infection endemic in over 88 countries. The *Leishmania* parasites are transmitted by sand fly vectors during bloodfeeding. Contingent on the species of *Leishmania* transmitted, the disease can take on two main pathologies: cutaneous leishmaniasis or visceral leishmaniasis. While cutaneous lesions remain in the skin and are generally self-healing, infection of the visceral organs is nearly always fatal if left untreated. The key determinants allowing some species to cause a fatal visceral infection while others are limited to a more benign cutaneous infection remain unclear. We hypothesized that genetic differences between parasites causing different pathologies are responsible for controlling disease progression.

As two highly related strains of *Leishmania donovani* were found to cause cutaneous and visceral infections in Sri Lanka, we first investigated whether we could improve genomic studies of *L. donovani* by generating an improved reference genome. Using new sequencing and assembly technologies, we were able to generate a complete genomic assembly which increased the number of annotated genes and increased the number of genetic variations identified between the cutaneous and visceral *L. donovani* strains.

As one of the genetic variations identified resulted in a non-conservative amino acid change in a Rag GTPase which signals to the TOR pathway in mammalian cells, we investigated this pathway and the contribution of this mutation in the atypical cutaneous strain seen in Sri Lanka. Through CRISPR gene editing and biochemical assays, we provided evidence that the RagC protein plays a role in *L. donovani* virulence during visceral infection.

As additional isolates were reported to cause cutaneous disease in Sri Lanka, but they did not appear related to the strains we previously identified, we investigated the etiological origins of all *L. donovani* strains sequenced to gain insight into the causative agents of the epidemic in Sri Lanka. Through our analysis, we identify evidence of multiple interspecies genomic hybridization events and introgression in some strains circulating in Sri Lanka that could account for the cutaneous disease phenotype.

Overall, these results show there are several ways in which *L. donovani* can modulate its pathology and that genomic hybridization may represent a major evolutionary adaptation mechanism in nature. These results highlight several independent genetic events in independent strains all culminating in the same geographical area with an identical phenotype which suggests a yet unidentified driving force for these changes.

RÉSUMÉ

La leishmaniose est une maladie tropicale négligée résultant d'une infection par *Leishmania* et est endémique dans plus de 88 pays. Les parasites *Leishmania* sont transmis par le vecteur phlébotome durant leur piqure. Tout dépendamment de l'espèce de *Leishmania*, la maladie peut prendre deux formes principales: la leishmaniose cutanée et la leishmaniose viscérale. Bien que les lésions cutanées guérissent généralement d'elles-mêmes, les lésions viscérales sont presque toujours mortelles sans traitement. Les facteurs déterminants qui permettent à certaines espèces de causer une maladie viscérale fatale tandis que d'autres sont limitées à une infection cutanée plus bénigne ne sont pas entièrement élucidés. Notre hypothèse est que certaines différences génétiques entre les parasites qui causent ces différentes pathologies contrôlent la progression de la maladie.

Puisque deux souches semblables de *Leishmania donovani* causent des lésions cutanées et viscérales au Sri Lanka, nous avons d'abord déterminé s'il serait possible d'améliorer les analyses génomiques portant sur *L. donovani* en générant un nouveau génome de référence. En utilisant de nouvelles technologies de séquençage et d'assemblage, nous avons produit un génome de référence complet ce qui a augmenté le nombre de gènes annotés et a augmenté le nombre de différences génétiques identifiées entre les souches cutanée et viscérale de *L. donovani*.

Comme une des variations identifiées se traduisant en un changement d'acide amine non-conservé dans une GTPase de type Rag qui signale la voie de signalisation TOR chez les mammifères, nous avons enquêtés sur la présence de cette voie de signalisation et l'effet de cette mutation identifiée dans la souche cutanée issue du Sri Lanka. En utilisant CRISPR et des analyses biochimiques, nous avons démontré le rôle central que joue la protéine RagC dans la virulence de *L. donovani* lors d'une infection viscérale.

Comme plusieurs nouvelles souches identifiées au Sri Lanka étant capables de cause de lésions cutaneous n'apparaissaient pas être apparentées aux souches que nous avons précédemment décrites, nous avons déterminés les origines étiologiques de toutes les souches de *L. donovani* séquencées afin de mieux comprendre l'agent causateur de l'épidémie présente au Sri Lanka. Dans notre analyse, nous avons identifié de multiples traces d'hybridation génomique et d'introgression entre espèces dans certaines souches au Sri Lanka.

Globalement, nos résultats démontrent que *L. donovani* peut moduler sa pathologie de plusieurs façons et que l'hybridation génomique peut représenter un mode d'adaptation évolutionnaire majeur dans la nature. Nos résultats démontrent que plusieurs mécanismes génétiques indépendants ont contribué au phénotype commun de plusieurs souches indépendantes qui culminent aux Sri Lanka, ce qui suggère l'existence d'une force motrice inconnue.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr Greg Matlashewski. More than a supervisor, he has been a true mentor throughout the years. His belief in me, his encouragement and feedback at every stage of my project are in large part responsible for the quality of work presented herein. I would also like to thank my advisory committee members, Dr Martin Olivier and Dr Martin Richer for their advice and critical insights.

I would like to thank Drs Laura-Isobel McCall and Wen-Wei Zhang for training me in the molecular techniques of *Leishmania* and for their support. I would like to thank the past and present undergraduate members of the lab for giving me an opportunity to in turn transfer this knowledge and for keeping me company during long experiments. I would like to thank Kayla Paulini for helping me with experiments when I had too many, getting me food when I had too little and keeping me sane when I was tired.

I would like to thank all past and current students at the Lyman-Duff building for sharing your reagents but also for your time and our discussions. Sacha, Felix, Michael, Anshul, Alex, Giustino, and Adam, you've made long days go by quicker and long meals go by slower. I would also like to thank the past executive members of MIGSA with who I had the pleasure to serve and the current members for their continued dedication to the graduate student body.

I would like to thank my wife Jasmin, for her never-ending support in every aspect of my career and life and for making every day brighter.

Finally, I would like to thank my family for always believing in me and supporting me in every endeavor I set my mind to. Everything I am today I owe to you.

CONTRIBUTION OF AUTHORS

Chapter 2:

Adapted from: **Patrick Lypaczewski**, Johanna Hoshizaki, Wen-Wei Zhang, Laura-Isobel McCall, John Torcivia, Vahan Simonyan, Amanpreet Kaur, Ken Dewar, Greg Matlashewski, 2018. A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. *Scientific reports*, *8*(1), p.16549.

PL wrote the main manuscript and performed the analysis. JH, JTR and AM performed analysis. WZ, LIM, VS and KD provided materials and insight. GM planned the experiment and wrote the manuscript. All authors reviewed the manuscript.

Chapter 3:

Adapted from: **Patrick Lypaczewski***, Wen-Wei Zhang* and Greg Matlashewski, 2021. Evidence that a naturally occurring single nucleotide polymorphism in the RagC gene of *Leishmania donovani* can reduce virulence. *PLoS Neglected Tropical Diseases,* 15(2): e0009079.

PL designed the study, performed experiments, data analysis and wrote the manuscript. WZ designed the study, performed experiments, data analysis and helped write the manuscript. GM designed the study, helped write and edited the manuscript. All authors reviewed the manuscript.

Chapter 4:

Adapted from: **Patrick Lypaczewski**, Greg Matlashewski, 2021. *Leishmania donovani* hybridization and introgression in nature: a comparative genomic investigation. *The Lancet Microbe*, (accepted).

PL designed the study, collected and analysed the data and wrote the manuscript. GM helped design the study, wrote and edited the manuscript. Both authors had access and verified the underlying data and have read and approved the final version of the manuscript.

CHAPTER 1: INTRODUCTION

1.1 Leishmaniasis

1.1.1 Clinical presentation - Diagnosis

Leishmaniasis encompasses the different diseases caused by the protozoan parasite *Leishmania,* transmitted through the bite of an infected sand fly. While the parasites are closely related, the pathologies resulting from infection can vary greatly (1, 2). Leishmaniasis is broadly classified in one of two main disease phenotypes: cutaneous leishmaniasis or visceral leishmaniasis.

1.1.1.1 Cutaneous Leishmaniasis

Cutaneous leishmaniasis (CL) presents itself as dry or wet cutaneous lesions on exposed areas of the skin at the site of the insect bite (3). As CL is caused by a diverse group of *Leishmania* species, the specific presentation of CL lesions can vary (2). The lesions can be focused or dispersed and in general cause no pain or discomfort with the tissue retaining sensation. Patients presenting with CL lesions are generally healthy and do not present with any systemic symptoms.

Generally within two to eight weeks, a red papule appears at the site of the sand fly bite (4). The lesions, often accompanied by raised edges, then progressively enlarge, ulcerate and crust over. Over the course of a few months to a few years, the lesions generally scab over and resolve without intervention. The lesions however are scarring which contributes to social stigmatisation and can lead to secondary bacterial or fungal infections (4, 5). A large number of unresolved CL lesions, in combination with the required species, can also contribute to the occurrence of secondary mucocutaneous leishmaniasis. While generally resulting in a single nodule per sand fly

bite, disseminated or diffuse cutaneous leishmaniasis can result in a large number of lesions across the body. The progression of CL to diffuse or disseminated disease has been linked to immunosuppression (2, 6).

CL lesions can sometimes be confused with leprosy or tuberculosis, requiring accurate diagnostics (7, 8). Current diagnostics rely mainly on microscopic observation, culture methods, or parasite targeted PCR using lesion content swabs (8). New methods consisting of immunochromatographic rapid diagnostic test strips (RDT) and loop mediated isothermal amplification (LAMP) aimed at increasing the availability of diagnostics to unspecialized personnel have more recently been developed but not yet routinely used (8–11).

1.1.1.2 Visceral Leishmaniasis

Visceral leishmaniasis (VL), unlike cutaneous leishmaniasis, does not cause visible symptoms until later stages of infection as asymptomatic infections are significantly more prevalent (2, 12). VL causing parasites migrate from the sand fly bite site to the visceral organs in infected individuals, notably the liver, spleen and bone marrow (13). The spread of the disease to those organs can lead to fever, lethargy, and hepatosplenomegaly. The pancytopenia associated particularly with late stage visceral infection increases the likelihood of fatal secondary bacterial infections and the presence of autoimmune antibodies can obscure proper diagnosis (2). Additionally, the disease can in some locations results in hyperpigmentation which resulted in the naming the disease ‘kala-azar’ or black fever in the Indian subcontinent.

While predominant in Ethiopia, co-infections of HIV/VL have been reported in 35 countries around the world including parts of southern Europe (14, 15). This co-infection, along

with other immunosuppressive conditions can sometimes result in severe disease manifestation several years following the initial infection (2). As both pathogens affect host macrophages, co-infection contributes to an acceleration of the progression of both diseases and poses challenges for successful treatment (2).

Once their visceral organs are affected, individuals do not recover without intervention and will succumb to the infection, usually within two years from the initial infection (2). If successfully diagnosed, the disease is treatable, and the several drug options are discussed further. Diagnosis of this disease can be performed using microscopic examination. Sample extraction is however complicated due to the nature of the affected organs. Parasite DNA targeted PCR or RDTs designed to detect antibodies against the parasite are routinely used but may be complicated by insufficient immune antibody responses due to immunosuppression linked to co-infections or malnutrition (1).

1.1.1.3 Mucosal Leishmaniasis

Although globally less prevalent than both CL and VL, a third disease type is the mucosal leishmaniasis (ML), which usually occurs years after a treated cutaneous infection and leads to acute inflammation of the mucosal membranes and their subsequent destruction due the host's heightened immune response (3). This damage is often disfiguring as it can also affect cartilage and can result in social stigmatization in addition to being potentially life-threatening due to the increased risk of aspiration pneumonia (2). As ML lesions have a low parasite concentration, diagnosis is most effective using sensitive molecular techniques such as PCR (2).

1.1.1.4 Post Kala-azar Dermal Leishmaniasis

A fourth possible type, also not as common as CL and VL, is post kala-azar dermal leishmaniasis (PKDL) (16). Although sometimes classified as a separate pathology, PKDL is (as the name implies) rather more of a possible complication following primary pathology. Kala-azar, meaning black fever is synonymous with VL caused by *Leishmania donovani* in India and treatment of the disease can sometimes result in a second wave of infection years later limited to, but dispersed across the skin in contrast to the primary visceral infection (2).

While most cases of PKDL in Africa present papular lesions, PKDL in the Indian subcontinent typically causes macular rashes and occasionally papular lesions. While spread across a large area of the body however, PKDL does not appear to contribute to the active transmission of VL as the incidence of VL in PKDL household contacts was not found to be higher than in control households (17) although xenodiagnoses in laboratory conditions showed that PKDL cases are able to transmit parasites to sand flies (18–20). The time interval between primary and secondary infections has been reported to be influenced by the type of treatment used during primary infection and the age of infected individuals, due to their influence on the immune response against the parasite (21). In addition to the frequency of PKDL being higher in immunocompromised individuals (2), the location of rashes being more frequent in areas exposed to the sun has been linked to the immunomodulatory effect of UV radiation, indicating a link between the immune response and the occurrence of PKDL lesions.

This infection can sometimes be confused with leprosy or fungal infections due to its appearance. Due to antibodies raised against the primary visceral infection remaining following treatment, serological diagnosis is often inconclusive. Diagnosis through microscopic observation

of parasites is possible but lacks sensitivity (2). While this secondary infection is treated in the Indian subcontinent, PKDL cases in East Africa are not as they are usually self-healing (2, 22).

1.1.2 Epidemiology

Of the leishmaniases, CL is by far the most prevalent. Across the 88 countries where leishmaniasis is endemic, over 400 million people are considered at risk of contracting CL (23). CL contributes an estimated 600,000 to 1,200,000 new cases each year with over 70 percent of new cases identified in Afghanistan, Algeria, Brazil, Colombia, Costa Rica, Ethiopia, Iran, Peru, Sudan and Syria (23). (**Figure 1.1**)

In many cases in the same tropical countries endemic for CL, VL is a threat to over 600 million inhabitants. Although not as common as CL, the 50,000 to 90,000 new cases of VL each year have a profound impact on the population due to the fatal nature of the infection. Of those cases, 90 percent are concentrated in India, Sudan, South Sudan, Kenya, Somalia, Ethiopia, and Brazil (23).

Figure 1.1 Geographical distribution of CL and VL

Map of the distribution of Leishmaniasis across the world. The disease is endemic in 88 countries.

Adapted from Soong, *Leishmaniasis,* 2009 (24).

1.1.3 Historical perspective/ Discovery

Reports of skin lesions consistent with CL have been found dating back to 700-2500 BC in the Middle East and genetic sequencing detected *Leishmania* genetic material in mummy tissues analyzed (25). It was not until the early 1900s that Leishman and Donovan independently identified the parasites in spleen samples that had been infected in India (25, 26) by staining and observing the samples under a microscope. These parasites were then confirmed to be carried by, and later transmitted by, the bite of infected sand flies by allowing infected sand flies to feed on volunteers (25). Although their discovery is fairly recent, these parasites and their vectors have been present for much longer as analysis of amber samples found sand flies infected with *Leishmania*-like parasites dating back to the cretaceous prehistoric period (25, 27).

1.1.4 Evolution

The evolution of *Leishmania* has been suggested to have occurred prior to the Pangea breakup resulting in ancient parasites across both the Old World (Europe, African and Asia) and New World (The Americas) (25, 27) with debate as to the location of the original evolution. *Leishmania* is suggested to originate from the sand fly monoxenous parasite *Leptomonas* which eventually adapted to a dixenous life cycle, originally through blood feeding on reptiles followed by rodents as mammals became more established, as evidence by comparative genomics and evidence of the possibility of mammalian infection (27, 28).

The most widely supported theory places the ancestral evolution event in South America (**Figure 1.2)**, with dispersion across America followed by Asia through the connecting Bering land

bridge (29). As the parasites spread through Central Asia, the parasite further diverged upon entering the Indian sub-continent (ISC), Europe, and Africa (25, 27, 29).

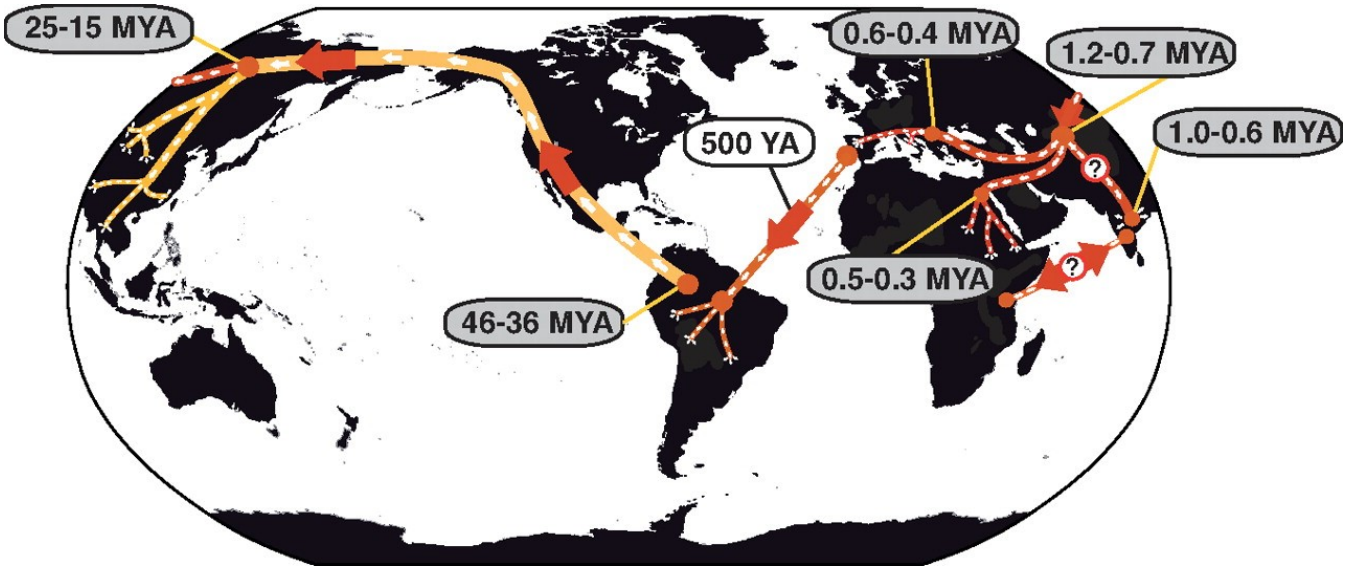


Figure 1.2 Evolution and spread of *Leishmania* across the world

Starting in South America, the ancestor to *Leishmania* spread northwards and colonized Asia (labelled in yellow) followed by speciation events (labeled in red). Estimated timing of events is labeled in millions of years ago (MYA). Adapted from Lukeš et al., *PNAS*, 2007 (29). Copyright (2007) National Academy of Sciences, U.S.A.

1.1.5 Sand flies

Sand flies are thought to have originally played a role as the host to the ancestor of the leishmania parasites rather than a vector (27). As the parasites evolved, they moved to the animal reservoirs the sand flies were feeding on, and the sand flies were relegated to vectors of this disease. Due to this co-evolution, there exists a preferential association between species of sand flies and the species of *Leishmania* they can carry (27). This close relationship has been suggested to involve the microbiome of the sand fly (30) and to provide the sand fly with colonization resistance from pathogenic bacterial infections (31). Around 166 different species of sand flies are thought to carry different species of *Leishmania* parasites (27) with the genus *Lutzomyia* responsible for transmission in the New World and *Phlebotomus* in the Old World. While there exists an ideal parasite species to vector species pairing, the vectors are not entirely restrictive to other species.

Sand flies are not however simply a vector for the parasite, they also contribute to the disease progression as the infection benefits from sand fly salivary antigens (32, 33) and compounds secreted by *Leishmania* during their growth inside this vector (34). Further, *Leishmania* infection of the sand fly has been shown to modulate the insect behavior in regards to blood feeding to enhance transmission to naïve hosts (35) and the sand fly stage of this parasite is responsible for a crucial stage of the evolution of this otherwise clonal parasite. Indeed, it is in the co-infected sand fly gut that genetic exchange between different strains or species of *Leishmania* occurs (36, 37).

1.1.6 Transmission reservoirs

Leishmania infected sand flies have been shown to exhibit some degree of preference for particular hosts but feed on a wide range of mammals, birds and lizards (38). As such, many species of *Leishmania* benefit from large animal reservoirs ranging from rodents to bats, through larger mammals (27, 39–42) and are defined as zoonotic *Leishmaniases*. Additionally, humans can serve as the major parasite reservoir in the case of anthroponotic VL in the Indian subcontinent and East Africa (13, 43).

1.1.7 Infection in mammalian host

As described earlier, *Leishmania* modulates the feeding behavior in infected sand flies. As it matures in the sand fly to metacyclic forms, the parasite migrates to the mouthpart. During subsequent blood meals the parasite can be injected in the naïve host as the sand fly regurgitates prior to bloodfeeding (44, 45). Investigation of the very early infection sequence revealed a role for rapidly recruited neutrophils in the course of *Leishmania* infection (46). These neutrophils scavenge the parasites in the skin at the bite site but do not destroy the phagocytosed parasites. Instead, the neutrophils shield the parasite from the immune response and contribute to the spread of the infection (46). These infected neutrophils have then been shown to enhance macrophage recruitment, which are the terminal host cells for *Leishmania* (44). Infected macrophages are then manipulated by the parasites to suit their needs through the activation or repression of macrophage signaling pathways (44, 47). Alterations can include the inhibition of apoptosis (48, 49), reduction in motility (50, 51), and upregulation of anti-inflammatory responses or down regulation of inflammatory responses (52).

Additionally, the parasites alters the environment of the macrophage phagosome in which it resides to make it more hospitable (53) (**Figure 1.3)**. Indeed, *Leishmania* infection can induce an arrest in phagosome maturation resulting in an increased accumulation of F-actin at the periphery (54), inhibition of radical oxygen species (ROS) production (55), and prevention of acidification by the exclusion of v-ATPase (56) resulting in the an inhibition to vesicle trafficking, a reduced oxidative response, and a more hospitable phagosome. These alterations are mediated by multiple virulence factors produced by *Leishmania* that are released into the parasitophorous vacuole and subsequently enter the host cytoplasm through the hosts' own SNARE machinery (57).

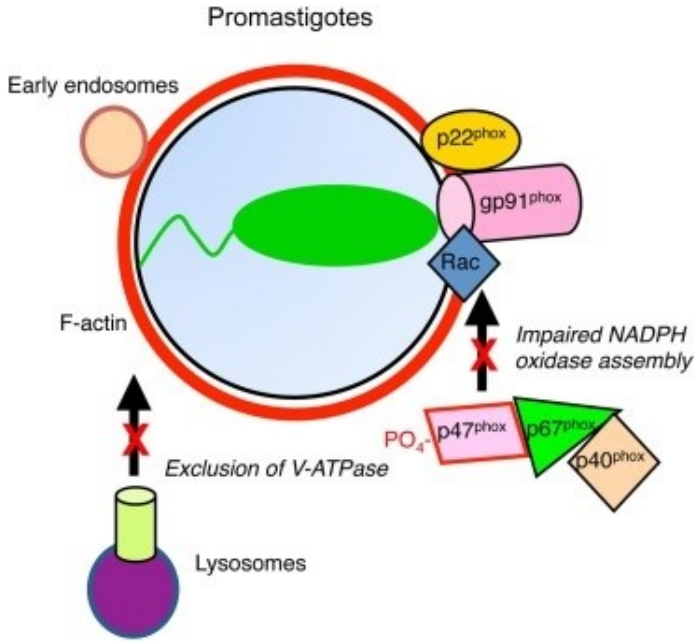


Figure 1.3 Alterations to the phagolysosome by internalized promastigotes

Accumulating F-actin is labeled in red at the periphery of the phagolysosome. The exclusion of the V-ATPase and the NADPH oxidase assembly are highlighted with a red X mark. Adapted from Moradin and Descoteaux, *Front Cell Infect Microbiol*, 2012 (54).

1.1.8 Treatments

Once diagnosed, various treatment options are available to infected individuals based on the type of infection, severity, and regulatory agency guidelines (58).

1.1.8.1 Antimony treatment

The most widely used compounds to treat *Leishmania spp.* infection historically are based on the toxic metal antimony, modified to produce derivates with less, but unfortunately not lacking, toxic side effects (most often as pentavalent antimony salts) (59, 60). Pentavalent antimony can be used to treat both CL and VL, however the existence of resistance has led to an increase in treatment dose and duration while successful treatment outcomes have continued to decline (61, 62). This increase in dosage and duration raises concerns due to the inherent toxicity of these compounds, and increases the length of time the patients must subject themselves to often painful injections (58, 63).

The mode of action of antimony compounds remains unclear with several possible modes of action. Suggested effects of pentavalent antimony are the disruption of the oxidoreductive homeostasis (64), disruption of Zinc-finger mediated interaction domains (65), direct inhibition of DNA replication machinery (66), interference with the purine salvage pathway (59, 67) as well as strengthening the host immune response (68). Due to the length of the this treatment and the rising widespread resistance, it is no longer recommended in many cases (62).

1.1.8.2 Antifungals

A particularity of *Leishmania spp*. is that it shares ergosterol as the main constituent of its cellular membrane with fungi (69). As such, antifungal agents which target this sterol or its synthesis can have antileishmanial effects (70, 71).

Of those, Amphotericin B has shown high cure rates and is recommended in the case of VL (62, 72, 73) as well as in some particular CL cases (74). Earlier data implicated amphotericin B in the direct attack of ergosterol containing *Leishmania* membranes leading to an increase in permeability and loss of cytoplasmic contents (69). More recently, data showed that Amphotericin B additionally impairs *Leishmania* entry into its host macrophage through its action on the macrophage membrane (75, 76).

Although used safely for decades to combat fungal infections, Amphotericin B is subject to numerous side-effects, most notably kidney toxicity (77). To reduce toxicity, this drug is often delivered as liposomal preparation (78, 79). Compared to a 40-day long courses of pentavalent antimony (61), a single dose of liposomal Amphotericin B (AmBisome) can be used to treat VL (80). Although *in vitro* resistant cell lines could be generated (81) and some clinical isolates were found to have a higher resistance to AmBisome (82), it remains rare but should nonetheless be the subject of continued surveillance (83).

Similarly to Amphotericin B although much more rarely used, itraconazole (71, 84, 85), posaconazole (86), ketoconazole (87) or fluconazole (88) ("azoles") are originally antifungal compounds that have antileishmanial activity. Unlike Amphotericin B however, the azoles target cellular machinery responsible for the synthesis of ergosterol instead of targeting ergosterol on the membrane surface (71, 89).

1.1.8.3 Repurposed drugs

More recently, the repurposed anti-cancer drug Miltefosine has been used to treat both VL and CL cases and benefits from an easy to administer oral route (90). Although the detailed targets are not fully known, Miltefosine was shown to induce DNA fragmentation and apoptosis-like cell death (91, 92). Although resistance was seen in laboratory settings (93) and is often used as a genetic editing marker (94), occurrence of naturally resistant isolates remains rare (95, 96).

Another repurposed drug used in treating leishmaniasis is pentamidine, originally developed for African trypanosomiasis (97). Pentamidine has been shown to target the parasite mitochondrion, leading to its disintegration (98). Resistance to this drug was however quickly generated in laboratory settings resulting in parasites that were able to exclude pentamidine from their mitochondria (99, 100) and a large decrease in treatment success rates was also seen in clinical cases (101).

Some broad spectrum antibiotics can also sometimes be used to treat leishmaniasis such as paromomycin with varying degrees of efficacy (102). Paromomycin was shown to target the leishmanial mitochondrion in addition to its known reduction of protein synthesis resulting from targeting of the ribosomal machinery based on bacterial studies (103, 104). This treatment suffers from variable sensitivity across *Leishmania* species (101) and laboratory strains with resistance against paromomycin have been generated and characterized (103, 105) indicating field isolate resistance is a real possibility with increased drug use.

1.1.8.4 Other treatments

Lastly, several combinations therapies are recommended using the above mentioned drugs in various countries in an effort to: increase efficacy, reduce the likelihood of resistant parasite generation, and decrease chances of patient relapse (62, 83, 101). However, further research is necessary to fully elucidate the safety and efficacy of combination therapies and parasites resistant to combinations of drugs have been experimentally generated (106).

In some cases of CL, an additional approach can be local treatment of the skin lesion (107). In those cases, the surface lesions are treated with heat to induce localized parasite death and lesion clearance with similar rates to systemic drug treatment (108, 109). These treatments have been associated with good cure rates and reduced toxicity, however it does suffer from limitations with respect to which areas of the skin can be treated and can result in some second degree burns (109).

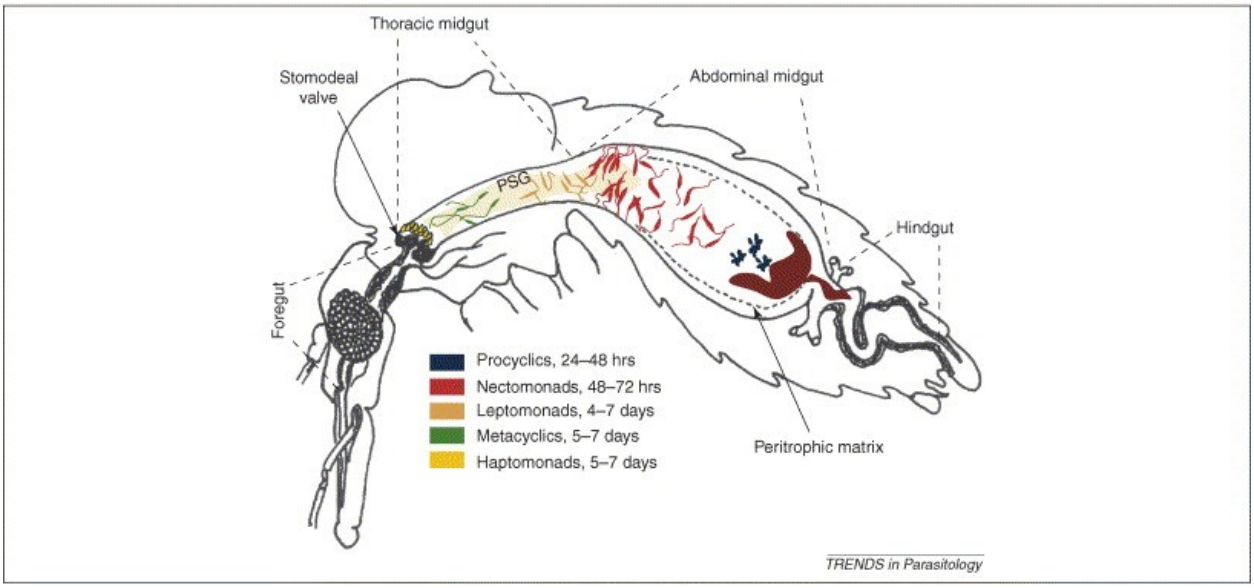
1.2 *Leishmania*

1.2.1 Life cycle

Concomitant with the changes in environment experienced by *Leishmania*, are changes to its morphology as it progresses through its life cycle (110) (**Figure 1.4**). Starting as **procyclic** promastigotes in the sand fly mid gut after a blood meal, *Leishmania* is in a highly replicative, flagellated but weakly mobile form (45). The parasite eventually replicates to high numbers and then spreads upwards and colonizes the gut by attaching to the epithelium (**nectomonad** promastigotes). Parasites that reach stomodeal valve at the foregut become **leptomonad** promastigotes, and begin secreting the promastigote secretory gel (PSG), responsible for forming

the gastric plug which causes the sand fly to regurgitate parasites during a blood meal and enhances the initial infection (34, 45). Finally, the **leptomonads** can differentiate into **metacyclic** promastigotes, the infective stage transmitted during the blood meal where it will encounter and subvert the host immune system (111). This change in host is also accompanied by a further change in life stage, where **metacyclic** promastigotes will differentiate into non-motile, non-flagellated **amastigotes** residing and multiplying inside macrophage phagolysosomes.

These **amastigotes** can be released during a subsequent blood meal by a non-infected sand fly where they will be draw into the gut along with host blood to initiate replication as **procyclic** promastigotes (45).



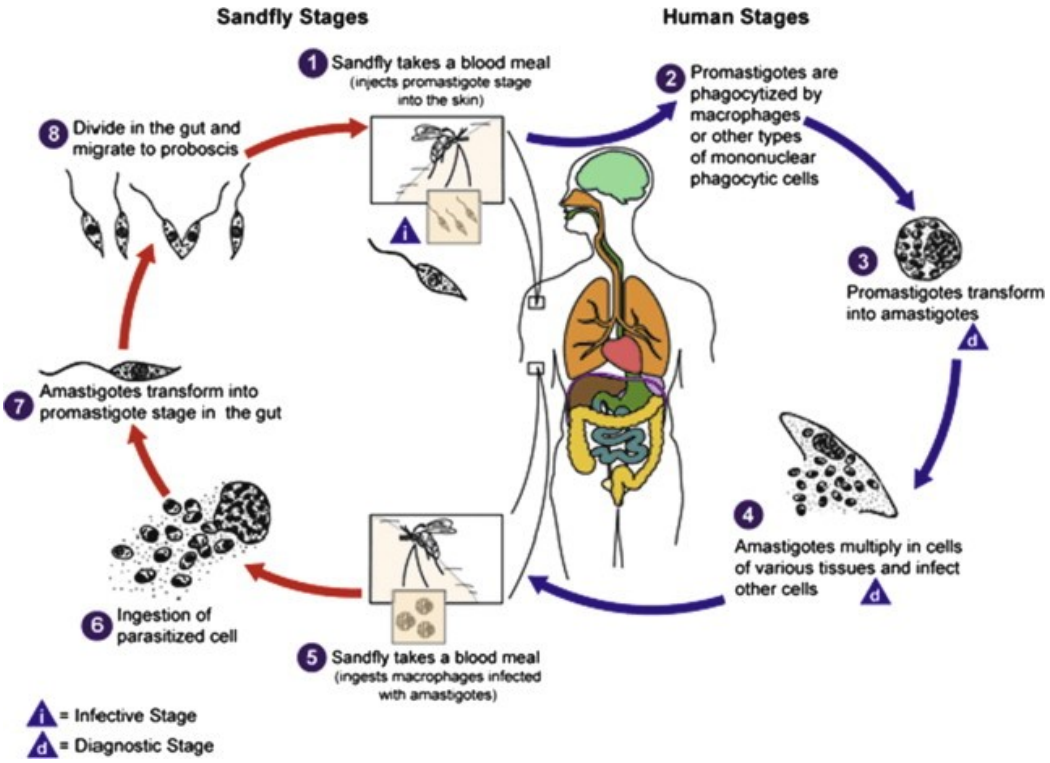


Figure 1.4 The Life cycle of *Leishmania*

Top: Developmental stages and their location in the sand fly gut. Adapted from Kamhawi, *Trends in Parasitology,* 2006 (112). Bottom: Digenic life cycle of *Leishmania*. Adapted from: CDC/ Alexander J. da Silva, *Public Health Image Library,* 2002 (ID#: 3400).

1.2.2 Disease speciation

Once in the infected host, the pathology caused by the *Leishmania* parasites is largely related to the species of the infecting parasites. Indeed, while certain exceptions arise and can be used to study the pathogenesis of this disease as will be discussed further, the disease progression is generally consistent across different hosts infected with the same parasite species with some species being responsible for CL, and others for VL (25).

In the Old World for example, *L. major* is considered the prototypical cutaneous parasite. Spread from Africa to Asia, this parasite causes CL in humans. Its polar opposite, *L. donovani*, shares a large geographical overlap with *L. major* but disseminates from the bite site at the skin into the visceral organs (113). Additionally, the Old World *Leishmania* species *tropica* and *aethiopica* also contribute to the burden of CL, although *L. tropica* can sometimes cause visceral disease (114–116). Meanwhile, VL in the Old World is largely limited to *L. donovani* and *L. infantum,* although *L. infantum* can contribute to CL, it is a major contributor to VL, especially in canines (25, 117–119).

In the New World, CL is largely due to *L. mexicana* and the *L. braziliensis* containing *Viannia* subgenus (120, 121). The parasites species within the *Viannia* subgenus differ from the *Leishmania* subgenus by the location of their development within infected sand flies (122). While the *Leishmania* species are limited to the sand fly midgut, *Viannia* species develop in the hindgut (112). While *L. braziliensis* is responsible for the most infections in the *Viannia* group, the subgenus also includes *L. panamensis, L. guyanensis,* and other lesser prevalent species. Conversely, VL in the New World is primarily caused by a single species, *L. chagasi*, a synonym for *L. infantum* introduced from the Old World to the Americas during colonization (123).

1.2.3 The Parasite Cell Surface

As components on the surface of the parasite cell are exposed to the sand fly vector and the infected host cells, many of those components carry the role of virulence factors. A particularity of *Leishmania* and other trypanosomes, is their heavy reliance on glycosylphosphatidylinositol (GPI) molecules on their cell surface and dense phosphoglycosylation of surface and secreted components (124). On this surface, GPI serves as an anchor for surface proteins, lipophosphoglycans (LPG), and proteophosphoglycan (PPG) but also unconjugated (GIPL) as a barrier to the extracellular milieu. GPI is a complex glycolipid anchored to membranes resulting from the addition of multiple sugars to bilayer associated phosphatidylinositol (PI) (125), which can then be used for protein attachment. The composition and length of sugar moieties used in phosphoglycosylation of surface and secreted components varies across *Leishmania* species but their presence is shared across the genus (124, 126).

1.2.3.1 GPI-anchored proteins

Unlike LPG and PPG, GPI anchored proteins do not contain phosphoglycan but are instead N-glycosylated and C-terminally conjugated to GPI during their ER transit (124, 127). The most abundant GPI-bound surface protein, the protease GP63 plays many roles in the parasite development and disease pathogenesis (128). Different isoforms of GP63 are expressed by different species and at different stages of the *Leishmania* life cycle, both in GPI-anchored and secreted forms (128). GP63 is a broad ranging endopeptidase capable of cleaving complement members (129), controlling macrophage signalling pathways (130, 131), inhibiting antigen presentation (132).

However, other surface proteins such as prohibitin have also been shown to contribute to virulence and host parasite interactions (133).

1.2.3.2 Proteophosphoglycan

PPG is comprised of heavily phosphoglycosylated protein backbones with close to 75% of their molecular weight originating from glycosylation. PPG can also lack its GPI anchor and in such cases is the major constituent of the PSG which induced blockage inside the sand fly as described earlier (134). GPI-free PPG is heavily secreted (> 1g/ml) and forms a filamentous aggregate (135) which can protect the parasites from enzyme digestion in the sand fly in addition to modulation of sand fly feeding as described earlier (35) and contribute to parasite pathogenesis (34).

1.2.3.3 Lipophosphoglycan

LPG is a protein-free complex of phosphoglycan and is the most abundant molecule on the parasite surface. Their size and composition have been shown to vary in relation to the parasite life cycle (136, 137). LPG contributes to the virulence of *Leishmania* by contributing to the establishment of the infection in the sand fly gut (impacting the sand fly species / *Leishmania* strain pairing), protection from human complement attack, protection from oxidation and modulation of immune responses (136, 138, 139).

1.2.3.4 Glycoinositolphospholipids

Small unconjugated glycoinositolphospholipids (GIPLs) are related to the anchors used to retain LPG, PPG and GP63 on the parasite surface, but also play an important role in the infection. GIPLs have been shown to accumulate in membrane microdomains (140), play a role in macrophage interactions and can also down regulate the macrophage immune response.

1.2.3.5 Molecular Import / Export Across the Cell Membrane

Due to its wide range of cellular environments, ranging from the digestive enzyme filled stomach to the intracellular phagosome, *Leishmania* has developed many ways to acquire the necessary nutrients for its survival and proliferation. Indeed, the surface of the *Leishmania* cell is also densely populated by nutrient transporters and permeases (141) adapted to the very different sand fly promastigote or host amastigotes stages of the parasite (142). Nutrients imported into the cell include: carbohydrates, purines, iron, amino acids, polyamines, folates and divalent cations (141). Further, traditional clathrin-mediated endocytosis (143) has been shown to contribute to the acquisition of required nutrients for *Leishmania* (144), proceeding in-line with the eukaryotic Rab controlled membrane traffic pathways (145, 146).

On the other hand, the parasite also needs to output several molecules to impart its effects on the sand fly and the host cells. For this purpose, exosomes have been shown to be a main source of protein export in the sand fly midgut (147) as well as within the host macrophage (148). Additionally, the traditional eukaryotic secretion pathway relying on N terminal signals, ER processing and Golgi mediated transport is functional in *Leishmania* (149). Interestingly, all

secretion in *Leishmania* is polarized towards the parasite flagellar pocket, an invagination at the attachment site of the flagellum on the anterior cell pole (149, 150).

1.2.4 Intra-cellular organization

1.2.4.1 Flagellum

Inside the cell, *Leishmania* adopts a polar sub cellular organization structure, albeit more prominently in promastigote stages (**Figure 1.5**). As described earlier, the anterior portion of the parasite is characterized by the presence of a single flagellum, used for swimming and attachment in the promastigote stages (151). In amastigotes, this flagellum either collapses or is rebuilt into a much shorter form and takes on an unclear function, although it is theorized to play a sensory role in monitoring the state of the infected host cell (152).

1.2.4.2 Exocytosis

At the base of the flagellum, the Golgi complex sits in close proximity to the flagellar pockets, the only site of endo and exocytosis in *Leishmania* (146). The transitional endoplasmic reticulum (tER) is located in the same pole, close to the Golgi network and mediates the production of proteins to be secreted or anchored at the cell surface (153).

1.2.4.3 Endocytosis

Unlike the exocytosis pathway, the endocytic pathway does not concentrate at the flagellar pole. Although import occurs in the flagellar pocket resulting in early endosomes at the anterior

pole, the endocytic pathway in *Leishmania* terminates in a multivesicular tubule lysosome (MVT) that transverses the cell on its entire length and is associated with cytoplasmic microtubules (153, 154).

1.2.4.4 The acidocalcisome

Additionally, an endocytic pathway related organelle termed the acidocalcisome is located at the posterior or aflagellated pole (155). The acidocalcisomes play a role in the homeostasis of calcium and other cations, phosphorous, intracellular pH and osmolarity (156). The biogenesis of the acidocalcisomes has been shown to be under the control of a TOR-kinase homologue in *Leishmania* termed TOR3, which is involved in environmental stress sensing as shown by knock-out studies resulting in virulence-attenuated parasites with a high glucose starvation and osmotic shock sensitivity (157).

1.2.4.5 The mitochondrion and kinetoplast

A particularity of the kinetoplastid parasites is the presence of a single mitochondrion within the cell, and the attached kinetoplast (110). The kinetoplast is a unique organelle located at the anterior portion of the cell and contains the mitochondrial genetic material. The mitochondrial genome is contained in maxi circles but contains mutations rendering it non-functional as encoded and transcribed RNA requires editing using minicircle encoded guide RNA to correct the mutations prior to translation (158).

1.2.4.6 The glycosome

Another distinguishing organelle feature of kinetoplastids is the glycosome, an organelle responsible for the major part of the glycolytic pathway (159, 160). This compartmentalization is thought to be due to the need for regulatory control over this pathway. Additionally, the glycosome is responsible for purine salvage, and contributes to pyrimidine metabolism. Being a unique feature of these parasites, the glycosomes have been the target of the hopeful development of new therapeutics (161).

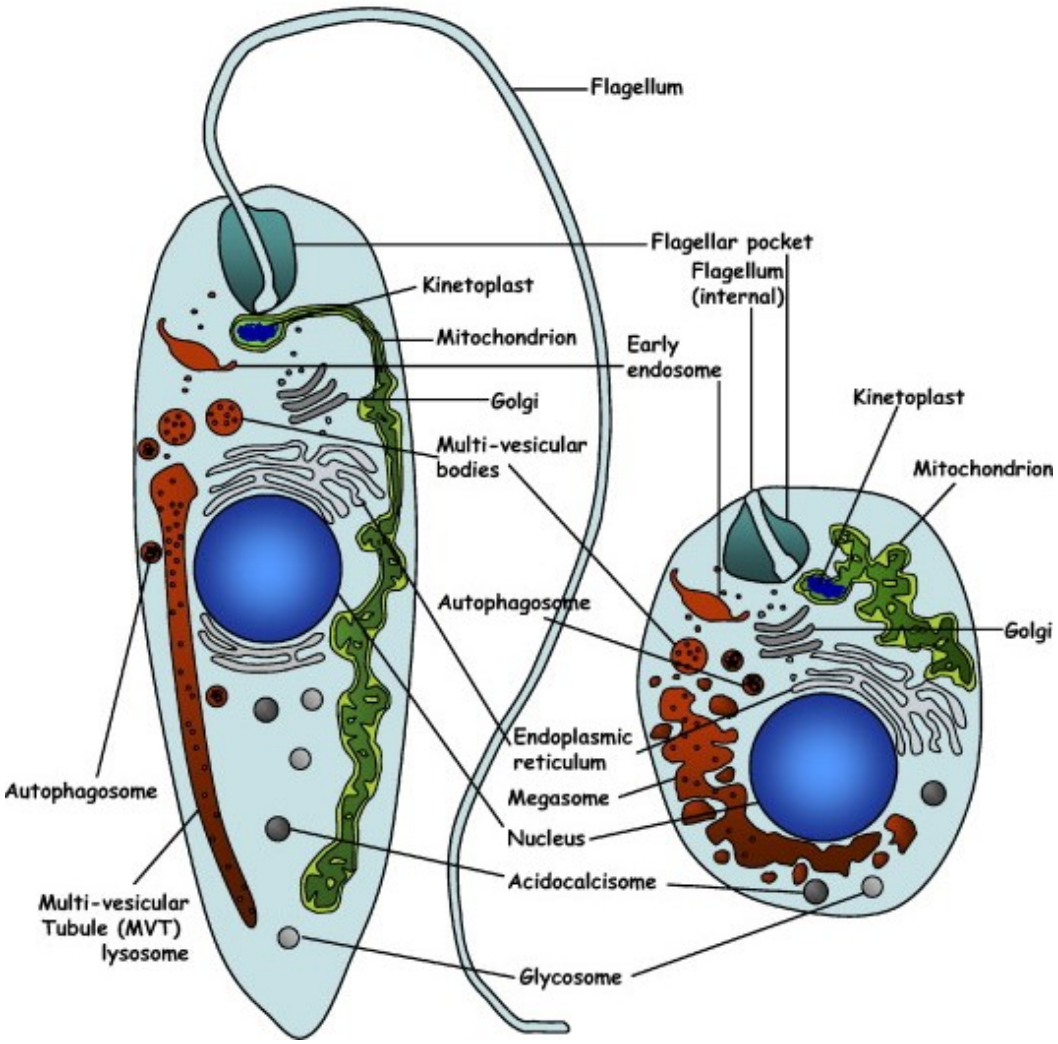


Figure 1.5 Sub-cellular organization of *Leishmania*

The sand fly promastigote stage is shown on the left. The macrophage intracellular amastigote stage is shown on the right. The main intracellular organelles are labeled. Adapted from Besteiro et al., *International Journal for Parasitology,* 2007 (162).

1.2.5 Genomic Organization

1.2.5.1 Chromosomal organization

At the center of the *Leishmania* cell, the nucleus contains approximately 35.5Mbp of genetic information contained in 34 to 36 chromosomes as some New World species contain fused chromosomes (163–165). While all Old World *Leishmania* contain 36 different chromosomes, *L. mexicana* contains a fused version of the equivalent of chromosomes 8 with 29, and 30 with 36 and *L. braziliensis* contains a fused version of chromosomes 20 with 34.

With the exception of chromosome 34 which is tetraploid, most chromosomes are diploid under normal conditions. However, extensive polyploidy in response to external stimuli occurs as regulation of protein levels by gene dosage has been shown to play a large role in *Leishmania* (166, 167) resulting a in a highly variable chromosome numbers at the strain level. Further, mosaic aneuploidy has been demonstrated in multiple experiments, where unequal segregation of chromosomes results in a "mosaic" of cells with varying chromosome numbers even in clonal populations (168).

Chromosome ends contain 6-mer telomeric repeat units followed by a 100bp telomeric associated sequence (169). In between the telomeres, the chromosomes contain long convergent or divergent polycistronic gene arrays (164, 170, 171). Additionally, the telomeric repeats also contain a non-canonical DNA base termed base J (172). This base arises from the hydroxylation and glycosylation of thymidine bases and has been shown to be a transcription termination site in *Leishmania.*

1.2.5.2 Genes

The genes of *Leishmania* are organized in large polycistronic arrays similar to those prokaryotic genomes. The transcription units can be either convergent or divergent, with DNA base J acting as a terminator in between different units to prevent transcriptional read-through from one cluster to another (172, 173). Transcription initiation is thought to occur at divergent strand-switch regions, where two polycistronic gene units face away from each other on opposing DNA strands (174).

Leishmania genes contain no intron and therefore undergo no internal splicing. As they are transcribed in polycistronic units however, each gene transcript requires processing prior to reaching a translation competent mRNA stage (175) (**Figure 1.6**). Intergenic sequences are processed to form monocystronic RNA by the addition of a 39 nucleotide long capped sequence (Spliced Leader, SL) that is trans-spliced 5' in relation to the gene and by polyadenylation of the 3' end (176). The SL sequence is added to an AG dinucleotide downstream of a ~200nt long polypyrimidine tract (C and U) (177). Although the precise location of the SL addition can vary due to the presence of multiple AG doublets, the specificity for an AG site is invariable (176, 178). On the other hand, there is no concrete evidence for any specific poly-adenylation site. Rather, as polyadenylation is coupled with the SL processing of the downstream gene, it is largely dependent on the location of trans-splicing (176).

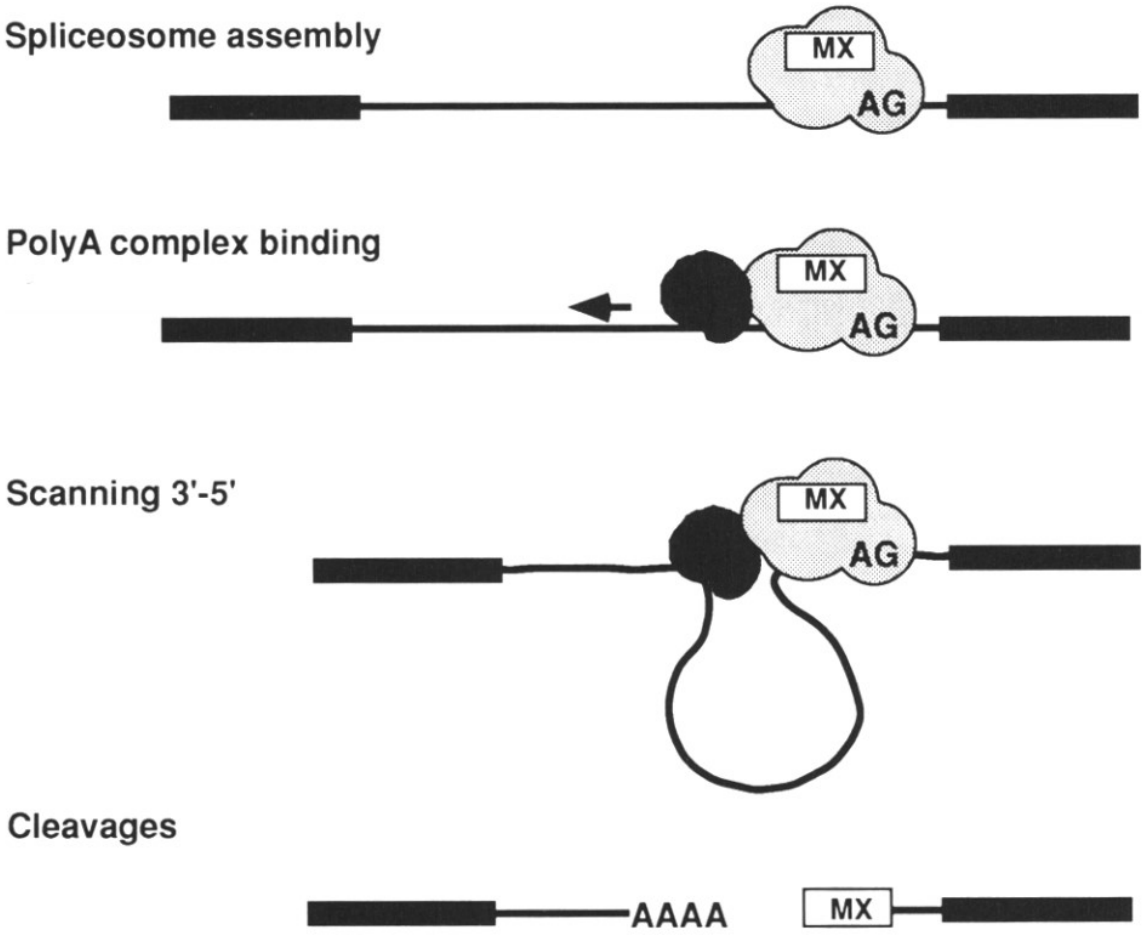


Figure 1.6 Trans-splicing and polyadenylation in *Leishmania*

Mechanism coupling trans-splicing of a 3' gene transcript with polyadenylation of a 5' gene transcript. The Splice Leader or miniexon is labeled MX, the acceptor AG nucleotides are highlighted at the spliceosome transcript boundary. Adapted from LeBowitz et al., *Genes & Development,* 1993 (176).

1.2.6 Gene regulation

As transcription occurs in large polycistronic units, with roughly the same amount of transcript being produced for each gene within the polycistronic unit, regulation of gene expression in *Leishmania* does not rely on transcriptional regulation. Instead, post-transcriptional regulation (179, 180) and gene dosage (166, 167, 181, 182) have been shown to contribute to the regulation of gene expression, resulting in multiple gene families being organized in tandem arrays that expand and contract to modulate gene expression.

1.2.6.1 Gene dosage

In response to external stimuli such as drug pressure (183) or different *in vivo* environments such as the change in host species (182), *Leishmania* has been shown to up or down-regulate its gene expression by altering its gene copy numbers. These alterations include the contraction or expansion of multi copy gene arrays, the generation or amplification of episomes containing gene coding sequences, and the modulation of ploidy of chromosomes containing relevant genes (184). As the genome of *Leishmania* is interspaced with repetitive elements, these serve as recombination points for DNA looping or excision (185, 186). Further, unequal crossing over events during chromosomal replication have been shown to naturally lead to a heterogenous population of parasites referred to as mosaic aneuploidy allowing for the pre-existence and selection of parasites with enhanced fitness in any given environment (168).

1.2.6.2 mRNA control

As all transcripts in a polycistronic unit are originally present at identical levels, the control over the relative abundance of different transcripts primarily relies on internal elements. Compositional codon bias, or the frequency of optimal codon usage as well as length of the mRNA has been shown to affect the abundance or stability of mRNA as well as protein expression (187, 188). Additionally, sequences in the intergenic regions, which are removed during the polycistronic processing described earlier, have been shown to influence these processing steps resulting in control over the levels of monocistronic transcripts (189, 190). As well, sequences present in the untranslated regions (UTR) of mRNA can affect their stability and translation levels such as for stage specific proteins or heat shock proteins (191, 192).

1.2.6.3 Translational control

While *Leishmania* possess several ways to regulate the relative abundance of various mRNAs, the transcript and protein levels have been shown to not always be consistent (193, 194). As mentioned earlier, this is partially due in to the control the 3' UTR can exert on the translation of specific mRNAs, however due to distinct elements in the 3'UTR (195). It was shown that 3' UTR elements, specifically in the stage-specific genes, promote translation initiation through an increase in polysome association.

1.3 The TOR pathway

The Target Of Rapamycin (TOR) pathway plays a central role in eukaryotic cell growth and metabolism as it integrates nutrient status and energy availability of the cell and governs key metabolic pathways (196). The TOR protein is a serine/threonine protein kinase and is a member of the phosphatidylinositol 3-kinase (PI3K)-related kinase family.

1.3.1 The Human TOR pathway

In humans, the TOR protein is present in two different protein complexes: the mammalian Target Of Rapamycin Complex 1 (mTORC1) and the mammalian Target Of Rapamycin Complex 2 (mTORC2). The mTORC1 complex is composed of the mTOR protein, Raptor, LST8, PRAS40 and DEPTOR. The mTORC1 complex is responsible for control of cell growth pathways such as autophagy as well as, protein, lipid and nucleotide synthesis (197) and responds to nutrients (198). The mTORC2 complex on the other hand responds to growth factor stimuli and is composed of mTOR, Rictor, LST8, DEPTOR, mSin1 and Protor1/2.

Sensing of nutrients and energy levels occurs upstream of mTORC1 through several sensing pathways. The AMPK regulator senses levels of energy (as ATP) and oxygen, the Rag GTPase can relay both glucose and amino acid levels, and GATOR1 and GATOR2 can signal leucine and arginine levels in the cytoplasm (199).

1.3.2 The TOR pathway components of *Leishmania*

Contrary to the human pathway, little is known about the *Leishmania* TOR pathway. In fact, prior to the identification of an expanded Target of Rapamycin family in 2010 (157), the TOR

acronym in *Leishmania* referred to TOxic nucleotide Resistance (200, 201). In the expanded family, a total of 3 TOR genes (*tor1*, *tor2*, *tor3*) were identified compared to a single TOR gene in humans, yeast or other eukaryotes (157). Interestingly, while two of the genes were deemed essential, the *tor3* gene was found to be dispensable and knockout parasites were found to have a severe defect in infectivity and acidocalcisome formation. These parasites were also found to be sensitive to glucose starvation but insensitive to low amino acid levels.

1.4 *Leishmania* sexual reproduction leading to hybrid parasites

While *Leishmania* predominantly reproduces asexually, evidence of a sporadic sexual reproduction cycle dates as far as 1994, when isoenzyme electrophoresis patterns were found to produce heterozygous patterns corresponding to both *L. braziliensis* and *L. panamensis* (202). Additional evidence by similar heterozygous patterns was later identified in several regions of the world (203–205) and even for phenotypically distant parasites such as *L. major* and *L. infantum* (206).

The presence of a sexual reproductive cycle within the sand fly was later confirmed experimentally through the generation of progeny bearing dual markers following co-infection of sand flies with parental strains bearing a single marker per genome (36). Further evidence for a complete meiosis-like sexual reproductive cycle was recently generated through the use of whole genome sequencing analysis to identify patterns of recombination and genetic crossing over in experimentally generated hybrid progeny (37).

While genetic hybrids have been identified in various foci around the globe, reports of hybrid parasites remain sporadic and the contribution of genomic hybridization to the progression of the disease of the evolution of the parasites remains unclear.

1.5 Sri Lanka

Originally faced mainly with imported cases of leishmaniasis, Sri Lanka has been the focus of an epidemic of cutaneous leishmaniasis since 2001, climbing ever since with an alarming expansion occurring in 2018 (207, 208). While a few rare cases of visceral leishmaniasis have been reported, the majority of cases are cutaneous leishmaniasis with diverse manifestations (207, 208). The causative agent for this outbreak has been classified as an *L. donovani* strain related to, but unique from, *L. donovani* strains circulating in neighboring India which cause visceral disease. Conflicting theories each supported by different evidence exist as to the origins of the circulating strains, whether recently imported or native to the island (208).

As closely related CL and VL samples have become available, preliminary research identified potential genetic elements that could explain the unusual phenotype of *L. donovani* in Sri Lanka (209) through comparative genomics. Notably, the cutaneous isolate analysed was found to contain less copies of the A2 virulence genes and a non-synonymous point mutation in the RagC GTPase. The A2 protein is a virulence factor found to contribute to the survival of visceral *L. donovani* but is non-functional in *L. major* (210, 211). The A2 protein was found to increase the parasite's survival when faced with heat stress (211, 212). The RagC GTPase as introduced earlier, is part of the amino acid sensing pathway upstream of mTORC1. A point mutation in the *RagC* gene resulting in a R231C change was identified when comparing cutaneous and visceral isolates

from Sri Lanka (209). Complementation of the CL isolate with an episomal copy of the wildtype gene was found to increase visceral organ survival in a mouse infection model (209).

1.6 Sequencing technologies

In order to analyze the composition of the genome of *Leishmania*, various sequencing technologies have been used throughout the years. No single solution exists, and new sequencing methods or improvements continue to be developed as every sequencing platform has intrinsic advantages and drawbacks. (**Figure 1.7**)

1.6.1 First generation

Used throughout the world from 1977 onwards (213), Sanger sequencing remains the most commonly used sequencing method due to its accuracy, robustness and low cost of operation. This sequencing methods otherwise known as sequencing by termination, relies on labelled nucleotides that stop the elongation of the nucleotide chain. Depending on the type of label, the sequence can then be determined based on the length of the stopped products using gel or capillary electrophoresis.

While Sanger sequencing delivers high base calling accuracy, large read length (up to 1,000 bases) and recent improvements benefiting from automation, it does suffer from a low overall throughput compared to more recently developed techniques (214).

1.6.2 Second generation

Aiming to increase the throughput of sequenced bases required to scale sequencing from short insert to whole genomes, many approaches were developed to parallelize sequencing reactions

1.6.2.1 Pyrosequencing

An early developed alternative to Sanger sequencing was pyrosequencing. In this method, rather than relying on reading the length of all the possible termination products after synthesis, pyrosequencing directly detects the added base. During synthesis, each of the four possible nucleotides are added in sequence and the release of pyrophosphate resulting from the incorporation one of those nucleotides is measured and recorded through a light producing reaction (215). This readout can be spread across a large number of wells each containing different fragments of sheared DNA to scale the total sequencing throughput as it is a single step process compared to Sanger sequencing where amplification and readout (electrophoresis) are uncoupled (216). This sequential addition and washing steps, coupled with enzyme stability limit the sequencing length of individual fragments to around 400 bases (214). Although shorter than Sanger sequencing, these reads are readily parallelized in as many as 1.6 million wells in the case of the most known commercial implementation of pyrosequencing, the 454 system (217).

1.6.2.2 IonTorrent

Similar to pyrosequencing, IonTorrent technology detects the addition of one of four complementary nucleotides in a stepwise manner. However, IonTorrent relies of the release of H^+

during incorporation rather than pyrophosphate coupled to luciferase, which allows the system to use electronic pH sensors instead of optical sensors (218). As the sensors are less complex, the number of wells and resulting reads in an IonTorrent system can reach 60-80 million increasing the throughput beyond that of pyrosequencing (219).

1.6.2.3 Sequencing by ligation

Based on the principle that proper base pairing is required for oligonucleotide primers to bind effectively to their target sequence, the method of sequencing by ligation was developed (220). In the SOLiD system commercialized by ABI, primers containing combination of the four nucleotides are linked to four different fluorophores allowing for detection of proper binding. Although this system can deliver over 1 billion reads, the successive binding steps limit the reads length to about 50bp (221).

1.6.2.4 Illumina

In the most commonly used second generation sequencing method, Illumina sequencing, the base-color encoding using different fluorophores from sequencing by ligation and base by base synthesis from pyrosequencing are combined to provide the largest throughput. In this method, rather than extending a nascent nucleotide chain by flooding and washing one type of nucleotide at a time, all four are present but coupled to different fluorophores (222). The fluorescent label also serves as a terminator, ensuring only one base is added per cycle. As such, the color peak at each extension position can be translated to a base call. To increase accuracy, single DNA fragments are first amplified into large colonies, as the final signal will reflect the consensus

incorporation rather than a single nucleotide. This process can be scaled up by synthesizing multiple DNA strands on a glass slide in parallel resulting in up to 20 billion reads with 250bp in length in the latest iteration (223).

1.6.3 Third generation

While multiple second-generation technologies benefited from an increased throughput compared to Sanger sequencing by leveraging massively parallel sequencing reactions, no single technology matched the read length provided by Sanger sequencing. As read length is particularly important to overcome low complexity regions and repetitive elements, several technologies were developed in an attempt to overcome these limitations (**Figure 1.8**).

1.6.3.1 PacBio

Originally able to resolve DNA fragments up to 1,500bp long (224), improvements to the polymerase and reaction chemistry have since enabled PacBio sequencers to routinely sequence fragments over 50kbp (225), with some reads extending beyond 175kbp (226). In order to achieve this, the technology anchors a polymerase to the bottom of a well in which a DNA template and color coded fluorescently labeled nucleotides are present and precise optical systems detect the individual signals of every nucleotide in real time without relying on chain terminators or artificial cycling such as in the case of second generation technologies (227).

As each signal originates from a single nucleotide and a single polymerase, this signal is not always clear enough above background which can result in inaccurate base calling (227) and

inherent polymerase errors can also occur (228) resulting in an error rate of approximately 15% in PacBio reads. To address this concern, several approaches have been developed.

As PacBio DNA fragments are circularized, the use of shorter inserts allows the polymerase to loop over and sequence the same fragment multiple times allowing for a consensus sequence to be generated (229). Additionally, alignments of different DNA fragments originating from an overlapping genomic location can be aligned to each other to generate an internal consensus (230). Both methods however tradeoff read length or total throughput for accuracy respectively.

A further approach to correct the error prone reads generated by PacBio instruments is referred to as hybrid correction (231). In this approach, reads originating from a high accuracy second generation sequencer (most often Illumina) of the same DNA source are mapped to the long PacBio reads and mismatches are corrected. This method however requires two separate sequencing pipelines which increases the cost of experiments.

1.6.3.2 Oxford Nanopore

An alternative to PacBio sequencing to produce long reads, Oxford Nanopore sequencing, is also classified as third generation technology. A major difference between Nanopore sequencing and all other methods, whether first second or third generation, is that Nanopore sequencing reads the DNA bases **directly** rather than inferring it by ligation or synthesis of a complementary base (232, 233).

Adapters are first ligated to the ends of a DNA molecule to mark the beginning and end of the fragment, and to help anchor the molecule to a nanopore. As the double stranded DNA

molecule is opened and one strand is pulled through the pore, the bases generate a unique ionic signal which is measured and converted to base calls electronically. To increase the accuracy of the read, the complementary strand can also be sequenced to increase base calling accuracy (227). Sufficient throughput is achieved by combining thousands of nanopores in a single flow cell, allowing concurrent sequencing of multiple DNA fragments.

In common with PacBio sequencing, as the sequencing signals are derived from a single molecule, the accuracy of Nanopore sequencing reads falls below that of second generating technologies. However, this accuracy tradeoff is offset by the ability of Nanopore sequencing to produce even longer reads than PacBio sequencing with some reads spanning multiple mega base pairs (225).

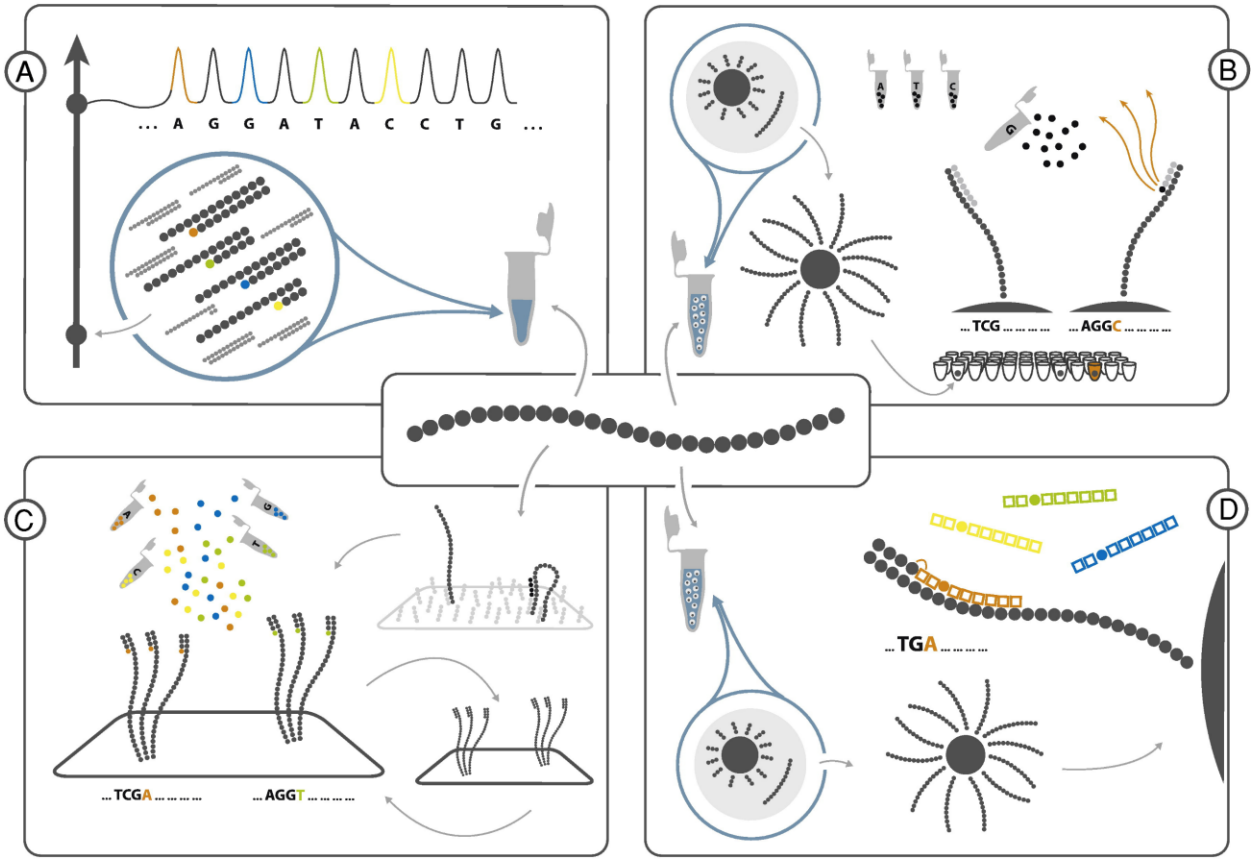


Figure 1.7 Comparison of sequencing technologies

A. Sequencing by chain termination (Sanger) results in a range of fragments with a terminal label that can be size separated and read to reconstitute the sequence. **B.** Pyrosequencing using sequential addition of nucleotides to detect a successful incorporation. **C.** Reverse termination (Illumina) generates large colonies of identical reads to increase signal output and terminator bases allow a single addition per strand from a pool of differentially labeled fluorescent bases. **D.** Sequencing by ligation (SOLiD) relies on fluorescently labeled primers to interrogate each base position. Adapted from Pettersson et al., *Genomics*, 2009 (214).

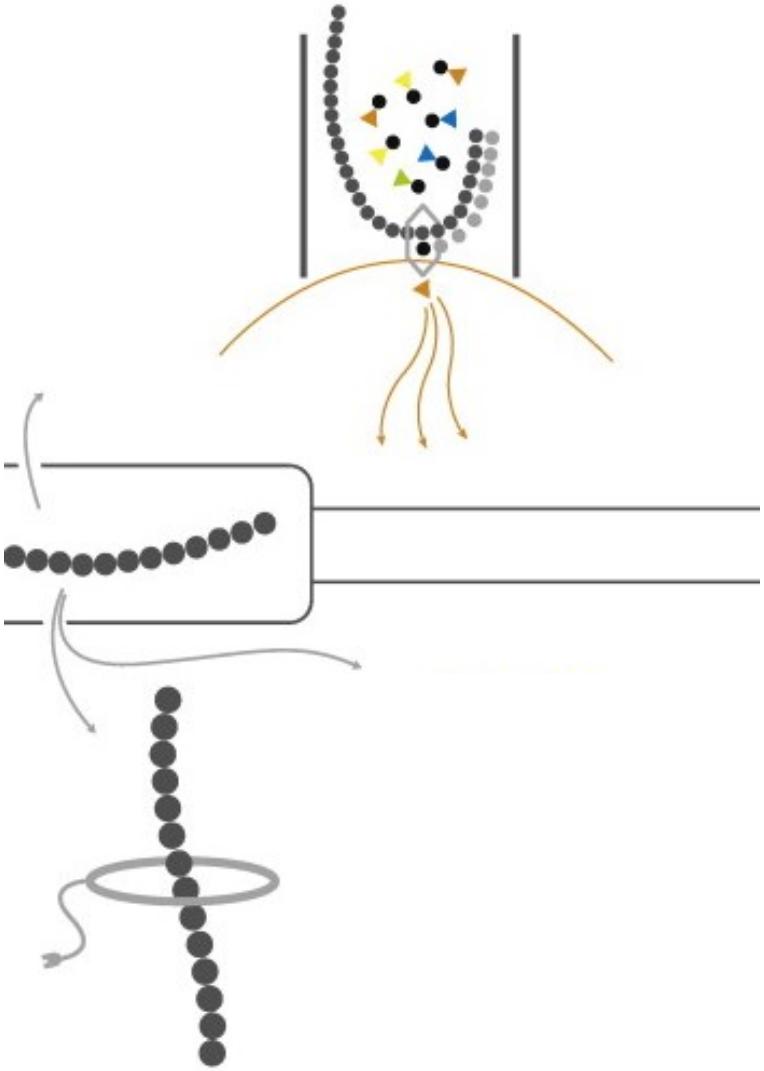


Figure 1.8 Third generation sequencing technologies

Top: PacBio sequencing uses an anchored DNA polymerase at the bottom of a well (hexagon). The fluorescent labels attached to each base is read in real-time as the complementary strand is synthesized. Bottom: NanoPore sequencing uses electronics to measure the charge of each base as it passes through a nanopore. Adapted from Pettersson et al., *Genomics*, 2009 (214).

1.7 Rationale and Objectives

While the outbreak of cutaneous leishmaniasis in Sri Lanka is a serious health concern, it also presents a unique and interesting research opportunity. As the Sri Lankan outbreak has been attributed to *L. donovani* which traditionally causes visceral disease, the study of *L. donovani* strains from Sri Lanka may help elucidate the factors contributing to disease progression.

As closely related parasite samples causing two different pathologies are available (CL vs VL), these represent ideal samples for comparative genomics studies. Genomic investigation, however, is contingent on obtaining high quality sequencing samples and a high-quality reference genome to serve in the analysis. The first objective of this thesis was therefore to generate a new reference genome for *L. donovani* with complete annotations to enable a comparative genomic study of cutaneous and visceral disease-causing isolates to identify potential virulence controlling factors.

Comparative genomic studies can identify potential virulence factors based on genetic differences occurring in strains with different phenotypes. However, these potential virulence factors require investigation to: first, confirm that the mutations identified play a role in tissue tropism and are not spontaneous mutations, and second, to characterize the function of the virulence factor which could contribute to the generation of new therapies. The characterization of the RagC GTPase was therefore carried out as the second objective of this thesis to better understand the role of this protein in *Leishmania* pathogenesis.

Finally, while comparative genomic studies are a powerful tool for identifying potential virulence factors, several samples in the analysis must have common polymorphism or affecting overlapping pathways to strengthen the analysis. As several samples originating from Sri Lanka

did not coincide with our findings, the final objective of this thesis was to perform a global genomic investigation of the *L. donovani* species to characterize the circulating strains in Sri Lanka in order to reconcile the conflicting studies.

1.8 References

1. Mcgwire BS, Satoskar AR. 2014. Leishmaniasis: Clinical syndromes and treatment. Qjm 107:7–14.
2. Burza S, Croft SL, Boelaert M. 2018. Leishmaniasis. Lancet 392:951–970.
3. David C V., Craft N. 2009. Cutaneous and mucocutaneous leishmaniasis. Dermatol Ther 22:491–502.
4. Markle WH, Makhoul K. 2004. Cutaneous Leishmaniasis: Recognition and Treatment. Am Fam Physician 69:1455–1460.
5. Layegh P, Ghazvini K, Moghiman T, Hadian F, Zabolinejad N, Pezeshkpour F. 2015. Bacterial contamination in cutaneous leishmaniasis: Its effect on the lesions' healing course. Indian J Dermatol 60:211.
6. Convit J, Pinardi ME, Rondón AJ. 1972. Diffuse cutaneous leishmaniasis: A disease due to an immunological defect of the host. Trans R Soc Trop Med Hyg 66:603–610.
7. G D. 2013. A Review on Biology, Epidemiology and Public Health Significance of Leishmaniasis. J Bacteriol Parasitol 04.
8. Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, Brooker S. 2007. Cutaneous leishmaniasis. Lancet Infect Dis 7:581–596.
9. Bennis I, Verdonck K, El Khalfaoui N, Riyad M, Fellah H, Dujardin JC, Sahibi H, Bouhout S, Van Der Auwera G, Boelaert M. 2018. Accuracy of a rapid diagnostic test based on antigen detection for the diagnosis of cutaneous leishmaniasis in patients with suggestive skin lesions in Morocco. Am J Trop Med Hyg 99:716–722.
10. De Silva G, Somaratne V, Senaratne S, Vipuladasa M, Wickremasinghe R, Wickremasinghe R, Ranasinghe S. 2017. Efficacy of a new rapid diagnostic test kit to diagnose Sri Lankan cutaneous leishmaniasis caused by Leishmania donovani. PLoS One 12.
11. Adams ER, Schoone GJ, Ageed AF, El Safi S, Schallig HDFH. 2010. Development of a reverse transcriptase loop-mediated isothermal amplification (LAMP) assay for the sensitive detection of Leishmania parasites in clinical samples. Am J Trop Med Hyg 82:591–596.
12. Hirve S, Boelaert M, Matlashewski G, Mondal D, Arana B, Kroeger A, Olliaro P. 2016. Transmission Dynamics of Visceral Leishmaniasis in the Indian Subcontinent – A Systematic Literature Review. PLoS Negl Trop Dis 10.
13. Ready PD. 2014. Epidemiology of visceral leishmaniasis. Clin Epidemiol 6:147–154.
14. Desjeux P, Alvar J. 2003. Leishmania/HIV co-infections: Epidemiology in Europe. Ann Trop Med Parasitol 97.
15. Lindoso JA, Cota GF, da Cruz AM, Goto H, Maia-Elkhoury ANS, Romero GAS, de Sousa-Gomes ML, Santos-Oliveira JR, Rabello A. 2014. Visceral Leishmaniasis and HIV

Coinfection in Latin America. PLoS Negl Trop Dis 8.

16. Zijlstra EE, Musa AM, Khalil EAG, El Hassan IM, El-Hassan AM. 2003. Post-kala-azar dermal leishmaniasis. Lancet Infect Dis 3:87–98.

17. Das VNR, Pandey RN, Siddiqui NA, Chapman LAC, Kumar V, Pandey K, Matlashewski G, Das P. 2016. Longitudinal Study of Transmission in Households with Visceral Leishmaniasis, Asymptomatic Infections and PKDL in Highly Endemic Villages in Bihar, India. PLoS Negl Trop Dis 10.

18. Le Rutte EA, Zijlstra EE, de Vlas SJ. 2019. Post-Kala-Azar Dermal Leishmaniasis as a Reservoir for Visceral Leishmaniasis Transmission. Trends Parasitol 35:590–592.

19. Mondal D, Bern C, Ghosh D, Rashid M, Molina R, Chowdhury R, Nath R, Ghosh P, Chapman LAC, Alim A, Bilbe G, Alvar J. 2019. Quantifying the infectiousness of post-kala-azar dermal leishmaniasis toward sand flies. Clin Infect Dis 69:251–258.

20. Molina R, Ghosh D, Carrillo E, Monnerat S, Bern C, Mondal Di, Alvar J. 2017. Infectivity of Post-Kala-azar Dermal Leishmaniasis Patients to Sand Flies: Revisiting a Proof of Concept in the Context of the Kala-azar Elimination Program in the Indian Subcontinent. Clin Infect Dis 65:150–153.

21. Zijlstra EE, Alves F, Rijal S, Arana B, Alvar J. 2017. Post-kala-azar dermal leishmaniasis in the Indian subcontinent: A threat to the South-East Asia Region Kala-azar Elimination Programme. PLoS Negl Trop Dis 11.

22. Zijlstra EE. 2016. The immunology of post-kala-azar dermal leishmaniasis (PKDL). Parasites and Vectors 9.

23. Global programme to, eliminate lymphatic filariasis. 2015. Relevé épidémiologique hebdomadaire: Programme mondial pour l'élimination de la filariose lymphatique: rapport de situation, 2014. OMSRelevé épidémiologique Hebd 38:489–504.

24. Soong L. 2009. Chapter 63 - LeishmaniasisVaccines for Biodefense and Emerging and Neglected Diseases.

25. Steverding D. 2017. The history of leishmaniasis. Parasites and Vectors 10:1–10.

26. Gibson ME. 1983. The identification of kala azar and the discovery of leishmania donovani. Med Hist 27:203–213.

27. Akhoundi M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, Sereno D. 2016. A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sand flies. PLoS Negl Trop Dis 10.

28. Kraeva N, Butenko A, Hlaváčová J, Kostygov A, Myškova J, Grybchuk D, Leštinová T, Votýpka J, Volf P, Opperdoes F, Flegontov P, Lukeš J, Yurchenko V. 2015. Leptomonas seymouri: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with Leishmania donovani. PLoS Pathog 11.

29. Lukeš J, Mauricio IL, Schönian G, Dujardin JC, Soteriadou K, Dedet JP, Kuhls K, Tintaya KWQ, Jirků M, Chocholová E, Haralambous C, Pratlong F, Oborník M, Horák A, Ayala

FJ, Miles MA. 2007. Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc Natl Acad Sci U S A 104:9375–9380.

30. Kelly PH, Bahr SM, Serafim TD, Ajami NJ, Petrosino JF, Meneses C, Kirby JR, Valenzuela JG, Kamhawi S, Wilson ME. 2017. The gut microbiome of the vector lutzomyia longipalpis is essential for survival of leishmania infantum. MBio 8.

31. Sant'Anna MRV, Diaz-Albiter H, Aguiar-Martins K, Al Salem WS, Cavalcante RR, Dillon VM, Bates PA, Genta FA, Dillon RJ. 2014. Colonisation resistance in the sand fly gut: Leishmania protects Lutzomyia longipalpis from bacterial infection. Parasites and Vectors 7.

32. Titus RG, Ribeiro JMC. 1988. Salivary gland lysates from the sand fly lutzomyia longipalpis enhance leishmania infectivity. Science (80-) 239:1306–1308.

33. Kamhawi S. 2000. The biological and immunomodulatory properties of sand fly saliva and its role in the establishment of Leishmania infections. Microbes Infect 2:1765–1773.

34. Rogers ME, Ilg T, Nikolaev A V., Ferguson MAJ, Bates PA. 2004. Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG. Nature 430:463–467.

35. Rogers ME, Bates PA. 2007. Leishmania manipulation of sand fly feeding behavior results in enhanced transmission. PLoS Pathog 3:0818–0825.

36. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, Dobson DE, Beverley SM, Sacks DL. 2009. Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. Science (80-) 324:265–268.

37. Inbar E, Shaik J, Iantorno SA, Romano A, Nzelu CO, Owens K, Sanders MJ, Dobson D, Cotton JA, Grigg ME, Beverley SM, Sacks D. 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in leishmania. PLoS Genet 15.

38. Ready PD. 2013. Biology of phlebotomine sand flies as vectors of disease agents. Annu Rev Entomol 58:227–250.

39. Kassahun A, Sadlova J, Dvorak V, Kostalova T, Rohousova I, Frynta D, Aghova T, Yasur-Landau D, Lemma W, Hailu A, Baneth G, Warburg A, Volf P, Votypka J. 2015. Detection of leishmania donovani and L. tropica in ethiopian wild rodents. Acta Trop 145:39–44.

40. Gramiccia M, Gradoni L. 2005. The current status of zoonotic leishmaniases and approaches to disease control. Int J Parasitol 35:1169–1180.

41. Kassahun A, Sadlova J, Benda P, Kostalova T, Warburg A, Hailu A, Baneth G, Volf P, Votypka J. 2015. Natural infection of bats with Leishmania in Ethiopia. Acta Trop 150:166–170.

42. Berzunza-Cruz M, Rodríguez-Moreno Á, Gutiérrez-Granados G, González-Salazar C, Stephens CR, Hidalgo-Mihart M, Marina CF, Rebollar-Téllez EA, Bailón-Martínez D, Balcells CD, Ibarra-Cerdeña CN, Sánchez-Cordero V, Becker I. 2015. Leishmania (L.) mexicana Infected Bats in Mexico: Novel Potential Reservoirs. PLoS Negl Trop Dis 9.

43. Al-Salem W, Herricks JR, Hotez PJ. 2016. A review of visceral leishmaniasis during the conflict in South Sudan and the consequences for East African countries. Parasites and Vectors 9.

44. Liu D, Uzonna JE. 2012. The early interaction of Leishmania with macrophages and dendritic cells and its influence on the host immune response. Front Cell Infect Microbiol 2:83.

45. Bates PA. 2007. Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies. Int J Parasitol 37:1097–1106.

46. Peters NC, Egen JG, Secundino N, Debrabant A, Kimblin N, Kamhawi S, Lawyer P, Fay MP, Germain RN, Sacks D. 2008. In vivo imaging reveals an essential role for neutrophils in leishmaniasis transmitted by sand flies. Science (80-) 321:970–974.

47. Moore KJ, Labrecque S, Matlashewski G. 1993. Alteration of Leishmania donovani infection levels by selective impairment of macrophage signal transduction. J Immunol 150:4457–65.

48. Moore KJ, Matlashewski G. 1994. Intracellular infection by Leishmania donovani inhibits macrophage apoptosis. J Immunol 152:2930–7.

49. Moore KJ, Turco SJ, Matlashewski G. 1994. Leishmania donovani infection enhances macrophage viability in the absence of exogenous growth factor. J Leukoc Biol 55:91–98.

50. de Menezes JPB, Koushik A, Das S, Guven C, Siegel A, Laranjeira-Silva MF, Losert W, Andrews NW. 2017. Leishmania infection inhibits macrophage motility by altering F-actin dynamics and the expression of adhesion complex proteins. Cell Microbiol 19.

51. Kamir D, Zierow S, Leng L, Cho Y, Diaz Y, Griffith J, McDonald C, Merk M, Mitchell RA, Trent J, Chen Y, Kwong Y-KA, Xiong H, Vermeire J, Cappello M, McMahon-Pratt D, Walker J, Bernhagen J, Lolis E, Bucala R. 2008. A Leishmania Ortholog of Macrophage Migration Inhibitory Factor Modulates Host Macrophage Responses . J Immunol 180:8250–8261.

52. Kane MM, Mosser DM. 2001. The Role of IL-10 in Promoting Disease Progression in Leishmaniasis. J Immunol 166:1141–1147.

53. Séguin O, Descoteaux A. 2016. Leishmania, the phagosome, and host responses: The journey of a parasite. Cell Immunol 309:1–6.

54. Moradin N, Descoteaux A. 2012. Leishmania promastigotes: building a safe niche within macrophages. Front Cell Infect Microbiol 2:121.

55. Podinovskaia M, Descoteaux A. 2015. Leishmania and the macrophage: A multifaceted interaction. Future Microbiol 10:111–129.

56. Vinet AF, Fukuda M, Turco SJ, Descoteaux A. 2009. The Leishmania donovani lipophosphoglycan excludes the vesicular proton-ATPase from phagosomes by impairing the recruitment of Synaptotagmin V. PLoS Pathog 5.

57. Arango Duque G, Jardim A, Gagnon É, Fukuda M, Descoteaux A. 2019. The host cell

secretory pathway mediates the export of Leishmania virulence factors out of the parasitophorous vacuole. PLoS Pathog 15.

58. WHO. 2010. Report of a meeting of the WHO Expert Committee on the Leishmaniasis control. Rep a Meet WHO Expert Comm 22–26.

59. Roberts WL, Berman JD, Rainey PM. 1995. In vitro antileishmanial properties of tri- and pentavalent antimonial preparations. Antimicrob Agents Chemother 39:1234–1239.

60. Herman JD. 1988. Chemotherapy for leishmaniasis: Biochemical mechanisms, clinical efficacy, and future strategies. Rev Infect Dis 10:560–586.

61. Haldar AK, Sen P, Roy S. 2011. Use of Antimony in the Treatment of Leishmaniasis: Current Status and Future Directions. Mol Biol Int 2011:1–23.

62. Matlashewski G, Arana B, Kroeger A, Battacharya S, Sundar S, Das P, Sinha PK, Rijal S, Mondal D, Zilberstein D, Alvar J. 2011. Visceral leishmaniasis: Elimination with existing interventions. Lancet Infect Dis 11:322–325.

63. Saheki MN, Lyra MR, Bedoya-Pacheco SJ, Antonio LDF, Pimentel MIF, Salgueiro MDM, Vasconcellos EDCFE, Passos SRL, Santos GPL Dos, Ribeiro MN, Fagundes A, Madeira MDF, Mouta-Confort E, Marzochi MCDA, Valete-Rosalino CM, Schubach ADO. 2017. Low versus high dose of antimony for American cutaneous leishmaniasis: A randomized controlled blind non-inferiority trial in Rio de Janeiro, Brazil. PLoS One 12.

64. Baiocco P, Colotti G, Franceschini S, Ilari A. 2009. Molecular basis of antimony treatment in Leishmaniasis. J Med Chem 52:2603–2612.

65. Demicheli C, Frézard F, Mangrum JB, Farrell NP. 2008. Interaction of trivalent antimony with a CCHC zinc finger domain: Potential relevance to the mechanism of action of antimonial drugs. Chem Commun 4828–4830.

66. Walker J, Saravia NG. 2004. Inhibition of Leishmania donovani promastigote DNA topoisomerase I and human monocyte DNA topoisomerases I and II by antimonial drugs and classical antitopoisomerase agents. J Parasitol 90:1155–1162.

67. Frézard F, Demicheli C, Ribeiro RR. 2009. Pentavalent antimonials: New perspectives for old drugs. Molecules 14:2317–2336.

68. Muniz-Junqueira MI, de Paula-Coelho VN. 2008. Meglumine antimonate directly increases phagocytosis, superoxide anion and TNF-α production, but only via TNF-α it indirectly increases nitric oxide production by phagocytes of healthy individuals, in vitro. Int Immunopharmacol 8:1633–1638.

69. Kumar Saha A, Mukherjee T, Bhaduri A. 1986. Mechanism of action of amphotericin B on Leishmania donovani promastigotes. Mol Biochem Parasitol 19:195–200.

70. Coukell AJ, Brogden RN. 1998. Liposomal amphotericin B: Therapeutic use in the management of fungal infections and visceral leishmaniasis. Drugs 55:585–612.

71. De Macedo-Silva ST, Urbina JA, De Souza W, Rodrigues JCF. 2013. In vitro activity of the antifungal azoles itraconazole and posaconazole against Leishmania amazonensis. PLoS

One 8.

72. Sundar S, Jaya J. 2010. Liposomal amphotericin B and leishmaniasis: Dose and response. J Glob Infect Dis 2:159.

73. Balasegaram M, Ritmeijer K, Lima MA, Burza S, Ortiz Genovese G, Milani B, Gaspani S, Potet J, Chappuis F. 2012. Liposomal amphotericin B as a treatment for human leishmaniasis. Expert Opin Emerg Drugs 17:493–510.

74. Yardley V, Croft SL. 1997. Activity of liposomal amphotericin B against experimental cutaneous leishmaniasis. Antimicrob Agents Chemother 41:752–756.

75. Chattopadhyay A, Jafurulla M. 2011. A novel mechanism for an old drug: Amphotericin B in the treatment of visceral leishmaniasis. Biochem Biophys Res Commun 416:7–12.

76. Paila YD, Saha B, Chattopadhyay A. 2010. Amphotericin B inhibits entry of Leishmania donovani into primary macrophages. Biochem Biophys Res Commun 399:429–433.

77. Lemke A, Kiderlen AF, Kayser O. 2005. Amphotericin B. Appl Microbiol Biotechnol 68:151–162.

78. Johnson PC, Wheat LJ, Cloud GA, Goldman M, Lancaster D, Bamberger DM, Powderly WG, Hafner R, Kauffman CA, Dismukes WE. 2002. Safety and efficacy of liposomal amphotericin B compared with conventional amphotericin B for induction therapy of histoplasmosis in patients with AIDS. Ann Intern Med 137:105–109.

79. Botero Aguirre JP, Restrepo Hamid AM. 2015. Amphotericin B deoxycholate versus liposomal amphotericin B: Effects on kidney function. Cochrane Database Syst Rev 2015.

80. Sundar S, Chakravarty J, Agarwal D, Rai M, Murray HW. 2010. Single-Dose Liposomal Amphotericin B for Visceral Leishmaniasis in India. N Engl J Med 362:504–512.

81. Mbongo N, Loiseau PM, Billion MA, Robert-Gero M. 1998. Mechanism of amphotericin B resistance in Leishmania donovani promastigotes. Antimicrob Agents Chemother 42:352–357.

82. Purkait B, Kumar A, Nandi N, Sardar AH, Das S, Kumar S, Pandey K, Ravidas V, Kumar M, De T, Singh D, Das P. 2012. Mechanism of amphotericin B resistance in clinical isolates of Leishmania donovani. Antimicrob Agents Chemother 56:1031–1041.

83. Ponte-Sucre A, Gamarro F, Dujardin JC, Barrett MP, López-Vélez R, García-Hernández R, Pountain AW, Mwenechanya R, Papadopoulou B. 2017. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. PLoS Negl Trop Dis 11.

84. Anversa L, Salles Tiburcio MG, Batista LR, Cuba MB, Nogueira Nascentes GA, Martins TY, Richini Pereira VB, Ruiz L da S, Dias da Silva VJ, Ramirez LE. 2017. Amiodarone and itraconazole improve the activity of pentavalent antimonial in the treatment of experimental cutaneous leishmaniasis. Int J Antimicrob Agents 50:159–165.

85. Consigli J, Danielo C, Gallerano V, Papa M, Guidi A. 2006. Cutaneous leishmaniasis: Successful treatment with itraconazole. Int J Dermatol 45:46–49.

86. Paniz Mondolfi AE, Stavropoulos C, Gelanew T, Loucas E, Perez Alvarez AM, Benaim G, Polsky B, Schoenian G, Sordillo EM. 2011. Successful treatment of old world cutaneous leishmaniasis caused by Leishmania infantum with posaconazole. Antimicrob Agents Chemother 55:1774–1776.

87. Navin TR, Arana BA, Arana FE, Berman JD, Chajón JF. 1992. Placebo-Controlled Clinical Trial of Sodium Stibogluconate (Pentostam) versus Ketoconazole for Treating Cutaneous Leishmaniasis in Guatemala. J Infect Dis 165:528–534.

88. Alrajhi AA, Ibrahim EA, De Vol EB, Khairat M, Faris RM, Maguire JH. 2002. Fluconazole for the Treatment of Cutaneous Leishmaniasis Caused by Leishmania major . N Engl J Med 346:891–895.

89. Chen CK, Leung SSF, Guilbert C, Jacobson MP, Mckerrow JH, Podust LM. 2010. Structural characterization of CYP51 from Trypanosoma cruzi and Trypanosoma brucei bound to the antifungal drugs posaconazole and fluconazole. PLoS Negl Trop Dis 4.

90. Sundar S, Jha TK, Thakur CP, Engel J, Sindermann H, Fischer C, Junge K, Bryceson A, Berman J. 2002. Oral Miltefosine for Indian Visceral Leishmaniasis. N Engl J Med 347:1739–1746.

91. Verma NK, Dey CS. 2004. Possible mechanism of miltefosine-mediated death of Leishmania donovani. Antimicrob Agents Chemother 48:3010–3015.

92. de Aquino Marinho F, da Silva Gonçalves KC, de Oliveira SS, de Siqueira Couto de Oliveira AC, Bellio M, D'Avila-Levy CM, dos Santos ALS, Branquinha MH. 2011. Miltefosine induces programmed cell death in leishmania amazonensis promastigotes. Mem Inst Oswaldo Cruz 106:507–509.

93. Pérez-Victoria FJ, Gamarro F, Ouellette M, Castanys S. 2003. Functional cloning of the miltefosine transporter: A novel p-type phospholipid translocase from leishmania involved in drug resistance. J Biol Chem 278:49965–49971.

94. Zhang W-W, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. mSphere 2.

95. Srivastava S, Mishra J, Gupta AK, Singh A, Shankar P, Singh S. 2017. Laboratory confirmed miltefosine resistant cases of visceral leishmaniasis from India. Parasites and Vectors 10:1–11.

96. Cojean S, Houzé S, Haouchine D, Huteau F, Lariven S, Hubert V, Michard F, Bories C, Pratlong F, le Bras J, Loiseau PM, Matheron S. 2012. Leishmania resistance to miltefosine associated with genetic marker. Emerg Infect Dis 18:704–706.

97. Bray PG, Barrett MP, Ward SA, De Koning HP. 2003. Pentamidine uptake and resistance in pathogenic protozoa: Past, present and future. Trends Parasitol 19:232–239.

98. Basselin M, Denise H, Coombs GH, Barrett MP. 2002. Resistance to pentamidine in Leishmania mexicana involves exclusion of the drug from the mitochondrion. Antimicrob Agents Chemother 46:3731–3738.

99. Mukherjee A, Padmanabhan PK, Sahani MH, Barrett MP, Madhubala R. 2006. Roles for mitochondria in pentamidine susceptibility and resistance in Leishmania donovani. Mol Biochem Parasitol 145:1–10.

100. Coelho AC, Beverley SM, Cotrim PC. 2003. Functional genetic identification of PRP1, an ABC transporter superfamily member conferring pentamidine resistance in Leishmania major. Mol Biochem Parasitol 130:83–90.

101. Croft SL, Sundar S, Fairlamb AH. 2006. Drug resistance in leishmaniasis. Clin Microbiol Rev 19:111–126.

102. Sosa N, Capitán Z, Nieto J, Nieto M, Calzada J, Paz H, Spadafora C, Kreishman-Deitrick M, Kopydlowski K, Ullman D, McCarthy WF, Ransom J, Berman J, Scott C, Grogl M. 2013. Randomized, double-blinded, phase 2 trial of wr 279,396 (paromomycin and gentamicin) for cutaneous leishmaniasis in panama. Am J Trop Med Hyg 89:557–563.

103. Jhingran A, Chawla B, Saxena S, Barrett MP, Madhubala R. 2009. Paromomycin: Uptake and resistance in Leishmania donovani. Mol Biochem Parasitol 164:111–117.

104. Maarouf M, Lawrence F, Croft SL, Robert-Gero M. 1995. Ribosomes of Leishmania are a target for the aminoglycosides. Parasitol Res 81:421–425.

105. Bhandari V, Sundar S, Dujardin JC, Salotra P. 2014. Elucidation of cellular mechanisms involved in experimental paromomycin resistance in leishmania donoVani. Antimicrob Agents Chemother 58:2580–2585.

106. Hendrickx S, Beyers J, Mondelaers A, Eberhardt E, Lachaud L, Delputte P, Cos P, Maes L. 2016. Evidence of a drug-specific impact of experimentally selected paromomycin and miltefosine resistance on parasite fitness in Leishmania infantum. J Antimicrob Chemother 71:1914–1921.

107. Aronson NE, Wortmann GW, Byrne WR, Howard RS, Bernstein WB, Marovich MA, Polhemus ME, Yoon IK, Hummer KA, Gasser RA, Oster CN, Benson PM. 2010. A randomized controlled trial of local heat therapy versus intravenous sodium stibogluconate for the treatment of cutaneous Leishmania major infection. PLoS Negl Trop Dis 4.

108. Lakhal-Naouar I, Slike BM, Aronson NE, Marovich MA. 2015. The Immunology of a Healing Response in Cutaneous Leishmaniasis Treated with Localized Heat or Systemic Antimonial Therapy. PLoS Negl Trop Dis 9.

109. Refai WF, Madarasingha NP, Sumanasena B, Weerasingha S, De Silva A, Fernandopulle R, Satoskar AR, Karunaweera ND. 2017. Efficacy, safety and cost-effectiveness of thermotherapy in the treatment of leishmania donovani-induced cutaneous leishmaniasis: A randomized controlled clinical trial. Am J Trop Med Hyg 97:1120–1126.

110. Wheeler RJ, Gluenz E, Gull K. 2011. The cell cycle of Leishmania: Morphogenetic events and their implications for parasite biology. Mol Microbiol 79:647–662.

111. Liévin-Le Moal V, Loiseau PM. 2016. Leishmania hijacking of the macrophage intracellular compartments. FEBS J 283:598–607.

112. Kamhawi S. 2006. Phlebotomine sand flies and Leishmania parasites: friends or foes?

Trends Parasitol 22:439–445.

113. McCall LI, Zhang WW, Matlashewski G. 2013. Determinants for the Development of Visceral Leishmaniasis Disease. PLoS Pathog 9.

114. Guessous-Idrissi N, Berrag B, Riyad M, Sahibi H, Bichichi M, Rhalem A. 1997. Short report: Leishmania tropica: Etiologic agent of a case of canine visceral Leishmaniasis in Northern Morocco. Am J Trop Med Hyg 57:172–173.

115. Magill AJ, Grogl M, Gasser RA, Sun W, Oster CN. 1993. Visceral Infection Caused by Leishmania tropica in Veterans of Operation Desert Storm. N Engl J Med 328:1383–1387.

116. Sacks DL, Kenney RT, Neva FA, Kreutzer RD, Jaffe CL, Gupta AK, Sharma MC, Sinha SP, Saran R. 1995. Indian kala-azar caused by Leishmania tropica. Lancet 345:959–961.

117. Saari S, Näreaho A, Nikander S. 2019. Introduction, p. 1–3. *In* Canine Parasites and Parasitic Diseases.

118. Ozbilgin A, Culha G, Zeyrek FY, Töz S, Gündüz C, Kurt Ö, Pratlong F, Ozbel Y. 2012. Cutaneous and visceral tropisms of Leishmania tropica/Leishmania infantum hybrids in a murine model: First report of hybrid Leishmania strains isolated in Turkey. Int J Infect Dis 16:e165.

119. Svobodová M, Alten B, Zídková L, Dvořák V, Hlavačková J, Myšková J, Šeblová V, Kasap OE, Belen A, Votýpka J, Volf P. 2009. Cutaneous leishmaniasis caused by Leishmania infantum transmitted by Phlebotomus tobbi. Int J Parasitol 39:251–256.

120. Grimaldi G, David JR, McMahon-Pratt D. 1987. Identification and distribution of new world Leishmania species characterized by serodeme analysis using monoclonal antibodies. Am J Trop Med Hyg 36:270–287.

121. Kevric I, Cappel MA, Keeling JH. 2015. New World and Old World Leishmania Infections: A Practical Review. Dermatol Clin 33:579–593.

122. Alexandre J, Sadlova J, Lestinova T, Vojtkova B, Jancarova M, Podesvova L, Yurchenko V, Dantas-Torres F, Brandão-Filho SP, Volf P. 2020. Experimental infections and co-infections with Leishmania braziliensis and Leishmania infantum in two sand fly species, Lutzomyia migonei and Lutzomyia longipalpis. Sci Rep 10.

123. Berman J. 2006. Visceral leishmaniasis in the New World & Africa. Indian J Med Res 123:289–294.

124. Ilgoutz SC, McConville MJ. 2001. Function and assembly of the Leishmania surface coat, p. 899–908. *In* International Journal for Parasitology.

125. Menon AK. 2013. Glycosylphosphatidylinositol Anchors, p. 476–478. *In* Encyclopedia of Biological Chemistry: Second Edition.

126. Ibraim IC, De Assis RR, Pessoa NL, Campos MA, Melo MN, Turco SJ, Soares RP. 2013. Two biochemically distinct lipophosphoglycans from Leishmania braziliensis and Leishmania infantum trigger different innate immune responses in murine macrophages. Parasites and Vectors 6.

127. Weise F, Stierhof YD, Kühn C, Wiese M, Overath P. 2000. Distribution of GPI-anchored proteins in the protozoan parasite Leishmania, based on an improved ultrastructural description using high-pressure frozen cells. J Cell Sci 113:4587–4603.

128. Yao C, Donelson JE, Wilson ME. 2003. The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. Mol Biochem Parasitol 132:1–16.

129. Brittingham A, Morrison CJ, McMaster WR, McGwire BS, Chang K-P, Mosser DM. 1995. Role of the Leishmania surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. Parasitol Today 11:445–446.

130. Isnard A, Shio MT, Olivier M. 2012. Impact of Leishmania metalloprotease GP63 on macrophage signaling. Front Cell Infect Microbiol 2:72.

131. Gomez MA, Contreras I, Hallé M, Tremblay ML, McMaster RW, Olivier M. 2009. Leishmania GP63 alters host signaling through cleavage-activated protein tyrosine phosphatases. Sci Signal 2.

132. Matheoud D, Moradin N, Bellemare-Pelletier A, Shio MT, Hong WJ, Olivier M, Gagnon É, Desjardins M, Descoteaux A. 2013. Leishmania evades host immunity by inhibiting antigen cross-presentation through direct cleavage of the SNARE VAMP8. Cell Host Microbe 14:15–25.

133. Jain R, Ghoshal A, Mandal C, Shaha C. 2010. Leishmania cell surface prohibitin: Role in host-parasite interaction. Cell Microbiol 12:432–452.

134. Rogers ME. 2012. The role of Leishmania proteophosphoglycans in sand fly transmission and infection of the mammalian host. Front Microbiol 3.

135. Ilg T, Stierhof YD, Craik D, Simpson R, Handman E, Bacic A. 1996. Purification and structural characterization of a filamentous, mucin- like proteophosphoglycan secreted by Leishmania parasites. J Biol Chem 271:21583–21596.

136. Forestier CL, Gao Q, Boons GJ. 2014. Leishmania lipophosphoglycan: How to establish structure-activity relationships for this highly complex and multifunctional glycoconjugate? Front Cell Infect Microbiol 4.

137. Silva LM Da, Owens KL, Murta SMF, Beverley SM. 2009. Regulated expression of the Leishmania major surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins. Proc Natl Acad Sci U S A 106:7583–7588.

138. Späth GF, Garraway LA, Turco SJ, Beverley SM. 2003. The role(s) of lipophosphoglycan (LPG) in the establishment of Leishmania major infections in mammalian hosts. Proc Natl Acad Sci U S A 100:9536–9541.

139. Kamhawi S, Ramalho-Ortigao M, Van MP, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG. 2004. A role for insect galectins in parasite survival. Cell 119:329–341.

140. Olivier M, Atayde VD, Isnard A, Hassani K, Shio MT. 2012. Leishmania virulence factors: Focus on the metalloprotease GP63. Microbes Infect 14:1377–1389.

141. Landfear SM. 2011. Nutrient transport and pathogenesis in selected parasitic protozoa. Eukaryot Cell 10:483–493.

142. Burchmore RJS, Barrett MP. 2001. Life in vacuoles - Nutrient acquisition by Leishmania amastigotes. Int J Parasitol 31:1311–1320.

143. Agarwal S, Rastogi R, Gupta D, Patel N, Raje M, Mukhopadhyay A. 2013. Clathrin-mediated hemoglobin endocytosis is essential for survival of Leishmania. Biochim Biophys Acta - Mol Cell Res 1833:1065–1077.

144. Ali HZ, Harding CR, Denny PW. 2012. Endocytosis and sphingolipid scavenging in leishmania mexicana amastigotes. Biochem Res Int https://doi.org/10.1155/2012/691363.

145. Patel N, Singh SB, Basu SK, Mukhopadhyay A. 2008. Leishmania requires Rab7-mediated degradation of endocytosed hemoglobin for their growth. Proc Natl Acad Sci U S A 105:3980–3985.

146. Rastogi R, Verma JK, Kapoor A, Langsley G, Mukhopadhyay A. 2016. Rab5 isoforms specifically regulate different modes of endocytosis in Leishmania. J Biol Chem 291:14732–14746.

147. Atayde VD, Aslan H, Townsend S, Hassani K, Kamhawi S, Olivier M. 2015. Exosome Secretion by the Parasitic Protozoan Leishmania within the Sand Fly Midgut. Cell Rep 13:957–967.

148. Silverman JM, Clos J, De'Oliveira CC, Shirvani O, Fang Y, Wang C, Foster LJ, Reiner NE. 2010. An exosome-based secretion pathway is responsible for protein export from Leishmania and communication with macrophages. J Cell Sci 123:842–852.

149. Lambertz U, Silverman JM, Nandan D, McMaster WR, Clos J, Foster LJ, Reiner NE. 2012. Secreted virulence factors and immune evasion in visceral leishmaniasis. J Leukoc Biol 91:887–899.

150. Wheeler RJ, Sunter JD, Gull K. 2016. Flagellar pocket restructuring through the Leishmania life cycle involves a discrete flagellum attachment zone. J Cell Sci 129:854–867.

151. Gluenz E, Ginger ML, McKean PG. 2010. Flagellum assembly and function during the Leishmania life cycle. Curr Opin Microbiol 13:473–479.

152. Sunter J, Gull K. 2017. Shape, form, function and Leishmania pathogenicity: from textbook descriptions to biological understanding. Open Biol 7.

153. Waller RF, McConville MJ. 2002. Developmental changes in lysosome morphology and function Leishmania parasites. Int J Parasitol 32:1435–1445.

154. Wang Z, Wheeler RJ, Sunter JD. 2020. Lysosome assembly and disassembly changes endocytosis rate through the Leishmania cell cycle. Microbiologyopen 9.

155. Moreno SNJ, Docampo R. 2009. The role of acidocalcisomes in parasitic protists, p. 208–213. *In* Journal of Eukaryotic Microbiology.

156. Miranda K, Docampo R, Grillo O, Franzen A, Attias M, Vercesi A, Plattner H, Hentschel

J, De Souza W. 2004. Dynamics of polymorphism of acidocalcisomes in Leishmania parasites. Histochem Cell Biol 121:407–418.

157. Madeira Da Silva L, Beverley SM. 2010. Expansion of the target of rapamycin (TOR) kinase family and function in Leishmania shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. Proc Natl Acad Sci U S A 107:11965–11970.

158. Blum B, Bakalara N, Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: RNA molecules transcribed from maxicircle DNA provide the edited information. Cell 60:189–198.

159. Michels PAM, Moyersoen J, Krazy H, Galland N, Herman M, Hannaert V. 2005. Peroxisomes, glyoxysomes and glycosomes. Mol Membr Biol 22:133–145.

160. Parsons M. 2004. Glycosomes: Parasites and the divergence of peroxisomal purpose. Mol Microbiol 53:717–724.

161. Moyersoen J, Choe J, Fan E, Hol WGJ, Michels PAM. 2004. Biogenesis of peroxisomes and glycosomes: Trypanosomatid glycosome assembly is a promising new drug target. FEMS Microbiol Rev 28:603–643.

162. Besteiro S, Williams RAM, Coombs GH, Mottram JC. 2007. Protein turnover and differentiation in Leishmania. Int J Parasitol https://doi.org/10.1016/j.ijpara.2007.03.008.

163. Britto C, Ravel C, Bastien P, Blaineau C, Pagès M, Dedet JP, Wincker P. 1998. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World Leishmania genomes. Gene 222:107–117.

164. Kazemi B. 2011. Genomic organization of Leishmania species. Iran J Parasitol 6:1–18.

165. Stiles JK, Hicock PI, Shah PH, Meade JC. 1999. Genomic organization, transcription, splicing and gene regulation in Leishmania. Ann Trop Med Parasitol 93:781–807.

166. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, Berriman M, Sacks DL, Cotton JA, Grigg ME. 2017. Gene expression in Leishmania is regulated predominantly by gene dosage. MBio 8.

167. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, De Doncker S, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Schönian G, Stark O, Sundar S, Vanaerschot M, Hertz-Fowler C, Dujardin JC, Berriman M. 2011. Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res 21:2143–2156.

168. Sterkers Y, Crobu L, Lachaud L, Pagès M, Bastien P. 2014. Parasexuality and mosaic aneuploidy in Leishmania: Alternative genetics. Trends Parasitol 30:429–435.

169. Fu G, Barker DC. 1998. Characterisation of Leishmania telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. Nucleic Acids Res 26:2161–2167.

170. 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 LRO, Barrell B, Cruz AK, Mottram JC, Smith DF, Berriman M. 2007. Comparative genomic analysis of three Leishmania species that cause diverse human disease. Nat Genet 39:839–847.

171. Worthey EA, Martinez-Calvillo S, Schnaufer A, Aggarwal G, Cawthra J, Fazelinia G, Fong C, Fu G, Hassebrock M, Hixson G, Ivens AC, Kiser P, Marsolini F, Rickell E, Salavati R, Sisk E, Sunkin SM, Stuart K, Myler PJ. 2003. Leishmania major chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. Nucleic Acids Res 31:4201–4210.

172. Van Luenen HGAM, Farris C, Jan S, Genest PA, Tripathi P, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramasamy G, Vainio S, Heidebrecht T, Perrakis A, Pagie L, Van Steensel B, Myler PJ, Borst P. 2012. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell 150:909–921.

173. Reynolds D, Cliffe L, Förstner KU, Hon CC, Siegel TN, Sabatini R. 2014. Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. Nucleic Acids Res 42:9717–9729.

174. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ. 2009. Histone acetylations mark origins of polycistronic transcription in Leishmania major. BMC Genomics 10.

175. Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ. 2004. Transcription initiation and termination on Leishmania major chromosome 3. Eukaryot Cell 3:506–517.

176. LeBowitz JH, Smith HQ, Rusche L, Beverley SM. 1993. Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes Dev 7:996–1007.

177. Curotto de Lafaille MA, Laban A, Wirth DF. 1992. Gene expression in Leishmania: Analysis of essential 5' DNA sequences. Proc Natl Acad Sci U S A 89:2703–2707.

178. Nourbakhsh F, Uliana SRB, Smith DF. 1996. Characterisation and expression of a stage-regulated gene of Leishmania major. Mol Biochem Parasitol 76:201–213.

179. Bringaud F, Müller M, Cerqueira GC, Smith M, Rochette A, El-Sayed NMA, Papadopoulou B, Ghedin E. 2007. Members of a large retroposon family are determinants of post-transcriptional gene expression in Leishmania. PLoS Pathog 3:1291–1307.

180. Kramer S. 2012. Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. Mol Biochem Parasitol 181:61–72.

181. Leprohon P, Fernandez-Prada C, Gazanion É, Monte-Neto R, Ouellette M. 2015. Drug resistance analysis by next generation sequencing in Leishmania. Int J Parasitol Drugs Drug Resist 5:26–35.

182. Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J, Pescher P, Vanaerschot M, Meehan CJ, Cuypers B, De Muylder G, Späth GF, Bussotti G, Vermeesch JR, Berriman M, Cotton JA, Volf P, Dujardin JC, Domagalska MA. 2017. Modulation of aneuploidy in

leishmania donovani during adaptation to different in vitro and in vivo environments and its impact on gene expression. MBio 8.

183. Patino LH, Imamura H, Cruz-Saavedra L, Pavia P, Muskus C, Méndez C, Dujardin JC, Ramírez JD. 2019. Major changes in chromosomal somy, gene expression and gene dosage driven by SbIII in Leishmania braziliensis and Leishmania panamensis. Sci Rep 9.

184. Bussotti G, Gouzelou E, Boité MC, Kherachi I, Harrat Z, Eddaikra N, Mottram JC, Antoniou M, Christodoulou V, Bali A, Guerfali FZ, Laouini D, Mukhtar M, Dumetz F, Dujardin JC, Smirlis D, Lechat P, Pescher P, Hamouchi A El, Lemrani M, Chicharro C, Llanes-Acevedo IP, Botana L, Cruz I, Moreno J, Jeddi F, Aoun K, Bouratbine A, Cupolillo E, Späth GF. 2018. Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. MBio 9.

185. Smith M, Bringaud F, Papadopoulou B. 2009. Organization and evolution of two SIDER retroposon subfamilies and their impact on the Leishmania genome. BMC Genomics 10.

186. Ubeda JM, Raymond F, Mukherjee A, Plourde M, Gingras H, Roy G, Lapointe A, Leprohon P, Papadopoulou B, Corbeil J, Ouellette M. 2014. Genome-Wide Stochastic Adaptive DNA Amplification at Direct and Inverted DNA Repeats in the Parasite Leishmania. PLoS Biol 12.

187. Subramanian A, Sarkar RR. 2015. Comparison of codon usage bias across Leishmania and Trypanosomatids to understand mRNA secondary structure, relative protein abundance and pathway functions. Genomics 106:232–241.

188. Jeacock L, Faria J, Horn D. 2018. Codon usage bias controls mRNA and protein abundance in trypanosomatids. Elife 7.

189. Ramamoorthy R, Swihart KG, McCoy JJ, Wilson ME, Donelson JE. 1995. Intergenic regions between tandem gp63 genes influence the differential expression of gp63 RNAs in Leishmania chagasi promastigotes. J Biol Chem 270:12133–12139.

190. Brooks DR, Denise H, Westrop GD, Coombs GH, Mottram JC. 2001. The Stage-regulated Expression of Leishmania mexicana CPB Cysteine Proteases Is Mediated by an Intercistronic Sequence Element. J Biol Chem 276:47061–47069.

191. Argaman M, Aly R, Shapira M. 1994. Expression of heat shock protein 83 in Leishmania is regulated post-transcriptionally. Mol Biochem Parasitol 64:95–110.

192. Larreta R, Soto M, Quijada L, Folgueira C, Abanades DR, Alonso C, Requena JM. 2004. The expression of HSP83 genes in Leishmania infantum is affected by temperature and by stage-differentiation and is regulated at the levels of mRNA stability and translation. BMC Mol Biol 5.

193. El Fakhry Y, Ouellette M, Papadopoulou B. 2002. A proteomic approach to identify developmentally regulated proteins in Leishmania infantum. Proteomics 2:1007–1017.

194. Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, Holland N, Kuzyk M, Borchers C, Zilberstein D, Myler PJ. 2011. Multiple levels of gene regulation mediate

differentiation of the intracellular pathogen Leishmania . FASEB J 25:515–525.

195. McNicoll F, Müller M, Cloutier S, Boilard N, Rochette A, Dubé M, Papadopoulou B. 2005. Distinct 3′-untranslated region elements regulate stage-specific mRNA accumulation and translation in Leishmania. J Biol Chem 280:35238–35246.

196. Jewell JL, Guan KL. 2013. Nutrient signaling to mTOR and cell growth. Trends Biochem Sci 38:233–242.

197. Kim J, Guan KL. 2019. mTOR as a central hub of nutrient signalling and cell growth. Nat Cell Biol 21:63–71.

198. Jhanwar-Uniyal M, Wainwright J V., Mohan AL, Tobias ME, Murali R, Gandhi CD, Schmidt MH. 2019. Diverse signaling mechanisms of mTOR complexes: mTORC1 and mTORC2 in forming a formidable relationship. Adv Biol Regul 72:51–62.

199. Shimobayashi M, Hall MN. 2016. Multiple amino acid sensing inputs to mTORC1. Cell Res 26:7–20.

200. Cotrim PC, Garrity LK, Beverley SM. 1999. Isolation of genes mediating resistance to inhibitors of nucleoside and ergosterol metabolism in Leishmania by overexpression/selection. J Biol Chem 274:37723–37730.

201. Detke S. 1997. Identification of a transcription factor like protein at the TOR locus in Leishmania mexicana amazonensis. Mol Biochem Parasitol 90:505–511.

202. Belli AA, Miles MA, Kelly JM. 1994. A putative Leishmania panamensis/Leishmania braziliensis hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. Parasitology 109:435–442.

203. Gelanew T, Kuhls K, Hurissa Z, Weldegebreal T, Hailu W, Kassahun A, Abebe T, Hailu A, Schönian G. 2010. Inference of population structure of leishmania donovani strains isolated from different ethiopian visceral leishmaniasis endemic areas. PLoS Negl Trop Dis 4.

204. Chargui N, Amro A, Haouas N, Schönian G, Babba H, Schmidt S, Ravel C, Lefebvre M, Bastien P, Chaker E, Aoun K, Zribi M, Kuhls K. 2009. Population structure of Tunisian Leishmania infantum and evidence for the existence of hybrids and gene flow between genetically different populations. Int J Parasitol 39:801–811.

205. Gelanew T, Hailu A, Scho'nian G, Lewis MD, Miles MA, Yeo M. 2014. Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural ethiopian isolates of leishmania donovani. Int J Parasitol 44:751–757.

206. Ravel C, Cortes S, Pratlong F, Morio F, Dedet JP, Campino L. 2006. First report of genetic hybrids between two very divergent Leishmania species: Leishmania infantum and Leishmania major. Int J Parasitol 36:1383–1388.

207. Siriwardana Y, Zhou G, Deepachandi B, Akarawita J, Wickremarathne C, Warnasuriya W, Udagedara C, Ranawaka RR, Kahawita I, Ariyawansa D, Sirimanna G, Chandrawansa PH, Karunaweera ND. 2019. Trends in Recently Emerged Leishmania donovani Induced Cutaneous Leishmaniasis, Sri Lanka, for the First 13 Years. Biomed Res Int 2019.

208. Karunaweera ND, Ginige S, Senanayake S, Silva H, Manamperi N, Samaranayake N, Siriwardana Y, Gamage D, Senerath U, Zhou G. 2020. Spatial epidemiologic trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerg Infect Dis 26:1–10.

209. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.

210. Zhang WW, Matlashewski G. 2001. Characterization of the A2-A2rel gene cluster in Leishmania donovani: Involvement of A2 in visceralization during infection. Mol Microbiol 39:935–948.

211. McCall LI, Matlashewski G. 2012. Involvement of the Leishmania donovani virulence factor A2 in protection against heat and oxidative stress. Exp Parasitol 132:109–115.

212. McCall LI, Matlashewski G. 2010. Localization and induction of the A2 virulence factor in Leishmania: Evidence that A2 is a stress response protein. Mol Microbiol 77:518–530.

213. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467.

214. Pettersson E, Lundeberg J, Ahmadian A. 2009. Generations of sequencing technologies. Genomics 93:105–111.

215. Harrington CT, Lin EI, Olson MT, Eshleman JR. 2013. Fundamentals of pyrosequencing. Arch Pathol Lab Med 137:1296–1303.

216. Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res 11:3–11.

217. Rothberg JM, Leamon JH. 2008. The development and impact of 454 sequencing. Nat Biotechnol 26:1117–1124.

218. Zhao J, F.A. Grant S. 2011. Advances in Whole Genome Sequencing Technology. Curr Pharm Biotechnol 12:293–305.

219. Marine RL, Magaña LC, Castro CJ, Zhao K, Montmayeur AM, Schmidt A, Diez-Valcarce M, Ng TFF, Vinjé J, Burns CC, Nix WA, Rota PA, Oberste MS. 2020. Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. J Virol Methods 280.

220. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Molecular biology: Accurate multiplex polony sequencing of an evolved bacterial genome. Science (80-) 309:1728–1732.

221. Thompson R, Drew CJG, Thomas RH. 2012. Next generation sequencing in the clinical domain: Clinical advantages, practical, and ethical challengesAdvances in Protein Chemistry and Structural Biology.

222. Braun P, LaBaer J. 2003. High throughput protein production for functional proteomics. Trends Biotechnol 21:383–388.

223. Illumina. 2016. NovaSeq 6000 Sequencing System. 770-2016-025-H 4:1–4.

224. Mason CE, Elemento O. 2012. Faster sequencers, larger datasets, new challenges. Genome Biol 13:314.

225. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21.

226. Pereira R, Oliveira J, Sousa M. 2020. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. J Clin Med 9:132.

227. Buck D, Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, Au KF. 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Research 6.

228. Hestand MS, Van Houdt J, Cristofoli F, Vermeesch JR. 2016. Polymerase specific error rates and profiles identified by single molecule sequencing. Mutat Res - Fundam Mol Mech Mutagen 784–785:39–45.

229. Jiao X. 2013. A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. J Data Mining Genomics Proteomics 04.

230. Salmela L, Walve R, Rivals E, Ukkonen E, Sahinalp C. 2017. Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics 33:799–806.

231. Fu S, Wang A, Au KF. 2019. A comparative evaluation of hybrid error correction methods for error-prone long reads. Genome Biol 20.

232. Mikheyev AS, Tin MMY. 2014. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 14:1097–1102.

233. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biol 17.

CHAPTER 2: A COMPLETE LEISHMANIA DONOVANI REFERENCE GENOME IDENTIFIES NOVEL GENETIC VARIATIONS ASSOCIATED WITH VIRULENCE

Patrick Lypaczewski, Johanna Hoshizaki, Wen-Wei Zhang, Laura-Isobel McCall, John Torcivia-Rodriguez, Vahan Simonyan, Amanpreet Kaur, Ken Dewar, Greg Matlashewski

2.1 Preface

As described in **Chapter 1**, the epidemic of cutaneous leishmaniasis in Sri Lanka is caused by an atypical strain of *L. donovani*. Comparative genomics of variant strains can help identify key genetic elements contributing to their phenotype. However, as the reference genome of *L. donovani* is incomplete, comparative genomic analyses fail to provide a complete picture. New sequencing technologies and improvements to existing ones have also become available since the first generation of the reference genome. We therefore investigated the possibility of generating a complete genome for *L. donovani* through the combination of second and third generation technologies. We further used this new reference genome to perform in depth comparative genomics and identify genetic polymorphisms in the cutaneous disease-causing Sri Lankan *L. donovani* isolates. Our results pave the way for future genomic investigations by providing a novel reference genome and a more complete annotation of genetic features.

- Adapted from: **Patrick Lypaczewski**, Johanna Hoshizaki, Wen-Wei Zhang, Laura-Isobel McCall, John Torcivia, Vahan Simonyan, Amanpreet Kaur, Ken Dewar, Greg Matlashewski, 2018. A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. *Scientific reports*, *8*(1), p.16549.

2.2 Abstract

Leishmania donovani is responsible for visceral leishmaniasis, a neglected and lethal parasitic disease with limited treatment options and no vaccine. The study of *L. donovani* has been hindered by the lack of a high-quality reference genome and this can impact experimental outcomes including the identification of virulence genes, drug targets and vaccine development. We therefore generated a complete genome assembly by deep sequencing using a combination of second generation (Illumina) and third generation (PacBio) sequencing technologies. Compared to the current *L. donovani* assembly, the genome assembly reported within resulted in the closure over 2,000 gaps, the extension of several chromosomes up to telomeric repeats and the re-annotation of close to 15% of protein coding genes and the annotation of hundreds of non-coding RNA genes. It was possible to correctly assemble the highly repetitive A2 and Amastin virulence gene clusters. A comparative sequence analysis using the improved reference genome confirmed 70 published and identified 15 novel genomic differences between closely related visceral and atypical cutaneous disease-causing *L. donovani* strains providing a more complete map of genes associated with virulence and visceral organ tropism. Bioinformatic tools including protein variation effect analyzer and basic local alignment search tool were used to prioritize a list of potential virulence genes based on mutation severity, gene conservation and function. This complete genome assembly and novel information on virulence factors will support the identification of new drug targets and the development of a vaccine for *L. donovani.*

2.3 Introduction

Visceral Leishmaniasis (VL) is the second most lethal parasitic disease following malaria and is prevalent throughout underdeveloped and tropical regions of the world. There are some 300,000 new cases each year (1) and *Leishmania donovani,* transmitted by the infected sand fly,

is the major causative agent of VL in the Indian and African continents. Although *L. donovani* is extensively studied, its genome remains poorly annotated because it is heavily fragmented and a complete assembly is crucial to understanding this parasite's biology, metabolic pathways, tissue tropism and disease pathology.

The pathology of leishmaniasis is predominantly parasite species-specific, such as for example *L. major* that causes cutaneous leishmaniasis (CL) whereas *L. donovani* typically causes lethal visceral leishmaniasis (VL). Previous studies have compared genomes of *L. major* and *L. donovani* parasites to study virulence and disease tropism and have identified a number of species specific genes including A2 present in *L. donovani* that is a pseudogene in *L. major* (2, 3). More recently, as cases of atypical CL caused by *L. donovani* have emerged, studies have compared cutaneous and visceral disease-causing strains of *L. donovani,* as these strains provide a unique opportunity to study the genetic determinants of disease pathogenesis using more recently diverged strains (4).

Second-generation sequencing technologies including Illumina, has made the sequencing of large genomes feasible through the mapping of short sequence reads of 50 to 250 nucleotides (nt) to a reference genome (5). While human and many other well studied higher vertebrates have better assembled reference genomes (6), the kinetoplastids suffer in this regard because most *Leishmania* species either lack sequencing information altogether or have incomplete reference genomes with sometimes thousands of sequence gaps. (7). The current *L. donovani* reference genome (ASM22713v2 from strain BPK282A (8)) was generated using second generation technologies and contains over 2,000 gaps and therefore there are many incomplete or inaccurate protein coding sequences. The first complete *Leishmania* genome generated is that of *L. major* by a consortium of laboratories employing large insert clone tiling paths to sequence each

chromosome individually (9, 10). This genome was later improved by the reassembly of complex collapsed loci that were incorrect in the original reference genome (11).

Since then however, advances in sequencing technologies have drastically reduced the cost of sequencing and eased genome assembly tasks by increasing the length of the generated sequences. Long read sequencing or "third-generation" sequencing refers to more recent technologies including Oxford NanoPore (12) and Pacific Biosciences (PacBio) (13) that can result in reads ranging up to 50kb or 100kb that are capable of generating more complete genomic assemblies, provided the read lengths traverse across repetitive elements. One such highly repetitive cluster is the A2 gene family from *L. donovani* considered to be an important virulence factor and is necessary for survival in visceral organs (14, 15) and protection against host response stress (16, 17). Due to its repetitive nature, the A2 gene cluster is misassembled in all *Leishmania* genomes generated using second generation sequencing, and only resolved in a recent resequencing effort targeted to *L. infantum* exploiting the long-read capabilities of PacBio sequencing which resulted in a complete genome assembly (18). The current *L. donovani* genome however was obtained from second generation sequencing and consequently, no precise DNA or complete protein sequences are available for any A2 protein in *L. donovani*, hindering the comparison of A2 genes in visceral disease-causing strains or using mass spectrometry to identify A2 proteins which relies on accurate genome sequences for protein identification.

In this study, we have combined second and third generation sequencing to generate a complete assembly of the *L. donovani* genome from the strain responsible for cutaneous leishmaniasis (CL) in Sri Lanka (4, 19). This new assembly enabled the generation of an improved genome annotation and an unbiased analysis of chromosome synteny comparing *L. donovani* and *L. major* genes and strand switch transcription units. We have used this complete assembly to re-

interrogate the genetic makeup of the visceral and cutaneous disease-causing *L. donovani* strains resulting in the identification of novel SNPs and indels generating a more complete and accurate chromosome map of the genetic differences between these phenotypically distinct *L. donovani* strains (4, 20). This study further enabled re-annotation of much of the genome highlighting the importance of a complete reference assembly to support future functional genomic and proteomic studies involving the *L. donovani* pathogen.

2.4 Results

2.4.1 A complete *L. donovani* genome assembly

The currently available assembly for *L. donovani* (ASM22713v2 from strain BPK282A (8)) contains over 2,000 gaps due to the presence of low complexity regions and the highly repetitive nature of the *Leishmania* genome (21). This incomplete assembly makes it difficult to compare *L. donovani* genomes from strains with different phenotypic properties. DNA was therefore isolated from the attenuated cutaneous disease-causing strain of *L. donovani* from Sri Lanka (4) and was subjected to deep sequencing using second and third generation sequencing. We reasoned that a complete assembly of the genome from this attenuated *L. donovani* strain will identify a more complete complement of genetic changes associated with loss of virulence of this strain. A total of 9 PacBio sequencing runs were performed generating 712,443 reads representing an estimated 107-fold coverage of the estimated 35 Mb genome. Importantly, there were 51,484 reads longer than 12 kb, representing a 20-fold coverage in very long reads. The long-read sequencing data was assembled using various assemblers as described in methods and merged using the longest chromosomes produced by each assembler followed by refinement using the high-quality short-read Illumina-generated data and iterative edge extension to close the remaining gaps.

The previous *L. donovani* reference assembly (ASM22713v2 from strain BPK282A) had over 2000 gaps spread across the 36 chromosomes. **Figure 2.1** depicts the location on each chromosome of the gaps that have been closed in the new assembly reported here. The new assembly now contains contiguous DNA sequences in all 36 chromosomes and a corresponding 22-fold increase in N50 indicating that a larger proportion of the data has been assembled into large contigs as 50% of the genome is contained in contigs >= N50, resulting in an N50 of over 1Mbp (**Table 2.1**). Further, using this completed assembly, we have generated annotations for more potential protein coding regions than previously annotated (8,633 compared to 7,969 proteins) and identified more transfer-RNA and ribosomal RNA genes as well as all 6 small nuclear RNA genes, all spliced leader RNA genes and close to a thousand small nucleolar RNA genes. An additional 13 genes were marked as pseudogenes due to multiple internal stop codons and/or frameshifts. (**Supplementary Table 2.S1**). Alignment of the second-generation Illumina reads to the PacBio generated assembly was used to cross-validate and correct the assembly at the nucleotide level. Graphs of the coverage from the alignment of Illumina and PacBio data across the 36 chromosomes are available in **Supplementary Figure 2.S1.** Taken together, we consider this new assembly to be contiguous and complete.

2.4.2 Assembly of the A2 virulence gene cluster and synteny comparison between *L. major* and *L. donovani*

A2 is a major virulence factor required for *L. donovani* survival in visceral organs (22). The A2 gene family cluster on chromosome 22 has recently been assembled for *L. infantum* (18), however has not been for *L. donovani.* We therefore investigated whether the structure of this region could be determined with this revised assembly. It was advantageous that the attenuated cutaneous *L. donovani* strain used in this assembly has fewer copies of the A2 genes than other

virulent strains of *L. donovani* (4). As shown in **Figure 2.2a**, the new assembly could read-through the entire cluster of highly repetitive A2 and flanking sequences and could position the A2 genes and interspersed flanking 3' A2-rel and 5' A2-rel genes. The A2 genes are contained in two opposite facing clusters on either side of a strand-switch locus consisting of one cluster of 3 copies and one cluster with a single A2 gene. The long sequence reads generated by the PacBio sequencing were crucial in generating the assembly of the A2 genes where reads of 11 kb and longer are shown spanning the repetitive cluster (**Figure 2.2b**).

To generate supporting evidence for this A2 gene assembly, Western blot analysis of the A2 proteins from this strain was performed to compare the number and sizes of the A2 proteins with the predicted molecular weights from this assembly (ORFs; **Supplementary Figure 2.S2**). As shown in **Figure 2.2c**, the apparent molecular weights from Western blotting correspond to the sizes predicted from the sequenced ORFs. The 3 bands on the Western blot are consistent with the molecular weights of the 4 gene products as the A2 gene copies 2 and 3 encode proteins of a similar size (**Supplementary Figure 2.S2**). This represents the first complete structure and sequence for A2 genes in *L. donovani*, a prototype virulence factor. The difficulty in assembling this complex region is demonstrated in **Figure 2.2d**, where a deviation from the average read coverage can be seen around the 300,000bp position, in and around the A2 cluster, due to difficulties in the aligner assigning a unique position to similar reads across a repetitive region.

Directly comparing synteny at the chromosomal level was not possible with the previous *L. donovani* assembly due to the heavy fragmentation of the genome. With the new *L. donovani* assembly, it was possible to accurately compare chromosome synteny between *L. donovani* and *L. major*. As shown in **Figure 2.3**, the genome of *L. donovani,* exhibited a very high level of synteny with the *L. major*. Chromosome 22 was highlighted here because this is the location of the A2

genes that have become pseudogenes in *L. major* and have therefore diverged between these old-world species (15). The level of synteny demonstrated here for chromosome 22 was maintained on all other chromosomes (**Supplementary Figure 2.S3**). These results indicate that evolution between cutaneous and visceral pathologies by different *Leishmania* species resulted largely from SNPs, pseudogenes and copy number variation and not from large changes such as chromosome rearrangements or complete gene deletions/insertions.

2.4.3 Identification of new genes and improvements in annotations

As this assembly was larger in terms of total number of bases covered and more contiguous due to the removal of sequence gaps, the impact this had on gene annotations was investigated. The genome from the new assembly was annotated using the Companion pipeline (23) and the new and previous annotations (GenBank: GCF_000227135.1) were then aligned together and overlapping annotations were removed. Remarkably, close to 15% of the *L. donovani* protein coding genes had new or edited annotations as shown in **Figure 2.4a**. Part of this increase in number of annotations resulted from the expansions of multi-copy gene families beyond the copy numbers in the previous annotation. An example is shown in **Figure 2.4b** where there are 10 amastin genes identified in this new assembly compared to the previously identified 2. These results support the use of this assembly as the reference for bioinformatic analysis as it provides a more complete and accurate annotation of the *L. donovani* genome ORFs.

2.4.4 Comparison of virulent and attenuated *L. donovani* parasites

As indicated above, there are 2 distinct strains of *L. donovani* in Sri Lanka where one is responsible for visceral leishmaniasis (VL) and the other for cutaneous leishmaniasis (CL) (4). Subsequently, the CL strain was experimentally passaged through the visceral organs of BALB/c mice to select for a gain-of-function strain with increased virulence (IV strain) for survival in

visceral organs where it was revealed through proteomic analysis that the resulting IV strain had an increase in stress response and antioxidant proteins (24). Illumina whole genome sequencing and comparative genomic analysis of the VL, CL and IV strains was performed to identify SNPs associated with change in virulence for survival in the visceral organs. As shown in **Figure 2.5**, all 70 of the previously identified homozygous SNP differences between the VL and CL strains (4) were confirmed in this new assembly and an additional 15 novel SNPs within protein coding genes were found using this complete assembly. In addition, there were 12 mutations associated with the IV strain with gain-of-function for increased survival in visceral organs not labeled in **Figure 2.5**; four were heterozygous but with frequencies changing towards the VL genotype (IV→VL), four were heterozygous but present only in the IV strain and four were homozygous deletions in the IV strain. The newly identified differences between the VL and CL strains and the ones contributed from the IV strain are summarized in **Table 2.2**.

Combining data from the previous and current analysis, all the genes with genetic differences were organized into priority clusters based on the likelihood to affect protein function and phenotype (**Figure 2.6**). A detailed list of the genes and cluster assignments is shown in **Table 2.3**. From the 66 genes containing 70 SNPs previously identified, 7 were previously experimentally assessed using gene replacement with a wildtype copy for virulence in visceral organs (4) and one gene was identified as a misannotation and was therefore removed from the list. In decreasing order of priority, 13 genes in the highest impact cluster (red) were characterized as potentially having a major effect on protein function due to large amino acid changes or co-occurrence of mutations in both the VL and IV strains. SNPs in common between the IV and VL strains (4 IV→VL in the red cluster) indicate a selection associated with survival in visceral organs during the experimental passaging of the CL strain in mouse visceral organ (24). Nine genes either

with multiple co-occurring SNPs or non-conservative amino acid changes in conserved domains with a high score as assessed by PROVEAN software were placed in the second highest priority cluster. As detailed in methods, PROVEAN is a bioinformatic tool that classifies the significance of specific genetic mutations with respect to protein function (25). Eighteen genes with non-conservative amino acid changes occurring in conserved domains but scored as unlikely to have a large effect on protein function by PROVEAN were placed in Cluster 3. Twenty-one genes with conservative amino acid changes in conserved domains were divided between Cluster 4 and Cluster 5 based on PROVEAN scores and 14 genes with mutations in domains with higher variability were placed in Cluster 6. Four mutations seen solely in the IV strain but not in the VL or CL strains are likely to be the result of random mutation or adaptation specific to the murine host were placed in Cluster 7.

A 25kb region on chromosome 36 containing 4 genes was found to be missing in the IV strain but present in the VL and CL strains. This deletion did not occur in a location previously identified on this chromosome where a fission can occur as seen in *L. alderi* (26). Upon experimental verification, this deleted region was present in a subpopulation of the parental CL strain (**Supplementary Figure 2.S4**). The enrichment of this deletion in the IV strain could therefore be a consequence of selection in the mouse and likely to be unrelated to human visceral disease because this region is present in wild type or VL strains of *L. donovani* as well as *L. major* and therefore classified in cluster 7.

The classification of genetic differences in the CL, VL and IV genomes summarized in **Figure 2.6** and **Table 2.3** represents a prioritization of genes to be empirically investigated for controlling the different phenotypes of these virulent and attenuated strains.

2.5 Discussion

It has been possible to generate a complete genome assembly for *L. donovani* through combining second and third generation sequencing technologies, similarly to a recent resequencing of the *L. infantum* genome resulting in a complete assembly, highlighting the usefulness of PacBio sequencing in regards to *Leishmania* genomes (18). This resulted in a more accurate annotation of the genome increasing the number of potential protein coding genes and identifying novel mutations/polymorphisms associated with virulence. It was remarkable that the present assembly resulted in annotation changes in close to 15% of the genome representing 1087 protein coding genes. Although 13 degenerate pseudogenes are identified in **Supplementary Table 2.S1** more do exist since our annotations derived from functional genes in *L. major* and therefore genes functional in other species were not identified. Through this updated genome annotation, additional SNPs have been identified including in genes potentially involved in visceral disease and several non-coding genes have been annotated allowing future *L. donovani* research beyond protein coding genes. It has also been possible to assemble known virulence factor gene families in *L. donovani* including the A2 and Amastin gene families. This version of the *L. donovani* genome assembly will significantly improve genomic, functional genomic and proteomic research outcomes and support the identification of drug targets and the development of vaccines. This assembly further provides a larger repertoire of target DNA sequences to identify diagnostic and prognostic disease progression markers. Given the recent interest in generating genetically modified live attenuated parasites as vaccine candidates (27), a complete genome assembly will permit the verification that genetic modifications target intended genes with no off target mutations.

Supported by a *de novo* assembly, this study provides the first direct evidence for synteny between chromosomes in *L. donovani* and *L. major,* two old world parasite species with different pathologies and reservoirs. Previously, due to the large number of gaps in the *L. donovani* genome, the segments were aligned to a reference assembly assumed to be syntenic and only gene synteny was possible. In contrast, the contiguous assembly presented within used an entirely reference-free and by extension, bias-free generation process. This assembly can be used in future sequencing efforts aimed at comparing genes and synteny of genomes of other *Leishmania* species with *L. donovani*. The strong gene level synteny further highlights the major phenotypic effects of SNPs and indel mutations when comparing genomes from *L. donovani* strains causing visceral and cutaneous pathologies. As no major chromosomal rearrangements or deletions are apparent between phenotypically different *Leishmania* species as previously reported (2, 4), and including this study, suggesting that virulence and tropism can be acquired or lost through relatively small coding changes at the amino acid level such as SNPs, indels and frameshifts without the need for chromosomal scale events.

This study reports the complete A2 gene continuous sequence and an assembly of entire A2 cluster including the A2-rel flanking genes in *L. donovani*. While the organization of this A2 gene cluster was previously theorized based on available sequences and Southern blot analysis, no sequencing technology could accurately read-through an entire cluster (4, 28) prior to third-generation sequencers. Interestingly, a similar organization or A2 and A2-rel flanking genes was obtained during the resequencing and assembly of the *L. infantum* genome (18), further supporting the genomic arrangement of this important virulence cluster. The present assembly contains entire A2 ORFs that are consistent with the corresponding protein sizes determined by Western blot analysis and provides novel insight into this elusive virulence factor through identifying 2 amino

acid insertions between the 10 amino acid repeats at geometric intervals as well as a defined C-terminal sequence (**Supplementary Figure 2.S2**). The deviations from the 10-amino acid repeat sequence could contribute to the proper folding and function of the A2 protein.

In an attempt to identify additional genes associated with survival in visceral organs, the attenuated cutaneous *L. donovani* strain was experimentally passaged continuously through the visceral organs of BALB/c mice over an 8 month period to generate a gain-of-function strain with increased survival in the liver and spleen and was termed the IV strain (24). Sequence analysis of the IV strain in this study did not identify homozygous SNP differences with the parental cutaneous strain but did identify four heterozygous SNPs with the same sequence in the virulent *L. donovani* strains, classified as high impact in **Figure 2.6**. The corresponding SNP-containing genes with unknown function are of high priority for future studies. Nevertheless, although the gain of function IV strain had significantly increased survival in visceral organs (24), it was surprising that this strain did not have more genetic differences associated with the increased virulence. It is possible that the selection process for survival in the visceral organs of mice is different from that in humans.

The Illumina sequence analysis of the cutaneous (CL) and visceral (VL) disease associated strains using the complete assembly identified 15 novel homozygous SNPs beyond the previously identified 70 SNPs (4) (**Figure 2.5**). One of these new SNPs was in the raptor gene that is part of the highly conserved Target Of Rapamycin (TOR) signaling pathway (29). There are three TOR gene homologs in the *Leishmania* genome (30) revealing this pathway is conserved in kinetoplastids. Interestingly, the RagC GTPase which is a binding partner of raptor in the TOR pathway is also mutated in the attenuated cutaneous *L. donovani* strain and restoration of the wildtype RagC GTPase increased virulence in visceral organs (4). Considering that there are two

mutated genes (RagC and Raptor) in the TOR pathway in the attenuated cutaneous *L. donovani* strain strongly highlights this pathway as playing a role in determining disease tropism and virulence.

As both HIVE and VarScan were used to identify SNPs and indels, we are confident that the expanded list of 83 variable genes shown in **Figure 2.6** contains most if not all the genes associated with visceral disease, with the exception of UTR mutations that may influence protein expression levels. Since this number of genes is relatively small, we are currently investigating all genes in clusters 1-4 with respect to their involvement in visceral organ virulence using CRISPR-Cas9 gene editing recently developed for use in *Leishmania* (31, 32). It is noteworthy that the correct selection of gRNA sequences for CRISPR-Cas9 gene editing requires a complete genome and accurate annotations for precise gene editing with no off-target mutations that is now possible with the complete assemble reported here.

2.6 Methods

2.6.1 Whole Genome sequencing

2.6.1.1 DNA Extraction

Leishmania DNA for both Illumina and PacBio sequencing was derived from the attenuated cutaneous strain of *L. donovani* from Sri Lanka (4) that was passaged through mice to increase survival in visceral organs (24). DNA was extracted following the previously described phenol-chloroform methods for isolation of Trypanosomatid genomic material. (33)

2.6.1.2 Illumina

Sequencing library preparation (Kapa HTP) and 250 nt paired-end sequencing (Illumina MiSeq) was performed using manufacturers' protocols.

2.6.1.3 PacBio Sequencing

A total of 9 sequencing cells were prepared. 7 cells were prepared using the DNA Template kit v2.0 (3kb-10kb) with DNA/Polymerase Binding Kit P4 and 2 using the DNA Template Prep Kit 3.0 with DNA/Polymerase Binding Kit P5. The cells were sequenced on a PacBio II RS instrument with BaseCaller v1 protocol.

2.6.2 Genome Assembly

2.6.2.1 HGAP Assembler

Raw reads from the 9 sequencing cells were loaded into the SMRT Analysis portal (Pacific Biosciences) in HD5 format. The Hierarchical Genome Assembly Process (HGAP) version 2 with Quiver polishing was chosen as version 3 is stated to improve speed at the detriment of assembly quality. Expected genome size was set to 36Mbp, minimum read length for pre-assembly was set to 500bp and minimum read length for full assembly was set to 100bp. Minimum Polymerase Read quality was set to 0.80, and the remainder of options remained at default settings.

2.6.2.2 Celera Assembler

The PacBio corrected Reads (PBcR) module of the celera-assembler version 8.3 was used to assemble the long reads data (34). The subreads were first extracted from the PacBio H5 files to FASTQ using bash5tools.py. The Bogart unitigger was used by specifying the "unitigger=bogart" option in the spec file. The consensus caller module was PBDAGCON. Due to the sequences originating from a non-clonal sample and the use of the DNA/Polymerase Binding kit P4 in some PacBio sequencing cells which produces lower quality data than P5 kits, error rate limits were relaxed for various variables, listed in the full spec file available in supplementary information (**Supplementary Methods 2.S1**).

2.6.2.3 Canu Assembler

The Canu v1.0 assembler is a modified version of the Celera Assembler designed to handle high noise data such as NanoPore and PacBio sequencing data. Canu has both the ability to assemble raw PacBio data by performing error correction using consensus sequence or assemble data in a hybrid mode where PacBio reads are pre-error corrected using short read Illumina data. In Raw mode, the trimmed PacBio reads were given to the assembler using default settings except for the expected genome size option which was set to 35Mbp using the option "genomeSize=35m". In hybrid mode, the Illumina reads were first error corrected by internal consensus using Pollux (35), the paired end reads were then merged together to form longer sequences with a high confidence core region using FLASh (36), and used to correct the PacBio reads using Proovread (37). The error-corrected PacBio data was then used by Canu to generate a draft assembly.

2.6.2.4 Pilon

The Pilon error correcting software was used to fix small errors present in the PacBio based assemblies using high depth and high accuracy Illumina data (38). The entire Miseq dataset in FASTQ format from the corresponding sample was aligned to the draft assembly using the Burrows-Wheeler Aligner (bwa) to generate SAM alignment files. Samtools was then used to convert and sort the files to a binary usable format as described in the Samtools section. This alignment was then passed to the Java Pilon executable for correction of small indels, SNPs, gap filling and assembly of unmapped reads using the command "java -Jar pilon.jar --genome [new-assembly.fasta] –frags [alignment.bam] –fix all,novel"

2.6.2.5 GMCloser

GMcloser was used to merge the assemblies generated using the different assemblers (39). Short read Illumina data was aligned to the contigs resulting from the different contigs from

different assemblers with identical reads mapped to them were assumed to be part of the same chromosome. When a contig from one assembly encompassed a gap present in another assembly, the gap was filled with the missing information to generate a merged assembly with the least number of gaps. All the alignment and merging steps are handled internally to GMcloser using the command "gmcloser -t [assembly1.fasta] -q [assembly2.fasta] -r miseq_R1.fastq miseq_r2.fastq -et "

2.6.2.6 IGV

The Broad Institute Integrative Genome Viewer (40, 41) was used to perform quality control on assemblies and manually inspect fragments in order to close gaps. The Pilon tools was used with the "-fix novel" option which assembled short contigs from unmapped data. The fragments were then placed on the appropriate likely chromosomes based on gene annotations and submitted to another round of gap filling using Pilon and GMCloser to find reads supporting this placement or were removed if no reads supported the join.

2.6.3 Annotations

2.6.3.1 Companion

The Companion webtool (https://companion.sanger.ac.uk/) was used to annotate genes on the assembly contigs and refine the assembly (23). The closest available reference organism was chosen (*L. major)* with the following options: contiguate pseudochromosomes, align reference proteins to target sequence, perform pseudogene detection, use RATT Species transfer type, and the *L. donovani* taxon ID. Additional *L. donovani* and *L. major* genes not automatically transferred were manually verified and appended if necessary. An additional 3 genes were manually added from a search of all ribosomal protein transcripts in trypanosomes. The snRNAs U1 through U6, ribosomal RNAs and the spliced leader RNA were manually annotated as necessary from the

sequences available for *L. major* on TriTrypDB (42). Sequences for H/ACA and C/D box snoRNA were manually mapped using published *L. major* snoRNA research (43).

2.6.3.2 Galaxy

The Galaxy webtool (https://usegalaxy.org/) (44) was used to perform file conversions and data extraction such as moving a chromosome's FASTA sequence from one assembly to another.

2.6.3.3 Identification of new genes

Genomic annotations from the Companion Pipeline were downloaded in General Feature Format (GFF) and gene annotations were extracted using the Galaxy tool "Extract features" set to look for the "CDS" keyword in column #3 of the GFF file. Known coding regions from the reference *L. donovani* strain BPK282A1, assembly ASM22713v2 were downloaded from GenBank and aligned to our improved assembly in BED format. Bedtools intersect intervals through Galaxy (45, 46) was used to identify annotations that were unique to our annotations or were not at least 95% covered previously using settings " -wa -f 0.95 -v -r"

2.6.3.4 Synteny

The online SynMap2 software (47) was used to generate the synteny dotplot across the entire genome using annotations from *L. major* and the annotations generated by Companion in this study. The chromosome to chromosome circular charts were generated by Companion as part of the annotation process.

2.6.4 Comparison of visceral (VL), cutaneous (CL) and increased virulence (IV) *L. donovani* strains

2.6.4.1 BWA

The Burrows-Wheeler Aligner (BWA) was used to process the FASTQ Illumina sequencing files obtained from Genome Quebec. The maximal exact match algorithm was used in paired-end mode using the command "bwa mem" and providing the matched pair read files and reference sequence as arguments in order to generate a SAM format alignment file of the reads on the reference (48).

2.6.4.2 Samtools

The samtools package was used for file manipulations and conversions (49). The commands "samtools view -b" was used to convert the BWA generated SAM file to the binary alignment BAM format. The file was then sorted by alignment location for compatibility with downstream analysis software using "samtools sort -@ 30 -o [output.file]". The alignment files where then prepared for analysis using the mpileup modules which tabulates the base distribution at every position using the command "samtools mpileup -B -f [reference assembly] [strain specific position sorted BAM file] > [output.file]"

2.6.4.3 VarScan

The VarScan v2 (50) mutation caller was used to generate a list of mutations in Variant Call Format (VCF) using the mpileup file generated by samtools as described above using the command "java -jar VarScan.jar mpileup2snp –output-vcf 1 [mpileup.file] > [output.VCF]". We also used VarScan to generate indel locations based on the same mpileup file using the command "java -jar VarScan.jar mpileup2indel –output-vcf 1 [mpileup.file] > [output.vcf]".

2.6.4.4 SnpEff

To filter the VCF files generated by VarScan to a list of non-synonymous SNP, we used the SnpEff software (51). The oriented and annotated assembly was downloaded from the Companion tool as described above along with the gene annotation file in GFF format containing the names, locations and amino acid sequences of identified genes. This GFF file was used to build a SnpEff database using the SnpEff.jar command "build" with argument "-gff" after installing the genome and GFF file in the appropriate locations according to the software instructions.

The SnpEff software was then used to annotate the 10^{th} column of the VCF file with mutation effect codes. All the mutations were then examined manually for accuracy using the Integrative Genomics Viewer (IGV) with all raw Illumina data loaded.

2.6.4.5 Classification

Non-synonymous mutations were clustered according to the mutation effect in order to prioritize further gene function studies. Each cluster was further broken down based on the mutation's PROVEAN score (25). The PROVEAN software was designed to predict the magnitude of a mutation's impact on protein function. To generate PROVEAN scores, we retrieved homologous sequences from other *Leishmania* species and kinetoplasts and generated a multiple sequence alignment (MSA). The MSA was then passed to the PROVEAN software which scored each SNP based on the alignment. We used PROVEAN scores below a threshold of -2.5 as an indication a SNP is likely to affect protein function. Cluster assignments were as follows:

1. Mutations likely to have the largest impact on protein function were included; non-sense, frameshift and amino acid insertion/deletions as well as all SNPs in the gain-

of-function IV strain that were the same in the virulent visceral strain (VL) allele, indicating a selection pressure on those genes for visceral organ survival.

2. Genes in with multiple SNPs and genes where non-conservative mutations occurred in highly conserved *Leishmania/Kinetoplastida* regions.
3. Due to the high number of genes in cluster 2, split off poor PROVEAN scoring genes.
4. Cluster 4 comprised genes with conservative amino acid changes but occurring in *Leishmania/Kinetoplastida* conserved regions
5. Due to the high number of genes in cluster 4, split off poor PROVEAN scoring genes.
6. Conservative and non-conservative amino acid changes in less conserved regions.
7. Changes present only in the gain of function IV strain. This cluster was considered low probability as it likely contains either random mutations or adaptations specific to survival in the murine host.

2.6.5 HIVE

HIVE (52) was used to perform differential profiling of genomes from visceral (VL), gain of function increased virulence (IV), and (cutaneous) CL strains.

1. Reads from all the samples were aligned to the assembly of the genome using HIVE-hexagon (53) parametrized for parasitic eukaryotic species and specifically adjusted to work with *Leishmania* analysis as demonstrated in previous studies (27)
2. Coverage and variant calling analysis was performed using HIVE-heptagon (54) to produce variant call frequencies and coverages for every genomic position.

3. HIVE differential profiler (52) was used to analyze relative differences in SNP calls and variant coverages for multiple samples.

2.6.6 A2 Immunoblotting

A2 Immunoblotting was performed as previously described (16). Briefly, $1x10^7$ cutaneous CL strain promastigotes were collected at mid log-phase and resuspended in 1mL fresh medium. The cells were then heat-shocked for 4h at 40°C to induce A2 protein expression, washed, lysed in SDS-PAGE loading buffer and loaded on a 10% (w/v) acrylamide gel. The proteins were transferred to nitrocellulose at 25V overnight at 4°C. The membrane was blocked for 1h in 10% (w/v) skim milk powder dissolved in PBS with 0.1% (v/v) Tween-20. The membrane was then incubated for 1h at RT with a 1:10,000 dilution of C9 Ascites fluid (anti-A2 Mab) in blocking solution followed by 6x 5min washes in PBS-T. Secondary HRP labeled anti-mouse IgG antibody (Thermo Fisher Scientific) was incubated at 1:10,000 in blocking buffer for 1h at RT followed by 6x 5min washes in PBS-T. The membrane was incubated in ECL reagent (Zm Tech) for 1min at RT before being exposed to x-ray film (Denville Scientific). Film images were captured using a Gel-Doc XR documentation system with Quantity One software (BioRad Laboratories).

2.7 References

1. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, de Boer M. 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7:e35671.
2. 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 LRO, Barrell B, Cruz AK, Mottram JC, Smith DF, Berriman M. 2007. Comparative genomic analysis of three Leishmania species that cause diverse human disease. Nat Genet 39:839–847.
3. Zhang WW, Matlashewski G. 2010. Screening Leishmania donovani-specific genes required for visceral infection. Mol Microbiol 77:505–517.

4. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.

5. Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351.

6. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, Murphy TD, Pruitt KD, Thibaud-Nissen F, Albracht D, Fulton RS, Kremitzki M, Magrini V, Markovic C, McGrath S, Steinberg KM, Auger K, Chow W, Collins J, Harden G, Hubbard T, Pelan S, Simpson JT, Threadgold G, Torrance J, Wood JM, Clarke L, Koren S, Boitano M, Peluso P, Li H, Chin CS, Phillippy AM, Durbin R, Wilson RK, Flicek P, Eichler EE, Church DM. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res 27:849–864.

7. Grisard EC, Teixeira SMR, de Almeida LGP, Stoco PH, Gerber AL, Talavera-López C, Lima OC, Andersson B, de Vasconcelos ATR. 2014. Trypanosoma cruzi Clone Dm28c Draft Genome Sequence. Genome Announc 2:2–3.

8. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, De Doncker S, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Sch??nian G, Stark O, Sundar S, Vanaerschot M, Hertz-Fowler C, Dujardin JC, Berriman M. 2011. Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res 21:2143–2156.

9. 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, Bianchettin G, Borzym K, Bothe G, Bruschi C V., Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, 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, Müller-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, Schäfer 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. 2005. The genome of the kinetoplastid parasite, Leishmania major. Science (80-) 309:436–442.

10. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin JC, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. Genome Res 21:2129–2142.

11. Alonso G, Rastrojo A, López-Pérez S, Requena JM, Aguado B. 2016. Resequencing and assembly of seven complex loci to improve the Leishmania major (Friedlin strain) reference

genome. Parasites and Vectors 9.

12. Mikheyev AS, Tin MMY. 2014. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 14:1097–1102.

13. Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma 13:278–289.

14. Zhang W-W, Matlashewski G. 1997. Loss of virulence in Leishmania donovani deficient in an amastigote-specific protein, A2. Proc Natl Acad Sci 94:8807–8811.

15. Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL, Matlashewski G. 2003. Comparison of the A2 gene locus in Leishmania donovani and Leishmania major and its control over cutaneous infection. J Biol Chem 278:35508–35515.

16. McCall LI, Matlashewski G. 2010. Localization and induction of the A2 virulence factor in Leishmania: Evidence that A2 is a stress response protein. Mol Microbiol 77:518–530.

17. McCall LI, Matlashewski G. 2012. Involvement of the Leishmania donovani virulence factor A2 in protection against heat and oxidative stress. Exp Parasitol 132:109–115.

18. González-De La Fuente S, Peiró-Pastor R, Rastrojo A, Moreno J, Carrasco-Ramiro F, Requena JM, Aguado B. 2017. Resequencing of the Leishmania infantum (strain JPCM5) genome and de novo assembly into 36 contigs. Sci Rep 7.

19. Karunaweera ND, Pratlong F, Siriwardane HVYD, Ihalamulla RL, Dedet JP. 2003. Sri Lankan cutaneous leishmaniasis is caused by Leishmania donovani zymodeme MON-37. Trans R Soc Trop Med Hyg 97:380–381.

20. Ranasinghe S, Zhang W-W, Wickremasinghe R, Abeygunasekera P, Chandrasekharan V, Athauda S, Mendis S, Hulangamuwa S, Matlashewski G, Pratlong F. 2012. *Leishmania donovani* zymodeme MON-37 isolated from an autochthonous visceral leishmaniasis patient in Sri Lanka. Pathog Glob Health 106:421–424.

21. Singh N, Chikara S, Sundar S. 2013. SOLiD^{TM} Sequencing of Genomes of Clinical Isolates of Leishmania donovani from India Confirm Leptomonas Co-Infection and Raise Some Key Questions. PLoS One 8:e55738.

22. McCall LI, Zhang WW, Matlashewski G. 2013. Determinants for the Development of Visceral Leishmaniasis Disease. PLoS Pathog.

23. Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto TD. 2016. Companion: a web server for annotation and analysis of parasite genomes. Nucleic Acids Res 44:W29–W34.

24. McCall LI, Zhang WW, Dejgaard K, Atayde VD, Mazur A, Ranasinghe S, Liu J, Olivier M, Nilsson T, Matlashewski G. 2015. Adaptation of leishmania donovani to cutaneous and visceral environments: In vivo selection and proteomic analysis. J Proteome Res 14:1033–1059.

25. Choi Y, Chan AP. 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747.

26. Coughlan S, Mulhair P, Sanders M, Schonian G, Cotton JA, Downing T. 2017. The genome of Leishmania adleri from a mammalian host highlights chromosome fission in Sauroleishmania. Sci Rep 7.

27. Gannavaram S, Torcivia J, Gasparyan L, Kaul A, Ismail N, Simonyan V, Nakhasi HL. 2017. Whole genome sequencing of live attenuated Leishmania donovani parasites reveals novel biomarkers of attenuation and enables product characterization. Sci Rep 7.

28. Zhang WW, Matlashewski G. 2001. Characterization of the A2-A2rel gene cluster in Leishmania donovani: Involvement of A2 in visceralization during infection. Mol Microbiol 39:935–948.

29. Sancak Y, Peterson TR, Shaul YD, Lindquist RA, Thoreen CC, Bar-Peled L, Sabatini DM. 2008. The rag GTPases bind raptor and mediate amino acid signaling to mTORC1. Science (80-) 320:1496–1501.

30. Madeira da Silva L, Beverley SM. 2010. Expansion of the target of rapamycin (TOR) kinase family and function in Leishmania shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. Proc Natl Acad Sci 107:11965–11970.

31. Zhang WW, Matlashewski G. 2015. CRISPR-Cas9-mediated genome editing in Leishmania donovani. MBio 6:e00861-15.

32. Zhang W-W, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. mSphere 2:1–15.

33. Medina-Acosta E, Cross GAM. 1993. Rapid isolation of DNA from trypanosomatid protozoa using a simple “mini-prep” procedure. Mol Biochem Parasitol 59:327–329.

34. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630.

35. Marinier E, Brown DG, McConkey BJ. 2015. Pollux: Platform independent error correction of single and mixed genomes. BMC Bioinformatics 16.

36. Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

37. Hackl T, Hedrich R, Schultz J, Förster F. 2014. Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 30:3004–3011.

38. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9.

39. Kosugi S, Hirakawa H, Tabata S. 2015. GMcloser: Closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. Bioinformatics 31:3733–3741.

40. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.

2011. Integrative genomics viewer. Nat Biotechnol.

41. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform 14:178–192.

42. Aslett M, Aurrecoechea C, Berriman M, Al. E. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38:D457-62.

43. Eliaz D, Doniger T, Tkacz ID, Biswas VK, Gupta SK, Kolev NG, Unger R, Ullu E, Tschudi C, Michaeli S. 2015. Genome-wide analysis of small nucleolar RNAs of leishmania major reveals a rich repertoire of RNAs involved in modification and processing of rRNA. RNA Biol 12:1222–1255.

44. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10.

45. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

46. Gruening BA. 2014. Galaxy wrapper.

47. Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E. 2017. SynMap2 and SynMap3D: Web-based whole-genome synteny browsers, p. 2197–2198. *In* Bioinformatics.

48. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.

49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

50. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

51. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6:80–92.

52. Simonyan V, Mazumder R. 2014. High-performance integrated virtual environment (hive) tools and applications for big data analysis. Genes (Basel) 5:957–981.

53. Santana-Quintero L, Dingerdissen H, Thierry-Mieg J, Mazumder R, Simonyan V. 2014. HIVE-hexagon: High-performance, parallelized sequence alignment for next-generation sequencing data analysis. PLoS One 9.

54. Simonyan V, Chumakov K, Donaldson E, Karagiannis K, Lam PVN, Dingerdissen H,

Voskanian A. 2017. HIVE-heptagon: A sensible variant-calling algorithm with post-alignment quality controls. Genomics 109:131–140.

2.8 Additional information

2.8.1 Data availability

All data used in this study have been deposited online with GenBank. Raw PacBio reads for the IV strains, Illumina Miseq reads for the CL, VL and IV, the new genome assembly and the annotations generated in this study can be found under the PRJNA450813 BioProject accession number.

2.8.2 Acknowledgements

GM acknowledges the support of the Canadian Institutes of Health Research and the Global Health Innovative Technologies Fund. PL acknowledges receiving a doctoral training award from the Fonds de Recherche du Quebec Santé. LIM was funded by a graduate fellowship from the Canadian Institutes of Health Research (#MOP235928). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

2.8.3 Author contributions

PL wrote the main manuscript and performed the analysis. JH, JTR and AM performed analysis. WZ, LIM, VS and KD provided materials and insight. GM planned the experiment and wrote the manuscript. All authors reviewed the manuscript.

2.8.4 Accession codes

Genbank BioProject PRJNA450813

2.8.5 Competing financial interests

The authors declare no conflict of interest.

2.9 Figures and Tables

Table 2.1 Quality assessment metrics of the previous and current assemblies.

	Contigs	N50 (bp)	Protein coding	tRNA	rRNA	snRNA	SLRNA	snoRNA	Genes mapped
Old Assembly	2,154	45,436	7,969	64	11	4	-	31	8,081
New Assembly	36	1,067,468	8,633	90	51	6	68	910	9,758

Table 2.1 Old assembly refers to ASM22713v2 from strain BPK282A, new assembly refers to the assembly presented in this work. Contigs denotes the number of genomic fragments uninterrupted by stretches of unknown bases (Ns) or chromosome ends. N50 is used as a measure of contiguity, 50% of the genome is contained in contigs of size N50 and above. Annotated genes were broken down into protein coding, transfer-RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), spliced leader RNA (SLRNA) and small nucleolar RNA (snoRNA) genes. The number of genes mapped indicates the number of annotated genes along the genome.

Table 2.2 Summary of novel mutations identified in this study

CHR	GENE	MUTATION	PROVEAN	PROTEIN NAME
7	*LdBPK_070700* LdCL_070011900	Ala282Val	-0.743	vacuolar-type Ca2+-ATPase, putative
12	*LdBPK_120275* LdCL_120008300	Glu1157Asp	-0.258	Myotubularin-related protein, putative
14	LdCL_140017600	Ser2919fs	N/A	kinesin k39
14	*LdBPK_141190* LdCL_140017700	Glu1034Asp	-1.06	kinesin K39
22	*LdBPK_220840* LdCL_220015800	Pro219F/S	N/A	hypothetical protein
23	LdCL_230017500	INS:446Glu[1]	-12.453	sucrose hydrolase-like protein
25	*LdBPK_250620* LdCL_250011400	Ala969Glu	0.736	Raptor N-terminal CASPase like domain containing protein
25	*LdBPK_250790* LdCL_250013200	INS :110 Ala, Asn, Ser, Ala, Ala, Ala, Ala	N/A	hypothetical protein
27	*LdBPK_270830* LdCL_270014900	Ala1493Thr	-0.25	ATP-binding cassette protein subfamily A
29	LdCL_290028400	Thr208Ala	0.4	VIT family putative
30[1]	*LdBPK_301640* LdCL_300021700	Gln334STOP [1]	N/A	hypothetical protein
31	*LdBPK_311390* LdCL_310020800	STOP1486Leu,Ser,His	0	hypothetical protein
31	*LdBPK_311470* LdCL_310021600	Thr498Ala[1]	-0.15	hypothetical protein
31	*LdBPK_311470* LdCL_310021600	His497Arg[1]	0.942	hypothetical protein
31	*LdBPK_311470* LdCL_310021600	Gly380Asp[1]	-0.383	hypothetical protein
	IV→ VL mutations			
23	LdBPK_230830 LdCL_230014900	Asp712Glu	-1.625	hypothetical protein, unknown function
31	*LdBPK_312870* LdCL_310037100	Met189Thr	-4	hypothetical protein, unknown function
31	*LdBPK_313290* LdCL_310041200	Val187Phe	-0.634	Hypothetical protein
34	*LdBPK_342210* LdCL_340029800	Thr116DEL	-1.098	hypothetical protein
	IV-Only mutations			

14	*LdBPK_140470* LdCL_140010000	Gln89Lys	-0.044	cystathionine beta-lyase-like protein
31	*LdBPK_312810* LdCL_310036400	Cys173Phe	-9.5	regulator of chromosome condensation (RCC1) repeat, putative
32	*LdBPK_312770* LdCL_310035800	Gly667Ser	-1.292	hypothetical protein
32	*LdBPK_324000* LdCL_320046000	Val250Ile	0	hypothetical protein, unknown function
36	*LdBPK_361580* LdCL_360021300	Gene deletion	N/A	Serine/Threonine Kinase, putative
36	*LdBPK_361590* LdCL_360021400	Gene deletion	N/A	Serine/Threonine Kinase, putative
36	*LdBPK_361600* LdCL_360021500	Gene deletion	N/A	Engulfment and cell motility domain 2, putative
36	*LdBPK_361610* LdCL_360021600	Gene deletion	N/A	Predicted tripartite motif protein

Table 2.2 Summary of the novel mutations and their potential effects on protein function. All mutations are annotated using VL as the wild type amino acids and CL as the mutated amino acids. Genes with annotations in the previous assembly list the previous gene ID in italic, genes annotated only in this assembly list only one gene ID. The top segment lists fifteen attenuated cutaneous strain specific mutations identified in this study. Mutations marked with [1] appear at 50% but also co-occur with gene duplication event and are therefore possibly homozygous on one copy. 'INS' denotes amino acid insertions, 'F/S' denotes frameshifts, 'DEL' denotes amino acid deletions. The middle segment lists four mutations where the gain-of-function IV strain's genotype changed towards that of the visceral genotype. The bottom segment lists eight mutations present only in the gain-of-function IV strain and likely represents adaptations specific to the murine host. Calculated PROVEAN scores are shown in the fourth column, scores below the -2.5 threshold for deleterious mutations are highlighted in red (25).

Table 2.3 Summary of all genes containing mutations in the cutaneous isolates and classification into clusters

Cluster Number	Cluster Mutation Type	New annotation	Equivalents (when available)
Cluster 1 (13)	Nonsense Frameshift Insertion Deletions IV to VL	LdCL_300021700	LdBPK_301640
		LdCL_310020800	LdBPK_311390
		LdCL_250013200	LdBPK_250790
		LdCL_310020100	LdBPK_311320
		LdCL_310022200	LdBPK_311510
		LdCL_080011700	LdBPK_080670
		LdCL_340029800	LdBPK_342210
		LdCL_230014900	LdBPK_230830
		LdCL_310037100	LdBPK_312870
		LdCL_220015800	-
		LdCL_140017600	-
		LdCL_230017500	-
		LdCL_310041200	LdBPK_313290
Cluster 2 (9)	Multiple SNPs in the same gene Non-conservative amino acid change in conserved region with good PROVEAN score	LdCL_270015000	LdBPK_270840
		LdCL_290026900	LdBPK_292100
		LdCL_310021600	LdBPK_311470
		LdCL_310028800	LdBPK_312080
		LdCL_340046300	LdBPK_343690
		LdCL_290022800	LdBPK_291720
		LdCL_310024300	LdBPK_311710
		LdCL_360006000	LdBPK_360120
		LdCL_360062000	LdBPK_365480
Cluster 3 (18)	Non-conservative amino acid change in conserved region with poor PROVEAN score	LdCL_070018300	LdBPK_071330
		LdCL_320013800	LdBPK_320820
		LdCL_250016900	LdBPK_251150
		LdCL_220022000	LdBPK_221470
		LdCL_250015300	LdBPK_251000
		LdCL_040011100	LdBPK_040560
		LdCL_360016300	LdBPK_361120
		LdCL_200014300	LdBPK_200960
		LdCL_090011700	LdBPK_090660
		LdCL_130016200	LdBPK_131090
		LdCL_340009000	LdBPK_340390
		LdCL_130017800	LdBPK_131230

		LdCL_230009900	LdBPK_230440
		LdCL_220018100	LdBPK_221070
		LdCL_340044900	LdBPK_343550
		LdCL_290028400	-
		LdCL_250011400	LdBPK_250620
		LdCL_270014900	LdBPK_270830
Cluster 4 (8)	Conservative amino acid change in conserved region with good PROVEAN score	LdCL_350013100	LdBPK_350830
		LdCL_360052700	LdBPK_364550
		LdCL_230026600	LdBPK_231940
		LdCL_230009400	LdBPK_230400
		LdCL_320031100	LdBPK_322560
		LdCL_020008200	LdBPK_020280
		LdCL_310027700	LdBPK_311990
		LdCL_320031200	LdBPK_322570
Cluster 5 (13)	Conservative amino acid change in conserved region with poor PROVEAN score	LdCL_330011900	LdBPK_330640
		LdCL_170010200	LdBPK_170470
		LdCL_070011900	LdBPK_070700
		LdCL_210025000	LdBPK_211930
		LdCL_290022900	LdBPK_291730
		LdCL_200006300	LdBPK_200140
		LdCL_290029000	LdBPK_292290
		LdCL_030007500	LdBPK_030250
		LdCL_250006200	LdBPK_250110
		LdCL_360015800	LdBPK_361070
		LdCL_310023400	LdBPK_311630
		LdCL_360062700	LdBPK_365540
		LdCL_140017700	LdBPK_141190
Cluster 6 (14)	Non-conservative amino acid change in less conserved region Conservative amino acid change in less conserved region	LdCL_340022100	LdBPK_341580
		LdCL_060011600	LdBPK_060650
		LdCL_210015400	LdBPK_211040
		LdCL_050010900	LdBPK_050580
		LdCL_230011600	LdBPK_230610
		LdCL_250024100	LdBPK_251840
		LdCL_070015100	LdBPK_071060
		LdCL_250005300	LdBPK_250040
		LdCL_310022100	LdBPK_311500
		LdCL_200006800	LdBPK_200200
		LdCL_250014400	LdBPK_250910
		LdCL_230010400	LdBPK_230500
		LdCL_120008300	LdBPK_120275
		LdCL_290028100	LdBPK_292210
Cluster 7 (4)	IV-only mutations	LdCL_140010000	LdBPK_140470

LdCL_310035800	LdBPK_312770
LdCL_310036400	LdBPK_312810
LdCL_320046000	LdBPK_324000
LdCL_360021300	LdBPK_361580
LdCL_360021400	LdBPK_361590
LdCL_360021500	LdBPK_361600
LdCL_360021600	LdBPK_361610

Entries were not repeated in multiple lists.

Table 2.3 Identified mutations were further classified into priority clusters for effect on protein function and future analysis for genes associated with survival in visceral organs. Mutations were prioritized by likelihood of contributing to visceral tissue tropism by severity of the coding change, accumulation of secondary mutations and conservation. Gene loci listed from the current assembly as well as previous ID numbers when available.

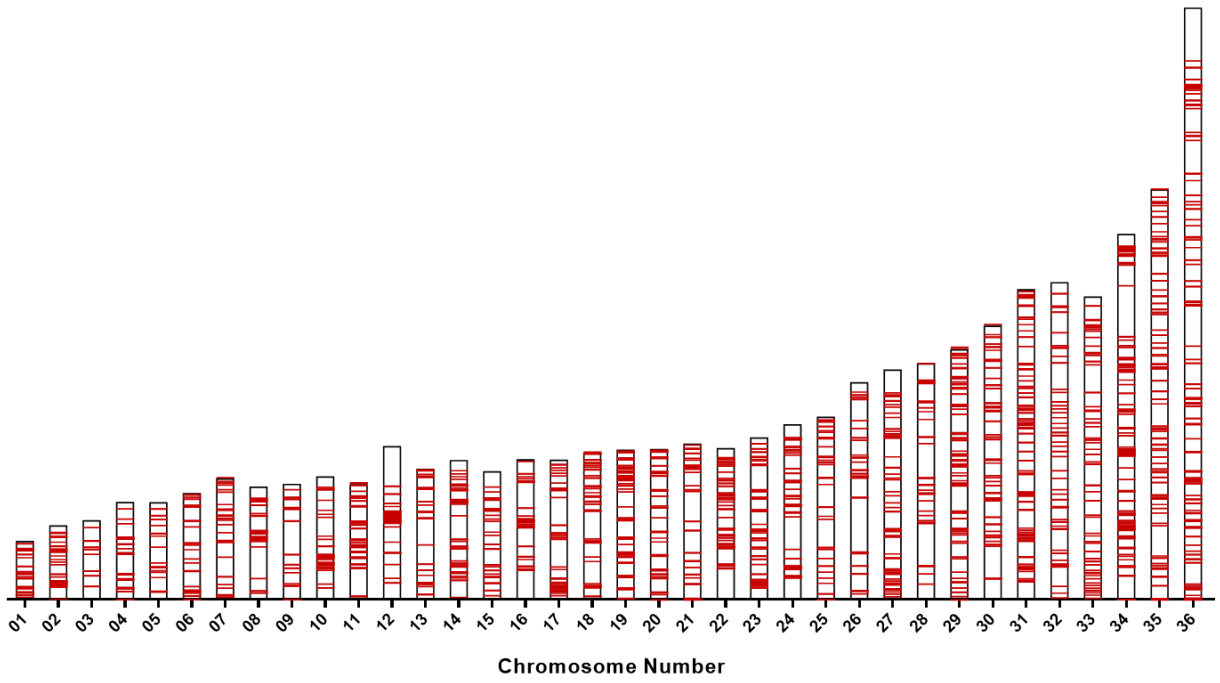


Figure 2.1 Location of the gaps along 36 chromosomes that have been closed in this new assembly

Chromosomal locations of gaps are indicated in red. No gaps remain in the current assembly.

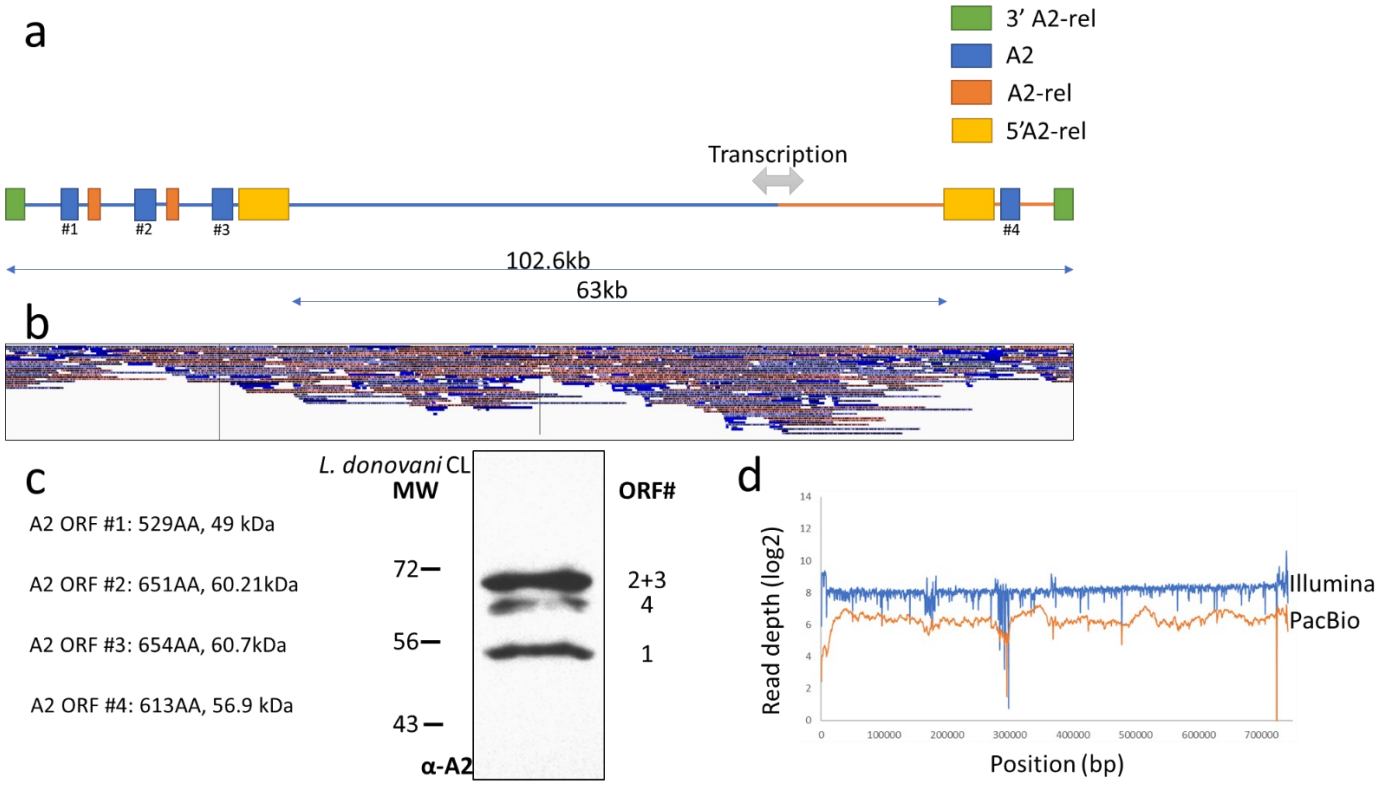


Figure 2.2 Organization of the 4 copies of the A2 gene on chromosome 22 in the attenuated cutaneous *L. donovani* strain

(a) Location of the 4 A2 genes are shown in blue and numbered 1- 4. Interspaced A2-rel genes are labeled in orange, 3' A2-rel genes are labeled in green and 5' A2-rel genes are labeled in yellow. A2-rel genes have no homology with A2 genes (15). Transcription direction is shown according to strandedness: blue represents reverse strand direction of transcription, red represents forward strand transcription. The genes located in the 63kb region between opposing A2 clusters are not depicted for clarity. (b) Alignment of the longest (~11kb+) PacBio reads to the A2 clusters. Reads in the 5' to 3' direction labeled in red; reads in the 3' to 5' direction labeled in blue. (c) Western blot analysis of A2 proteins in the attenuated cutaneous *L. donovani* strain. The sizes of the A2 proteins are consistent with the ORFs and number of A2 genes identified in this assembly. (d) Coverage graph of chromosome 22 using Illumina (blue) and PacBio (orange) reads.

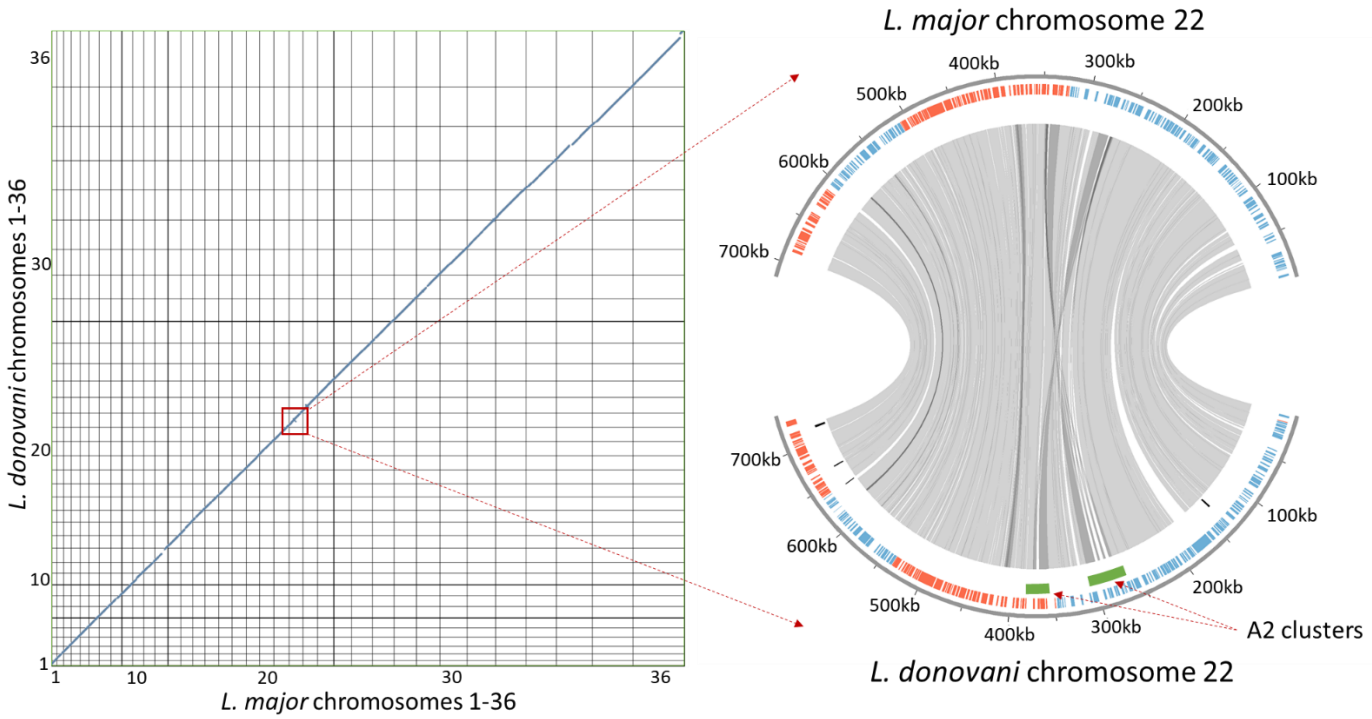


Figure 2.3 *L. donovani* maintains high levels of synteny with *L. major* including chromosome 22 where the A2 genes are located

Left: Dot plot of the coding DNA sequences of *L. major* compared to those of *L. donovani* generated from our assembly across the entire genome. Right: Synteny comparison of chromosome 22. The outer most circle represents the chromosomal location. The second circle is labelled with genes on the forward strand (blue) and genes on the reverse strand (red). The third circle represents genes that are only present in one of the two compared species. The inner association lines join syntenic genes between the two species.

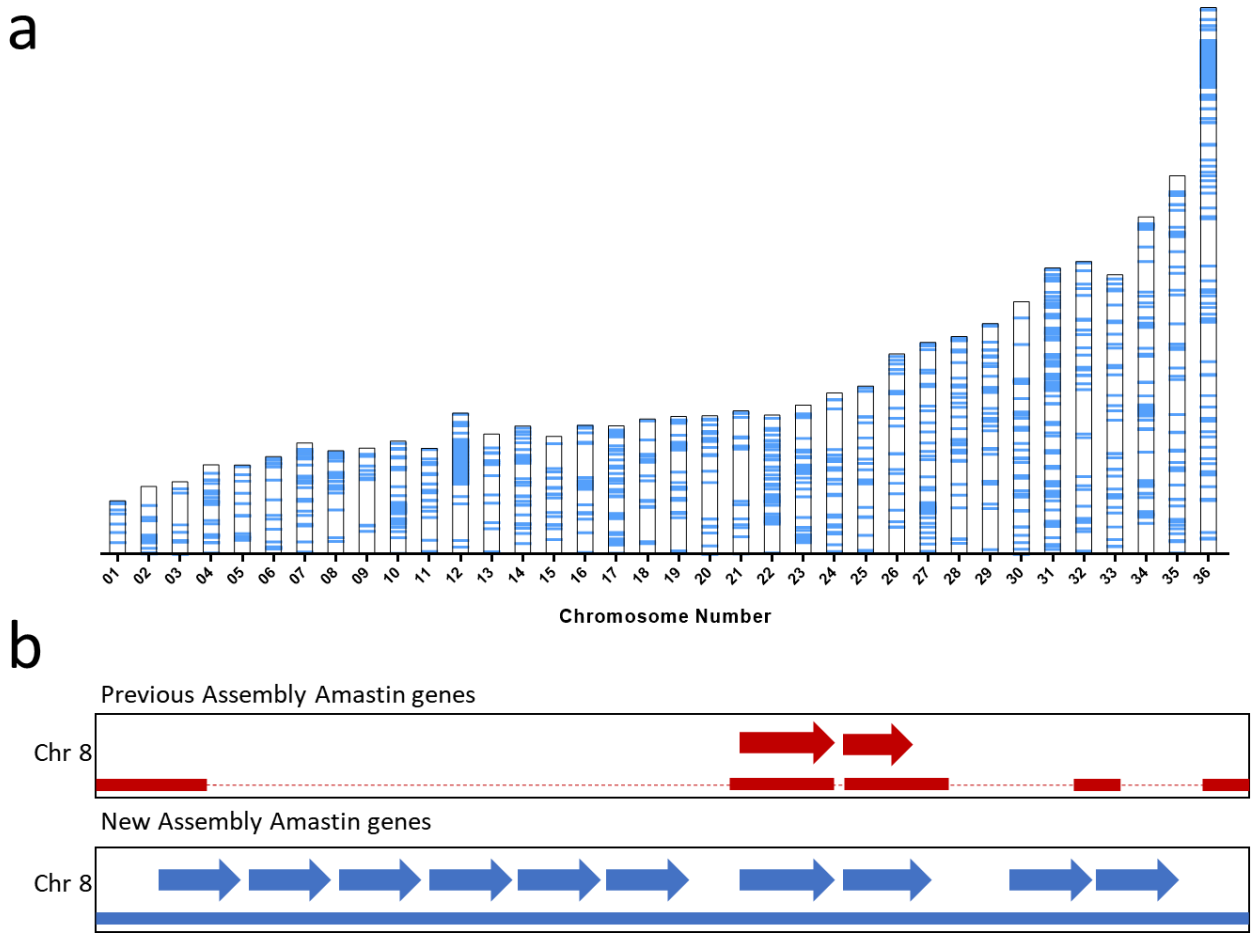


Figure 2.4 The new *L. donovani* genome assembly results in a significant change in gene annotations

(a) New or improved gene annotations are highlighted in Blue along the 36 chromosomes. Compared to the previous *L. donovani* reference assembly (ASM22713v2 from strain BPK282A1), there were 1,087 protein coding genes unannotated or differently annotated in the current assembly. Unannotated or differently annotated genes were obtained by removing all annotations generated from our assembly that shared 95% or greater similarity to those previously available (8). (b) Expansion of the amastin gene cluster on chromosome 8. Top track contains the previously two known coding sequences aligned to the previous *L. donovani* reference assembly (ASM22713v2 from strain BPK282A1). Gaps in the previous assembly depicted as dotted lines.

Bottom track contains 10 amastin genes identified in the updated assembly. One previously identified Amastin gene has been aligned, 1 has been expanded and 8 have been annotated *de novo*.

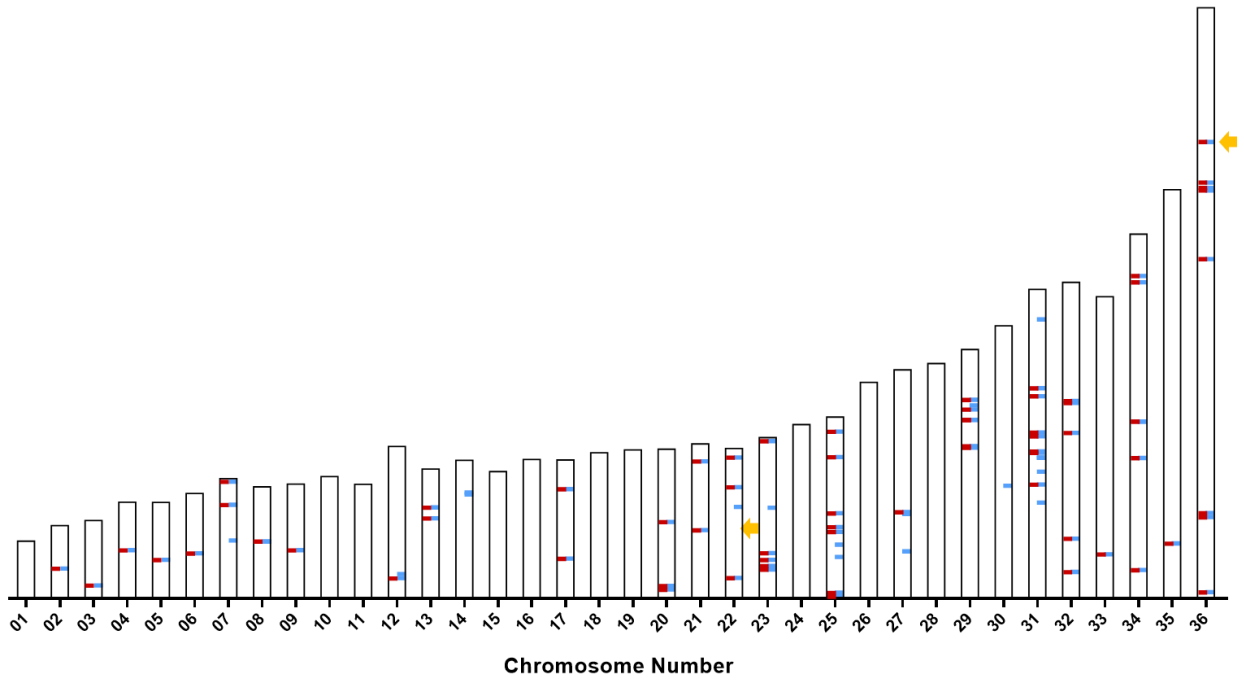


Figure 2.5 Verification of previously identified SNPs and location of new SNPs that differ between the virulent VL and attenuated CL strains of *L. donovani*

Chromosomal location of previously identified homozygous non-synonymous SNPs between the cutaneous and visceral disease derived *L. donovani* strains (Red) (4) compared to the novel SNPs identified only in this study (Blue) (synonymous and heterozygous codon changes identified are not labeled). Note that all the previously identified SNPs were also identified, or confirmed, in this study. 70 SNPs were previously identified across 66 genes. The same 70 SNPs were identified in this study, with an additional 15 novel SNPs not previously seen specific to the cutaneous strain. Genomic locations of SNPs identified in the previous study were translated to new genomic coordinates based on the new assembly for consistency. Arrows in yellow highlight the position of the previously identified RagC SNP on chromosome 36 and the A2 copy number difference on chromosome 22.

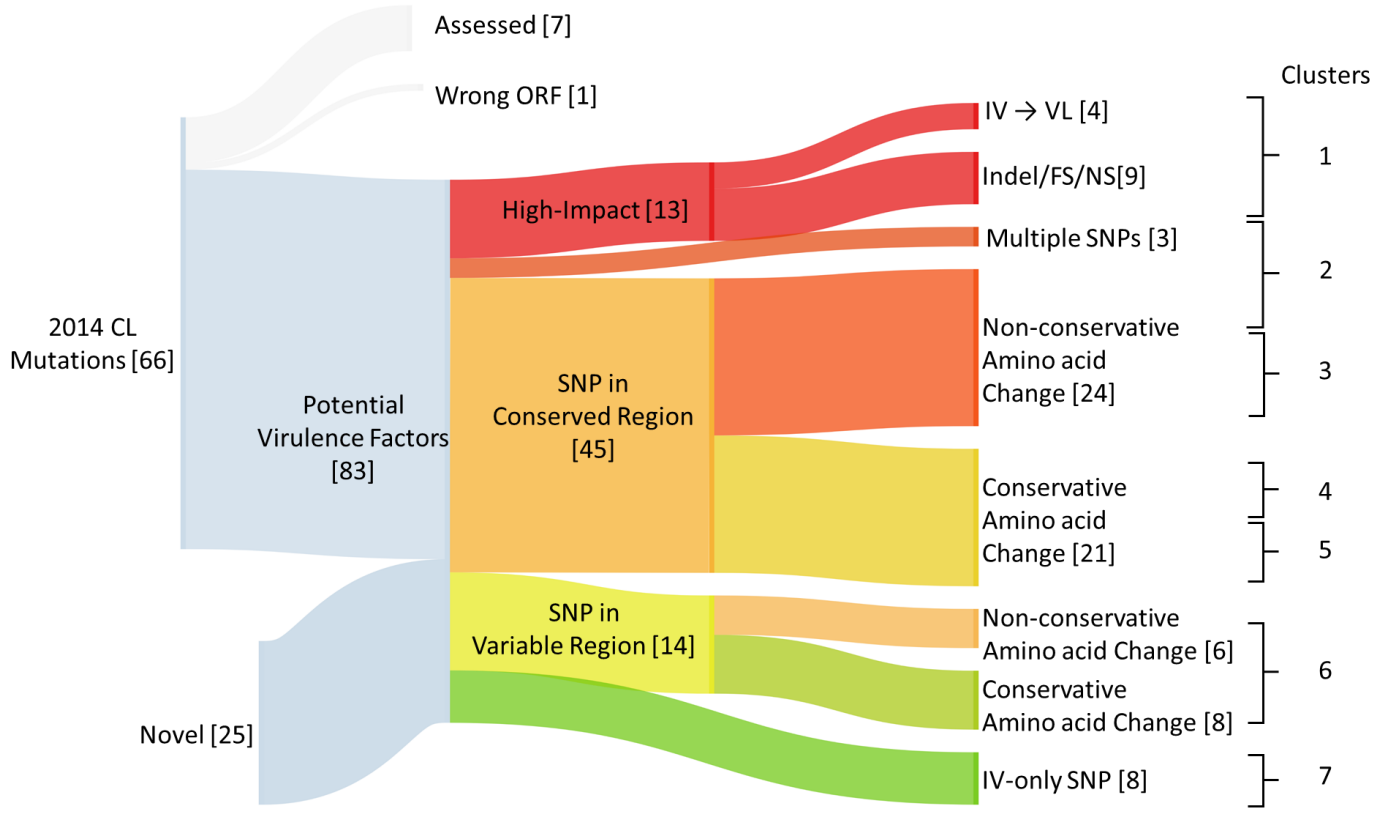


Figure 2.6 Summary of all genes with non-synonymous mutations between the cutaneous, visceral, and gain-of-function strains of *L. donovani*

All non-synonymous SNPs and Indels were classified as common to our previous study (2014 CL (4)) or identified in this study (Novel), as well as by their effect on amino acid changes from top to bottom, colored red to green in descending order of likelihood to affect the phenotype of the parasite. 66 genes were common to the previous data set. Of those genes, 7 were previously investigated (4) and 1 was rejected due to an open reading frame misannotation. 25 genes were only listed in this study (Novel). Diagram created using SankeyMATIC (http://sankeymatic.com).

2.10 Supplementary Material

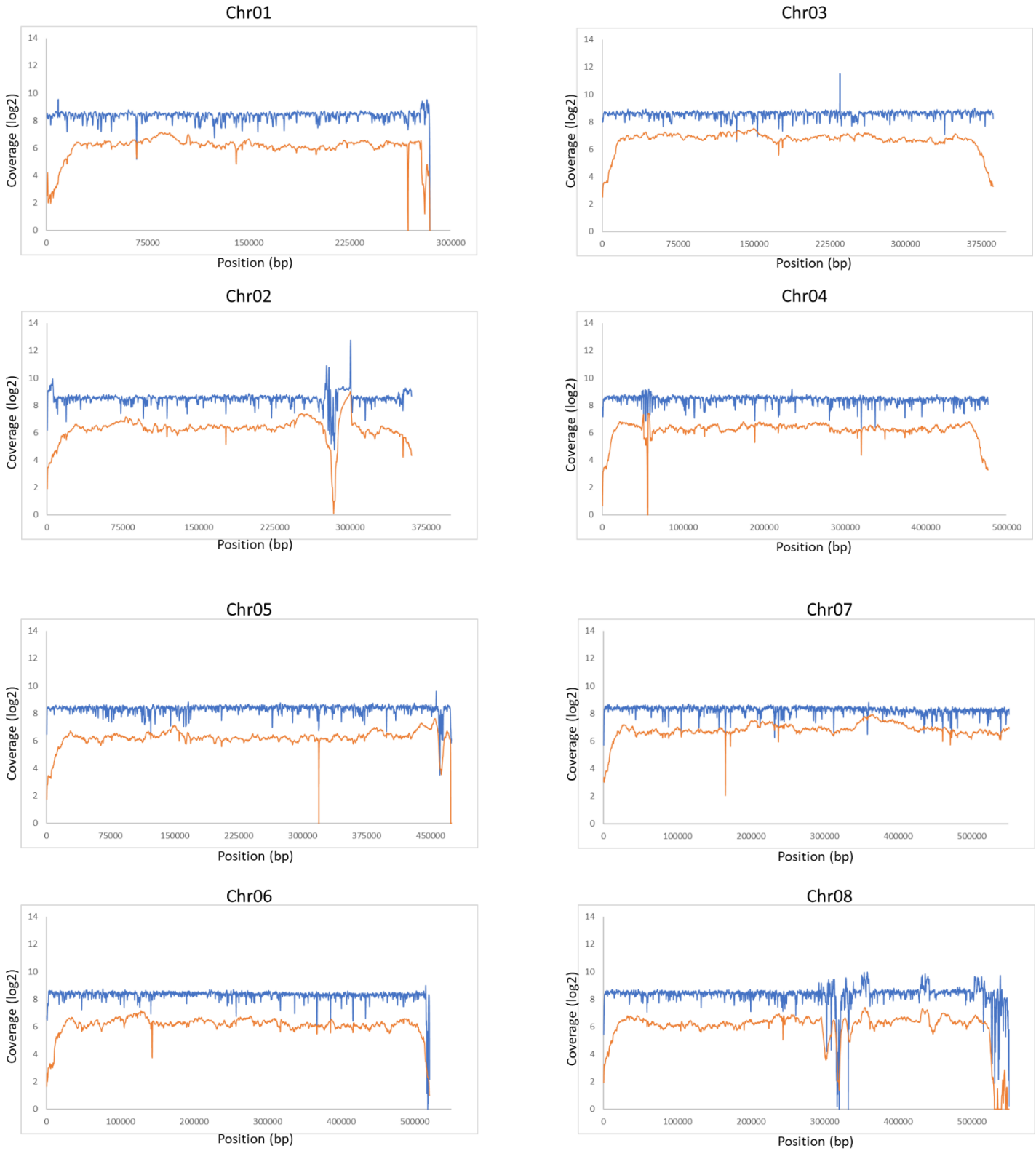
Supplementary Table 2.S1 List of *L. donovani* pseudogenes derived from *L. major*

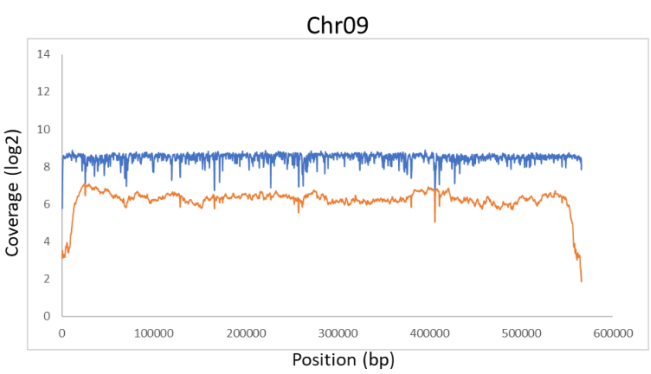
Chromosome	Start	Stop	Strand	Gene ID	Internal Stop	Frameshift	L. Major origin
LdCL_06	518294	520041	+	LdCL_060018900		XXXX	LmjF.22.0010
LdCL_08	550785	551865	+	LdCL_080018700		XXXXX	LmjF.08.1265
LdCL_10	79199	79729	+	LdCL_100006900	XX	XX	LmjF.10.0185
LdCL_12	603220	604085	+	LdCL_120018300		XXXXXX	LmjF.12.0880
LdCL_16	316885	317791	-	LdCL_160014100	XXXX	XXXX	LmjF.16.0880
LdCL_26	580729	582519	+	LdCL_260021400	XXXX	XXX	LmjF.26.1590
LdCL_27	527810	529301	-	LdCL_270017800	XXXX	XXXXXX	LmjF.14.0180
LdCL_27	732393	732680	-	LdCL_270023400*	X		LmjF.27.1740
LdCL_31	1179486	1180414	-	LdCL_310031700	XXXXX	XXX	LmjF.31.2310
LdCL_31	1334630	1335706	-	LdCL_310035400	XXX	XX	LmjF.29.1570
LdCL_32	784327	785612	-	LdCL_320026000	XX	X	LmjF.32.1950
LdCL_33	702441	703568	+	LdCL_330023800	XXXXX	XXXXX	LmjF.32.2640
LdCL_35	1479577	1480153	-	LdCL_350044600	XX	XX	LmjF.35.3910

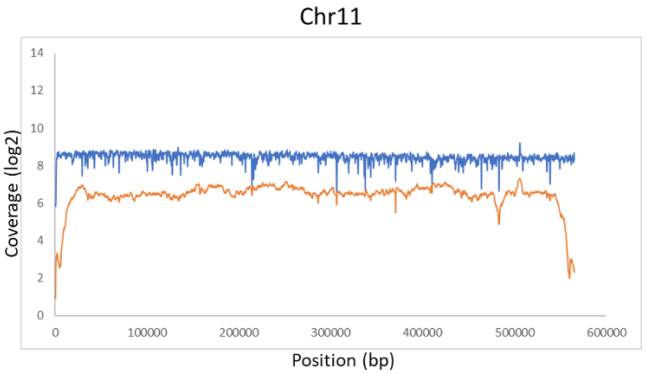
13 pseudogene annotations in *L. donovani* originated from 13 intact functional genes from *L. major* in the current assembly. The 13 pseudogenes were common across all 3 isolates sequences in this study. The locations of the pseudogenes are given in chromosomal position from start to stop. Strand indicates direction of transcription. Each stop codon and frameshift are marked with an X *L. major* origin denotes the equivalent functional gene ID in *L. major*.

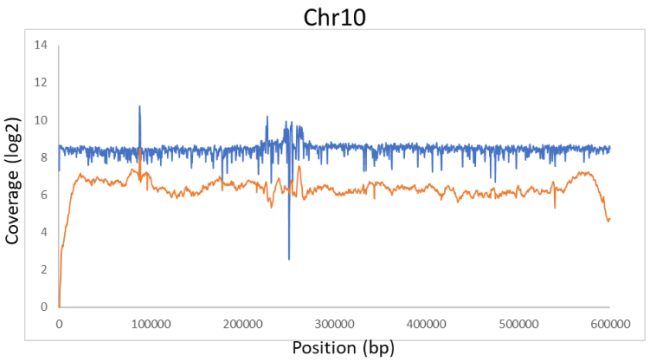
*One gene contains only one stop codon; this gene however is also missing a start codon.

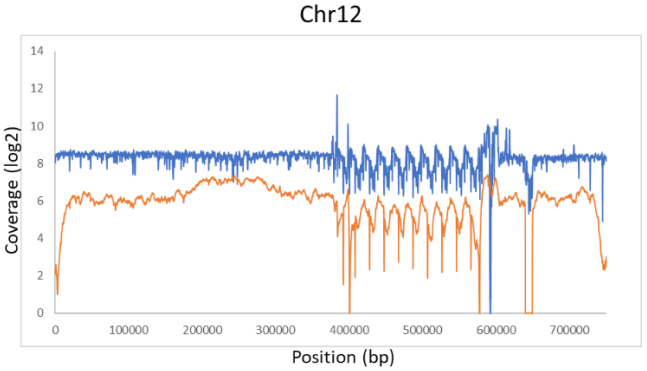
Supplementary Figure 2.S1 Coverage graphs of the new *L. donovani* assembly

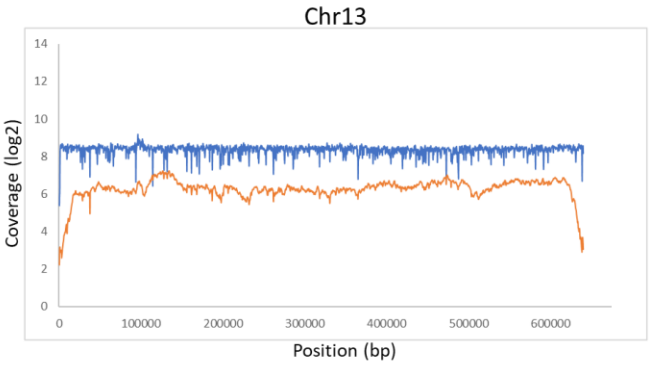


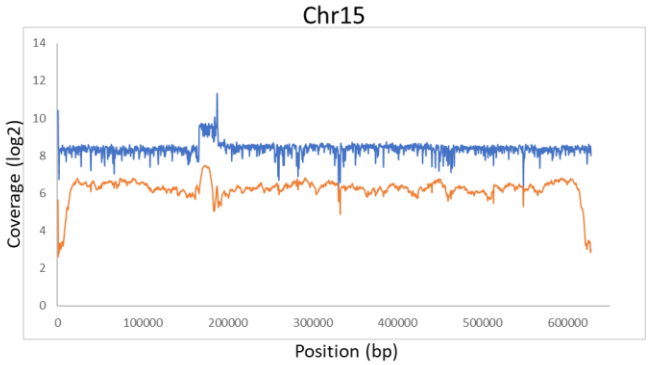

Chr09
Coverage (log2)
Position (bp)

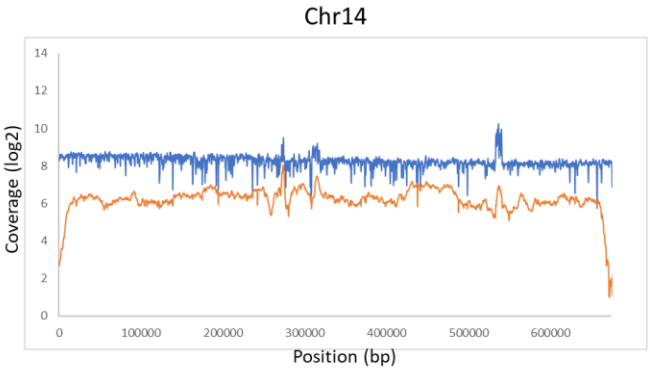

Chr11
Coverage (log2)
Position (bp)

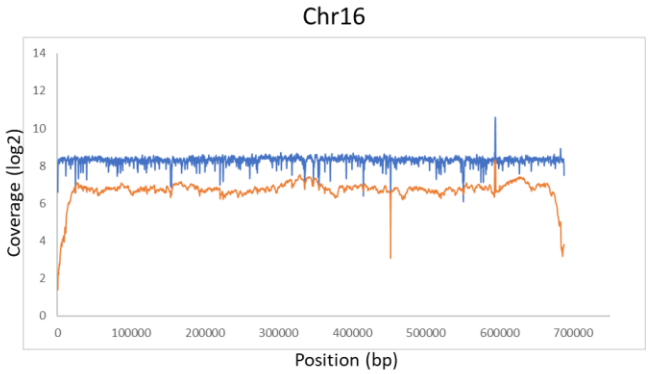

Chr10
Coverage (log2)
Position (bp)

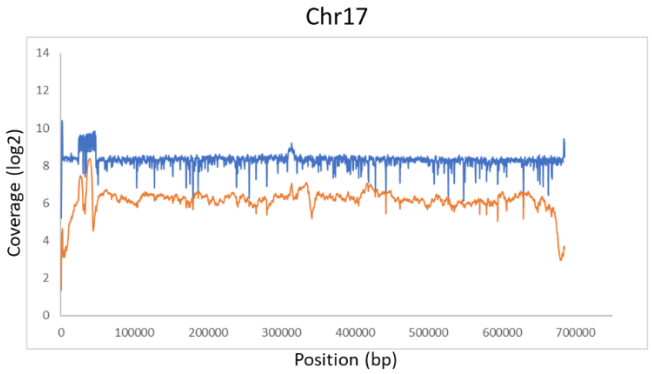

Chr12
Coverage (log2)
Position (bp)

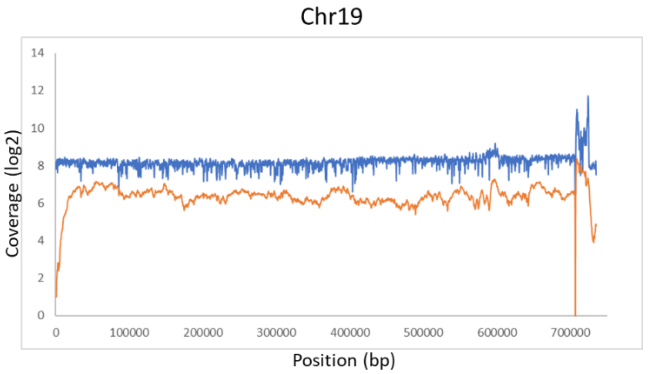

Chr13
Coverage (log2)
Position (bp)

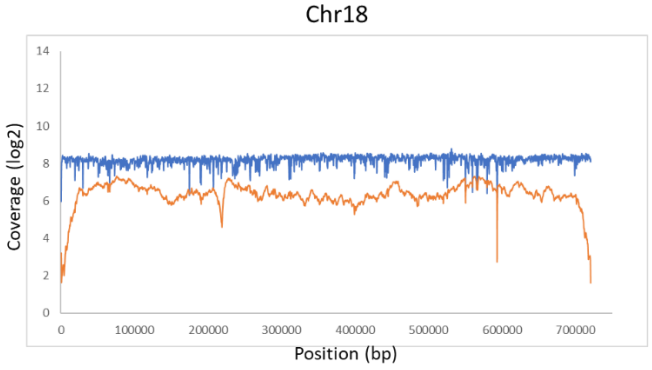

Chr15
Coverage (log2)
Position (bp)

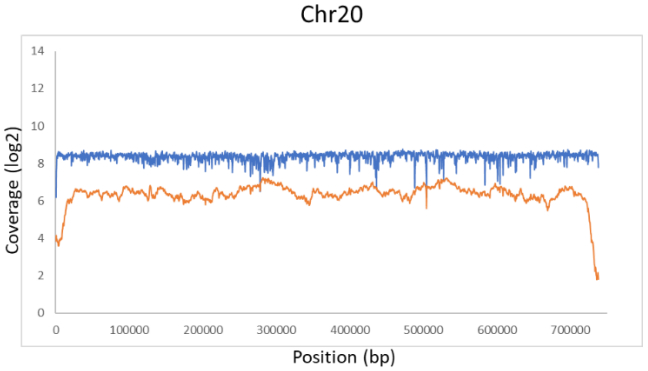

Chr14
Coverage (log2)
Position (bp)

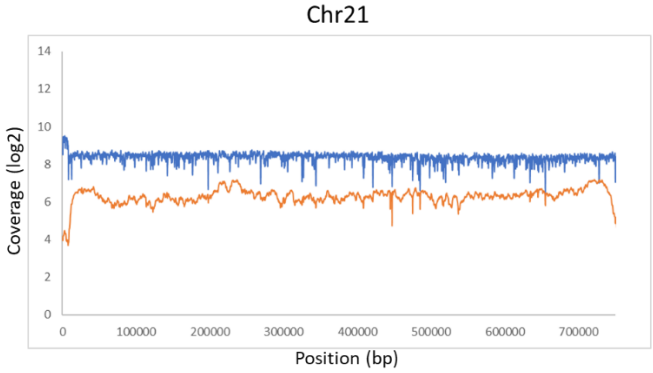

Chr16
Coverage (log2)
Position (bp)

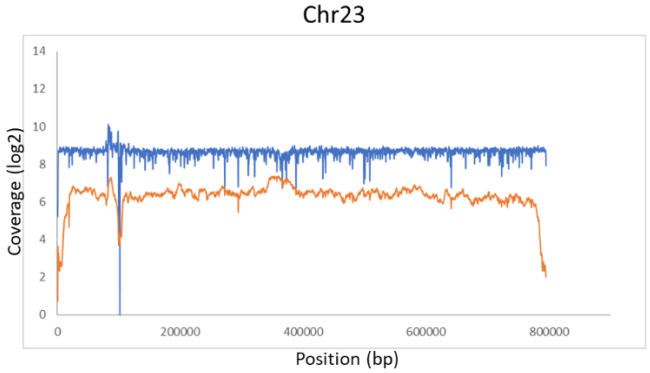

Chr17
Coverage (log2)
Position (bp)

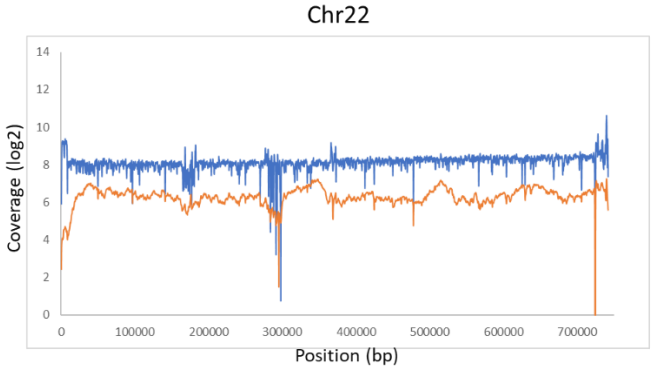

Chr19
Coverage (log2)
Position (bp)

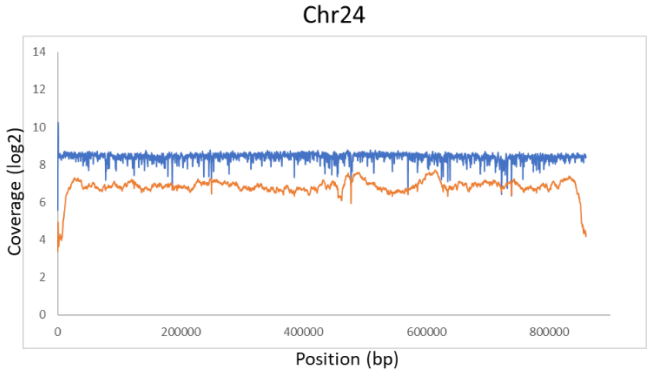

Chr18
Coverage (log2)
Position (bp)

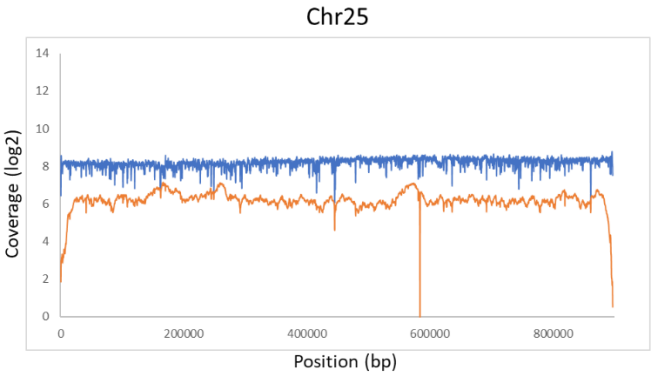

Chr20
Coverage (log2)
Position (bp)

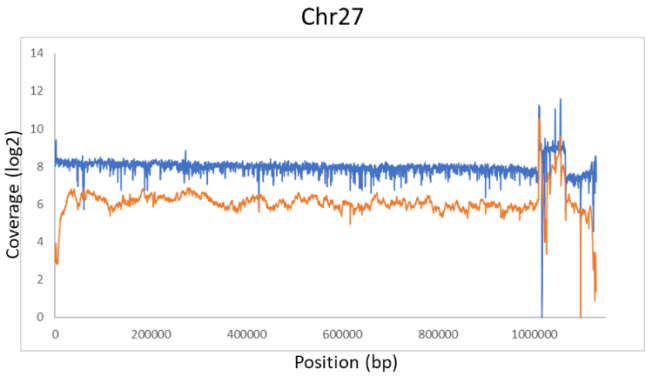

Chr21
Coverage (log2)
Position (bp)

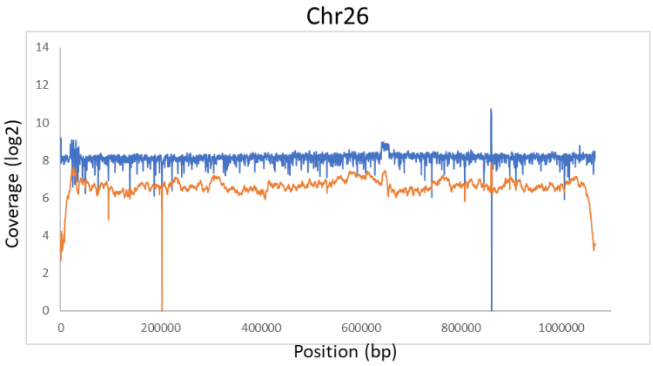

Chr23
Coverage (log2)
Position (bp)

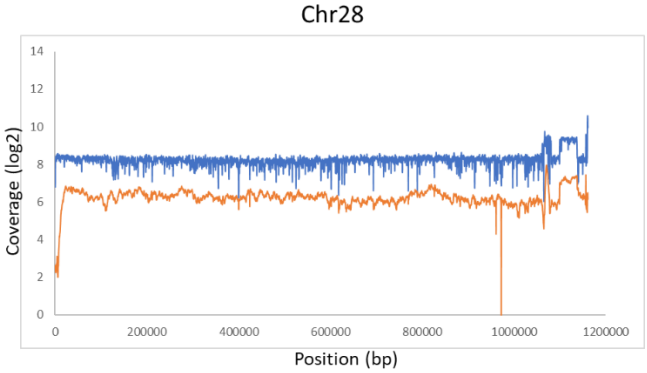

Chr22
Coverage (log2)
Position (bp)

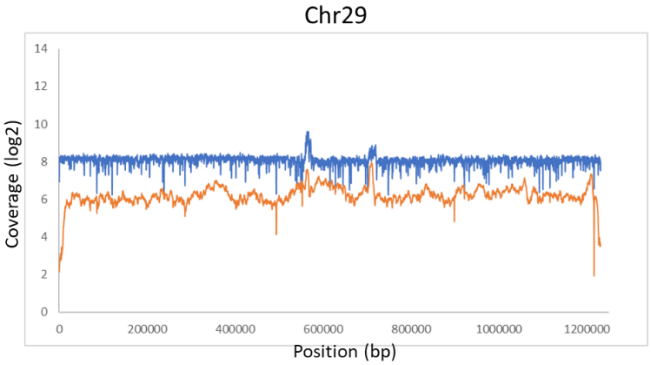

Chr24
Coverage (log2)
Position (bp)

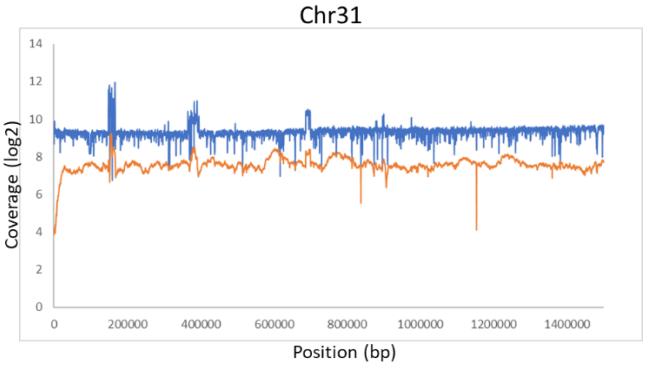

Chr25
Coverage (log2)
Position (bp)

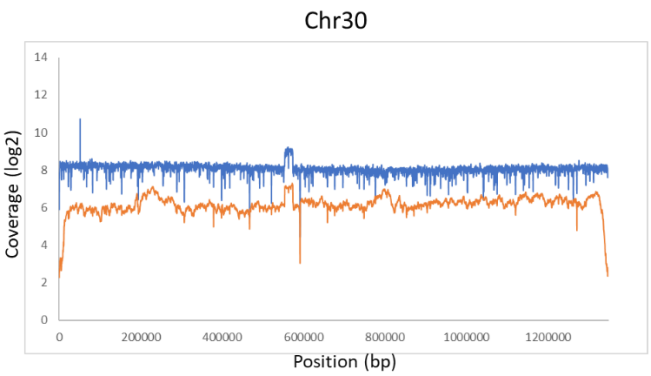

Chr27
Coverage (log2)
Position (bp)

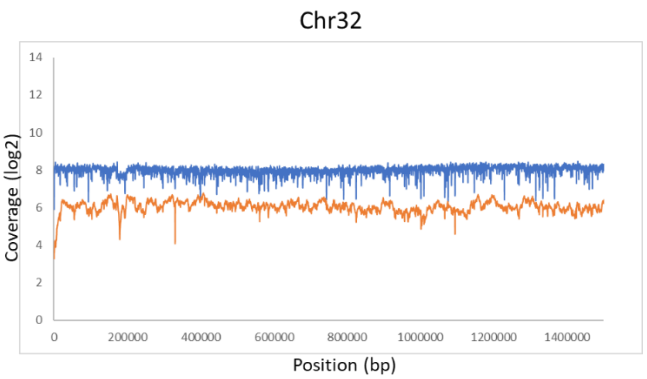

Chr26
Coverage (log2)
Position (bp)

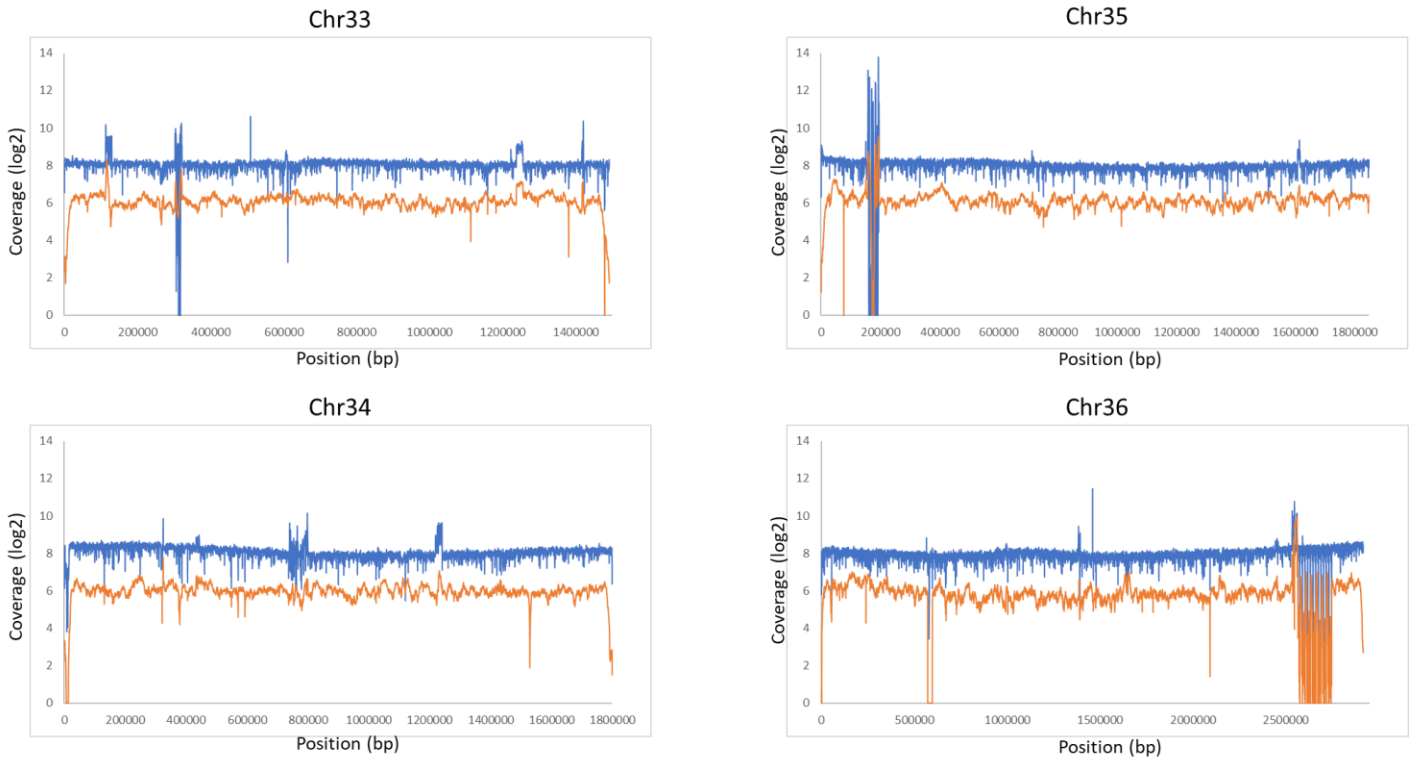

Chr28
Coverage (log2)
Position (bp)


Chr29
Coverage (log2)
Position (bp)

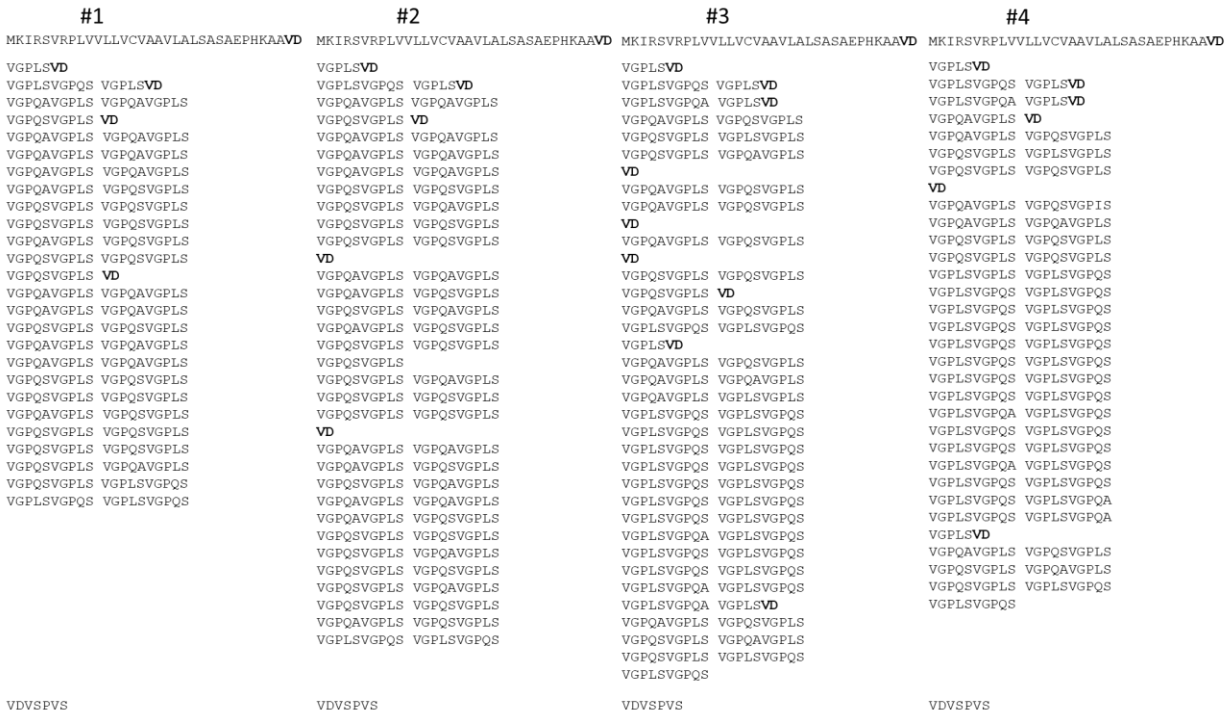

Chr31
Coverage (log2)
Position (bp)


Chr30
Coverage (log2)
Position (bp)


Chr32
Coverage (log2)
Position (bp)



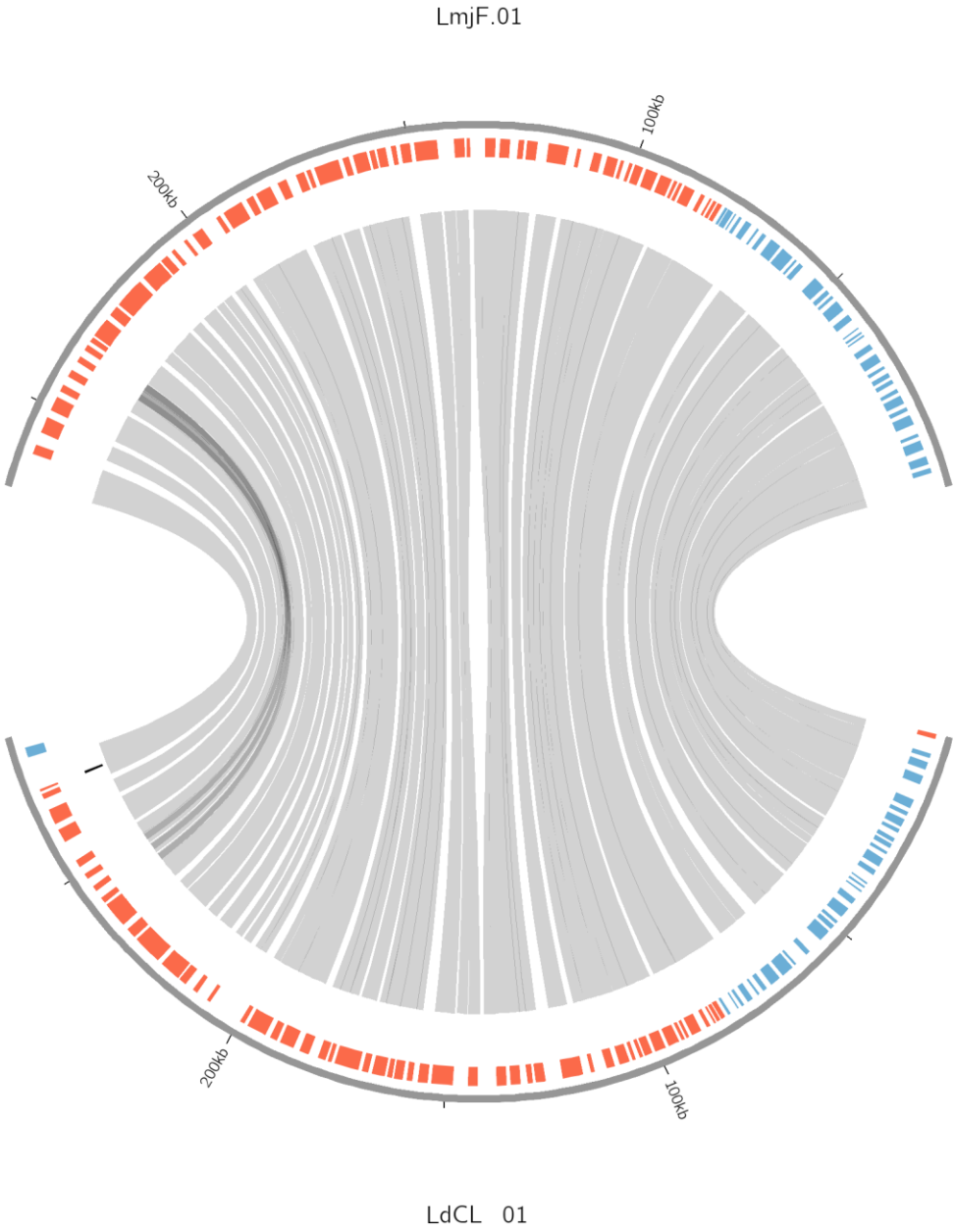
Read coverage of each chromosome in the new assembly. Graphs were generated from the moving average of the raw coverage at each position with a window size of 450bp, equivalent to the average Illumina read insert size. Illumina data plotted in blue, PacBio data plotted in orange.

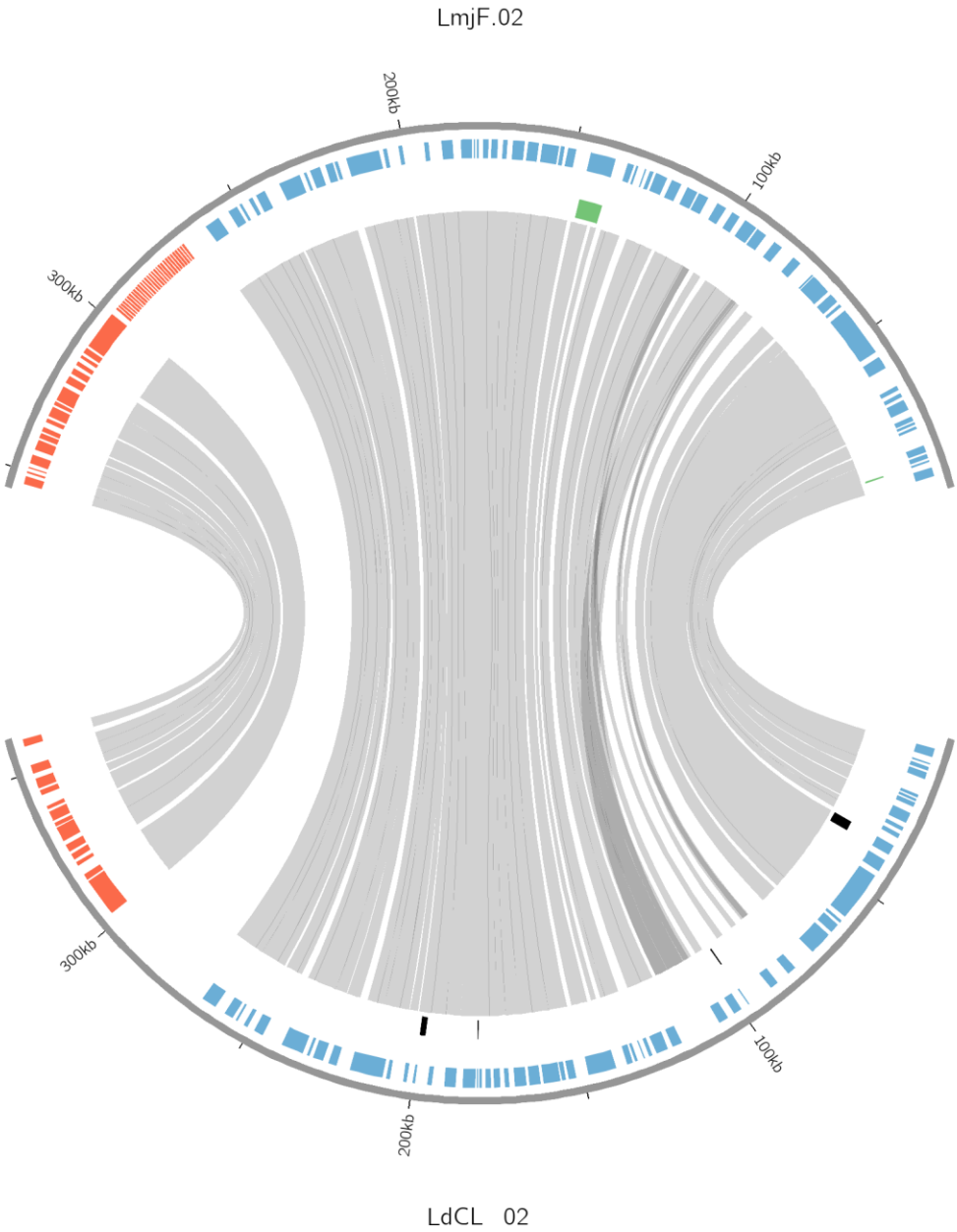


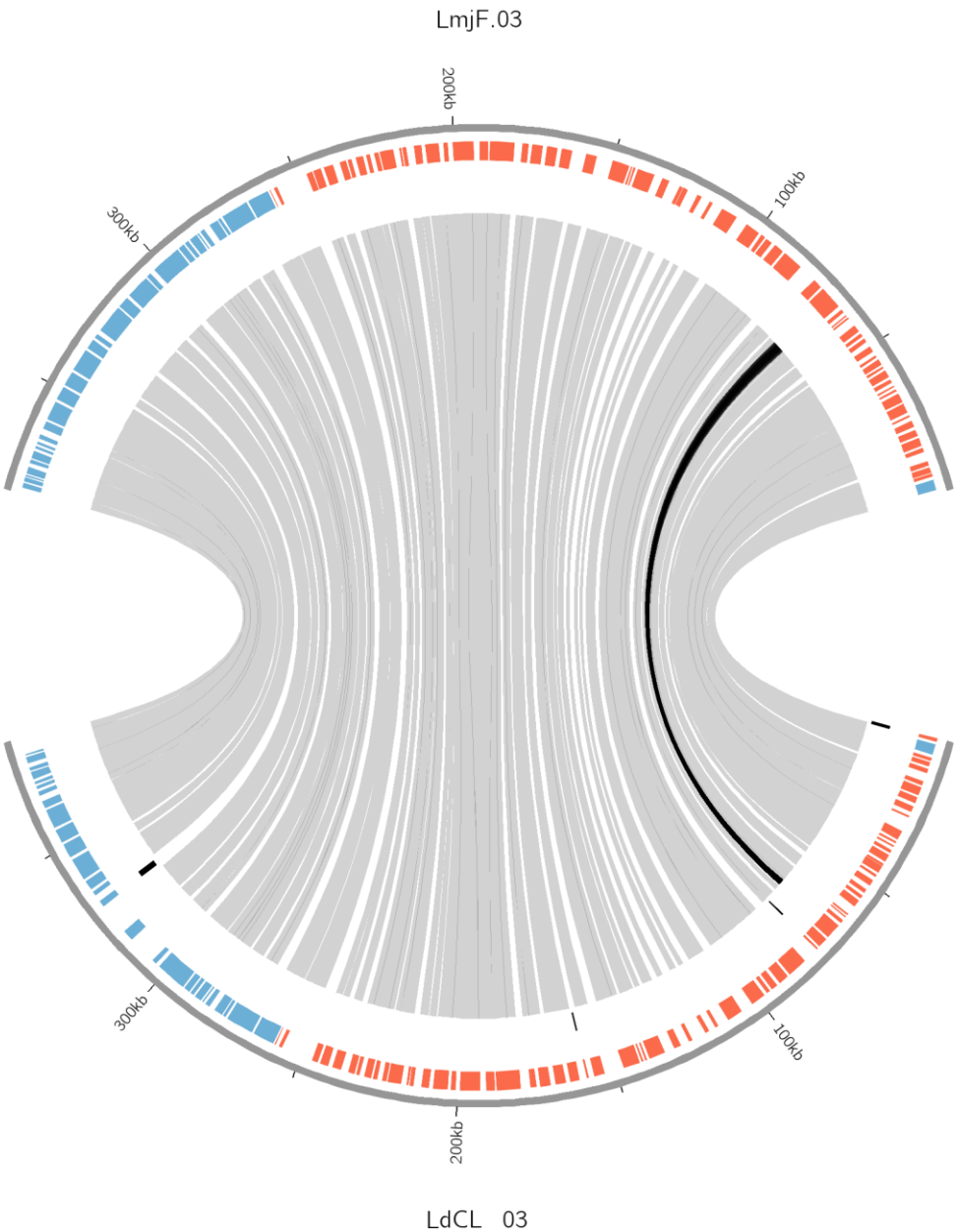
Supplementary Figure 2.S2 Complete amino acid sequences of A2 genes from the attenuated cutaneous *L. donovani* strain from Sri Lanka

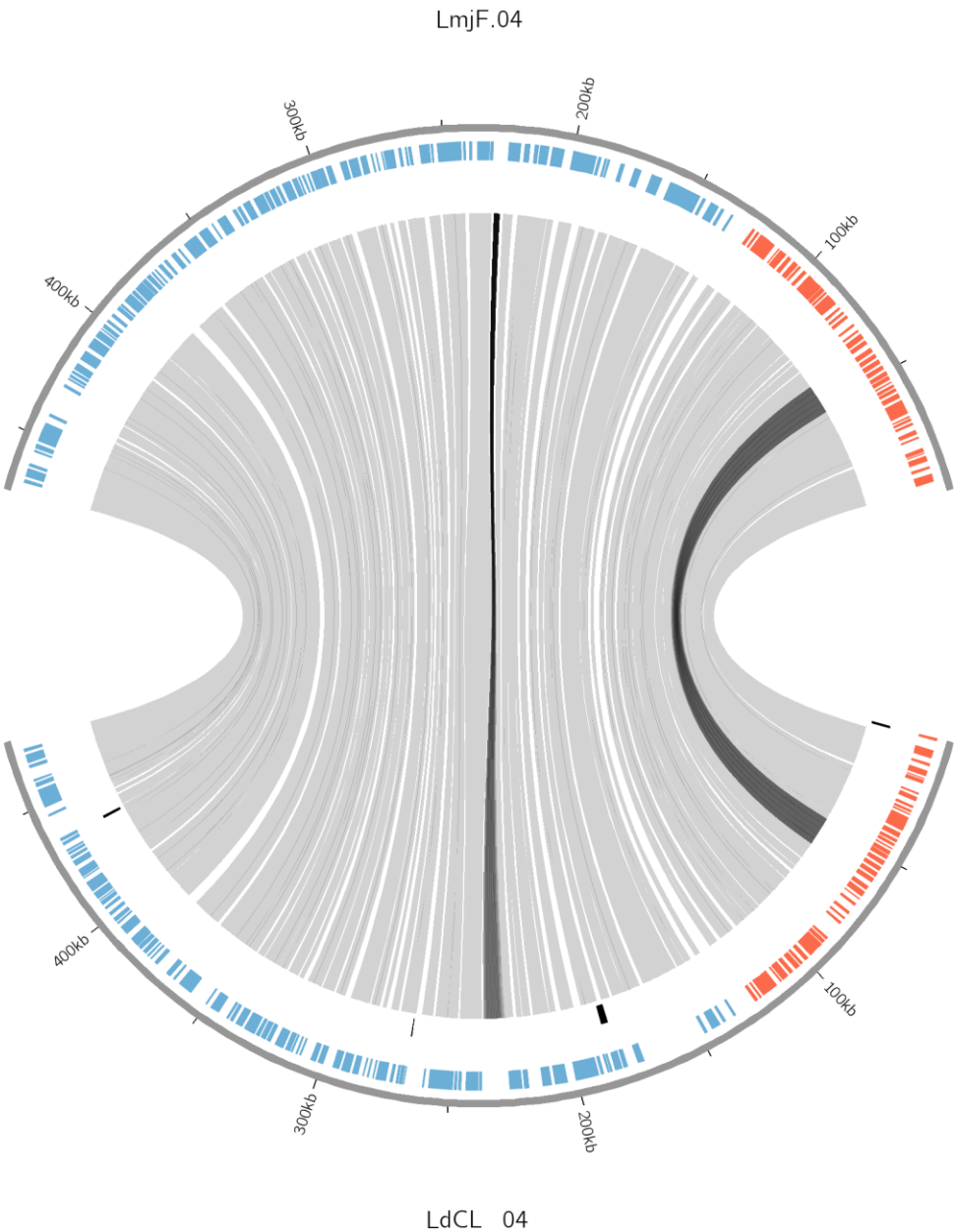
The DNA sequences obtained from the assembled A2 cluster were translated into their corresponding amino acids. All four A2 proteins share a common leader and terminal sequence. Deviations from the VGPLS VGPQS/A by the insertion of valine-aspartate are highlighted in bold.

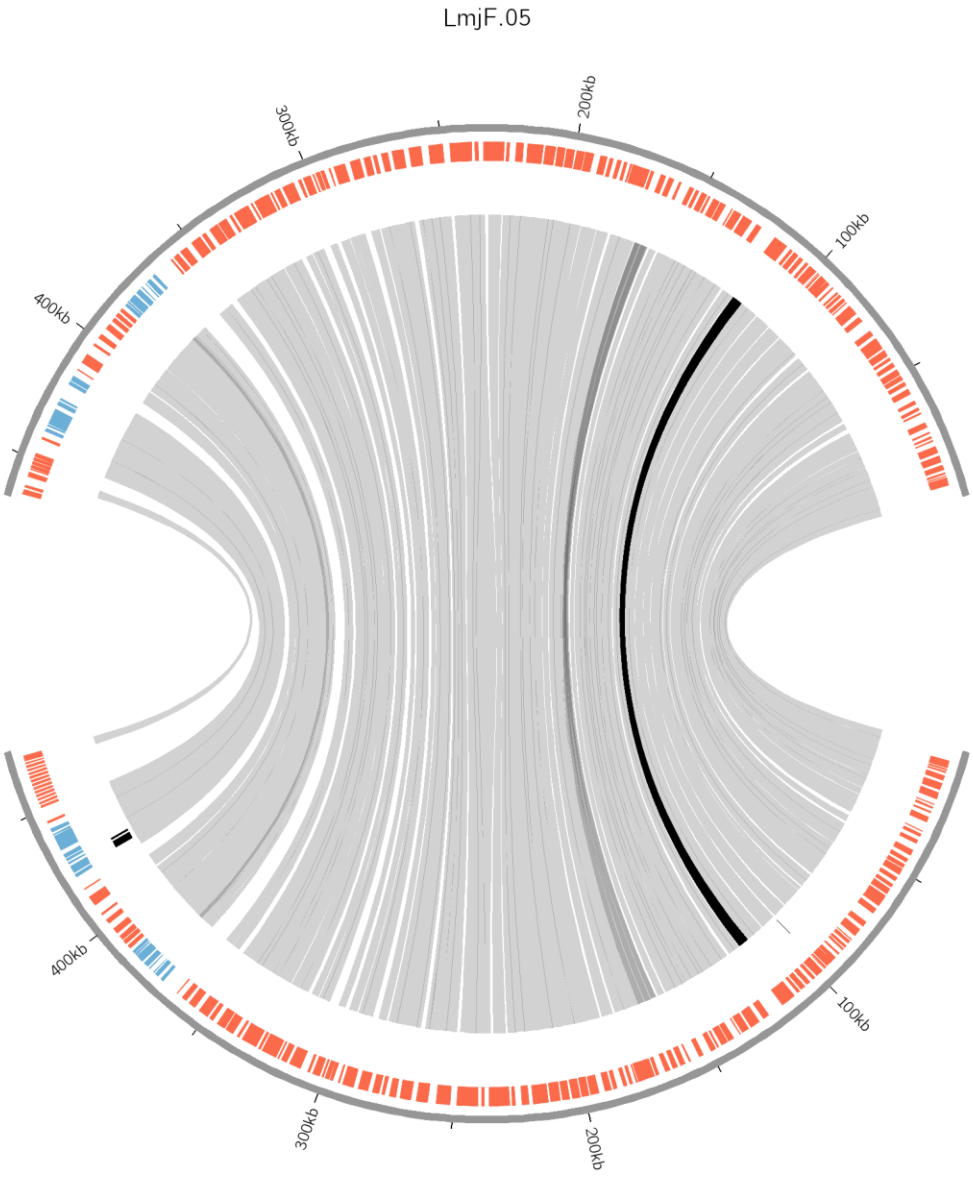
Supplementary Figure 2.S3 Synteny map of the *L. major* and *L. donovani* chromosomes.

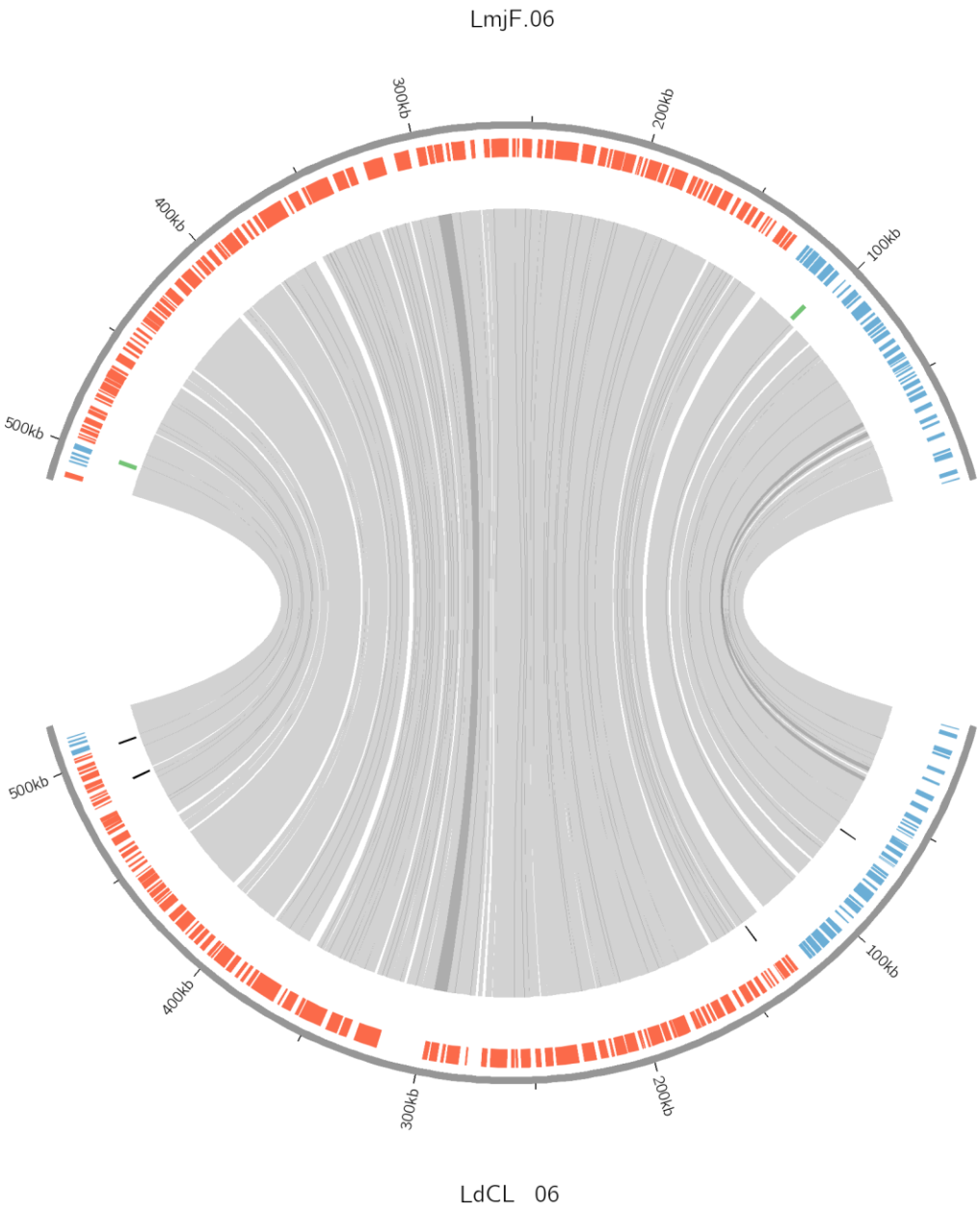


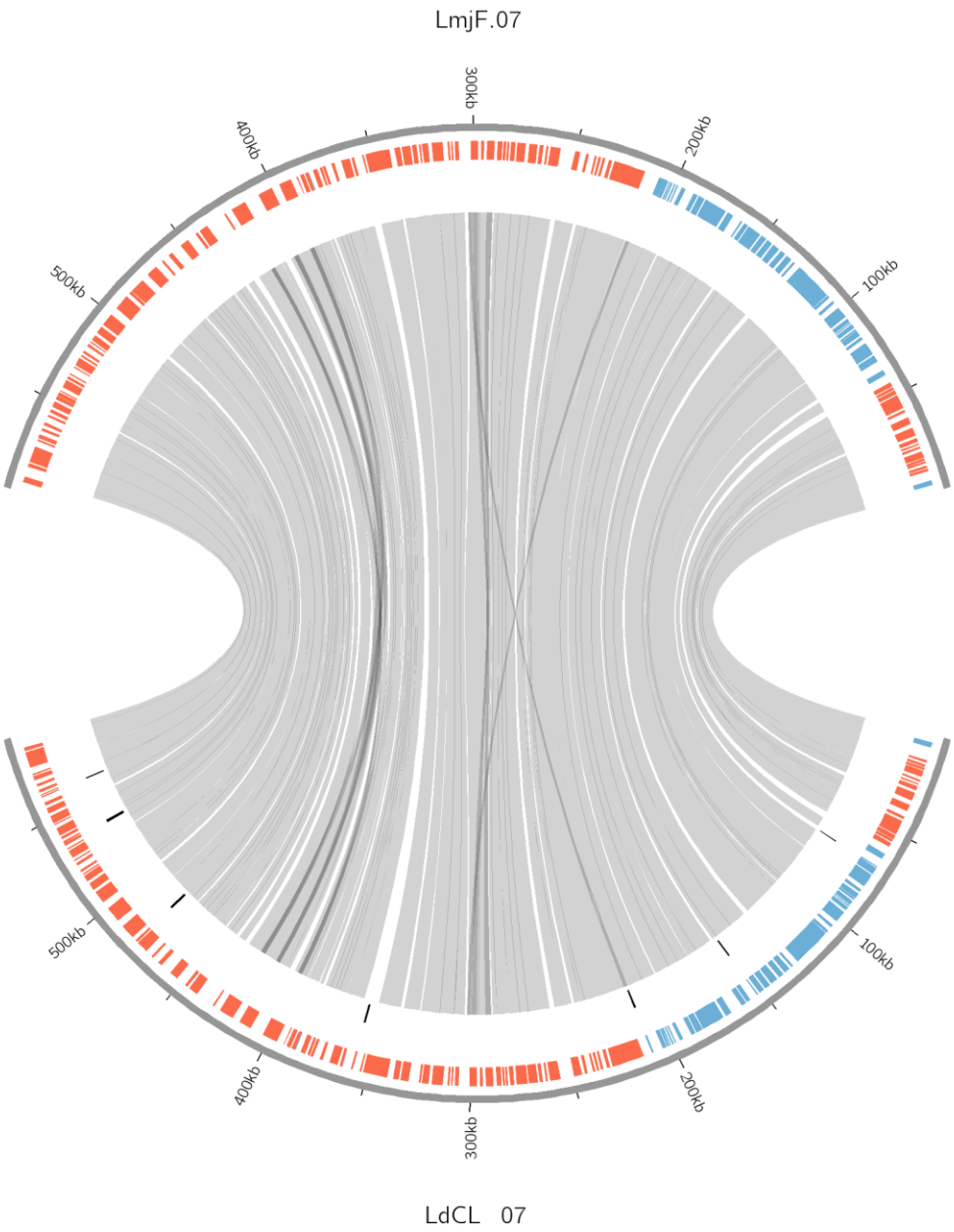

LmjF.02
200kb
100kb
300kb
300kb
100kb
200kb
LdCL 02

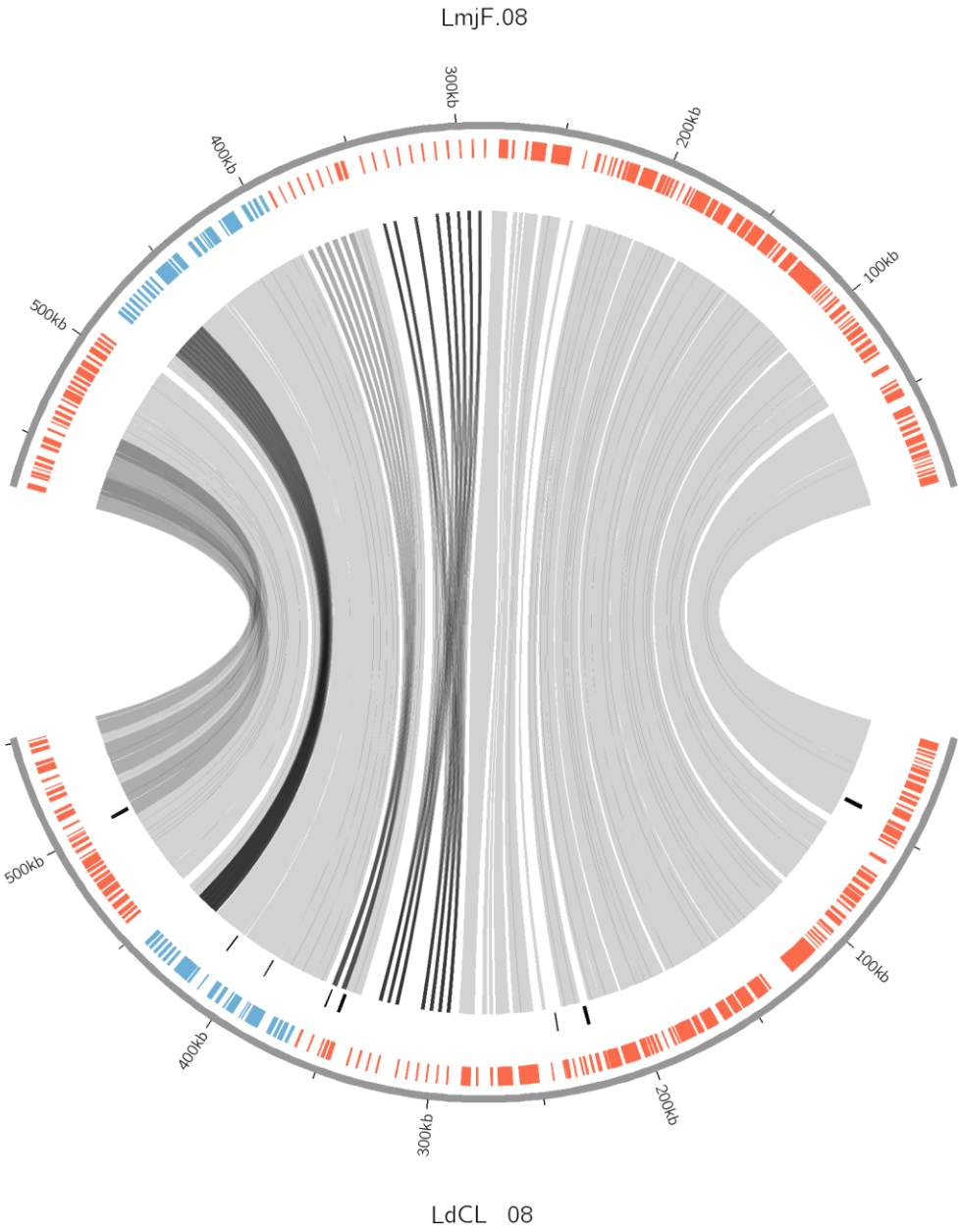

LmjF.03
200kb
100kb
300kb
300kb
100kb
200kb
LdCL 03

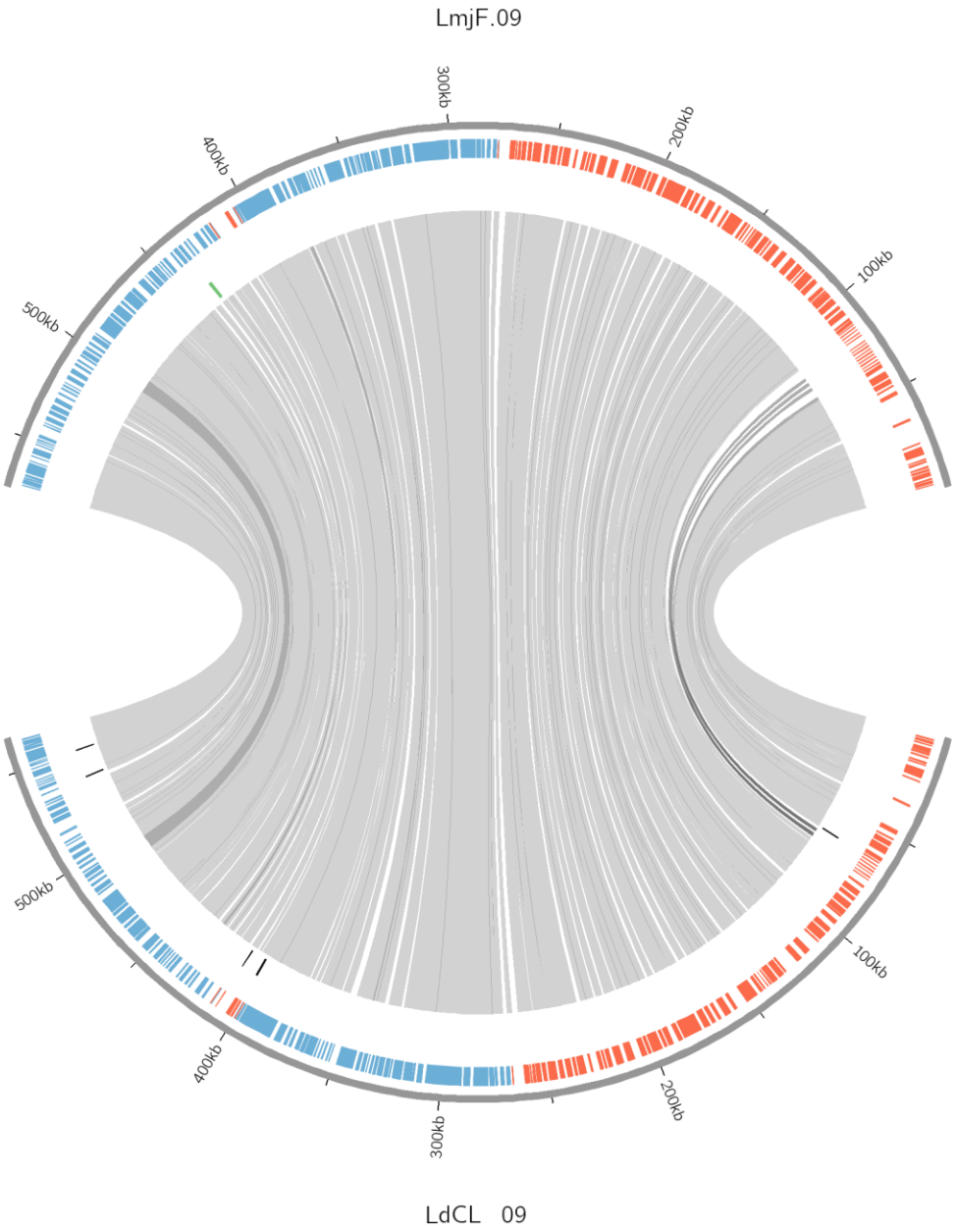

LmjF.04
300kb
200kb
100kb
400kb
400kb
100kb
300kb
200kb
LdCL 04

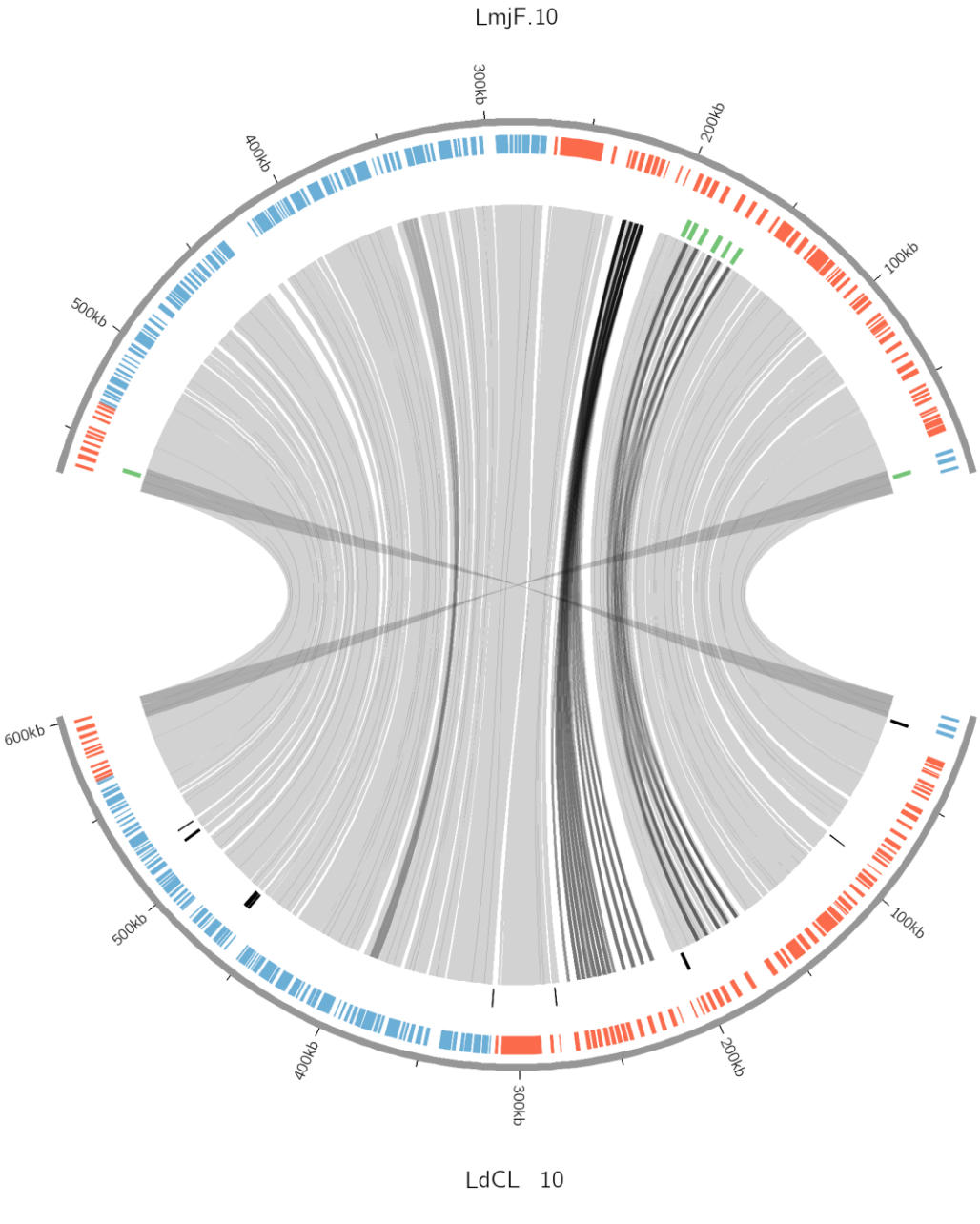

LmjF.05
200kb
100kb
400kb
300kb
LdCL 05

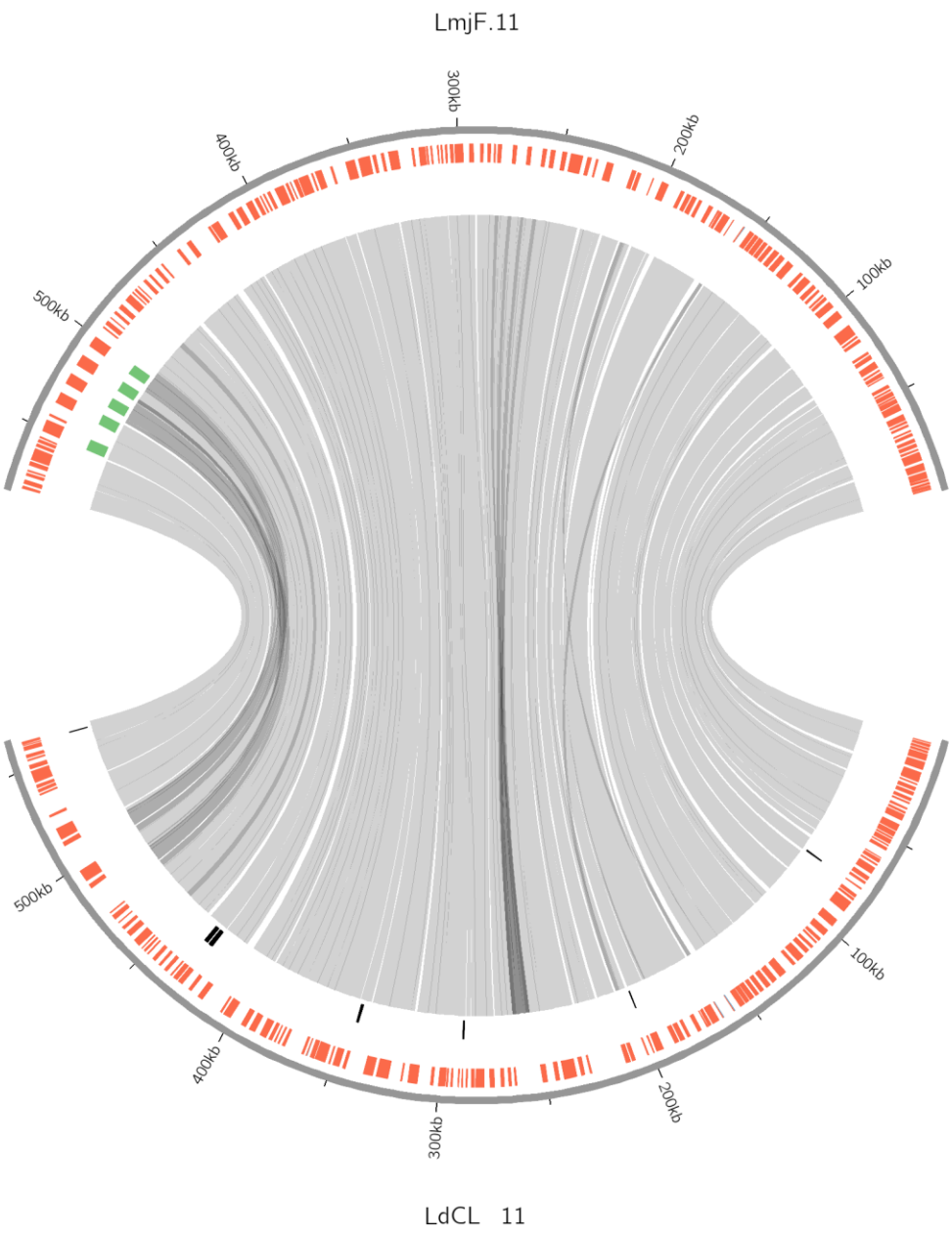

LmjF.06
300kb
200kb
100kb
400kb
500kb
500kb
400kb
300kb
200kb
100kb
LdCL 06

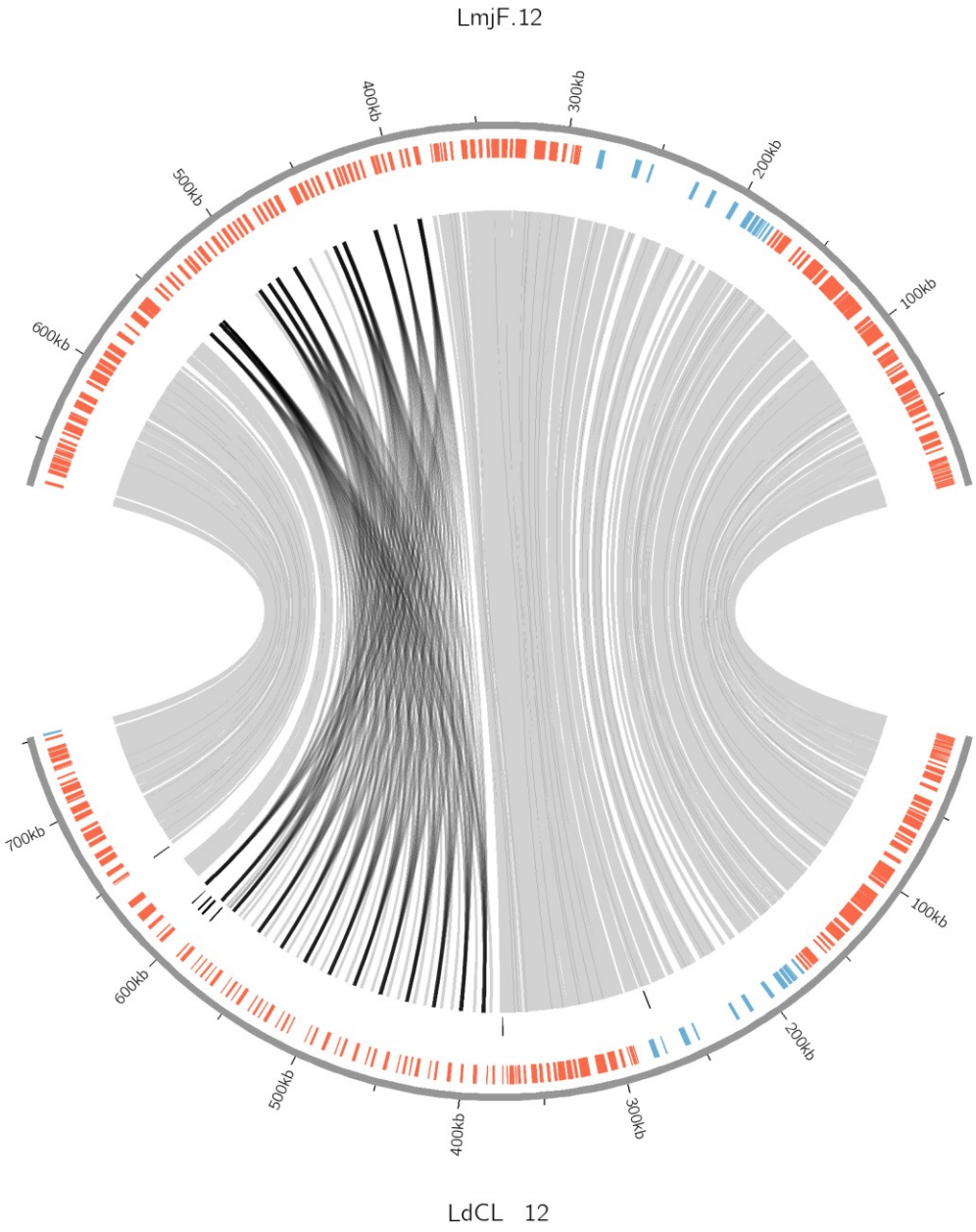

LmjF.07
300kb
200kb
100kb
500kb
400kb
LdCL 07

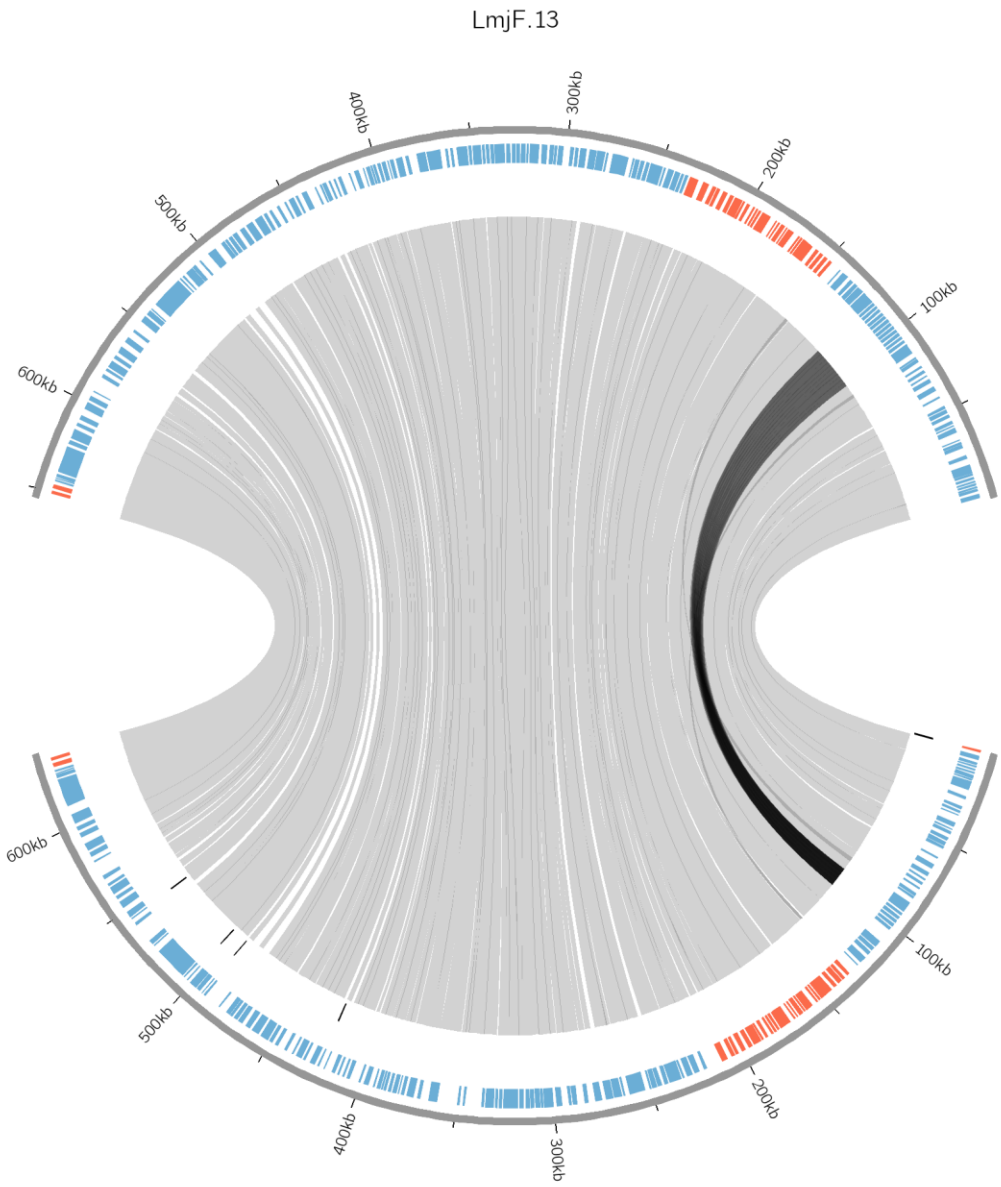

LmjF.08
300kb
200kb
100kb
500kb
400kb
500kb
100kb
400kb
300kb
200kb
LdCL 08

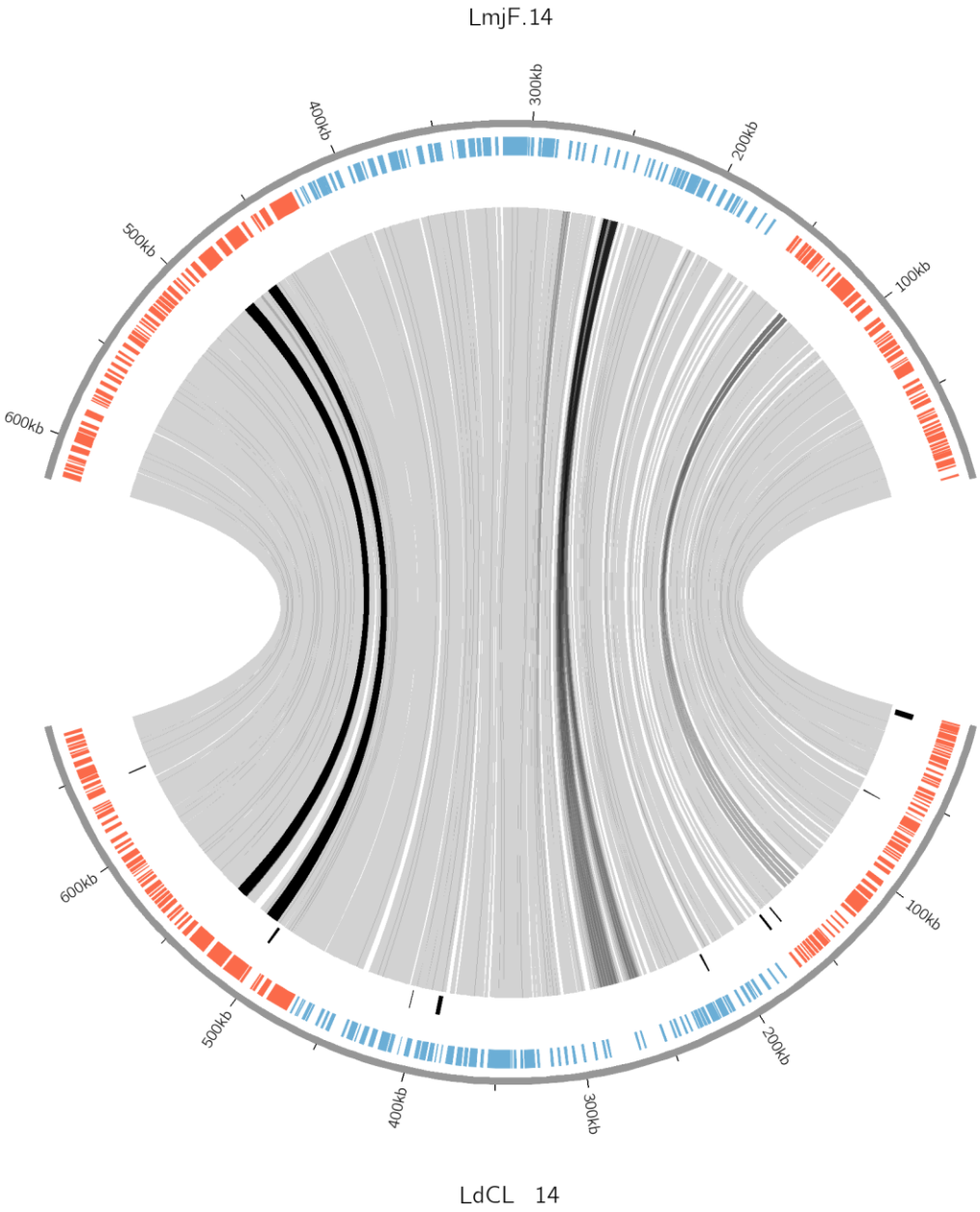

LmjF.09
300kb
200kb
100kb
400kb
500kb
500kb
100kb
400kb
300kb
200kb
LdCL 09

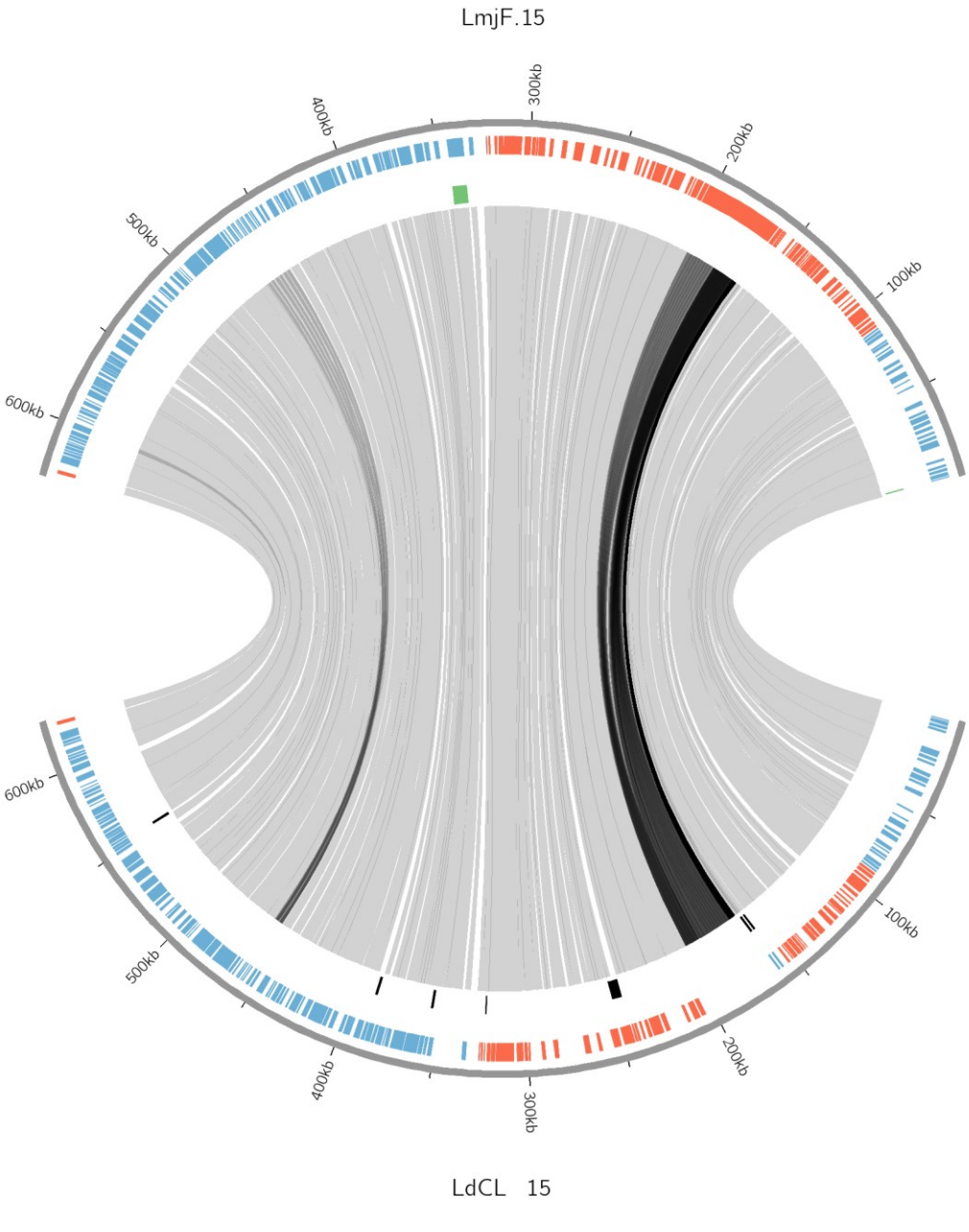

LmjF.10
300kb
200kb
100kb
400kb
500kb
600kb
LdCL 10

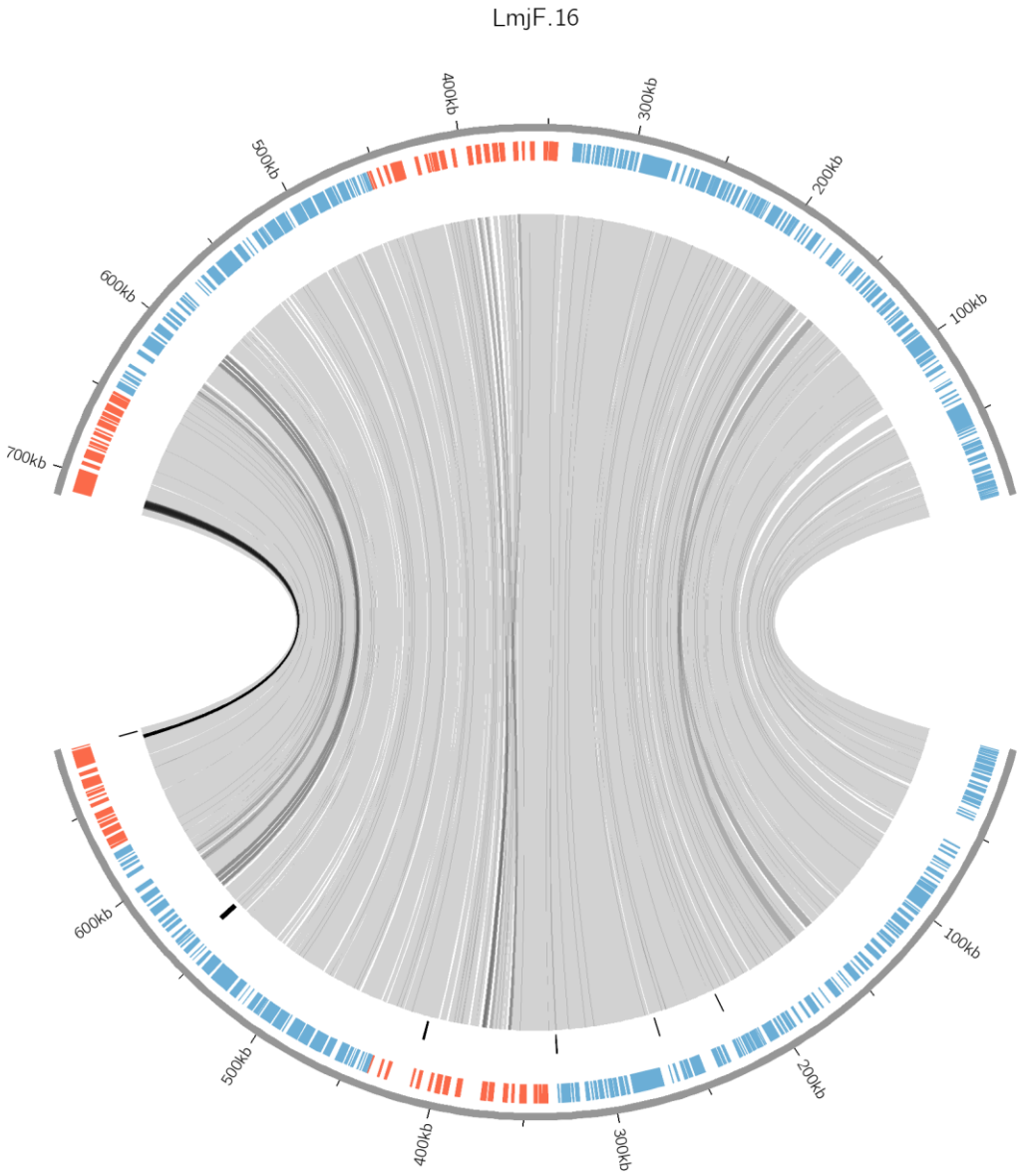

LmjF.11
300kb
200kb
100kb
500kb
400kb
500kb
100kb
400kb
300kb
200kb
LdCL 11

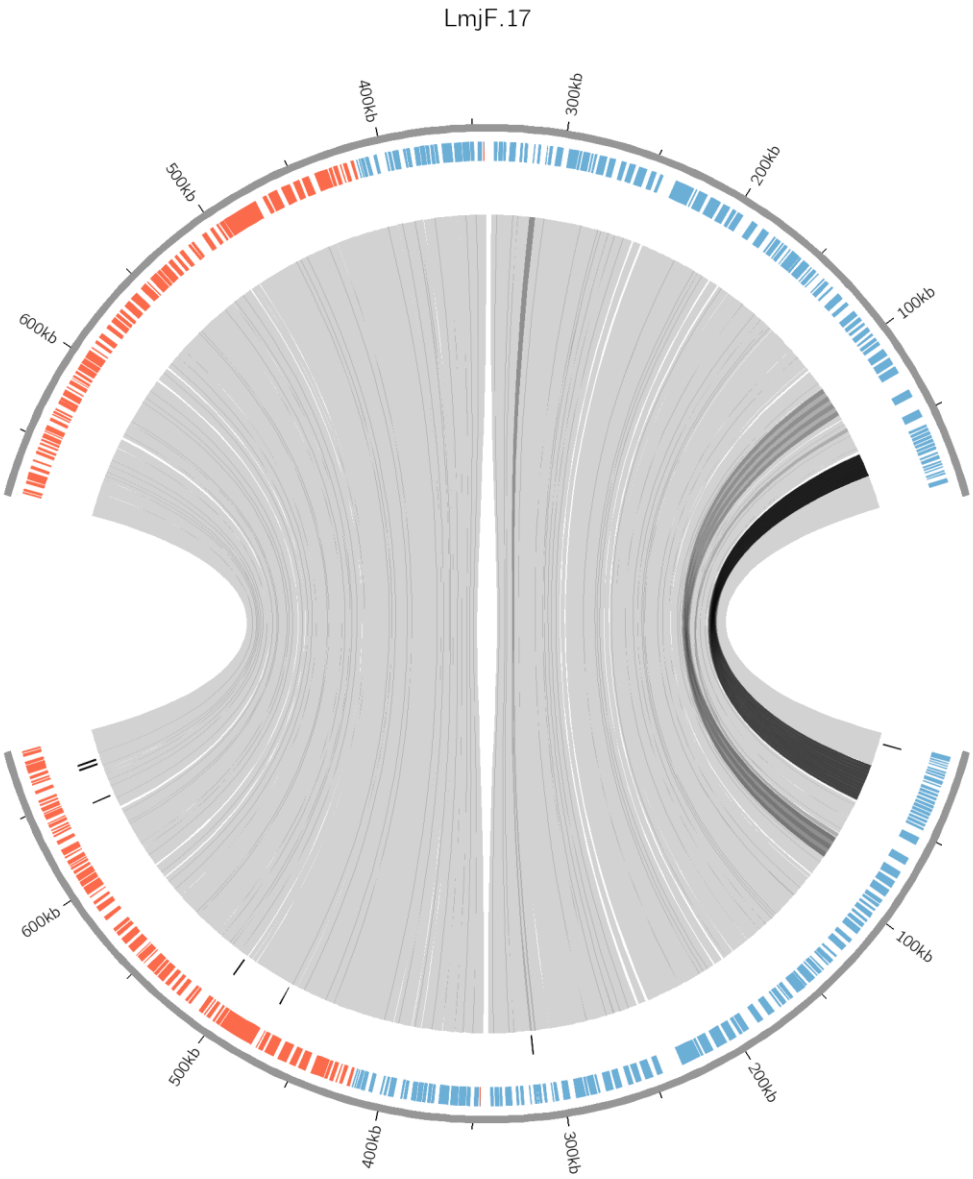

LmjF.12
400kb
300kb
500kb
200kb
600kb
100kb
700kb
100kb
600kb
200kb
500kb
300kb
400kb
LdCL 12

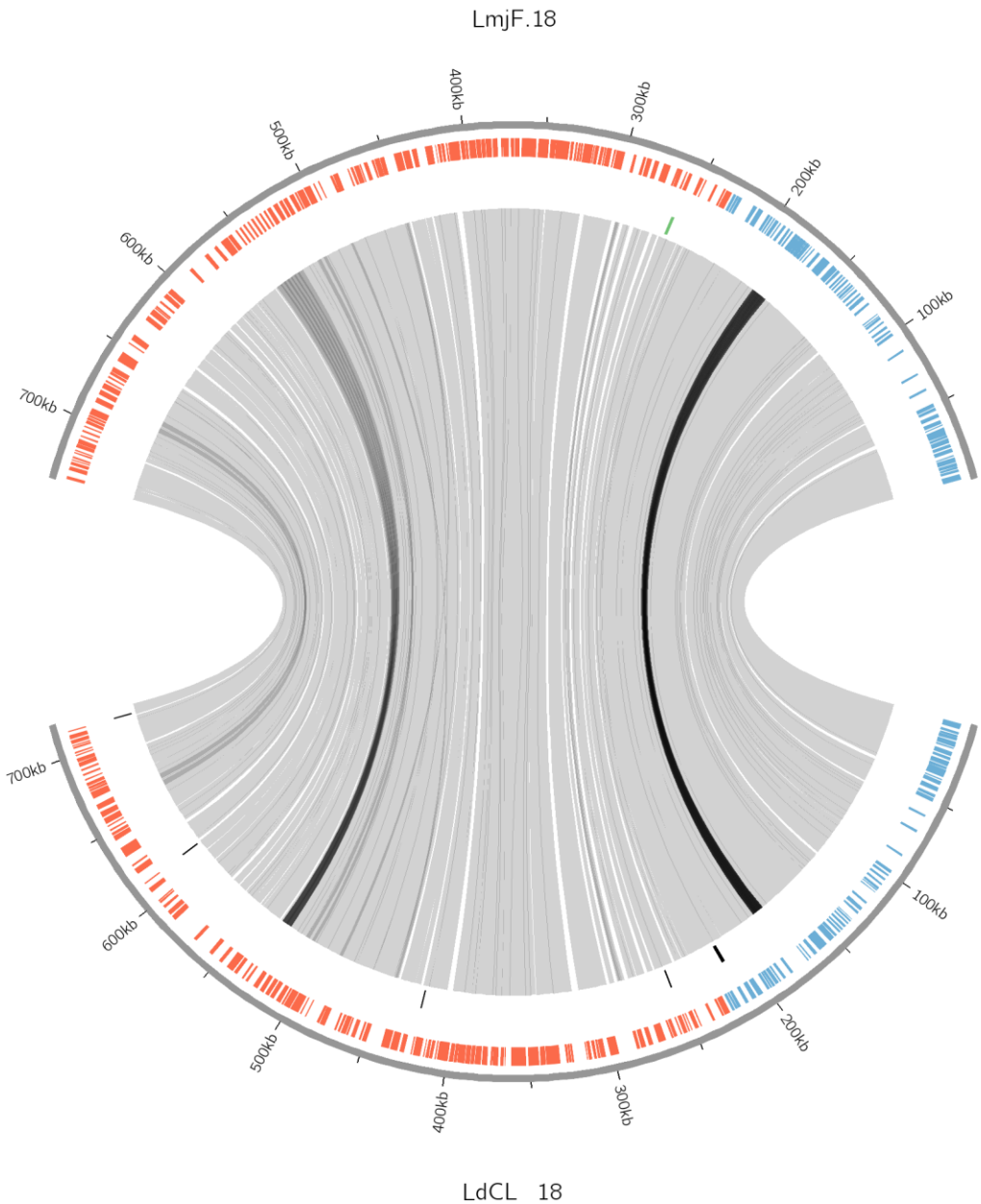

LmjF.13
300kb
200kb
100kb
400kb
500kb
600kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 13

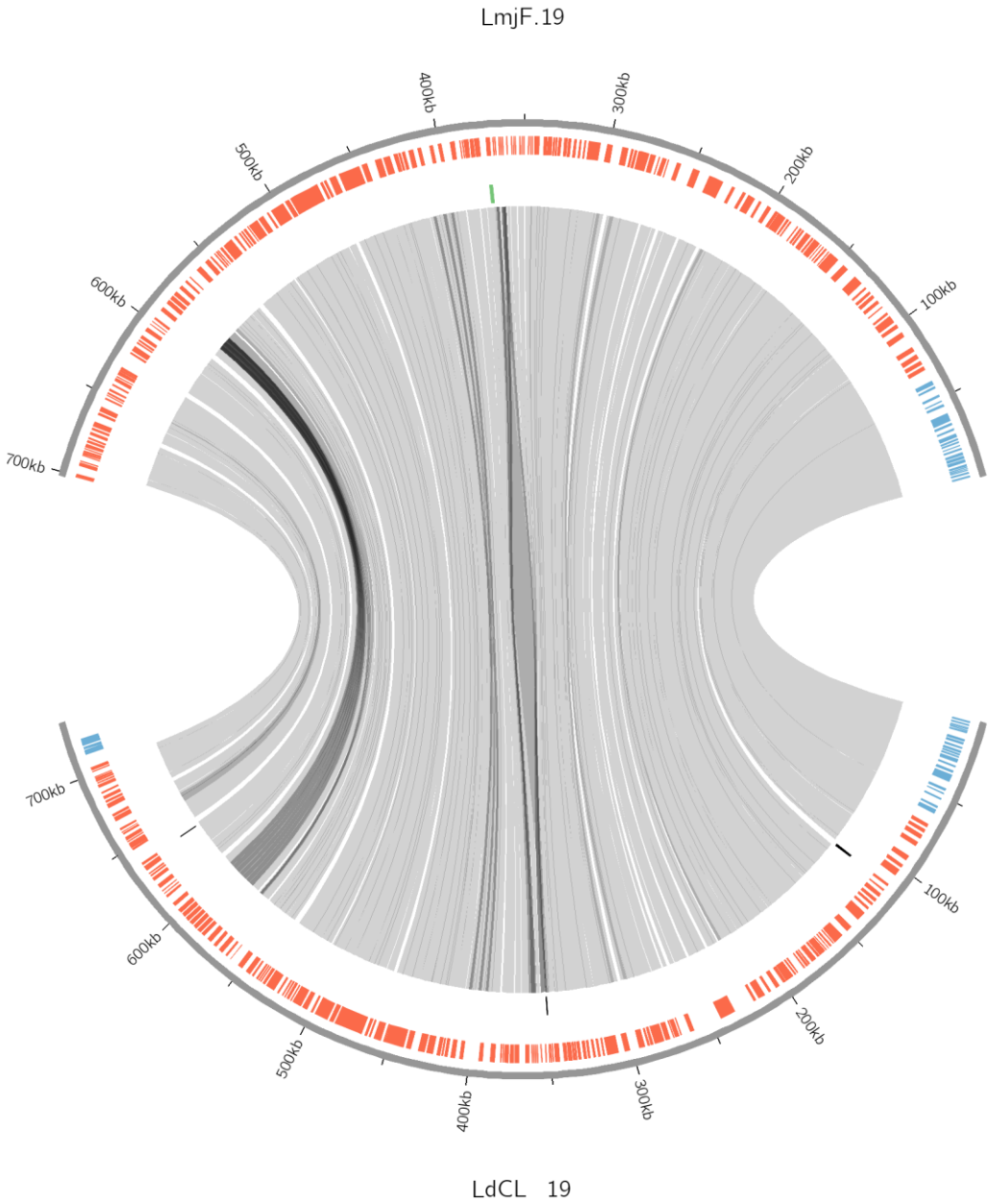

LmjF.14
300kb
200kb
100kb
400kb
500kb
600kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 14

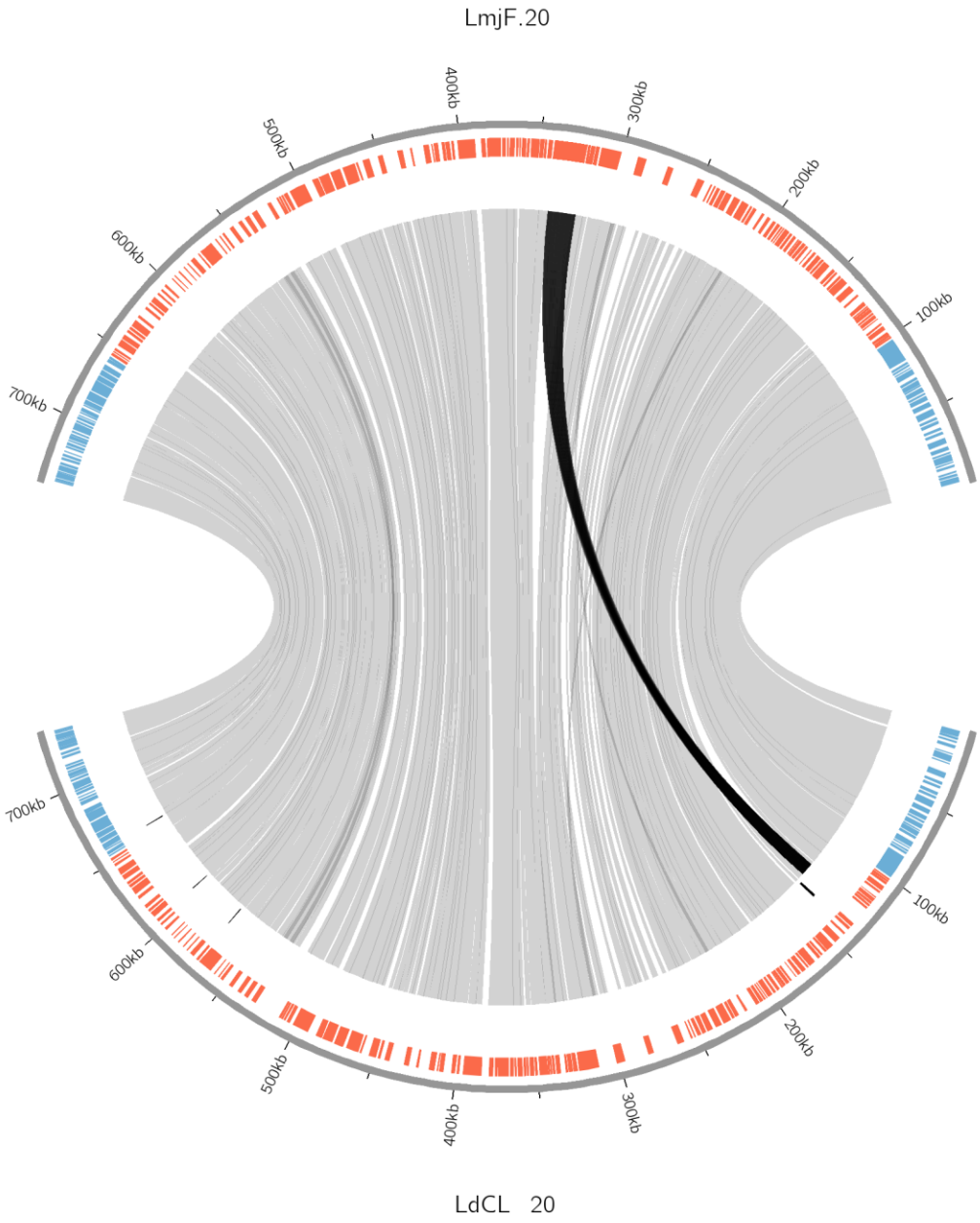

LmjF.15
300kb
200kb
100kb
400kb
500kb
600kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 15

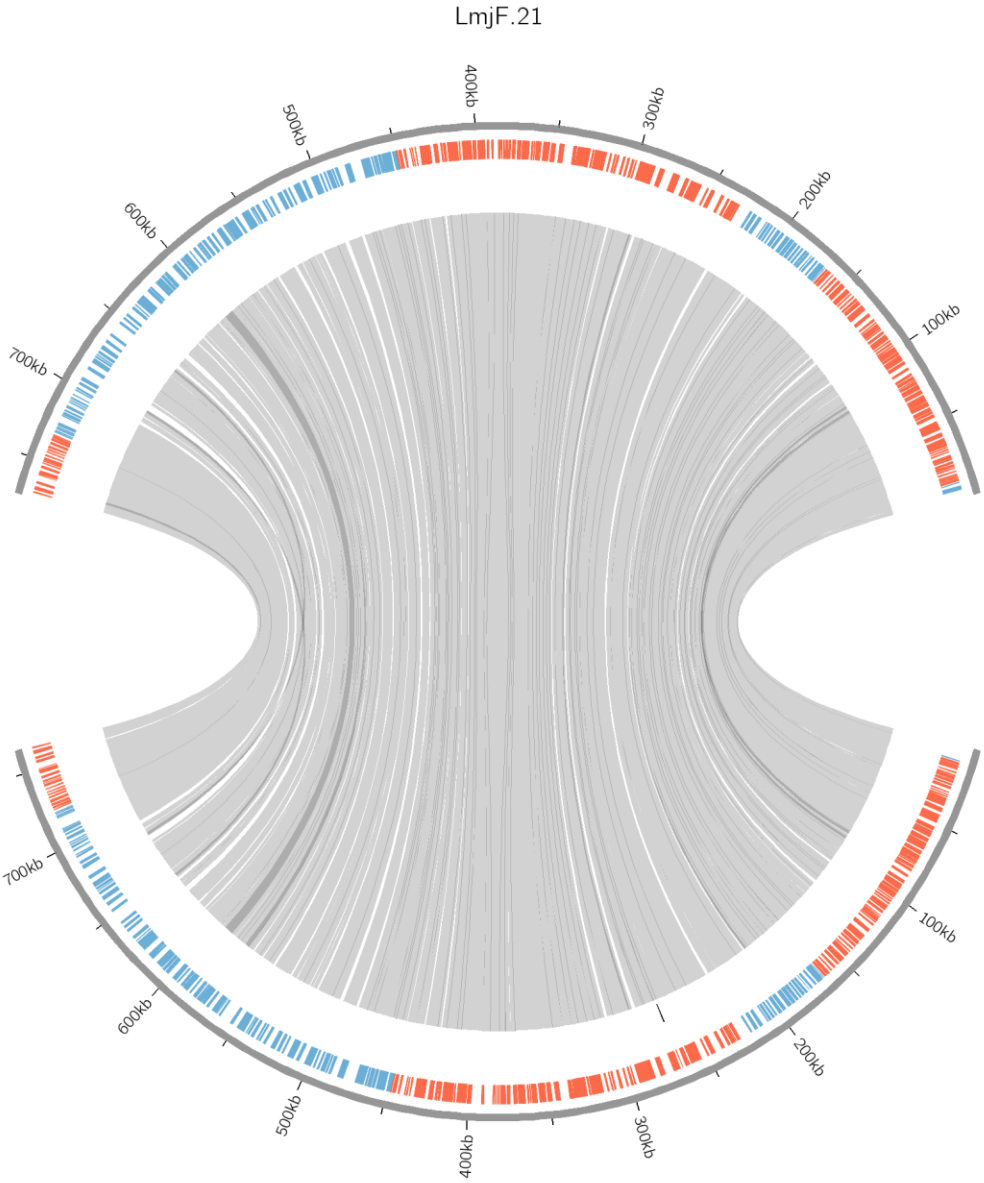

LmjF.16
400kb
300kb
500kb
200kb
600kb
100kb
700kb
600kb
100kb
500kb
200kb
400kb
300kb
LdCL 16

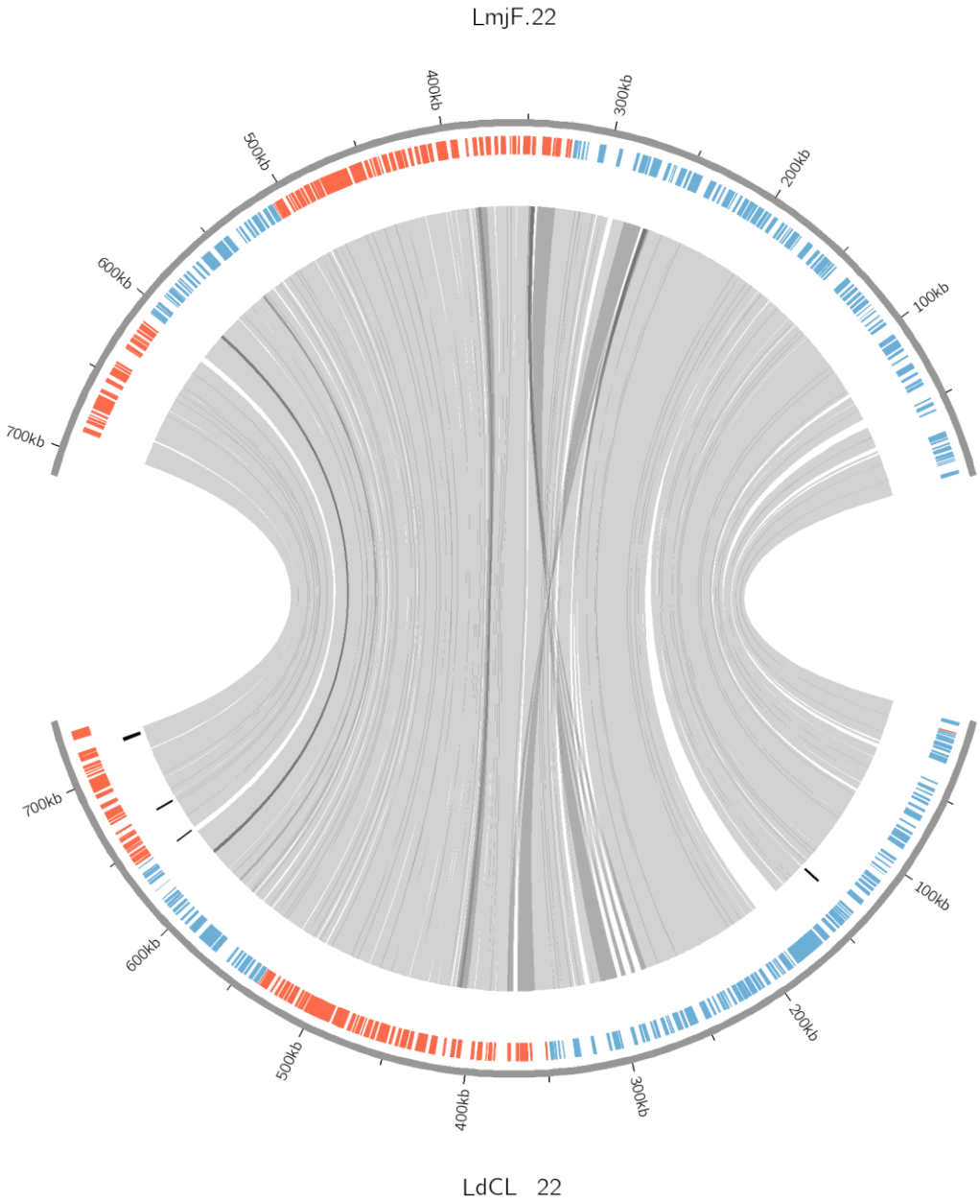

LmjF.17
400kb
300kb
200kb
100kb
500kb
600kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 17

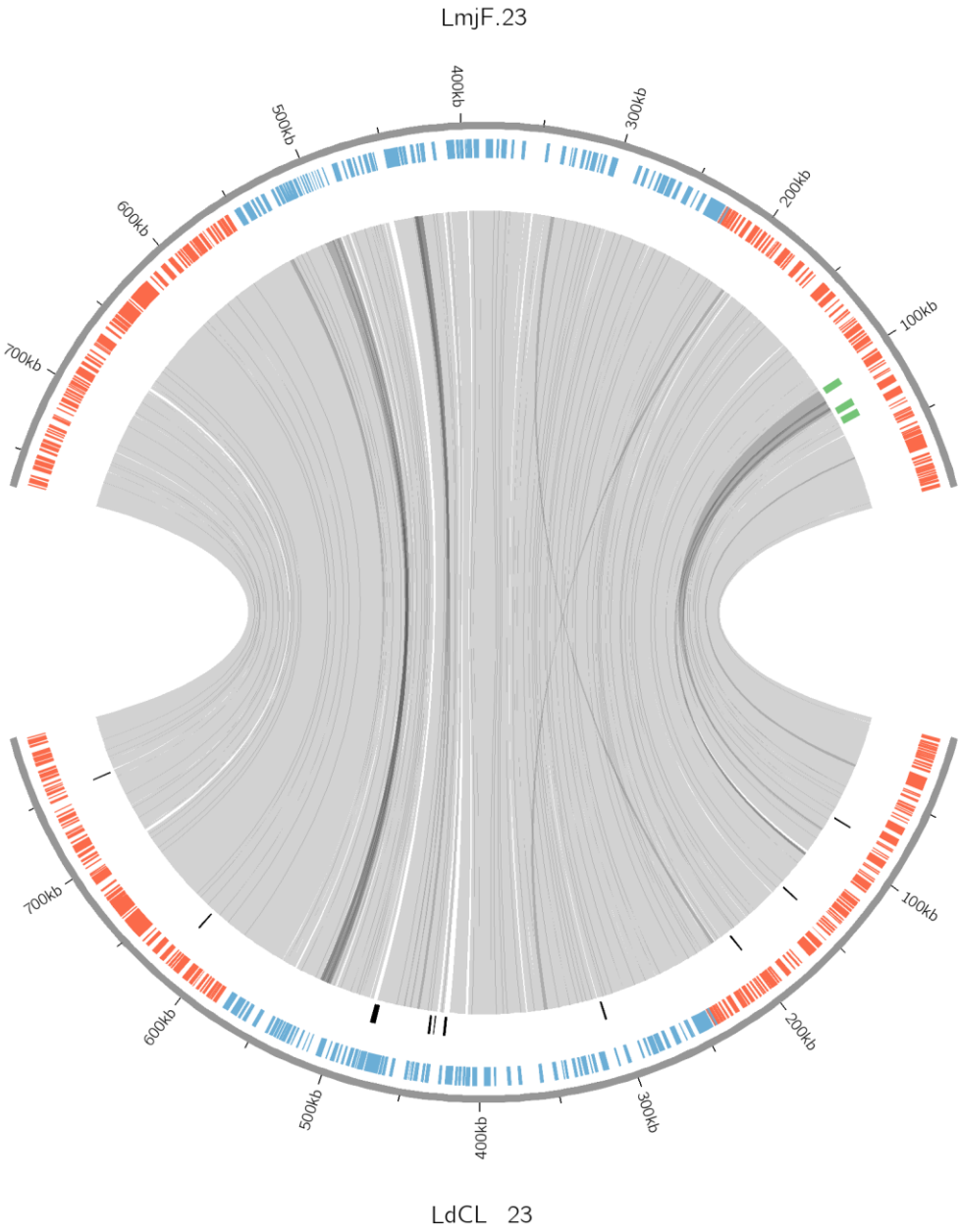

LmjF.18
400kb
300kb
200kb
100kb
500kb
600kb
700kb
700kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 18

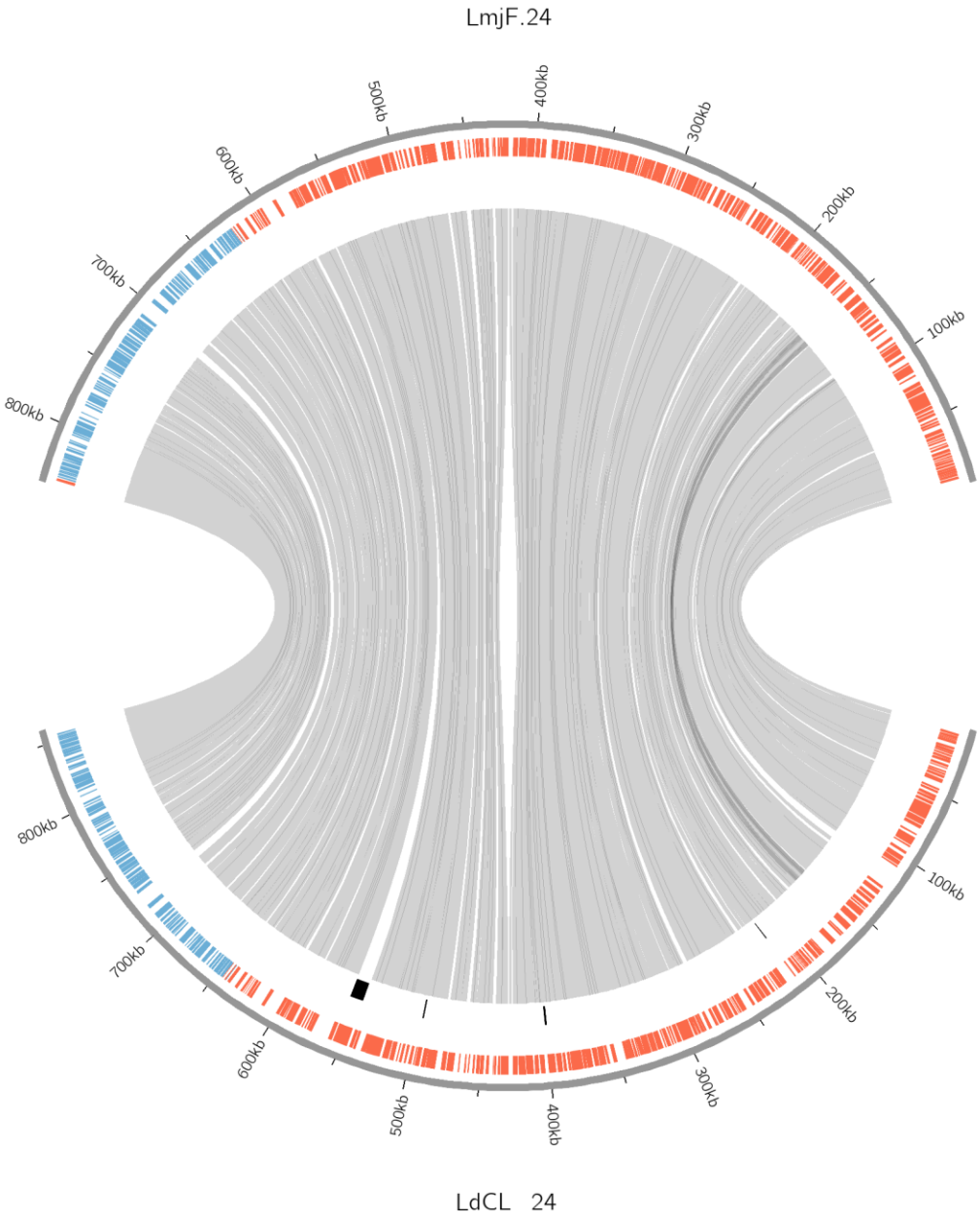

LmjF.19
400kb
300kb
500kb
200kb
600kb
100kb
700kb
700kb
600kb
100kb
500kb
200kb
400kb
300kb
LdCL 19

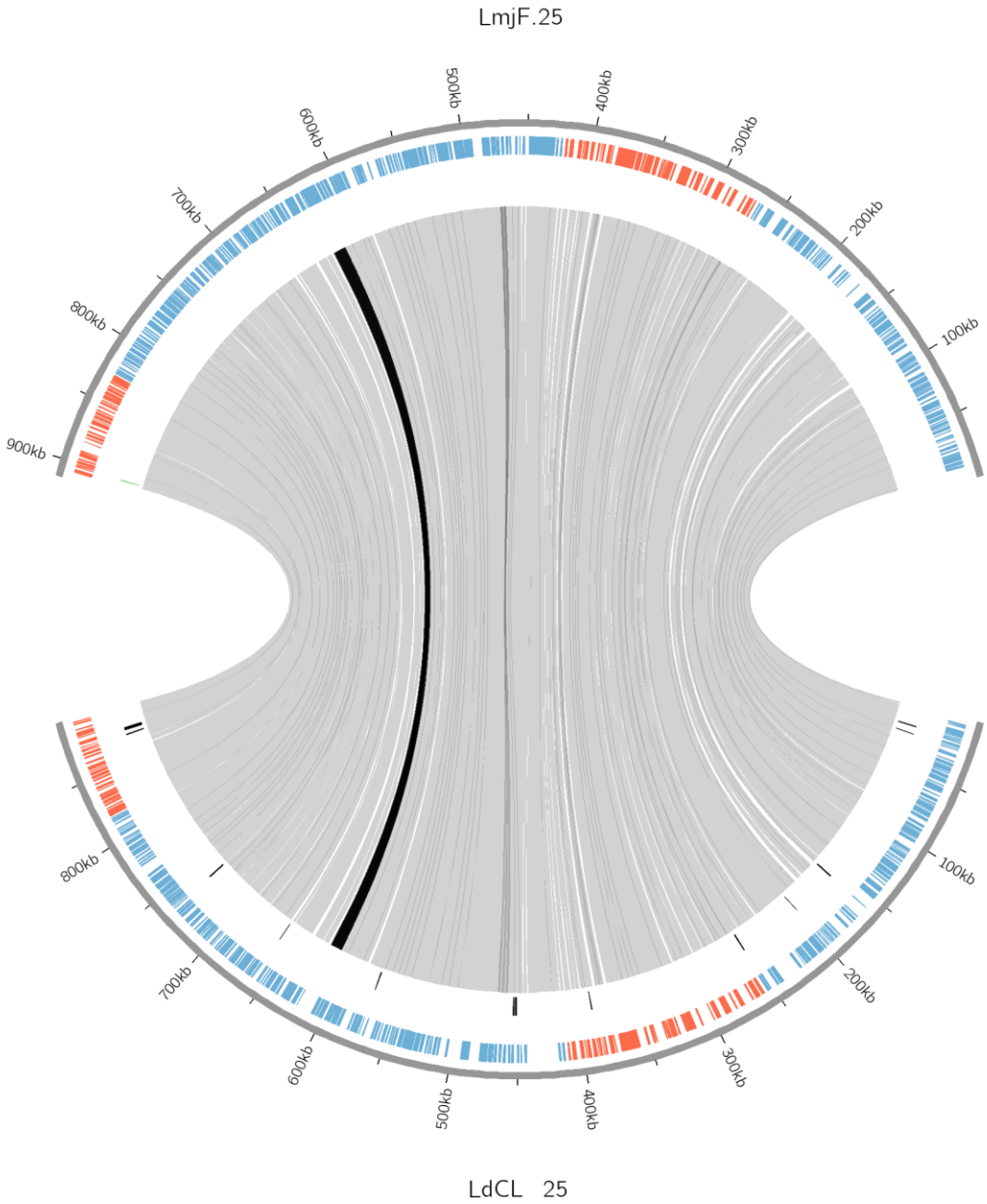

LmjF.20
400kb
300kb
500kb
200kb
600kb
100kb
700kb
700kb
100kb
600kb
200kb
500kb
300kb
400kb
LdCL 20

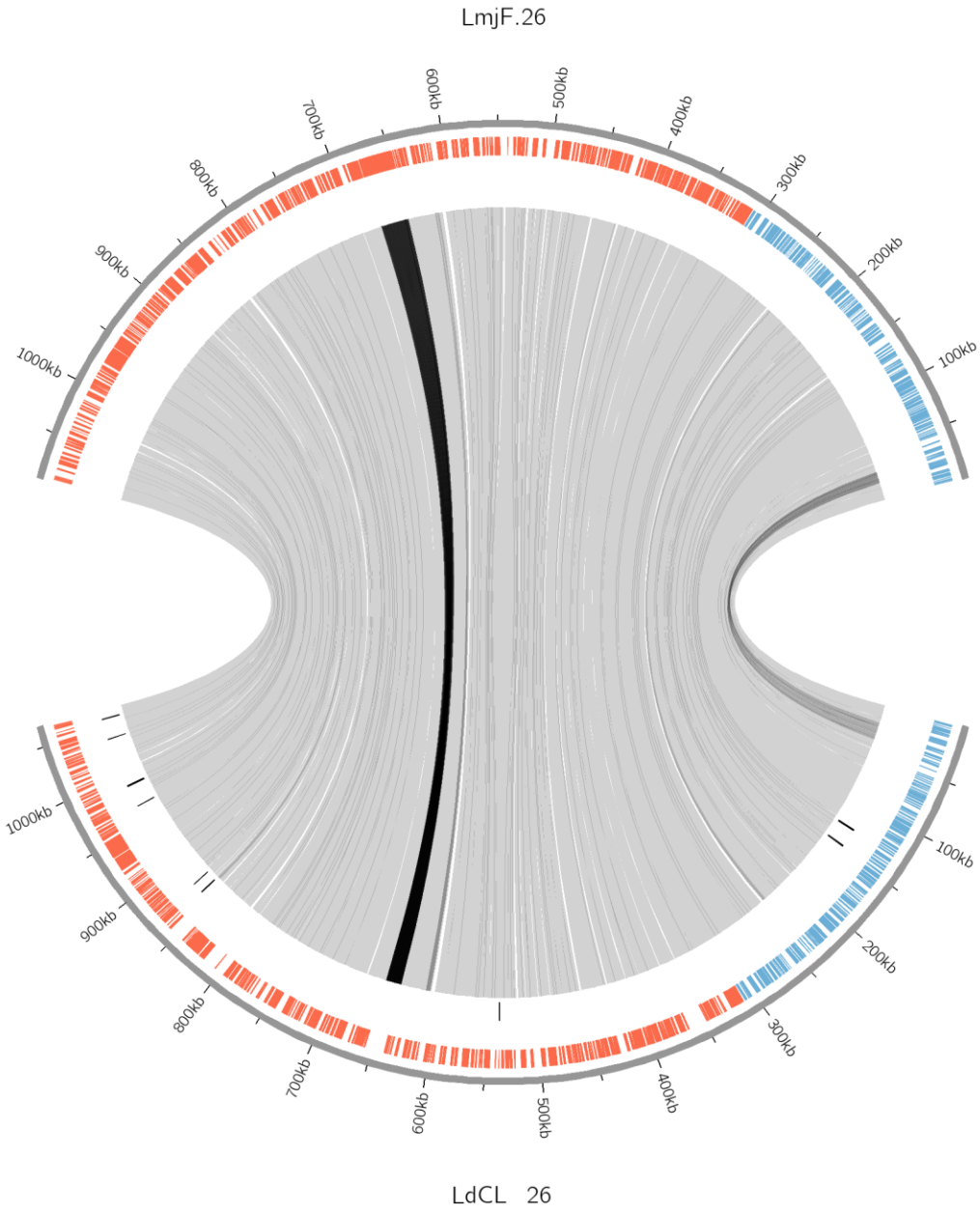

LmjF.21
400kb
300kb
200kb
100kb
500kb
600kb
700kb
700kb
600kb
500kb
400kb
300kb
200kb
100kb
LdCL 21

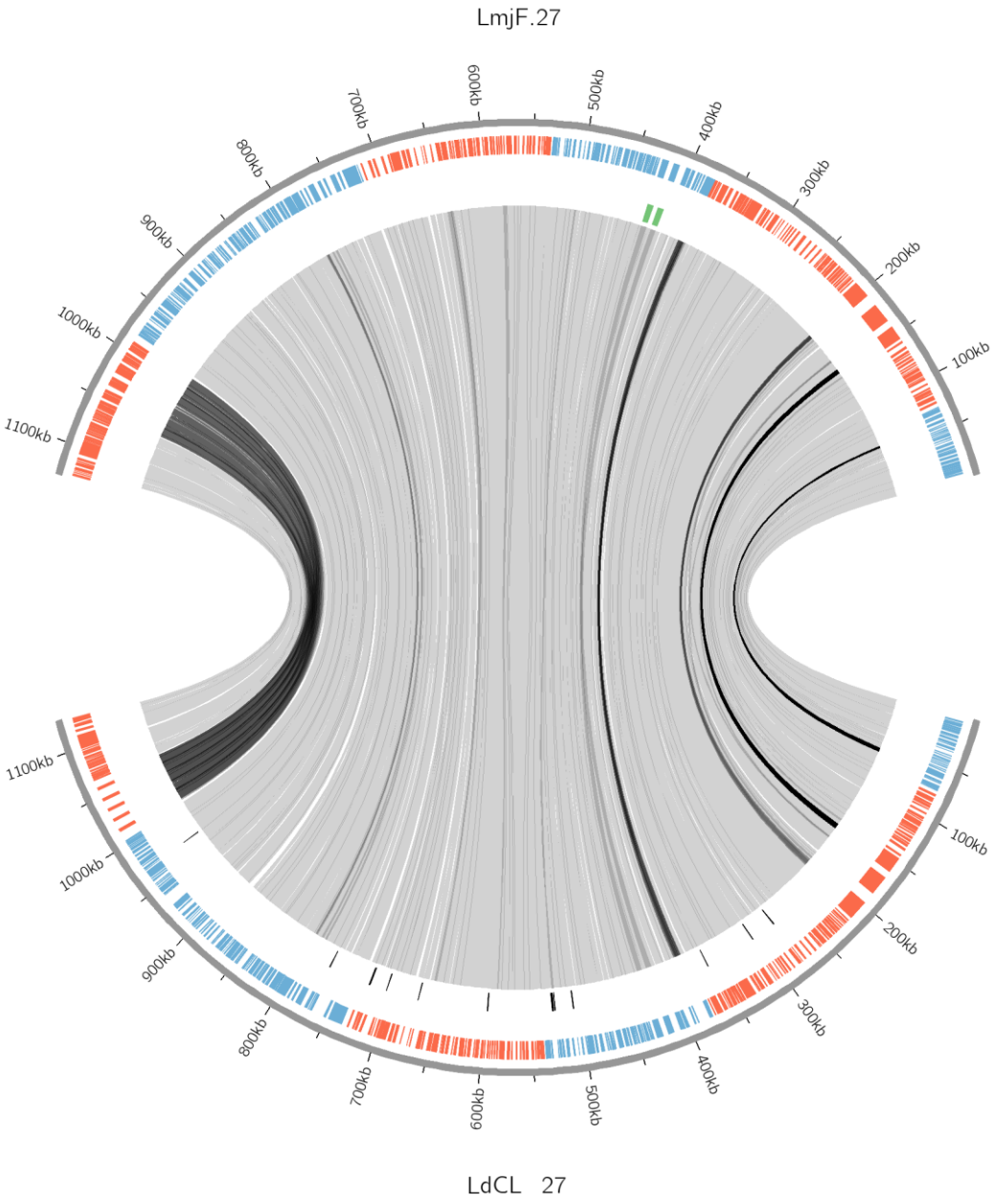

LmjF.22
400kb
300kb
500kb
200kb
600kb
100kb
700kb
700kb
100kb
600kb
200kb
500kb
400kb
300kb
LdCL 22

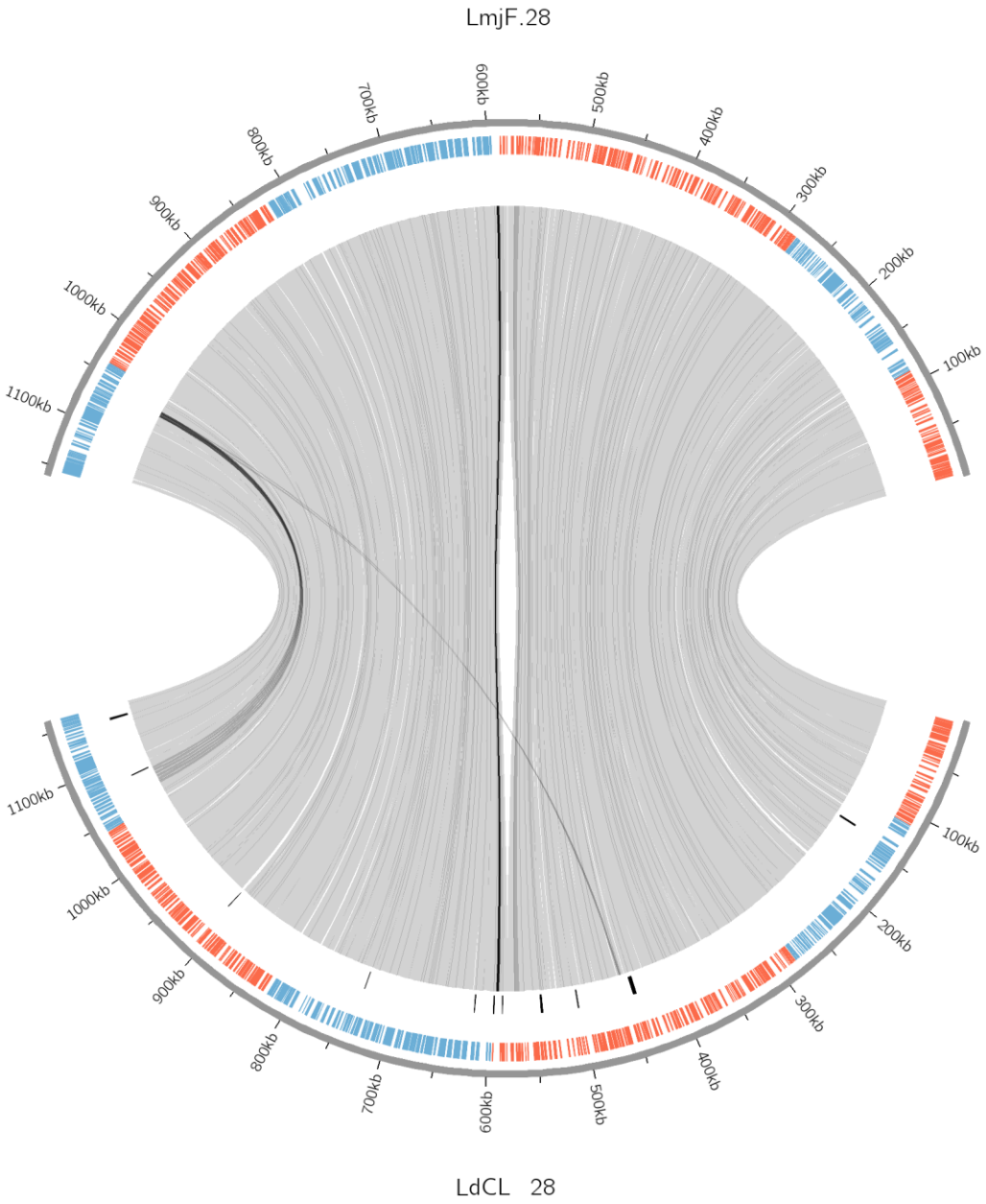

LmjF.23
400kb
300kb
200kb
100kb
500kb
600kb
700kb
LdCL 23

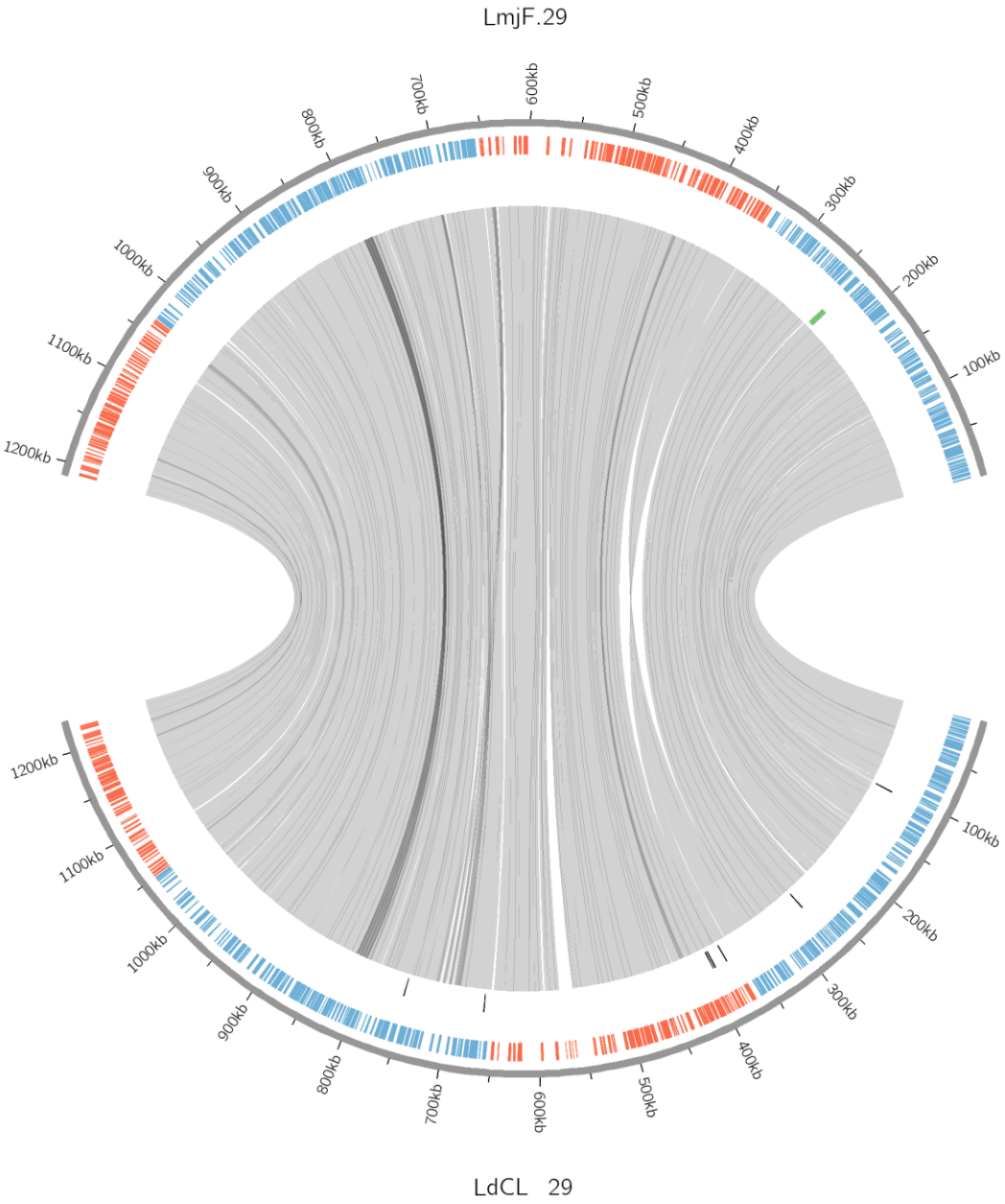

LmjF.24
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
LdCL 24

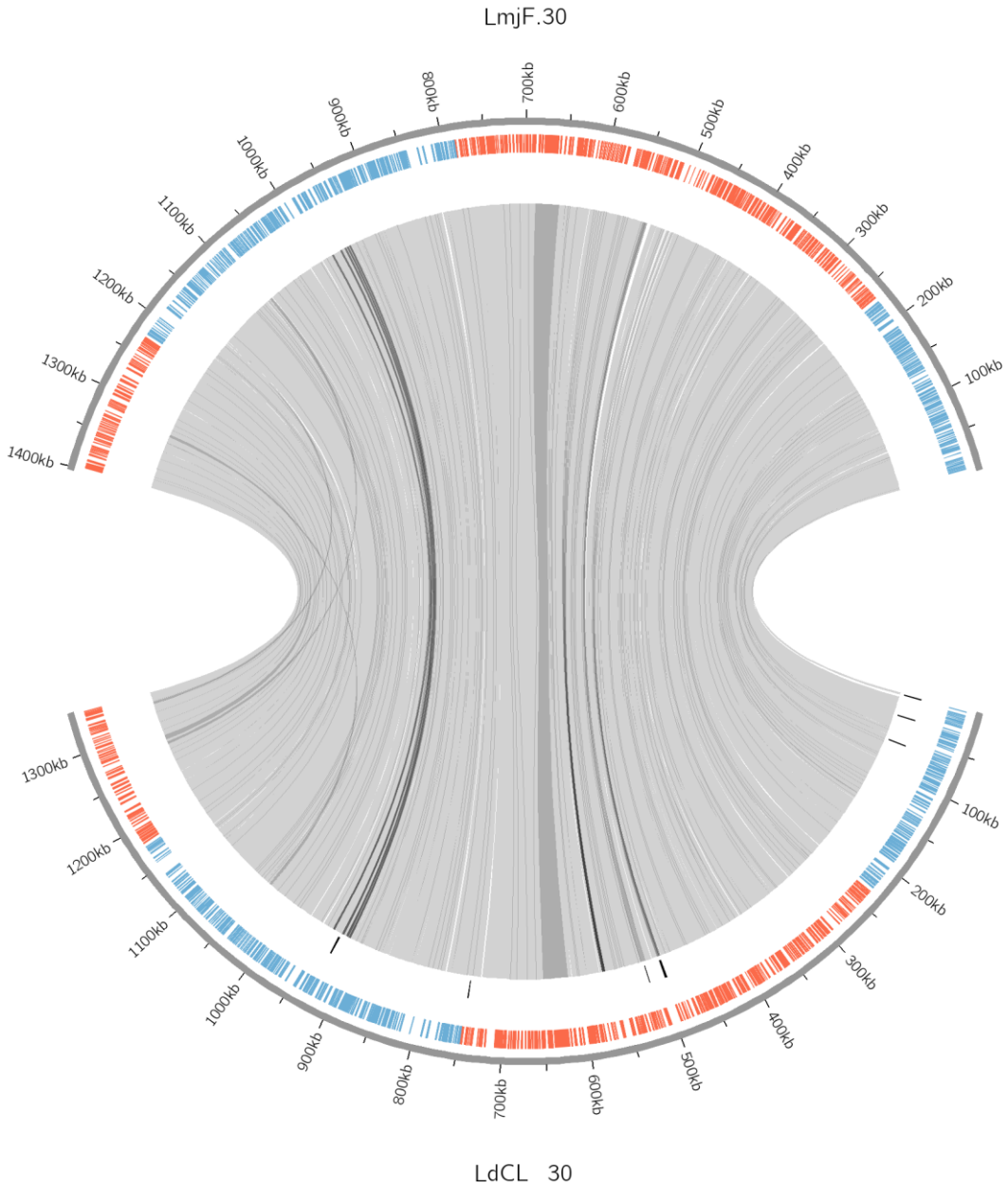

LmjF.25
500kb
400kb
300kb
200kb
100kb
600kb
700kb
800kb
900kb
LdCL 25

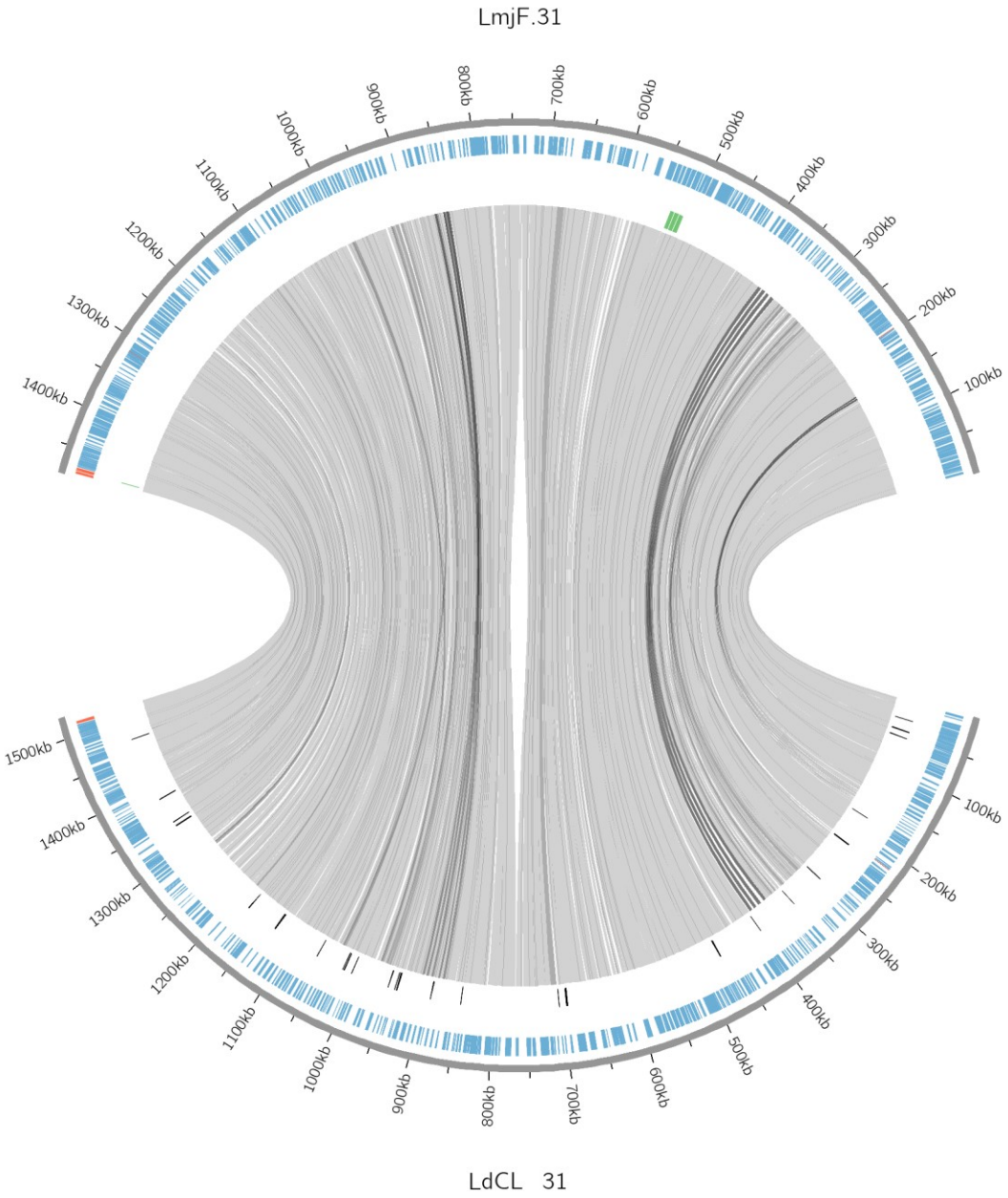

LmjF.26
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
LdCL 26

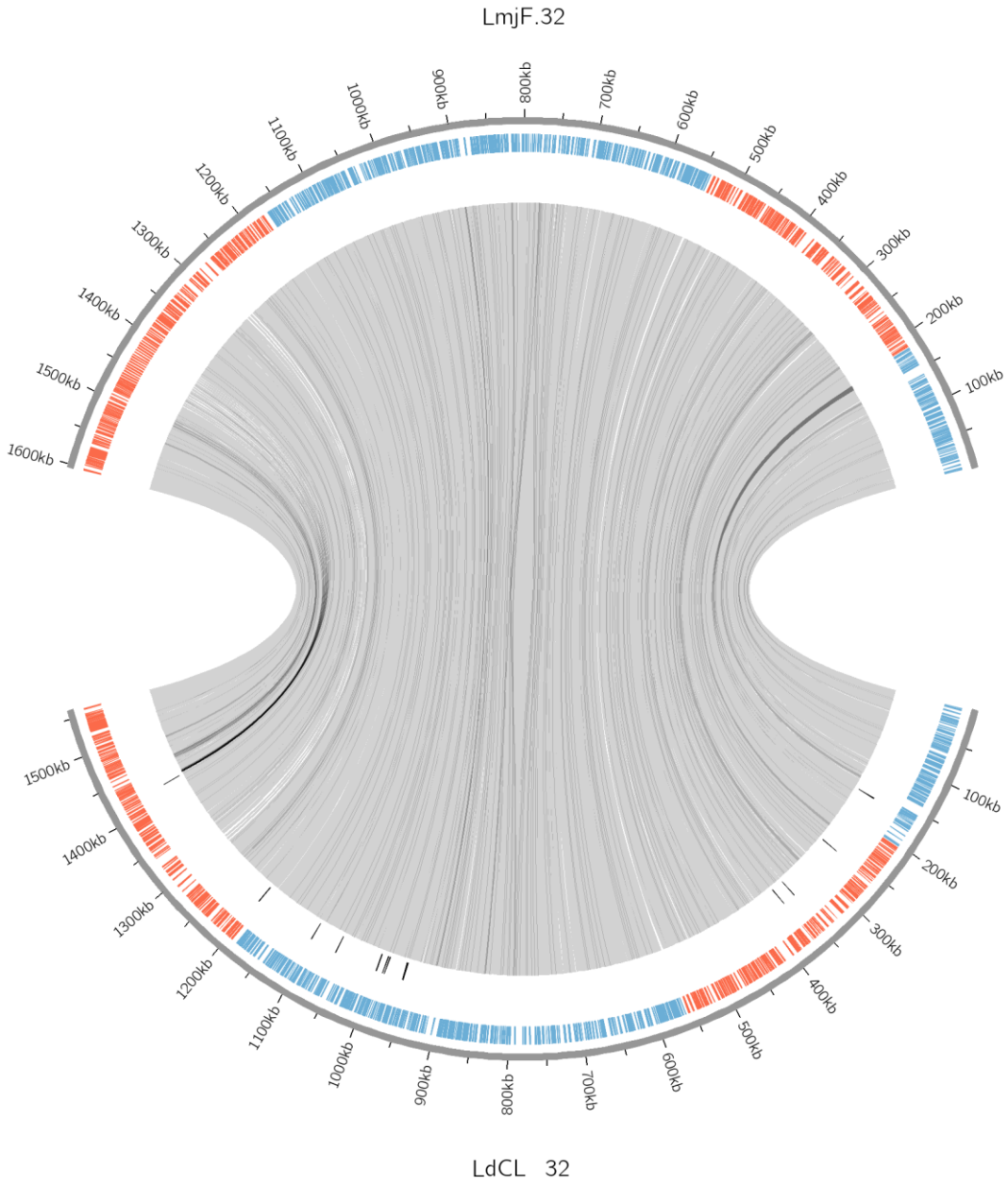

LmjF.27
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
LdCL 27

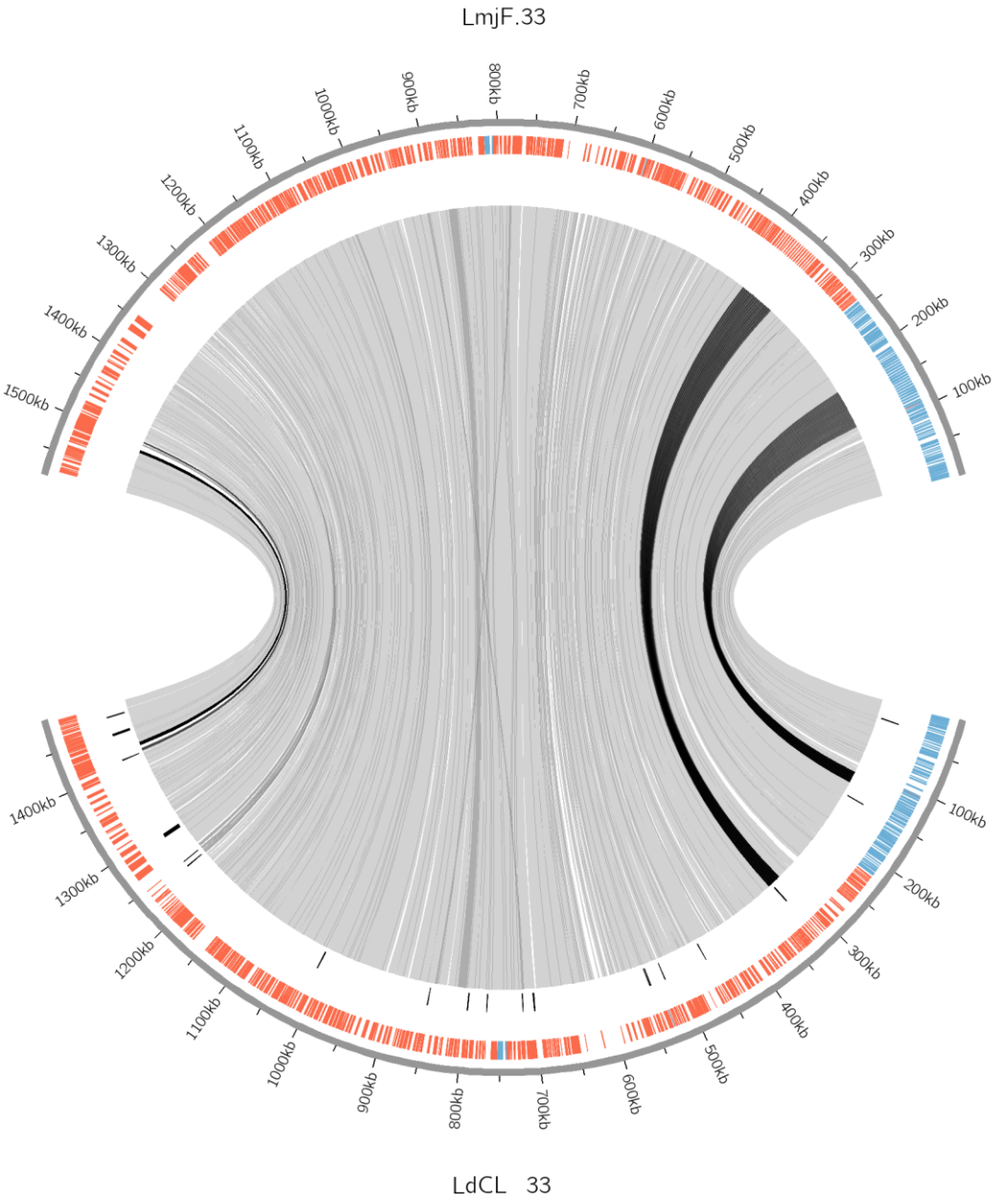

LmjF.28
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
LdCL 28

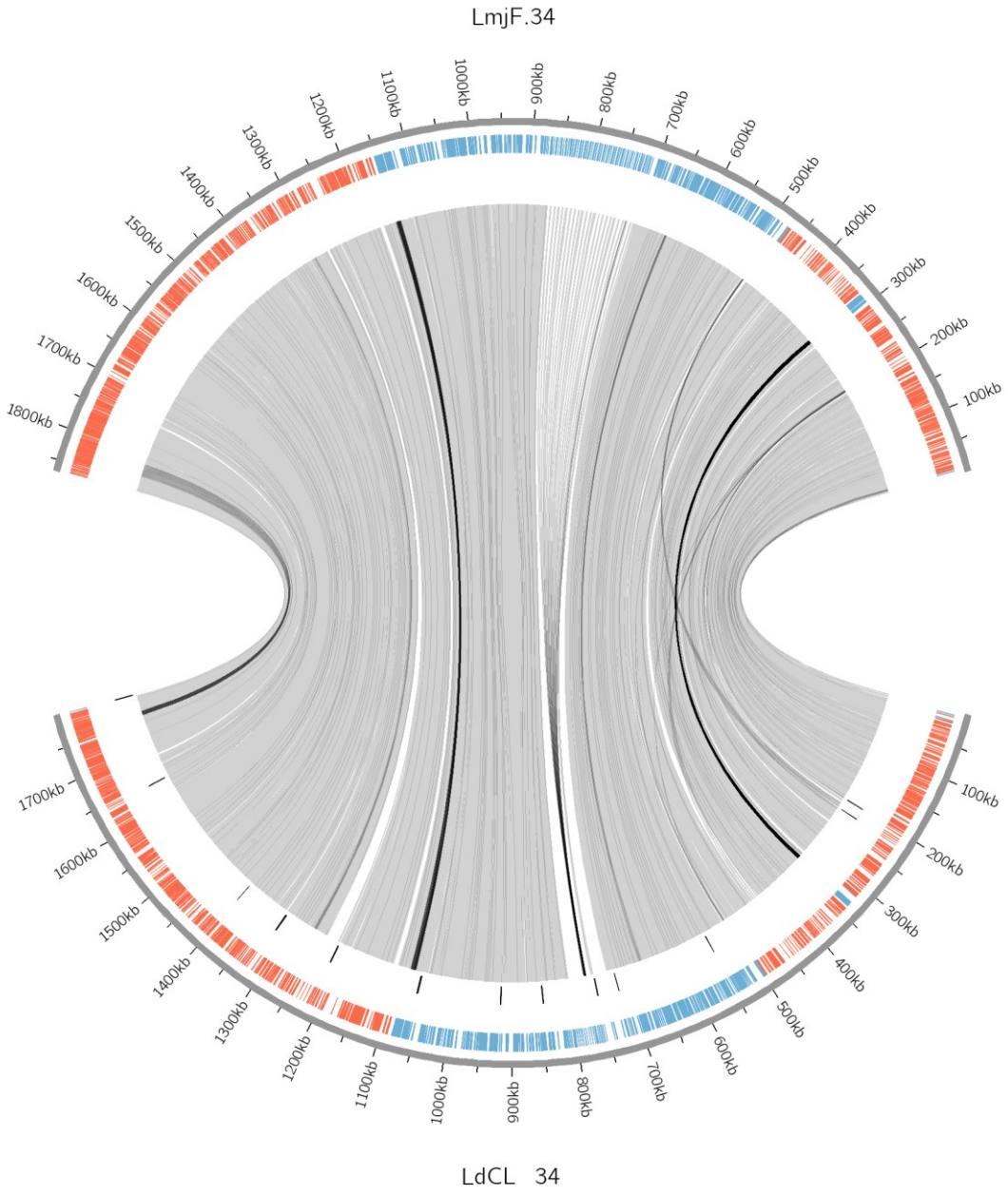

LmjF.29
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
LdCL 29

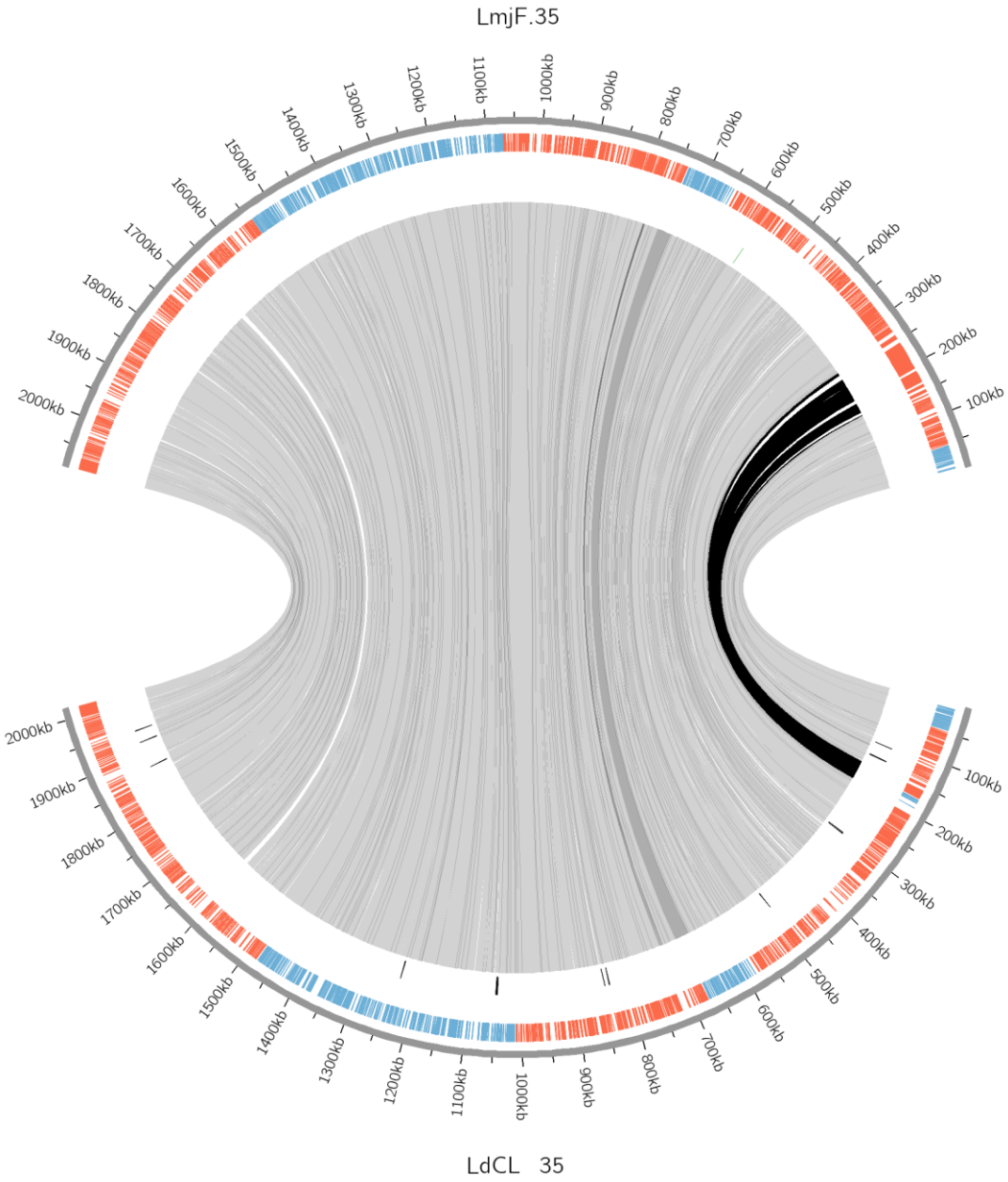

LmjF.30
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
LdCL 30

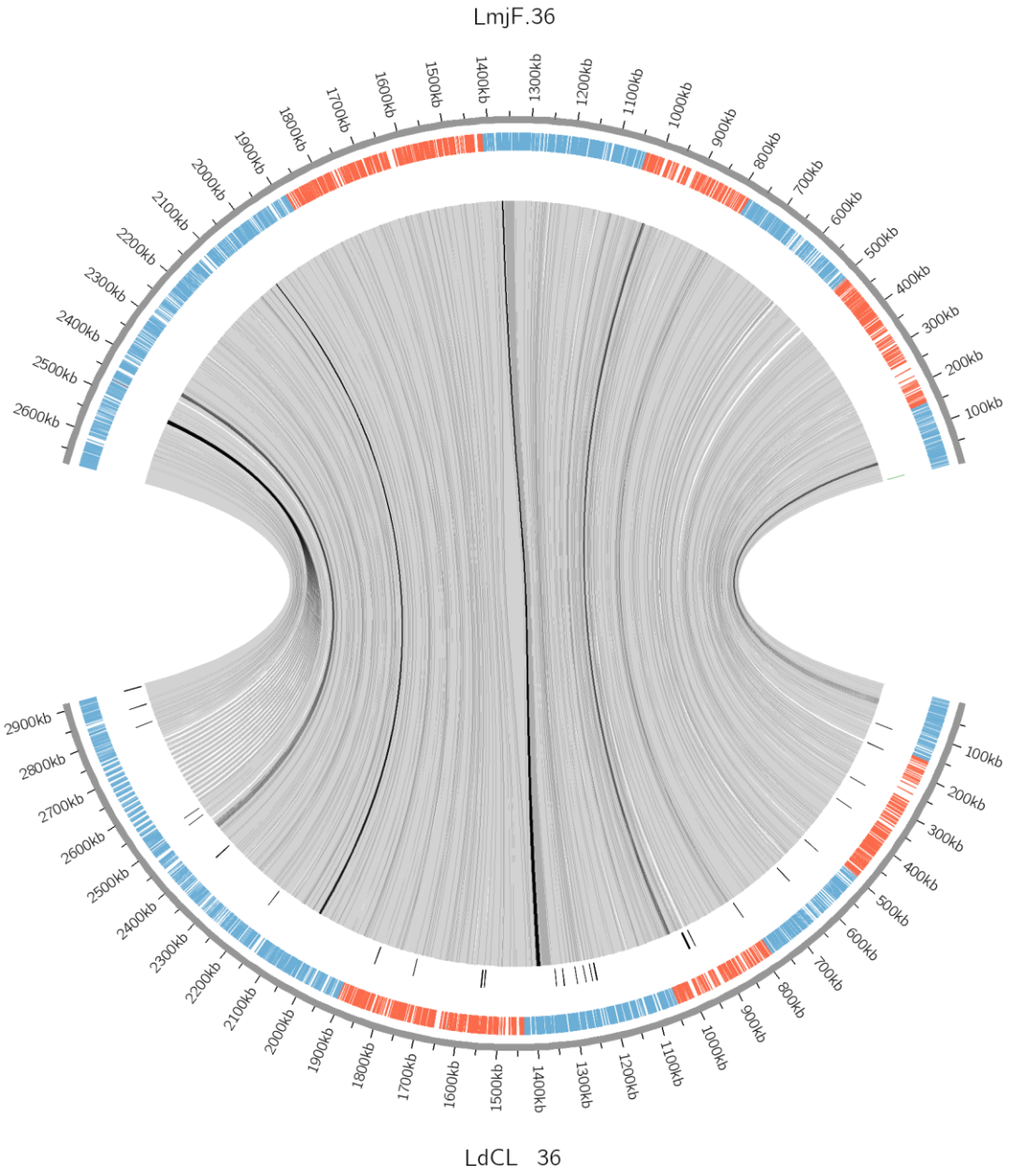

LmjF.31
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
1500kb
LdCL 31


LmjF.32
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
1500kb
1600kb
LdCL 32

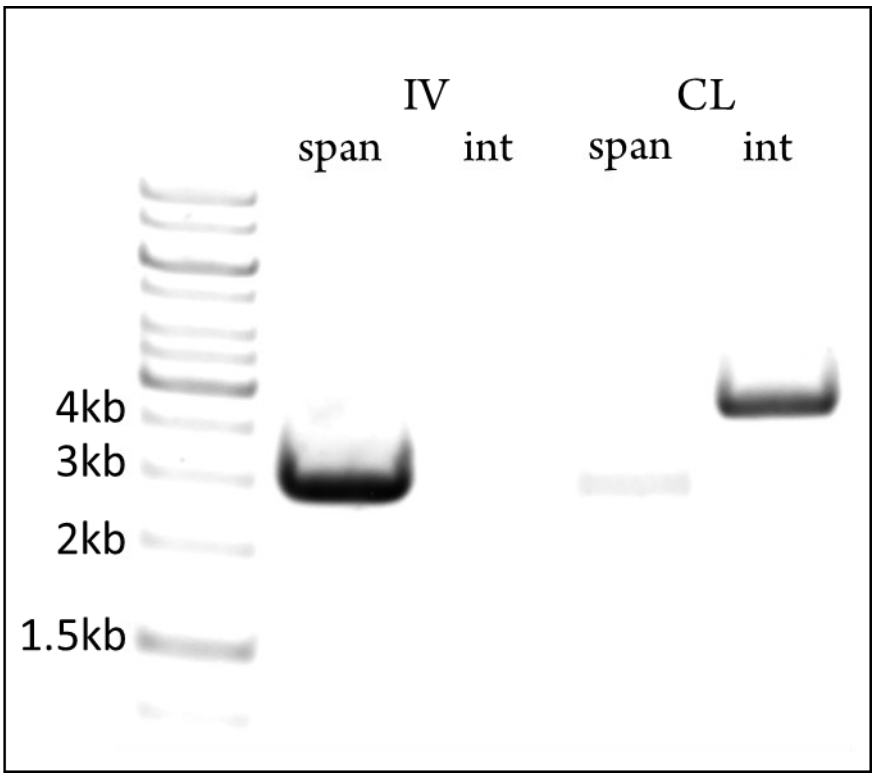

LmjF.33
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
1500kb
LdCL 33


LmjF.34
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
1500kb
1600kb
1700kb
1800kb
LdCL 34


LmjF.35
100kb
200kb
300kb
400kb
500kb
600kb
700kb
800kb
900kb
1000kb
1100kb
1200kb
1300kb
1400kb
1500kb
1600kb
1700kb
1800kb
1900kb
2000kb
LdCL 35



Annotations from both species were aligned to their respective genome and compared to each other. Each chromosome is represented as a half circle. ORFs are colored according to their coding strands. Black lines represent pairs of homologous genes between *L. major* (top) and attenuated cutaneous *L. donovani* (bottom).



Supplementary Figure 2.S4 Confirmation of a 25kb deletion on chromosome 36

Primers were designed to amplify a fragment spanning across the deletion (span) and to amplify a region internal to the deletion (int). The IV strain only produced an amplicon from either end of the deletion indicating the 28kb region collapsed to 3kb but was unable to produce any amplicon internal to the deleted region, indicating the 25kb region was absent rather than translocated. The CL strain mainly produced the internal region as the entire 28kb fragment is too long to be efficiently resolved by PCR and a faint band also appeared in the span reaction indicating the presence of some cells within the population that also contained the deletion.

Supplementary Methods 2.S1

```
Relaxed.spec:
merSize=14
asmOvlErrorRate=0.10
asmUtgErrorRate=0.10
asmCnsErrorRate=0.10
asmCgwErrorRate=0.10
asmOBT=1
asmObtErrorRate=0.08
asmObtErrorLimit=4.5
utgGraphErrorRate=0.05
utgMergeErrorRate=0.05
ovlHashBits=26
ovlHashLoad=0.80
sensitive = 1
maxCoverage=60
gridOptionsOverlap = -pe threads 16 -l mem=2GB
gridOptionsConsensus = -pe threads 16
gridOptionsScript = -pe threads 5
merylThreads = 60
ovlThreads = 60
merylMemory = 12000
ovlStoreMemory = 12000
ovlConcurrency = 1
mbtThreads=60
merOverlapperThreads=60
frgCorrThreads=60
batThreads=60
```

CHAPTER 3: EVIDENCE THAT A NATURALLY OCCURRING SINGLE NUCLEOTIDE POLYMORPHISM IN THE *RAGC* GENE OF *LEISHMANIA DONOVANI* CONTRIBUTES TO REDUCED VIRULENCE

Patrick Lypaczewski*, Wen-Wei Zhang* and Greg Matlashewski

*These authors contributed equally to this work

3.1 Preface

As described in **Chapters 1** and **2**, the RagC GTPase is a key protein of the nutrient sensing pathway upstream of the central TORC1 regulator complex in eukaryotic cells. In *Leishmania,* the role and function of this protein remain largely unknown. Previous studies identified a polymorphism in the gene encoding this protein between cutaneous and visceral strains of *L. donovani* in Sri Lanka. In **Chapter 1**, an additional polymorphism was identified in a gene encoding a member of the downstream TORC1 complex: *Raptor*. In this study, we aimed to elucidate the role of the RagC protein and its cutaneous associated variant in *Leishmania.* Our results shed light on the conservation of the RagC GTPases and TORC1 pathways in *Leishmania,* and suggest these pathways play a key role in determining the progression of the parasitic infection.

- Adapted from: **Patrick Lypaczewski***, Wen-Wei Zhang* and Greg Matlashewski, 2021. Evidence that a naturally occurring single nucleotide polymorphism in the RagC gene of *Leishmania donovani* can reduce virulence. *PLoS Neglected Tropical Diseases,* 15(2): e0009079.

3.2 Abstract

Leishmaniasis is a widespread neglected tropical disease transmitted by infected sand flies resulting in either benign cutaneous infection or fatal visceral disease. *Leishmania donovani* is the principal species responsible for visceral leishmaniasis, yet an atypical *L. donovani* has become attenuated in several countries including Sri Lanka and causes cutaneous leishmaniasis. Previous studies have identified 91 genes altered in the atypical cutaneous *L. donovani* compared to typical visceral disease associated *L. donovani* including mutations in the *RagC* and *Raptor* genes that are part of the eukaryotic conserved TOR pathway and its upstream sensing pathway. In the present study, we investigate whether the RagC R231C mutation present in atypical cutaneous *L. donovani* introduced into the virulent *L. donovani* 1S2D chromosome by CRISPR gene editing could affect virulence for survival in visceral organs. Through bioinformatic analysis, we further investigated the presence of sensing pathway components upstream of TOR in *L. donovani* including RagC complexing proteins, RagA and Raptor. *L. donovani* 1S2D edited to express mutant RagC R231C were viable in promastigote but had reduced visceral parasitemia in infected BALB/c mice. The RagC R231C mutant retained the ability to interact with RagA and gene knockout experiments revealed that although the *RagA* gene was essential, the *RagC* gene was not essential under promastigote culture conditions but was essential for survival in the liver of experimentally infected mice. These results provide evidence that the TOR associated sensing pathway plays a prominent role in *L. donovani* visceral disease and the RagC R231C mutation contributed to the atypical pathology of cutaneous *L. donovani* in Sri Lanka.

Author Summary

The *Leishmania donovani* parasite is transmitted by infected sand flies and is the principal causative agent of visceral leishmaniasis, a neglected tropical disease with no available vaccine.

The infection normally spreads to the liver, spleen and bone marrow and is fatal if untreated. In Sri Lanka however, *L. donovani* causes only skin pathology at the site of the sand fly bite and does not cause visceral disease. Previous genetic analysis of this atypical attenuated strain identified 91 gene mutations including single amino acid mutations in the *RagC* and *Raptor* genes that are part of the conserved eukaryotic TOR regulatory pathway and upstream sensing complexes. This study investigated whether members of the TOR and associated sensing pathways could be identified by bioinformatic analysis and the potential role a RagC mutation plays in the attenuation of *L. donovani* in Sri Lanka for causing visceral disease. This study provides novel insight into the role of the TOR and associated sensing pathways and the *RagC* single nucleotide polymorphism in *L. donovani* virulence.

3.3 Introduction

Leishmaniasis is a neglected tropical disease present throughout developing countries and is caused by *Leishmania* protozoa parasites that are transmitted by infected sand flies (1). There are over 20 species of *Leishmania* that infect humans of which most remain in the dermal layer of the skin at the site of the sand fly bite resulting in cutaneous leishmaniasis that usually self-cures (2). The *Leishmania donovani* species however typically disseminates from the dermis to the liver, spleen and bone marrow resulting in visceral leishmaniasis with persistent fever, hepatosplenomegaly, internal bleeding, anemia and is fatal if not treated. Humans are the only known reservoir for *L. donovani* while virtually all cutaneous leishmaniasis causing species have animal reservoirs (3). Parasite tropism plays a fundamental role in leishmaniasis virulence and pathogenesis because parasites that visceralize are deadly compared to parasites that remain in the skin that generally self-heal. Visceral leishmaniasis is the second most deadly vector borne

parasitic infection after malaria. What controls disease pathology and tropism during infection remains poorly understood although this is largely dependent on the species of *Leishmania* (4).

L. donovani is endemic in the Indian subcontinent and East Africa where most cases of visceral leishmaniasis occur. Notably, *L. donovani* has recently evolved to also cause cutaneous leishmaniasis in some geographic locations (5). For example, the number of cutaneous leishmaniasis cases in Sri Lanka continues to increase with over 3000 cases in 2018 while there is no transmission of visceral leishmaniasis (6). Whole genome sequencing of a *L. donovani* cutaneous leishmaniasis strain in Sri Lanka identified a variety of single nucleotide polymorphisms (SNPs), indels and copy number variations including a non-conservative arginine to cysteine codon change at position 231 (R231C) in the *RagC* gene (7, 8). Determining the specific genetic changes responsible for the *L. donovani* conversion from a visceral to a cutaneous leishmaniasis strain would help define the molecular basis for disease tropisms and virulence associated with visceral leishmaniasis.

The life cycle of *Leishmania* includes the promastigote stage that replicates in the midgut of the sand fly vector and the amastigote stage that replicates in the phagolysosome of mammalian host macrophage cells. *Leishmania* must adapt to different environmental conditions including temperature, pH and nutrient availability to differentiate and proliferate in the different life cycle stages. In eukaryotes, the Mechanistic Target Of Rapamycin Complex 1 (mTORC1) within the TOR cell signaling pathway has multiple regulatory mechanisms controlling protein synthesis and autophagy (9). Deregulation of the TOR pathway is implicated in the pathogenesis of various human diseases including cancer and immunological defects (10, 11). Three distinct TOR kinase genes have been identified in *Leishmania* (12) and it is largely unknown what other components of the TOR or TOR upstream sensing pathways are present in *Leishmania* and what role the

mTORC1 complex plays in the *Leishmania* cell cycle, tropism and virulence. One of the *Leishmania* TOR kinases (TOR3) has been shown to be necessary for formation of the ancient acidocalcisome organelle associated with storage of calcium and the other 2 members (TOR1, TOR2) are essential for promastigote stage survival but their function is unknown (12).

The mTORC1 associated pathway is well described in mammalian cells involving multiple regulatory proteins and protein complexes that respond to different kinds of nutrients and growth signals (13, see also **Table 3.1** and **Figure 3.1**). Accurate regulation of mTORC1 activity to control protein synthesis and autophagy is essential for cell growth, replication and survival (14). Nutrients utilized by eukaryotic cells include carbohydrates, lipids, and amino acids. The Rag GTPases (Ras-related guanosine triphosphatases) have been identified as important mediators of amino acid signaling to mTORC1 in mammals (**Figure 3.1**). The Rag GTPases consist of RagA, RagB, RagC, and RagD. RagA and RagB with 90% sequence identity are functionally redundant. RagC and RagD are also functionally redundant and share 81% sequence identity. RagA or RagB bind to RagC or RagD to form a stable and functional heterodimeric complex. Like other GTP-binding proteins, the nucleotide loading state of the Rag GTPases regulates their function. Under amino acid-sufficient conditions, the RagA-RagC heterodimer localizes to the lysosomal surface and is transformed into the active form $RagA^{GTP}$-$RagC^{GDP}$ through the Ragulator complex (9, 15). The active $RagA^{GTP}$-$RagC^{GDP}$ then binds to Raptor resulting in the recruitment of mTORC1 to the lysosome (16, 17). Once mTORC1 is recruited to the lysosome, the TOR serine-threonine kinase is activated by Rheb (18) and phosphorylates its canonical substrates S6K1, 4EBP1, ULK1 and TFEB. Phosphorylation of S6K1 and 4EBP1 promotes protein synthesis and cell growth, while phosphorylation of ULK1 and TFEB inhibits autophagy. When amino acids are scarce, the Rag heterodimer is inactivated to form $RagA^{GDP}$-$RagC^{GTP}$ which detaches from Raptor resulting in

dislocation of mTORC1 from lysosome and inactivation of mTORC1, thus suppresses further protein synthesis, and induces autophagy if necessary to maintain cell survival (9, 18).

In the present study, we performed a bioinformatic analysis to identify potential members of the TOR and upstream sensing pathways in *L. donovani* to investigate how a naturally occurring mutation in the *RagC* (R231C) gene may contribute to reduced visceral pathogenesis. It was previously shown that introduction of a wildtype *RagC* gene in the cutaneous *L. donovani* strain by plasmid transfection resulted in an incremental but significant increased parasite survival in the visceral organs in experimentally infected mice providing evidence that the (R231C) mutation may have contributed to the attenuation for survival in visceral organs of this *L. donovani* strain in Sri Lanka (8). To specifically investigate the role of the naturally occurring RagC R231C mutation in pathogenesis in the absence of other mutations present in the atypical cutaneous *L. donovani*, CRISPR gene editing was used to engineer the RagC R231C single amino acid mutant and a *RagC* gene disruption mutant in the wildtype virulent *L. donovani* 1S2D strain. We further undertook to disrupt the partner *RagA* gene and assessed whether the RagC/RagA interaction is conserved in *Leishmania* and if the RagC R231C point mutation affected the interaction with RagA. The results of this study provide evidence that the RagC R231C mutation in *L. donovani* plays a role in the attenuation of the atypical cutaneous *Leishmania* strain present in Sri Lanka and that several components of the TOR and upstream sensing pathways are conserved in *Leishmania*.

3.4 Results

3.4.1 Conservation of the amino acid sensing arm of the mTOR pathway

A complete *L. donovani* genome sequence from a cutaneous strain from Sri Lanka has recently closed over 2000 sequence gaps resulting in the identification of over 600 novel open reading frames (7). Using sequence analysis (BLASTP) of this complete genome sequence, it was

possible to identify conserved components of the TOR and upstream sensing pathways (9, 14–18). As shown in **Figure 3.1**, **Supplementary Figure 3.S1**, several components of this pathway are well conserved (BLASTP E-value < 0.001) between human cells and *L. donovani* (**Table 3.1, Supplementary Figure 3.S1**). Homologous components include sensors such as the v-ATPase or Leucyl-tRNA synthetase (19), signaling proteins including the RagA GTPase which is the binding partner for the RagC GTPase and cellular regulators such as the Raptor protein that acts as a bridge between the RagA/RagC heterodimer and the TOR kinase complex. However, several components present in mammalian eukaryotes are missing entirely or partially: The RagB/RagD complex, Ragulator complex, Folliculin complex, KICSTOR and CASTOR complexes were not found to any degree of homology in *Leishmania*. While the GATOR2 complex was present, only 1 member of the 3 protein GATOR1 complex was identified. Several of the TORC1 complex components were identified that had homology to the Raptor, TOR and LST8 proteins but not to the Deptor and PRas40 proteins (**Supplementary Figure 3.S1**). These observations demonstrate that several components of the TOR and upstream sensing pathways are conserved between mammalian cells and *L. donovani* suggesting that a TOR signaling pathway is present in *Leishmania* but distinct from mammalian cells. Note also that mutations identified in the atypical cutaneous *L. donovani* strain are highlighted in the RagC and Raptor proteins (7) (**Supplementary Figure 3.S1**).

3.4.2 Generation of the RagC R231C mutant and RagC disruption mutant from the virulent *L. donovani* 1S2D

We initially focused on the RagC single amino acid mutation (R231C) since this mutation is naturally occurring in the atypical cutaneous strain of *L. donovani* in Sri Lanka (7, 8). Using CRISPR gene editing, an identical single amino acid mutation was introduced into both

chromosomal copies of the *RagC* gene in the virulent *L. donovani* 1S2D strain as outlined in **Figure 3.2 A, B**. To engineer the missense mutation, a gRNA with its targeting site close to the polymorphism site was designed. A single strand oligonucleotide donor DNA flanking the Cas9 cleavage site was synthesized with the desired C to T point mutation along with several synonymous mutations to protect edited genomes from further Cas9 cleavage (**Figure 3.2A & B**). RagC missense mutation containing clones were isolated after transfection of the donor DNA as previously described (20, 21). The heterozygous ($231^{R/C}$) clones were first identified by using donor specific PCR. Since the Cas9 continued scanning the genome and generated the specific double-strand DNA break, the homozygous RagC R231C mutant was subsequently isolated from one of these heterozygous clones and confirmed by DNA sequencing (**Figure 3.2C**). Interestingly, the RagC R231C homozygous mutant isolated was the slowest growing clone among 10 sub-clones examined which include one homozygous R231C mutant clone and nine heterozygous ($231^{R/C}$) clones (not shown). In addition to generating the RagC R231C single amino acid mutant, the *RagC* gene was functionally disrupted by the insertion of a bleomycin resistance gene into the Cas9 cut site as also shown in **Figure 3.2A** and **3.2D**.

3.4.3 Effects of the RagC mutations on *L. donovani* promastigote proliferation and infection in BALB/c mice

As the TOR pathway plays a fundamental role in regulating cell proliferation in mammalian eukaryotes, we examined whether the RagC R231C point mutation and the *RagC* gene disruption mutation affected *L. donovani* 1S2D proliferation under culture conditions. As shown in the sand fly stage promastigote culture (27 ºC, pH 7.0), both the RagC R231C and *RagC* gene disrupted mutants were viable but could not reach the same cell density at day 4 as the wildtype *L.*

donovani 1S2D (**Figure 3.3A**). Re-introducing a plasmid derived wildtype *RagC* gene to the mutant strains restored their proliferation to a level similar to that of the wildtype *L. donovani* 1S2D promastigotes. These results show the point mutation of the *RagC* gene was similar to disruption of the *RagC* gene and generated viable promastigotes that were unable to reach the same cell density as wildtype promastigotes. Interestingly, although both the RagC mutants showed reduced growth in culture, their ability to generate metacyclic like cells was not impaired (**Supplementary Table 3.S1**).

We next determined whether the RagC R231C mutation or gene disruption affected *L. donovani* parasite survival during visceral infection. The RagC R231C mutant and gene disruption mutant were injected into the tail vein of BALB/c mice. The liver parasite burdens usually peak at 4 weeks post infection in the BALB/c mouse model for visceral *Leishmania* infection and decrease subsequently due to development of immunity in the infected mice. The liver parasite burdens were therefore determined 4 weeks post tail vein infection. As shown in **Figure 3.3B**, both the RagC R231C mutant and the *RagC* gene disruption mutant had dramatically reduced parasite burden in the liver compared to wildtype *L. donovani* 1S2D. Addback of the wildtype RagC to the *RagC* gene disruption mutant fully restored parasite numbers in the liver. Addback of wildtype RagC to the RagC R231C mutant only partially restored the infection levels in the liver possibly because of competition with the chromosomal derived mutant R231C RagC. Taken together, these data demonstrate that similar to the *RagC* gene disruption mutant, the RagC R231C single amino acid mutant was attenuated for visceral infection in the mouse liver compared to wildtype *L. donovani* 1S2D.

3.4.4 Leishmania wildtype RagC and mutant RagC R231C interact with RagA

Considering the bioinformatic analysis of the *Leishmania* TOR and upstream sensing pathways (**Supplementary Figure 3.S1**), and the phenotypic impact of the RagC mutation (**Figure 3.3**), it was necessary to determine whether RagC interacts with RagA and whether the R231C single amino acid mutation alters this interaction. Plasmid vectors expressing epitope tagged versions of wildtype RagC, mutant RagC and RagA were transfected into *L. donovani* 1S2D and co-immunoprecipitation analysis carried out as detailed in Methods.

Although it was possible to express the tagged wildtype RagC, it was difficult to obtain stable expression of the tagged RagC R231C mutant in *L. donovani* 1S2D. Further analysis revealed this was due to a selection against the transfected plasmid retaining the insert DNA encoding the mutant R231C RagC, although there was no selection against expressing wildtype RagC (**Supplementary Figure 3.S2**). Plasmids expressing wildtype and mutant RagC were identical except for the R231C codon. Attempts were then made to express the RagC R231C mutant in the cutaneous *L. donovani* strain where this RagC mutant was first identified (8). As shown in **Figure 3.4**, at 20- and 30-days post-transfection, the cutaneous *L. donovani* strain (cutaneous SL-Ld) could support the expression of the mutant RagC R231C but selected against expression of the wildtype RagC and conversely, wildtype *L. donovani* 1S2D could support the expression of the wildtype RagC but not the mutant RagC R231C. We further attempted to express the RagC R231C mutant in *RagC* gene disrupted 1S2D, however as shown in **Figure 3.4B**, *L. donovani* 1S2D devoid of a genomic copies of wildtype *RagC* selected against expressing the plasmid derived RagC R231C mutant. This is consistent with the above observation that expression of the RagC R231C mutant from the edited chromosome showed reduced cell density compared to wildtype *L. donovani* 1S2D (**Figure 3.3A**). It is possible that mutant RagC R231C

produces a dominant negative effect resulting in the strong selection against plasmids expressing mutant RagC R231C in *L. donovani* 1S2D.

Considering the above observations, the interaction between wildtype RagC and RagA was initially investigated in transfected *L. donovani* 1S2D promastigotes. As shown in **Figure 3.5A**, lane 9, following immunoprecipitation of FLAG-tagged RagC, it was possible to detect RagA by immunoblotting with HA antibodies. Control lanes showed that RagA was not pulled down in the absence of RagC (**Figure 3.5A**, Top Panel, lane 7) or if anti-FLAG antibodies were not present (**Figure 3.5A**, Top Panel lane 10). In the reverse, following immunoprecipitation of HA-tagged RagA, it was possible to detect RagC following immunoblotting with FLAG antibodies (**Figure 3.5A**, Bottom Panel, lane 9). Control lanes showed that RagC was not detectable in the absence of RagA (**Figure 3.5A**, Bottom Panel, lane 8) or if anti-HA antibodies were not present (**Figure 3.5A**, lane 10). Further, epifluorescence microscopy using GFP tagged RagA and mCherry tagged RagC revealed that both proteins localized to overlapping nuclear-adjacent foci on the kinetoplast pole of the promastigote as shown in **Supplementary Figure 3.S3A**, indicating the proteins co-localize *in vivo*. These results show that wildtype RagC and RagA can form a heterodimer or are present in the same complex consistent with the *in silico* bioinformatic analysis shown in **Supplementary Figure 3.S1**.

We next determined whether the R231C mutation in RagC impaired the interaction with RagA when both proteins were co-expressed in the cutaneous strain of *L. donovani* that was permissive for expression of transfected RagC mutant R231C. As shown in **Figure 3.5B**, following immunoprecipitation of the RagC R231C mutant with FLAG antibodies, it was possible to detect RagA following immunoblotting with HA antibodies (**Figure 3.5B**, Top Panel, Lane 10) at comparable levels to immunoprecipitation of wildtype RagC expressed in *L. donovani* 1S2D

(**Figure 3.5B**, Top Panel, Lane 12). Control lanes showed that RagA was not pulled down in the absence of RagC R231C mutant (**Figure 3.5B**, Top Panel, lane 8) or in the absence of FLAG antibodies (**Figure 3.5B**, Top Panel, lane 11). In the reverse, following immunoprecipitation of HA-tagged RagA, it was possible to detect the RagC R231C mutant following immunoblotting with FLAG antibodies at comparable levels to wildtype RagC expressed in *L. donovani* 1S2D (**Figure 3.5B**, Bottom Panel, lanes 10 and 12). Control lanes showed that the RagC R231C mutant was not detectable in the absence of RagA (**Figure 3.5B**, Bottom Panel, lane 9) or in the absence of HA antibodies (**Figure 3.5B**, Bottom Panel, lane 11). Further, epifluorescence microscopy showed GFP tagged RagA and mCherry tagged RagC R231C mutant also localized to overlapping nuclear-adjacent foci (**Supplementary Figure 3.S3B**). These results provide evidence that the R231C mutation in RagC did not impair its ability to interact with RagA or alter the cell location of the RagA-RagC heterodimer under these experimental conditions.

As the RagC R231C mutant was able to interact with RagA, we used *in silico* analysis to further investigate how the mutation could affect the structure of RagC. The amino acid sequences for mutant RagC were subjected to homology modelling using a human template as detailed in methods. As shown in **Supplementary Figure 3.S4A**, the R231C mutation (red) is located away from the RagA (yellow) and RagC (blue) proposed interaction site based on human RagA and RagC crystal structures (22), indicating this mutation is unlikely to significantly affect the interaction between these proteins, consistent with the results shown in **Figure 3.5**. In addition, the point mutation is not in the GTPase domain suggesting this does not affect enzyme activity. We did however identify an extra loop of amino acids forming a new motif on the surface of RagC in *Leishmania* not present in human and yeast proteins as shown in the alignment in

Supplementary Figure 3.S4B and in the ribbon representation the new motif is highlighted in green (**Supplementary Figure 3.S4C**).

Bioinformatic analysis of mTORC1 complex proteins revealed not only the conservation of the *Raptor* gene but also a single amino acid mutation (A969E) between the cutaneous and visceral strains of *L. donovani* (**Supplementary Figure 3.S1**) (7). Since Raptor interacts with the RagC/RagA heterodimer in mammalian cells and plays a key role in mTORC1, we investigated whether introducing this *Raptor* gene mutation in wildtype *L. donovani* 1S2D (**Supplementary Figure 3.S5A**) could make these cells permissive for the expression of the R231C RagC. Although this single amino acid mutation was successfully edited in the chromosome using CRISPR, this mutation alone could not reverse the selection against expression of the R231C RagC in the *L. donovani* 1S2D strain (**Supplementary Figure 3.S5B**).

3.4.5 The *RagA* gene is essential for *Leishmania* promastigotes

In yeast and mammalian cells, it is possible to generate single and double *RagA* and *RagC* null mutants showing both these proteins are non-essential in these eukaryotic cells (9, 16, 23, 24). We therefore investigated whether it was possible to disrupt the *RagA* gene in *L. donovani* 1S2D and to compare the resulting phenotype to the *RagC* disruption mutant. The experimental approach to disrupt the *RagA* gene in *L. donovani* 1S2D is shown in **Figure 3.6A**. While it was possible to generate promastigotes with a single *RagA* allele disrupted, one copy of the wild type *RagA* allele persisted in the surviving *Leishmania* promastigotes (**Figure 3.6B**, lower band). Following single cell cloning of surviving promastigotes to isolate double allele knockout cells, the *RagA* double null mutant clones continued replicating to approximately one hundred cells before dying while

single allele *RagA* disrupted parasites continued to proliferate (**Figure 3.6C**). It was therefore not possible to establish a culture with a double *RagA* gene disruption and thus was not possible to isolate DNA for PCR analysis. Together, this demonstrates that *RagA* is essential for *Leishmania* promastigotes survival.

3.5 Discussion

A major objective of this study was to investigate the role of a naturally occurring single amino acid mutation, R231C in RagC, in contributing to the attenuation of *L. donovani* in Sri Lanka for causing visceral disease. CRISPR gene editing was used to change this single amino acid in the RagC protein in the background of wildtype virulent *L. donovani* 1S2D. The observations presented in experimentally infected mice provide evidence that this naturally occurring RagC mutation in the cutaneous *L. donovani* strain from Sri Lanka (7, 8)contributed to its attenuation for infection in visceral organs. It is however unknown whether this mutation also contributes to the ability of this atypical strain to cause cutaneous infections as is widespread for *L. donovani* in Sri Lanka (6). It is likely that in addition the *RagC* mutation, some of the additional 91 mutations identified in the cutaneous strain (7, 8) are necessary for survival in cutaneous sites and stable expression of the RagC R231C protein isoform.

Using CRISPR gene editing, it was possible to investigate the biological outcome of the RagC R231C mutation against a wildtype virulent *L. donovani* genetic background in the absence of the other 91 gene mutations in the atypical *L. donovani* cutaneous strain (7). It is noteworthy that the independent *RagC* gene disruption mutation had a similar phenotype in promastigotes and in infected mice as the RagC R231C mutation and in both mutants this was reversed in the add back clones further supporting the argument for the importance of the *RagC* gene for *L. donovani*

visceral infection. It is remarkable that a single homozygous SNP resulting in RagC R231C can have such a dramatic effect on parasite numbers in the liver in experimentally infected mice. It is unknown what effect these mutations have on differentiation in the sand fly such as for example the development of metacyclics. It is however unlikely that a *Leishmania* mutant with impaired metacyclic development would be viable in nature. This is consistent with the observation that, if anything, there was an increase in the number of metacyclics in both the R231C RagC point mutant and the *RagC* gene disruption mutant promastigote stationary phase cultures. Future studies are needed to investigate the increased metacyclic formation in the RagC mutants.

A secondary objective of this study was to investigate components of the *Leishmania* TOR signaling and upstream sensing pathways that includes RagC. *In silico* bioinformatic analysis suggested that similar to mammalian cells, *L. donovani* retains several key components of the TOR sensing pathway including RagC, RagA and several components of the TORC1 complex including Raptor (**Supplementary Figure 3.S1**). Immunoprecipitation analysis demonstrated that RagC can form heterodimers with or is in the same complex as RagA which experimentally supports the *in silico* analysis. Attempts to identify the cellular defect resulting from the RagC mutations were however unsuccessful since the mutant RagC retained the ability to interact with RagA (**Figure 3.5**) and introduction of the *Raptor* A969E gene mutation by gene editing did not make *L. donovani* 1S2D permissive for expression of transfected mutant RagC (**Supplementary Figure 3.S5**). Further, The RagC mutation did not appear to affect the localization of the heterodimer complex as both wildtype and mutant proteins were shown to co-localize in perinuclear foci (**Supplementary Figure 3.S3**). This localization pattern is consistent with the localization of the *Leishmania* homologue to Rab7 (25, 26), which in mammalian cells is present in the same lysosomal compartment as the RagA/RagC (or RagB/RagD) complexes (17). One possibility

consistent with these observations is that the mutant RagC/RagA interaction abnormally controls the activity of downstream components of the TOR pathway (perhaps mTORC1 itself) resulting in attenuation of *L. donovani* for survival in visceral organs.

Several sensor proteins upstream of and other regulators of the RagA/RagC complex present in mammalian cells appear to be missing from *Leishmania* including RagB or RagD proteins (**Figure 3.1, Supplementary Figure 3.S1, Table 3.1**), suggesting the RagA/RagC complex plays a more central role in *Leishmania* signaling than in mammalian cells. This is consistent with the observation that it was not possible to disrupt both alleles of the *RagA* gene in *L. donovani* (**Figure 3.6**) although this is possible in mammalian and yeast cells (9, 16, 23, 24). Although the R231C RagC retained the ability to interact with RagA, wildtype *L. donovani* 1S2D could not tolerate the expression of the mutant R231C RagC (**Figure 3.4**) likely due to the mutant having a dominant effect. Presumably, other mutations or alterations in the cutaneous *L. donovani* strain enable the expression of the R231C RagC through epistasis. Indeed, there is also a *Raptor* A969E gene mutation in the atypical cutaneous *L. donovani* strain although introduction of this mutation in *L. donovani* 1S2D did not enable expression of the R231C RagC (**Supplementary Figure 3.S5**). One or more of the 91 altered genes previously identified in the cutaneous *L. donovani* strain (7) may be necessary to enable expression of the R231C RagC protein. Future studies could include performing pull down immunoprecipitation and mass spectrometry to compare mutant R231C RagC and wildtype RagC protein interactions. It will also be interesting to investigate whether the R231C RagC mutant or the Raptor A969E mutant affect the response to different nutrients or the recruitment of mTORC1 proteins to the lysosomes and activation of downstream signaling pathways. Interpretation of these experiments could however be

compromised because of the necessity to express wildtype and mutant epitope-tagged RagC proteins in different *L. donovani* strain backgrounds as demonstrated here.

In mammals, leucine, arginine, and glutamine have been shown to trigger mTORC1 activity. The CASTOR1/2 protein complex and solute carrier family 38 member 9 (SLC38A9) (a lysosomal amino acid transporter) were identified as intracellular arginine sensors (14, 18, 27).Our bioinformatic analysis has shown that *Leishmania* has the homologs for SLC38A9 and components for the GATOR2 complex, normally downstream of CASTOR1/2 but we did not identify any proteins with significant CASTOR homology in *Leishmania*. It has been recently reported that in order to survive arginine deprivation in macrophages, *Leishmania* upregulates expression of its arginine transporter (AAP3) through a MAPK2-mediated arginine deprivation response (ADR) pathway (28). As the identity of the arginine sensor that triggers the ADR pathway is unknown, it will be interesting to investigate whether the ADR pathway shares the same potential arginine sensors (SLC38A9 and a possibly unidentified protein upstream of GATOR2) with mTORC1 signaling pathway and whether the ADR pathway is downstream of or parallel to the mTORC1 pathway. It will also be of interest to determine whether the potential arginine sensors can directly interact with the RagA-RagC heterodimer and whether the RagC R231C mutation interferes with interactions between these upstream amino acid sensors and the RagA-RagC heterodimer.

It was not possible to detect wildtype RagC protein expression in the transfected cutaneous *L. donovani* promastigotes. This is not consistent with our previous observation where transfection of the wildtype *RagC* gene in the cutaneous *L. donovani* isolate increased parasite numbers in the spleen of BALB/c mice following tail vein injection, albeit at a much lower level than wildtype *L. donovani* (8). Since RT-PCR did show the presence of wildtype *RagC* transcripts in the transfected

cutaneous *L. donovani* isolate (8), it is possible that although the transfected cutaneous strain promastigotes in culture had selected against expression of wildtype RagC, a small minority of promastigotes retained expression and these promastigotes then gained a survival advantage once present in the mouse visceral organs.

Although it was possible to generate a disruption mutation in the *RagC* gene, disruption of the *RagA* gene in *L. donovani* 1S2D resulted in promastigote death. *RagA* can therefore be considered an essential gene and plays an important role in *L. donovani* viability while RagC potentially plays a more regulatory role. Indeed, it has been observed that in human cells, the RagA protein is responsible for the initial interaction with the TORC1 complex and that RagC stabilizes this interaction (29) and that RAPTOR is preferentially sensitive to the GTP/GDP state of RagA and not RagC (9, 16). This could help explain our observations that RagC but not RagA was dispensable in *L. donovani* promastigotes.

A recent genome sequence analysis of 8 additional *Leishmania* parasites from Sri Lanka revealed that all were indeed *L. donovani* but were genetically distinct from each other (30) and different from the cutaneous *L. donovani* strain initially sequenced (7, 8) and under investigation in this study. This demonstrates that there is a high level of diversity in the atypical *L. donovani* strains currently circulating in Sri Lanka. Analysis of the deposited GenBank sequences for these additional genomes (30) revealed that none had the RagC R231C or Raptor A969E mutations described here nor did they have any of the other nonsynonymous SNPs previously identified in the original cutaneous *L. donovani* strain (7, 8). This is interesting because it reveals that there are multiple genotypes of *L. donovani* in Sri Lanka yet virtually no cases of visceral leishmaniasis(6). Although the RagC R231C mutation may contribute to the attenuated visceral phenotype in the *L. donovani* strain under investigation here (7, 8), it appears that other atypical strains in Sri Lanka

became attenuated through alternative mechanisms (30). The observations from this study nevertheless provide evidence that perturbation of the conserved sensing and TOR pathways in *L. donovani* can influence human pathogenesis.

3.6 Materials and Methods

3.6.1 BLAST searches

Comparison of the human and *L. donovani* TOR pathways was generated using human protein sequences obtained from UniProt in FASTA format (31). The protein sequences were then searched for using the BLASTP (32) application accessed through TriTrypDB (33) using the *Leishmania donovani* strain LdCL (7) database. The sequences were also compared to the available protein sequences for all *L. major, L. infantum, L. tropica, L. mexicana* and *L. braziliensis* strains available on TriTrypDB. For all searches, an E-value cut-off of 1.0 was used to filter out false positive low scoring alignments.

3.6.2 *Leishmania* strain and culture media

L. donovani 1S2D strain used in this study were routinely cultured at 27°C in M199 medium (pH 7.4) supplemented with 10% heat-inactivated fetal bovine serum, 40 mM HEPES (pH 7.4), 0.1 mM adenine, 5 mg l^{-1}hemin, 1 mg l^{-1} biotin, 1 mg l^{-1} biopterin, 50 U ml^{-1} penicillin and 50 µg ml^{-1} streptomycin. Cultures were passaged to fresh medium at a 40-fold dilution once a week. The growth curves of *L. donovani* RagC (R231C) mutant and null mutant in promastigote culture medium were obtained by inoculating the parasites in 1 x 10^6 / ml into the 96 well plate (180 µl/well) in quadruplicate, the OD values were measured once a day for eight days.

3.6.3 Negative selection of metacyclic promastigotes with peanut agglutinin (PNA)

The metacyclic promastigotes were isolated with PNA as described (34). Briefly, the densities of healthy *L. donovani* stationary phase promastigotes was first determined by light microscopy. The promastigotes were then harvested at 2,000 g for 5 min and resuspended at a cell density of 2 x 10^8 cells/ml in complete medium containing 50 µg/ml PNA. Promastigotes were allowed to agglutinate at room temperature for 30 min. The supernatants were then carefully transferred to new tubes. The PNA negative promastigotes from these supernatants were counted using microscopy.

3.6.4 CRISPR Plasmid Construction

The *L. donovani* RagC R231C mutant and the gene disrupted *RagC* and *RagA* using CRISPR/Cas9 gene editing technique were generated as described (20, 35). The addback RagC expression plasmid was generated by cloning the full RagC coding sequence into the *Hind* III and *BamH* I sites of the *Leishmania* expression plasmid pLphyg2 (36). The RagC expression plasmid with a FLAG tag was generated by cloning the coding sequence with C-terminal tag into the *Hind* III and *BamH* I sites of *Leishmania* expression plasmid pLphyg2 (36) The RagA-HA tag expression plasmid was generated by first cloning the full RagA coding sequence into pLpneo2, followed by the insertion of an HA tag in frame.

The oligos and primers used in this study are listed below.

The following primers and oligos were used for generating the *L. donovani* RagC R231C mutant, the disruption mutant and for constructing the addback RagC expression plasmid with FLAG tag at its C terminus.

Oligos used to insert the *RagC* targeting gRNA into the pLdCN CRISPR vector:

Ld366140+ 5’ TTGTGGTCGTAGGTGCGCGACTTG

Ld366140- 5’ AAACCAAGTCGCGCACCTACGACC

R231C mutation containing repair donor oligonucleotide

Ld366140donor 5’

ATCTACGTCGCCGTGGACGAGCGCAACTGTCTGAAAAGCCGCACCTACGACCTCTGC

AGCGACGC

Primers used to generate the Bleomycin resistance donor:

Ld366140BleF1 5’

CGTCGCCGTGGACGAGCGCAACCGCCTCAAGATCTTCATCGGATCGGGTA;

Ld366140BleR1 5’

GCGTCGCTGCAGAGGTCGTAGGTGCGCGACGTCGGTCAGTCCTGCTCCT.

Primers for PCR to verify the CRISPR generated point mutations and the disrupted mutants:

Ld366140F2(F2 in Fig3.2) 5’ CTGCTGCAGATGCTCAACTC

Ld366140R2(R2 in Fig3.2) 5’ CACGTTACGGTCAATCAACG

Ld366140dF 5’ AGCGCAACTGTCTGAAAAGC

Primers used to construct the RagC (WT or R231C) expression plasmids with FLAG tag at its C terminus:

Ld366140F 5’ CCCAAGCTTCACTACTTGCTCGCCCTTT

Ld366140F1 5’ CCCAAGCTTCCGAGCATGTCCAACAACATGCTCGCACT

Ld366140R 5’ CCGAGATCTTGCGTGCCTCTCTCTCTCT

Ld366140R1(3xFLAG)5’

GATCCTTGTAGTCTCCGTCGTGGTCCTTATAGTCTGGATCCCGGGATGCACTCGTGTT

GAAGATG

Bgl2FLAG 5’

CCGAGATCTACTTATCGTCATCGTCTTTGTAATCAATATCATGATCCTTGTAGTCTCC

GTCGTGG

The following primers and oligos were used for generating the *L. donovani* Raptor (LdCL_250011400) A969E mutant:

Oligos used to insert the gRNA guide coding sequences into pLdCN vector:

LdRptM+ 5’ TTGTCGACGGTGTGCTGGTGAACA

LdRptM- 5’ AAACTGTTCACCAGCACACCGTCG

LdRptMdonor 5’

TCTACGCAAACTGCGACGGTGTGCTCGTCAATAAAGAACTGCGCTACCTCTCAAACG

GC GCTC.

Primers for PCR to verify the CRISPR generated point mutations:

LdRptMdonF 5’ GGTGTGCTCGTCAATAAAGAA

LdRptML 5’ TTGACAGCCGATGATCCACT

LdRptMR 5’ TGACTGCTCTGAAAGGTCGT

The following primers and oligos were used for generating the *L. donovani RagA* disruption mutant and for constructing the RagA expression plasmid with an HA tag at its C terminus.

Oligos used to insert the gRNA guide coding sequences into pLdCN vector

Ld131620+ 5’TTGTGGCAGCTTCCAATCCGTCG

Ld131620- 5’ AAACCGACGGATTGGAAGCTGCC

Primers used to generate the Bleomycin resistance donor:

Ld131620BleF1 5’

GTTGGTGGCGAGGGCAGCAGCGCCACGATCTTCATCGGATCGGGTA

Ld131620BleR1 5’

CTCCAGCATCTCCGGCAGCTTCCAATCCGTCGGTCAGTCCTGCTCCT

Primers used to construct the RagA expression plasmids with an HA tag at its C terminus:

Ld131620F (F in Figure 3.6) 5’ CCCAAGCTTCCACGCATGATTCTTCCGCGCTA

Ld131620R (R in Figure 3.6) 5’ CCGGGATCCAGAACTTCGCGCATTGCACAC

HAtag+ 5’ GATCTGTACCCATACGATGTTCCAGATTACGCTT

HAtag- 5’ GATCAAGCGTAATCTGGAACATCGTATGGGTACA

Ld131620F2 5’GCATGAACAACACGACCAGT

3.6.5 *Leishmania* transfection and single cell cloning assay

Leishmania transfections were performed as previously described (20). Briefly, 10 µg pLdCN CRISPR plasmid DNA encoding the specific gRNA was electroporated into 1 x 10^8 early stationary phase *L. donovani* promastigotes. The transfected cells were then selected with G418 (50-100 µg /ml). To generate the RagC (R231C) mutant, once the pLdCN plasmid transfected *L. donovani* culture was established, the cells were subjected to three rounds of sequential oligonucleotide donor transfection, 10 µl (100 µM) single strand oligonucleotide donor was used

per transfection, once every three days. After the third oligonucleotide donor transfection, the *Leishmania* promastigotes were counted and inoculated into 96 well plates in one promastigote per 100 µl medium per well. The growth of *Leishmania* cells in 96 well plates were monitored under microscope. After culture for three weeks in 96 well plates, parasites from the relatively slow growing clones were expanded in 24 well plates. The genomic DNA extracted from these slow growth clones were subjected to PCR and sequencing analysis. To generate the *RagC* or *RagA* disruption mutant, the *L. donovani* cells containing the corresponding pLdCN plasmid were subjected to transfection of the Bleomycin resistance gene donor PCR product and selected with 100 µg/ml phleomycin. The phleomycin resistant cells were then cloned into 96 well plates and monitored for growth under a microscope. After culture for three weeks in 96 well plates, the surviving parasites were expanded in 24 well plates and subjected to PCR analysis.

3.6.6 Infection of mice with *L. donovani* RagC mutants

To determine how the RagC R231C amino acid change and disruption of the *RagC* gene would affect *L. donovani* virulence in visceral infection in BALB/c mice, mice were infected with both *L. donovani* RagC mutants and wildtype *L. donovani* 1S2D by tail vein injection with the same number of stationary phase promastigotes ($1x10^8$ pro/mouse). Mice were examined for the liver parasite burden (*L. donovani* units, LDU) four weeks after infection as previously described (8).

3.6.7 Co-immunoprecipitation

For each lane, $1x10^8$ cells were isolated and lysed on ice for 30 mins in 100 µL of lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.25% NP-40) supplemented with one tablet of

cOmplete ULTRA protease inhibitor cocktail (Roche). Note: As the cutaneous *Leishmania donovani* strain expressed plasmid encoded proteins to a lower level than the wildtype strain 1S2D, 5X more lysate was used as input in cutaneous *Leishmania* co-IP experiments. Lysate was clarified by centrifugation at 21,000x g for 20 mins. 10 µL of clarified lysate was mixed with SDS Sample Buffer and used as 'Input' lanes. 1 µL of FLAG or HA antibody was added to each tube of clarified lysate and the reaction volume adjusted to 500 µL with Lysis Buffer. Lysate:Antibody mixture were incubated for 2h at RT with gentle agitation on a tube revolver. For each tube, 25 µL of Protein A/G Magnetic beads (Pierce) were washed in Lysis Buffer for 5 mins and added to the mixture. Tubes were incubated for an additional hour at RT on a tube revolver. Beads were isolated from the tubes using a DynaMag2 magnetic stand (Life Technologies). The beads were then washed three times with 1 mL of Lysis Buffer and resuspended fully each time. Captured and bait proteins were eluted off the beads by incubating in 2X SDS Sample Buffer for 10 mins followed by SDS-PAGE.

3.6.8 Immunoblotting

RagC-FLAG: Input and Co-IP samples were prepared as described above subjected to SDS-PAGE. Primary mouse IgG against FLAG tag (M2 clones, Sigma) was used at 1:10,000 diluted in PBS-T with 5% skim milk powder. Secondary HRP labelled goat IgG against mouse IgG (GE Healthcare) was used at 1:10,000 diluted in PBS-T with 5% skim milk powder. ECL reagent (ZmTech Scientific) was incubated for 2 minutes and the membranes were exposed to X-Ray film (Denville). Resulting film was imaged on a FluorChem FC8800 gel imager (Alpha Innotech).

RagA-HA: Immunoblotting was carried out as described above with rabbit HA antibodies (EMD Millipore) and HRP conjugated donkey anti-rabbit IgG antibodies (ThermoFisher Scientific)

3.6.9 Epifluorescence Microscopy

The RagA-GFP expression plasmid was created by cloning the full *L. donovani* RagA coding sequence into the *Hind* III and *BamH* I restriction sites of the pLGFPN plasmid (37). The RagC-mCherry expression plasmid was created by using NEBuilder HiFi (New England Biolabs) to insert the mCherry coding sequence into the *Hind* III restriction site of the plphyg2-RagC plasmids. WT and Sri Lanka cutaneous *L. donovani* parasites were transfected with both plasmids and selected with a combination of hygromycin and neomycin. 10 μL of cell suspension was diluted into 100 μL of PBS and cytospun on poly-lysine coated cover slips. The cells were fixed with 4% paraformaldehyde for 10 minutes and blocked with 5% BSA in PBS for 30 mins. The coverslips were mounted on glass slides with a drop of Prolong Antifade Glass (Invitrogen) and 2 μL of 10 μM DRAQ5 (ThermoFisher Scientic) and left to cure overnight at RT in the dark. Slides were imaged the following day on an ImageXpress Micro Confocal high content screener (Molecular Devices) in widefield acquisition with MetaXpress software.

3.6.10 Homology Modelling

The RagA and RagC protein coding sequences were used to generate the homology models using the SWISS-MODEL protein homology server (38). The best scoring human template was selected (ID 6CES (22)) for modeling. The resulting *Leishmania* RagA and RagC proteins were aligned over the human proteins in complex for orientation. The modeled proteins were then

further relaxed using the Relax application of the Rosetta software (39–42). Protein models were visualized and color coded using UCSF Chimera (43).

3.7 References

1. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, de Boer M. 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7:e35671.
2. Arenas R, Torres-Guerrero E, Quintanilla-Cedillo MR, Ruiz-Esmenjaud J. 2017. Leishmaniasis: A review. F1000Research 6.
3. Gramiccia M, Gradoni L. 2005. The current status of zoonotic leishmaniases and approaches to disease control. Int J Parasitol 35:1169–1180.
4. McCall LI, Zhang WW, Matlashewski G. 2013. Determinants for the Development of Visceral Leishmaniasis Disease. PLoS Pathog 9.
5. Thakur L, Singh KK, Shanker V, Negi A, Jain A, Matlashewski G, Jain M. 2018. Atypical leishmaniasis: A global perspective with emphasis on the Indian subcontinent. PLoS Negl Trop Dis 12.
6. Karunaweera ND, Ginige S, Senanayake S, Silva H, Manamperi N, Samaranayake N, Siriwardana Y, Gamage D, Senerath U, Zhou G. 2020. Spatial epidemiologic trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerg Infect Dis 26:1–10.
7. Lypaczewski P, Hoshizaki J, Zhang WW, McCall LI, Torcivia-Rodriguez J, Simonyan V, Kaur A, Dewar K, Matlashewski G. 2018. A complete Leishmania donovani reference genome identifies novel genetic variations associated with virulence. Sci Rep 8.
8. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.
9. Kim E, Goraksha-Hicks P, Li L, Neufeld TP, Guan KL. 2008. Regulation of TORC1 by Rag GTPases in nutrient response. Nat Cell Biol 10:935–945.
10. Elkabets M, Vora S, Juric D, Morse N, Mino-Kenudson M, Muranen T, Tao J, Campos AB, Rodon J, Ibrahim YH, Serra V, Rodrik-Outmezguine V, Hazra S, Singh S, Kim P, Quadt C, Liu M, Huang A, Rosen N, Engelman JA, Scaltriti M, Baselga J. 2013. MTORC1 inhibition is required for sensitivity to PI3K p110α inhibitors in PIK3CA-mutant breast cancer. Sci Transl Med 5.
11. Hu S, Cheng M, Fan R, Wang Z, Wang L, Zhang T, Zhang M, Louis E, Zhong J. 2019. Beneficial effects of dual TORC1/2 inhibition on chronic experimental colitis. Int

Immunopharmacol 70:88–100.

12. Madeira Da Silva L, Beverley SM. 2010. Expansion of the target of rapamycin (TOR) kinase family and function in Leishmania shows that TOR3 is required for acidocalcisome biogenesis and animal infectivity. Proc Natl Acad Sci U S A 107:11965–11970.

13. Sabatini DM. 2017. Twenty-five years of mTOR: Uncovering the link from nutrients to growth. Proc Natl Acad Sci U S A 114:11818–11825.

14. Shimobayashi M, Hall MN. 2016. Multiple amino acid sensing inputs to mTORC1. Cell Res 26:7–20.

15. Powis K, De Virgilio C. 2016. Conserved regulators of Rag GTPases orchestrate amino acid-dependent TORC1 signaling. Cell Discov 2.

16. Sancak Y, Peterson TR, Shaul YD, Lindquist RA, Thoreen CC, Bar-Peled L, Sabatini DM. 2008. The rag GTPases bind raptor and mediate amino acid signaling to mTORC1. Science (80-) 320:1496–1501.

17. Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. 2010. Ragulator-rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. Cell 141:290–303.

18. Yang H, Jiang X, Li B, Yang HJ, Miller M, Yang A, Dhar A, Pavletich NP. 2017. Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. Nature 552:368–373.

19. Han JM, Jeong SJ, Park MC, Kim G, Kwon NH, Kim HK, Ha SH, Ryu SH, Kim S. 2012. Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. Cell 149:410–424.

20. Zhang W-W, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. mSphere 2:1–15.

21. Zhang WW, Lypaczewski P, Matlashewski G. 2020. Application of CRISPR/Cas9-Mediated Genome Editing in Leishmania, p. 199–224. *In* Ginger, M, Zilberstein, D, Michels, P (eds.), Methods in Molecular Biology. Springer Nature.

22. Shen K, Huang RK, Brignole EJ, Condon KJ, Valenstein ML, Chantranupong L, Bomaliyamu A, Choe A, Hong C, Yu Z, Sabatini DM. 2018. Architecture of the human GATOR1 and GATOR1-Rag GTPases complexes. Nature 556:64–69.

23. Valbuena N, Guan KL, Moreno S. 2012. The Vam6 and Gtr1-Gtr2 pathway activates TORC1 in response to amino acids in fission yeast. J Cell Sci 125:1920–1928.

24. Kim J, Kim E. 2016. Rag GTPase in amino acid signaling. Amino Acids 48:915–928.

25. Patel N, Singh SB, Basu SK, Mukhopadhyay A. 2008. Leishmania requires Rab7-mediated degradation of endocytosed hemoglobin for their growth. Proc Natl Acad Sci U S A 105:3980–3985.

26. Denny PW, Lewis S, Tempero JE, Goulding D, Ivens AC, Field MC, Smith DF. 2002. Leishmania RAB7: Characterisation of terminal endocytic stages in an intracellular parasite. Mol Biochem Parasitol 123:105–113.

27. Kim J, Guan KL. 2019. mTOR as a central hub of nutrient signalling and cell growth. Nat Cell Biol 21:63–71.

28. Goldman-Pinkovich A, Kannan S, Nitzan-Koren R, Puri M, Pawar H, Bar-Avraham Y, McDonald J, Sur A, Zhang WW, Matlashewski G, Madhubala R, Michaeli S, Myler PJ, Zilberstein D. 2020. Sensing host arginine is essential for leishmania parasites' intracellular development. MBio 11:1–13.

29. Gong R, Li L, Liu Y, Wang P, Yang H, Wang L, Cheng J, Guan KL, Xu Y. 2011. Crystal structure of the Gtr1p-Gtr2p complex reveals new insights into the amino acid-induced TORC1 activation. Genes Dev 25:1668–1673.

30. Samarasinghe SR, Samaranayake N, Kariyawasam UL, Siriwardana YD, Imamura H, Karunaweera ND. 2018. Genomic insights into virulence mechanisms of Leishmania donovani: Evidence from an atypical strain. BMC Genomics 19.

31. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Bye-AJee H, Cowley A, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Castro LG, Figueira L, Garmiri P, Georghiou G, Gonzalez D, Hatton-Ellis E, Li W, Liu W, Lopez R, Luo J, Lussi Y, MacDougall A, Nightingale A, Palka B, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Speretta E, Turner E, Tyagi N, Volynkin V, Wardell T, Warner K, Watkins X, Zaru R, Zellner H, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, ArgoudPuy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, De Castro E, Coudert E, Cuche B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J. 2017. UniProt: The universal protein knowledgebase. Nucleic Acids Res 45:D158–D169.

32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

33. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H. 2009. TriTrypDB: A functional genomic resource for the Trypanosomatidae. Nucleic Acids Res

38:D457-62.

34. Alcolea PJ, Alonso A, Degayón MA, Moreno-Paz M, Jiménez M, Molina R, Larraga V. 2016. In vitro infectivity and differential gene expression of Leishmania infantum metacyclic promastigotes: Negative selection with peanut agglutinin in culture versus isolation from the stomodeal valve of Phlebotomus perniciosus. BMC Genomics 17.

35. Zhang WW, Matlashewski G. 2015. CRISPR-Cas9-mediated genome editing in Leishmania donovani. MBio 6:e00861-15.

36. Zhang WW, Charest H, Matlashewski G. 1995. The expression of biologically active human p53 in Leishmania cells: A novel eukaryotic system to produce recombinant proteins. Nucleic Acids Res 23:4073–4080.

37. Zhang WW, Matlashewski G. 2010. Screening Leishmania donovani-specific genes required for visceral infection. Mol Microbiol 77:505–517.

38. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.

39. Nivón LG, Moretti R, Baker D. 2013. A Pareto-Optimal Refinement Method for Protein Design Scaffolds. PLoS One 8.

40. Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D. 2014. Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci 23:47–55.

41. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popović Z, Baker D, Players F. 2011. Algorithm discovery by protein folding game players. Proc Natl Acad Sci U S A 108:18949–18953.

42. Tyka MD, Keedy DA, André I, Dimaio F, Song Y, Richardson DC, Richardson JS, Baker D. 2011. Alternate states of proteins revealed by detailed energy landscape mapping. J Mol Biol 405:607–618.

43. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

3.8 Additional Information

3.8.1 Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

3.8.2 Acknowledgements

This research was support by grant from the Canadian Institutes of Health Research (CIHR) to GM and a doctoral training award from the Fond de Recherche du Quebec en Santé (FRQS) to PL.

The funding agencies had no role in the design, collection, analysis, or decision to publish this study.

3.8.3 Author Contributions

PL designed the study, performed experiments, data analysis and wrote the manuscript. WZ designed the study, performed experiments, data analysis and helped write the manuscript. GM designed the study, helped write and edited the manuscript. All authors reviewed the manuscript.

3.8.4 Ethics Statement

The animal protocol was approved by the McGill University Faculty of Medicine Animal Care Committee, Protocol Number: 2012-7112

3.8.5 Competing Interest

The authors declare no conflict of interest.

3.9 Figures and tables

Table 3. 1 List of *Leishmania donovani* proteins homologous to humans in the RagC axis of the TOR pathway

Protein Name	Protein Complex	UniProt Accession	*L. donovani* Protein ID	Present in *Leishmania spp*	Coverage (Match/query)	Identity	BLAST score	E-value
CASTOR		Q8WTX7	ABSENT	None				
SEC13	**GATOR2**	P55735	LdCL_320005400	I, Mj, Tr, Mx, Br	310/322	40%	198	1.00E-58
WDR59		Q6PJI9	LdCL_300009300	I, Mj, Tr, Mx, Br	123/974	31%	70.1	7.00E-11
WDR24		Q96S15	LdCL_200015700	I, Mj, Tr, Mx, Br	69/790	43%	70.9	5.00E-11
SEH1L		Q96EE3	LdCL_320005400	I, Mj, Tr, Mx, Br	302/360	28%	99.8	1.00E-22
MIOS		Q9NXC5	LdCL_250018400	I, Mj, Tr, Mx, Br	82/875	40%	60.1	5.00E-08
NPRL2	**GATOR1**	Q8WTW4	ABSENT	None				
NPRL3		Q12980	ABSENT	None				
DEPDC		O75140	LdCL_070015500	I, Mj, Tr, Mx, Br	80/1603	34%	56.2	2.00E-06
Sestrin 1	**Sestrin**	Q9Y6P5	LdCL_180011700	I, Mj, Tr, Mx, Br	109/492	25%	49.7	3.00E-05

Sestrin 2		P58004	LdCL_180011700	I, Mj, Tr, Mx, Br	149/480	25%	54.3	1.00E-06
LeuRS		Q9P2J5	LdCL_130015300	I, Mj, Tr, Mx, Br	1089/1176	42%	843	0.00E+00
SH3BP4		Q9P0V3	ABSENT	None				
RagA		Q7L523	LdCL_130020300	I, Mj, Tr, Mx, Br	338/313	36%	218	2.00E-66
RagB		Q5VZM2	ABSENT	None				
RagC		Q9HB90	LdCL_360068800	I, Mj, Tr, Mx, Br	356/399	37%	231	3.00E-70
RagD		Q9NQL2	ABSENT	None				
LAMTOR1	**RAGULATOR**	Q6IAA8	ABSENT	None				
LAMTOR2		Q9Y2Q5	ABSENT	None				
LAMTOR3		Q9UHA4	ABSENT	None				
LAMTOR4		Q0VGL1	ABSENT	None				
LAMTOR5		O43504	ABSENT	None				
SLC38A9		Q8NBW4	LdCL_270011600	I, Mj, Tr, Mx, Br	161/562	22%	36.2	0.056
v-ATPase (subunit H)		Q9UI12	LdCL_210021400	I, Mj, Tr, Mx, Br	312/483	33%	153	1e-40
Rab7		P51149	LdCL_180014100	I, Mj, Tr, Mx, Br	227/207	52%	224	3.00E-72
KPTN	**KICSTOR**	Q9Y664	ABSENT	None				
SZT2		Q5T011	ABSENT	None				

C12Orf66		Q96MD2	ABSENT	Mj	136/445	26%	37	0.12
ITFG2		Q969R8	ABSENT	Mj	83/447	30%	34.3	0.99
Rheb		Q15382	LdCL_320027000	I, Mj, Tr, Mx, Br	156/184	37%	90.5	3.00E-21
FLCN		Q8NFG4	ABSENT	None				
FNIP1		Q8TF40	ABSENT	None				
SAMTOR		Q1RMZ1	LdCL_230020100	I, Mj, Tr, Mx, Br	88/405	35%	55.1	2.00E-07
Arf1		P84077	LdCL_310037200	I, Mj, Tr, Mx, Br	170/181	75%	281	9.00E-96
RuvB-like 1	**RuvB-like**	Q9Y265	LdCL_340041700	I, Mj, Tr, Mx, Br	455/456	70%	661	0.00E+00
RuvB-like 2		Q9Y230	LdCL_340032000	I, Mj, Tr, Mx, Br	449/463	70%	673	0.00E+00
Raptor	**mTORC1**	Q8N122	LdCL_250011400	I, Mj, Tr, Mx, Br	491/1335	37%	328	2.00E-91
Deptor		Q8TB45	ABSENT	None				
LST8		Q9BVC4	LdCL_100015100	I, Mj, Tr, Mx, Br	314/326	20%	70.1	2.00E-12
mTOR		P42345	LdCL_360073500	I, Mj, Tr, Mx, Br	2395/2549	29%	858	0.00E+00
			LdCL_340051700	I, Mj, Tr, Mx, Br	688/2549	41%	529	3.00E-152
			LdCL_340046900	I, Mj, Tr, Mx, Br	1211/2549	28%	478	8.00E-136

Pras40		Q96B36	ABSENT	None				

Table 3.1 Proteins part of the RagC arm of the TOR pathway in humans and their parent complex names are listed along with UniProt accession numbers used for sequence retrieval. Matching *L. donovani* proteins resulting from a BLASTP search are listed when available along with their respective matching sequence length, percent identity, BLAST score and E-values. Homologues were marked as ABSENT when no protein was found. The 5th column indicates the presence or absence of homologous proteins in *L. infantum (I), L. major (Mj), L. tropica (Tr), L. mexicana (Mx) or L. Braziliensis (Br).* When matching proteins were found in species other than *L. donovani* only, the BLASTP metrics are reflective of the top scoring hit in the indicated species.

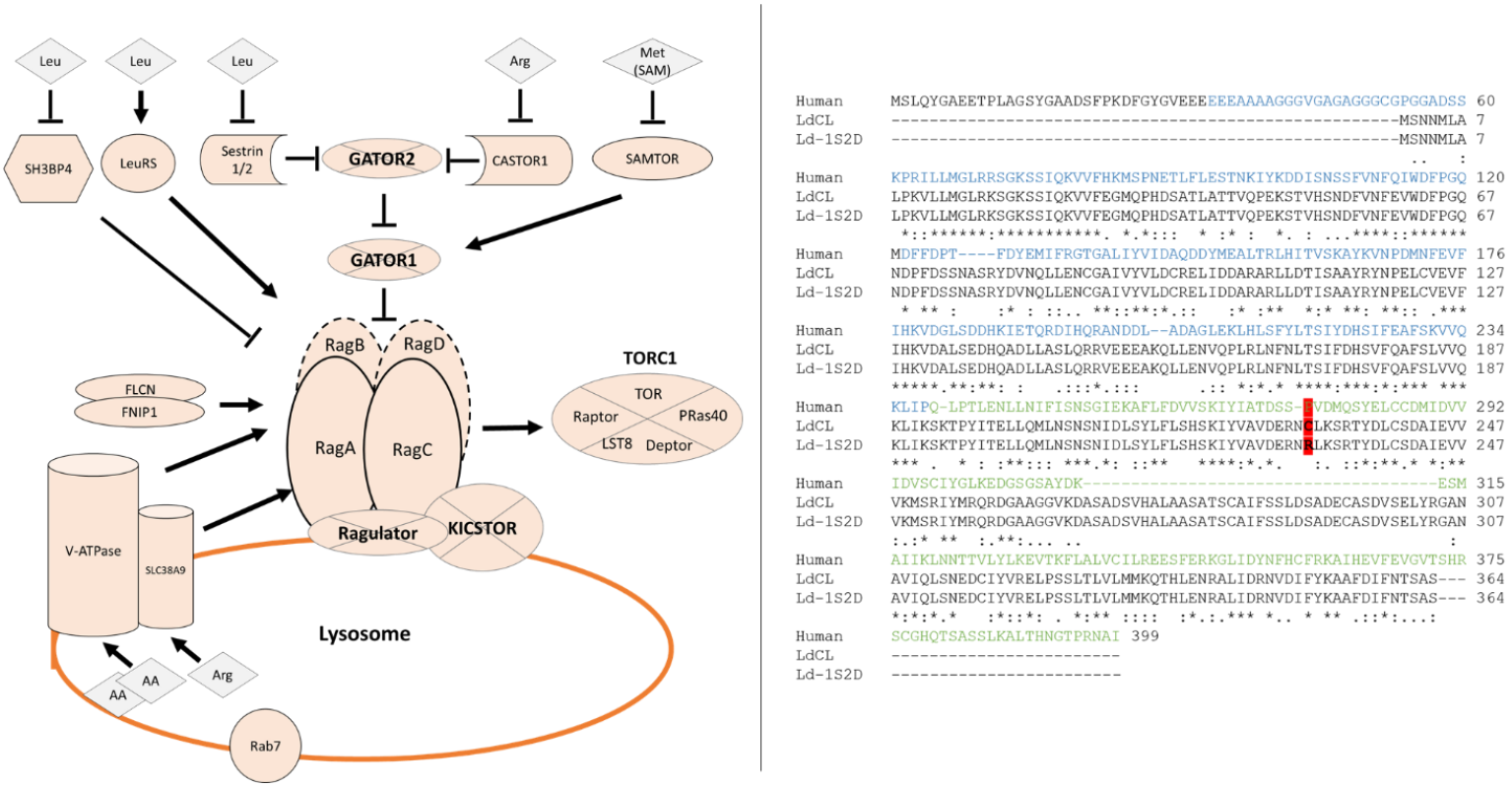


Figure 3.1 Comparison of the Rag pathway and sequence between humans and *Leishmania*

Left: The Rag sensing pathway upstream of mTORC1 signaling based on human studies. Cytosolic amino acid sensors responding to levels of Leucine, Arginine or Methionine are shown on top and inhibit Rag activation repressors or activate activators. Multi-protein complexes labeled in bold on segmented circles. Lysosomal amino acid sensors (vATP-ase and SLC38A9) and Rag heterodimer anchoring complex (Ragulator) shown on the lysosome. The Rag complex integrates signals from nutrient sensors in addition to growth factors (FLCN/FNIP1) resulting in modifications to its GTP/GDP bound state. Active state RagA-RagC bind Raptor effectively recruiting TORC1. *Note: The Rag protein dimer are overlapping on the human diagram as they perform the same function. This dimer consists either of RagA with RagC or of RagB with RagD.* Right: Multiple sequence alignment of human, wildtype and cutaneous *Leishmania* RagC proteins with domain structure labeled on the human RagC sequence with GTPase domain (blue) and C-terminal Roadblock domain (green). The R231C polymorphism identified in the cutaneous isolate is highlighted in red. Sequence identity is marked with (*), sequence similarity is marked with (:)

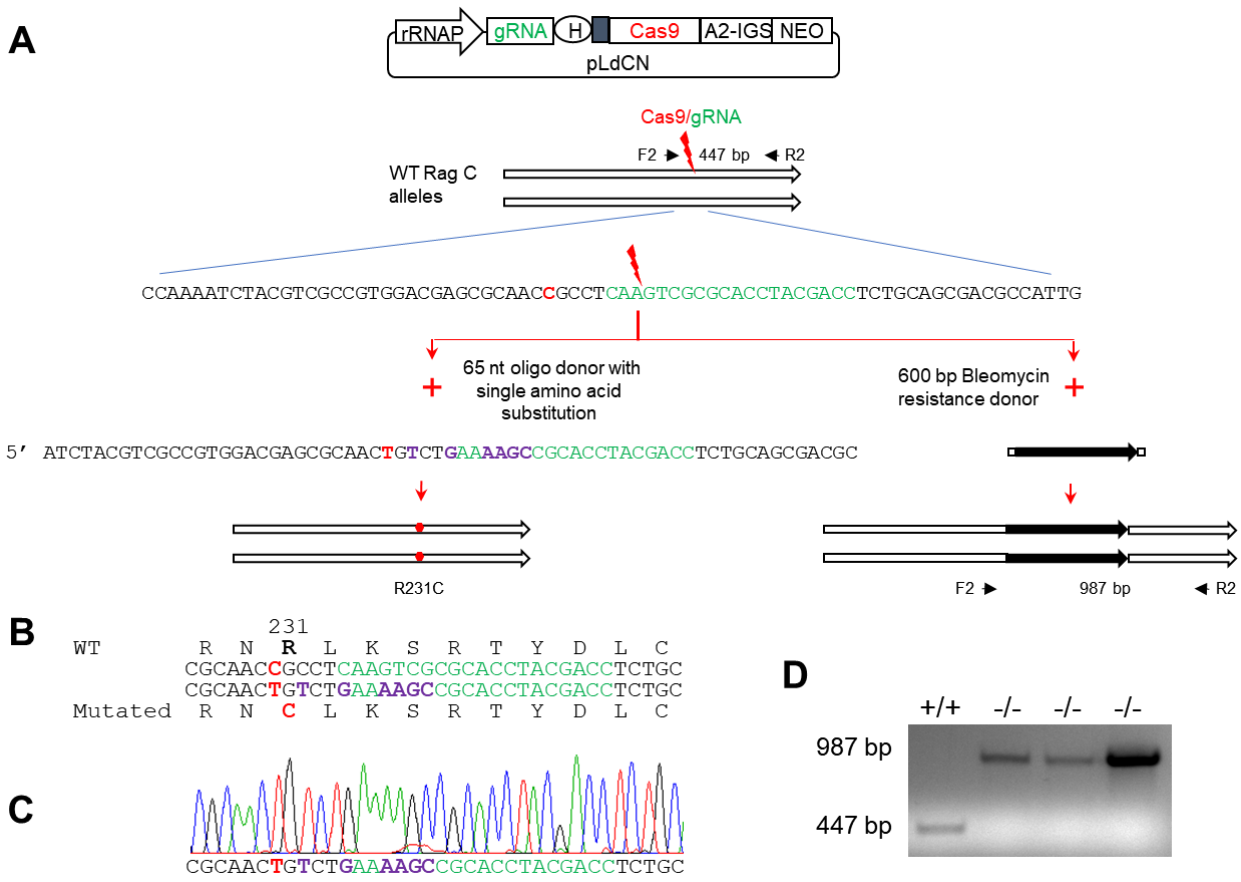


Figure 3.2 Generation of the RagC single amino acid substitution (R231C) mutant and RagC disruption mutant by CRISPR/Cas9

A. Strategy used to generate the RagC R231C mutant and *RagC* gene disrupted strains. A gRNA was designed to target a site (green) close to the R231C polymorphism (red) identified in the *RagC* gene of the Sri Lankan cutaneous *L. donovani* isolate. *L. donovani* 1S2D promastigotes were transfected with a CRISPR vector (pLdCNld366140) expressing this *RagC* specific gRNA, followed by transfection of the donor repair template which contained either: the targeted point mutation (C/T, red) and an additional six nucleotides resulting in silent mutations (purple) to protect the repaired genome from subsequent Cas9 cleavage, or a bleomycin selection marker (black). Genomic DNA from these *L. donovani* cells clones was subjected to PCR and sequencing analysis. **B.** Partial sequence of the oligo donor repair induced mutations resulting in a single amino acid substitution in RagC protein (R231C, shown in red) and inactivation of the gRNA targeting site (shown in green, interspaced with disrupting silent mutations in purple). **C.** Direct sequencing of a PCR product amplified from a *L. donovani* clone showing both alleles of the *RagC* gene have

been edited to the sequence of the oligo donor (see A & B) repair template. **D.** PCR analysis of *RagC* double allele gene disrupted mutants. PCR analysis of three phleomycin resistant clones with primers F2 and R2 show the Bleomycin resistance gene has been inserted into the target site as expected resulting in a 987 bp band, and no 447 bp WT F2R2 band was detected in these *RagC* disruption mutants.

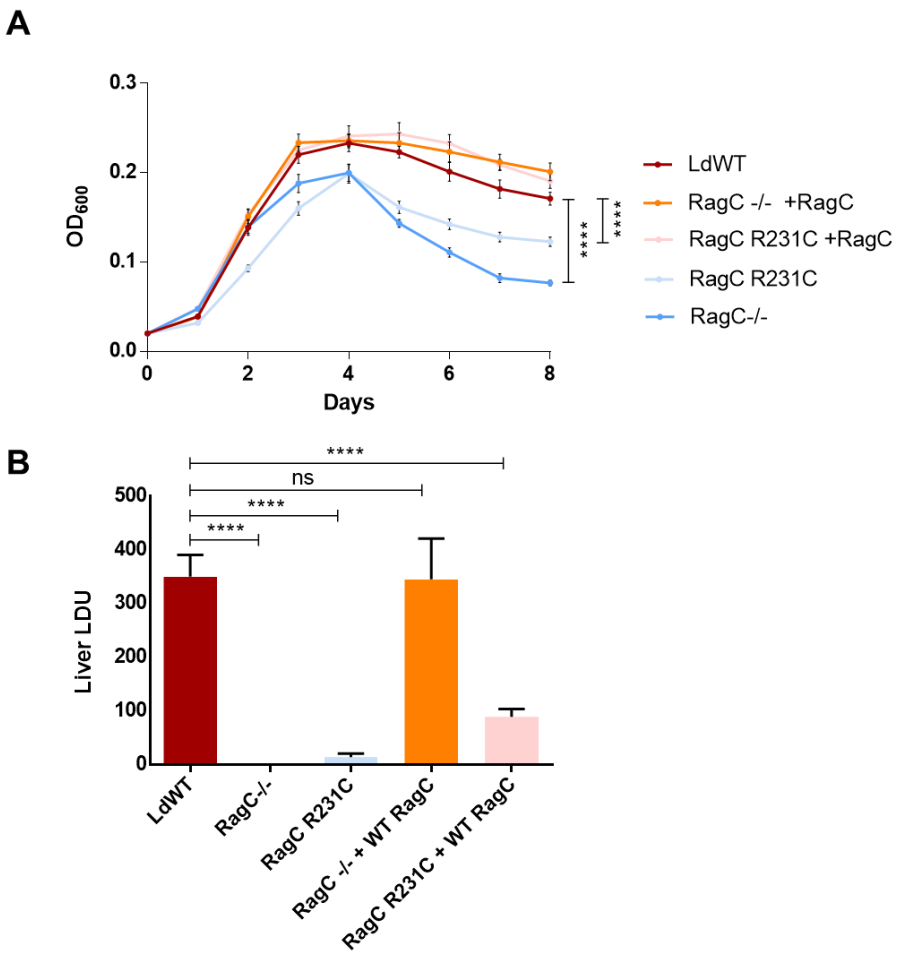


Figure 3.3 Biological effects of a RagC single amino acid change (R231C) and disruption of the *RagC* gene

A. Proliferation of RagC WT (red), gene disrupted (-/-) (blue) and gene edited (R231C) strains (light blue), and RagC WT addback transfections in gene disrupted (orange) and gene edited (pink) strains. Data was measured in quadruplicate and statistical significance calculated using 2way ANOVA with multiple comparisons using RagC WT as the control group and marked for significance if consistent for every time point past day 2. This is the representative data of four repeat experiments. **B.** Effects of a single amino acid change (R231C) and disruption of *RagC* on *L. donovani* 1S2D infection in mice. BALB/c mice were infected by tail vein injection ($1x10^8$ pro/mouse) with *L. donovani* WT, gene edited RagC R231C, gene disrupted *RagC* (RagC-/-), and their corresponding RagC WT addback strains. Mice were examined for liver parasite burden

four weeks after infection. Data was measured using 4 mice per group and statistical significance calculated using 2way ANOVA with multiple comparisons using RagC WT as the control group.

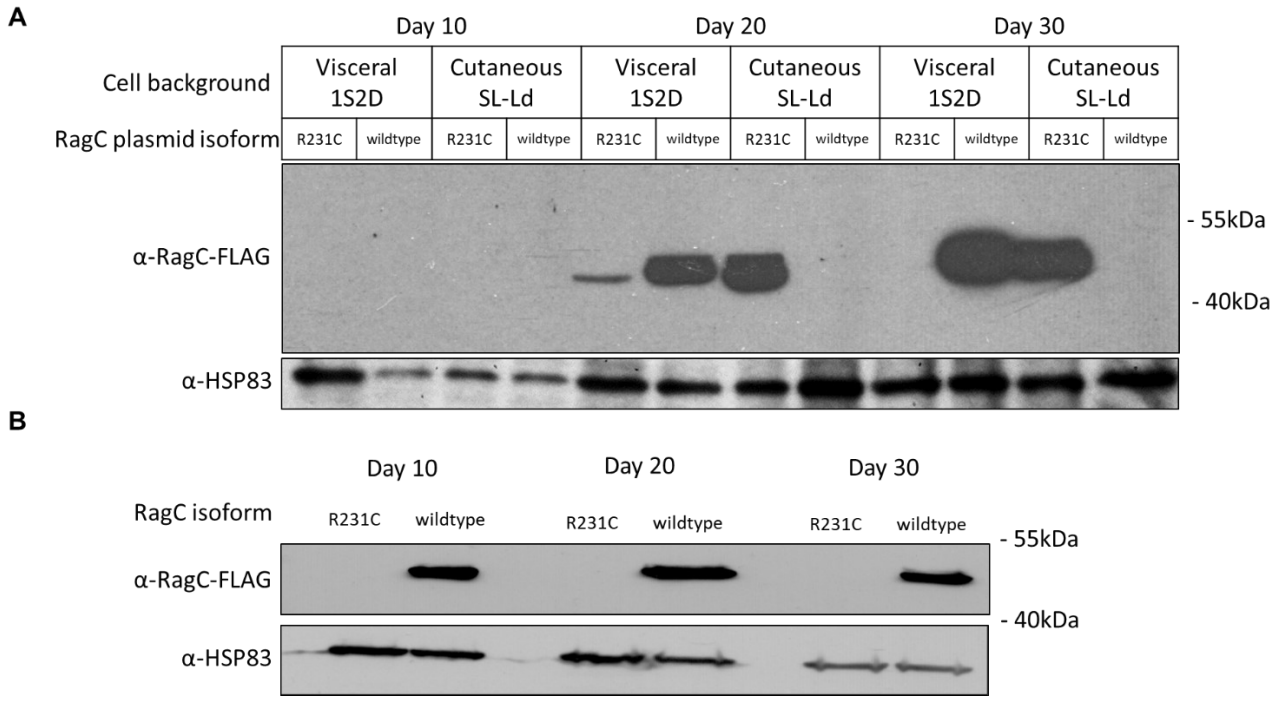


Figure 3.4 Incompatibility between the co-expression of wildtype and mutant RagC due to dominant negative effect

A. Immunoblotting of FLAG-tagged RagC wildtype and R231C mutant transfected in *L. donovani* 1S2D (Ld 1S2D) and cutaneous *L. donovani* from Sri Lanka (Cutaneous SL-Ld) strains followed over a period of 30 days. **B.** Immunoblotting of FLAG-tagged RagC wildtype and R231C transfected in *L. donovani* 1S2D cell following disruption of the endogenous *RagC* gene. The results are representative of three independent experiments.

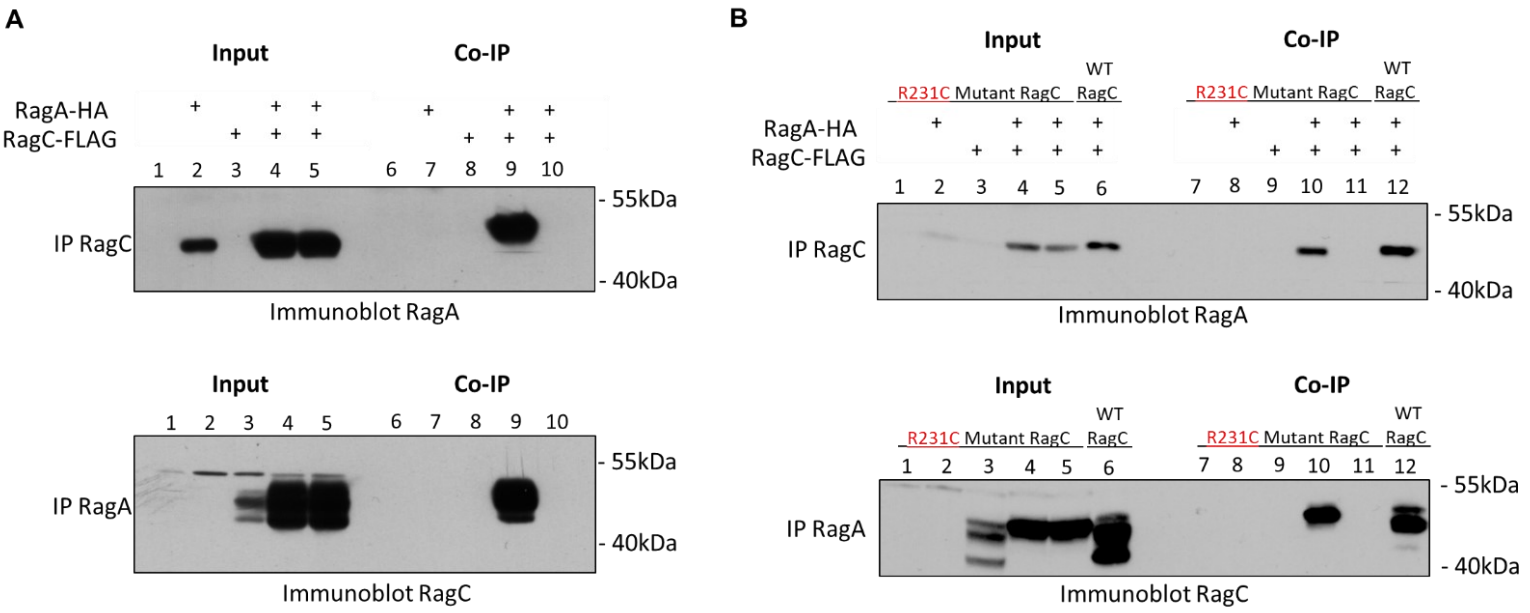


Figure 3.5 ***Leishmania*** **RagA and RagC proteins form a heterodimer complex**

A. Heterodimer formation between wildtype RagC and RagA**.** Top Panel: RagC was immunoprecipitated with FLAG antibodies and immunoblotting with HA antibodies shows the presence of HA-tagged RagA (Lanes 9) specifically in RagC co-transfected cells. Western blot analysis of input lanes 1-5 are also shown. Bottom Panel: Co-immunoprecipitation showing RagC is captured following immunoprecipitation of RagA. HA-tagged RagA was immunoprecipitated followed by immunoblotting with FLAG antibodies shows the presence of RagC (Lane 9) specifically in Rag A co-transfected cells. Western blot analysis of input lanes 1-5 are also shown. **B.** Heterodimer formation between mutant R231C RagC and RagA. Top panel: RagC R231C was immunoprecipitated using FLAG antibodies followed by immunoblotting with HA antibodies shows the presence of HA-tagged RagA in RagC R231C co-transfected cells (Lane 10 and positive control Lane 12). Western blot analysis of input lanes 1-6 are shown. 5B Bottom Panel: HA-tagged RagA was immunoprecipitated using HA antibodies followed by immunoblotting with FLAG antibodies shows the presence of FLAG-tagged RagC (Lane 10 and positive control Lane 12). Western blot analysis of input lanes 1-6 are shown. The results are representative of three independent experiments.

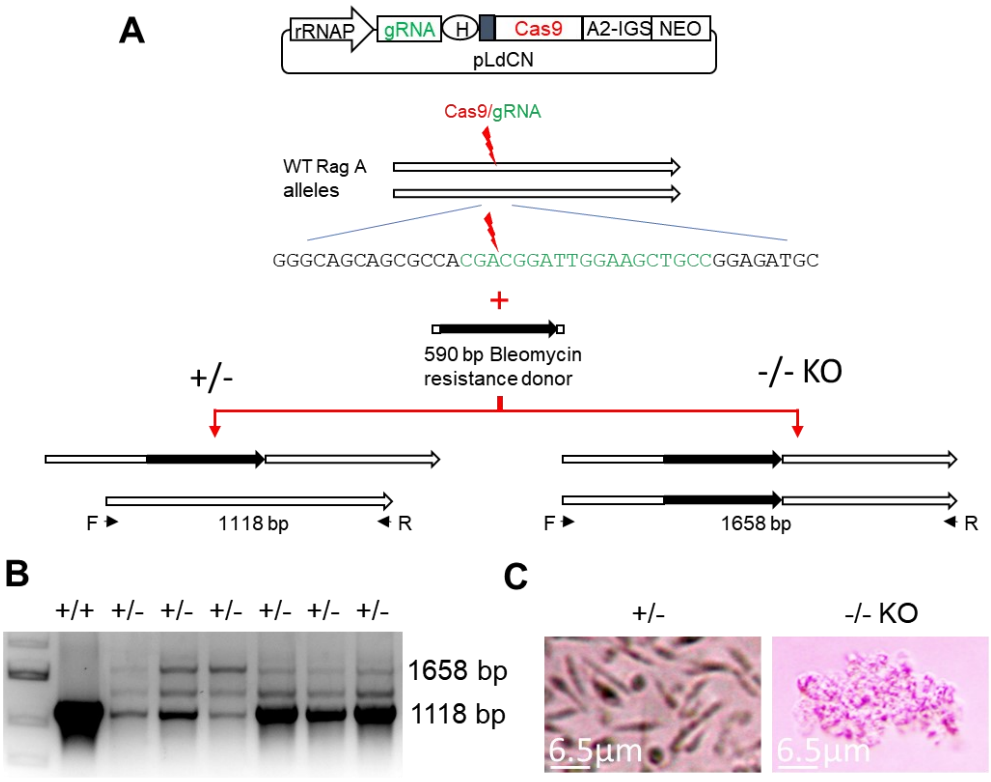


Figure 3.6 *RagA* is essential for *Leishmania*

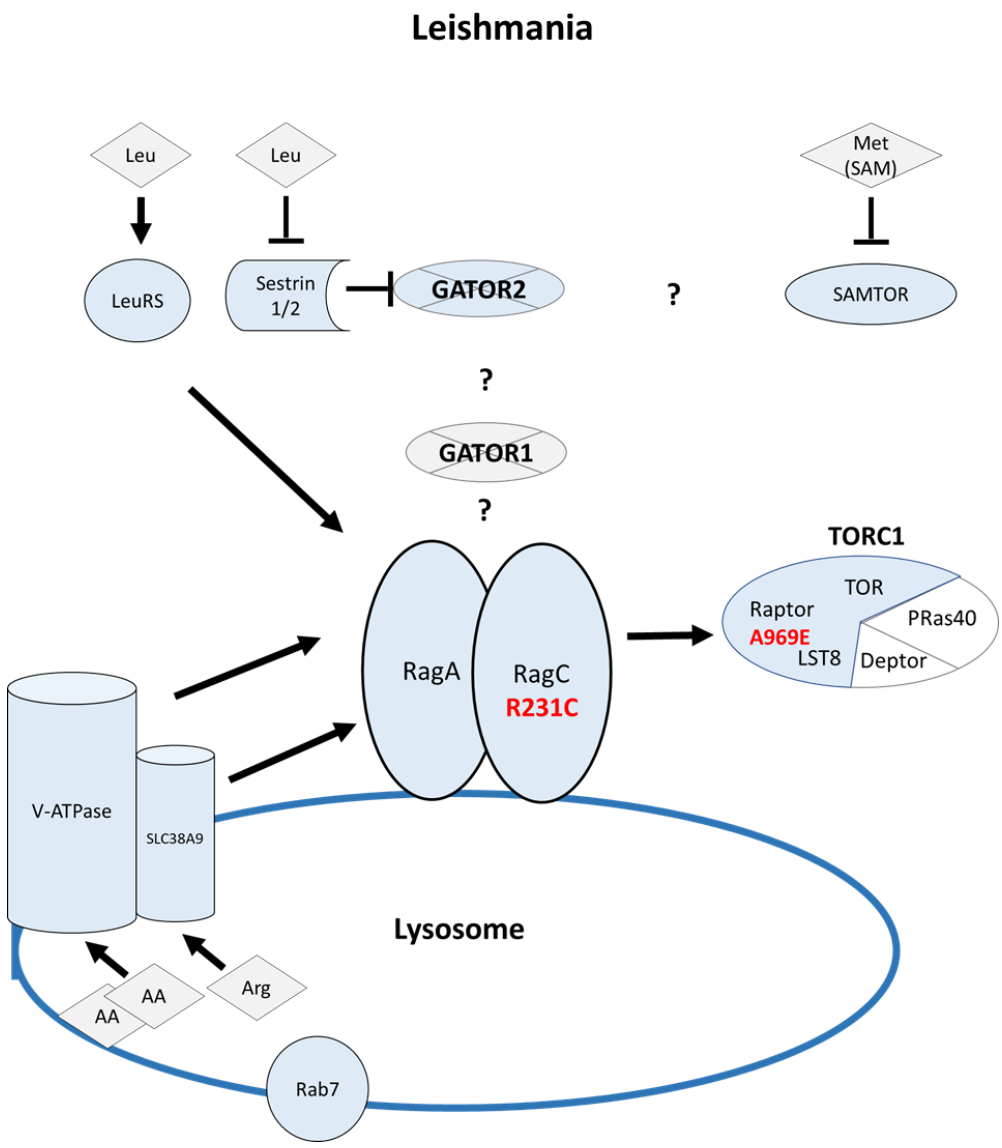
A. Strategy used to generate the single allele or double allele *L. donovani RagA* disruption mutants. A gRNA was designed to target the first half of the RagA coding sequence. *L. donovani* 1S2D cells were transfected with a CRISPR vector (pLdCNld131620) expressing this *RagA* specific gRNA, followed by transfection of the bleomycin selection marker donor with 25 nt flanking sequences to integrate into the Cas9 cut site. **B.** PCR analysis of the surviving phleomycin resistant clones showing the bleomycin resistance gene was inserted into the target site of one *RagA* allele as expected, but the 1118 bp WT *RagA* band was still detected in all these phleomycin resistant clones. *Note: the middle band is the annealing product during PCR between the 1118 bp WT RagA band and the 1658 bp disruption band.* **C**. Microscope images showing the disruption of both *RagA* alleles is lethal for *L. donovani*. The *RagA*+/- partial mutant cells expressing the *RagA* targeting gRNA were cloned in a 96 well plate and cell growth was monitored by microscopy. The image for *RagA* +/- cells was taken one week after cloning; The image for *RagA*-/- cells was taken four weeks after cloning.

3.10 Supplementary Material

Supplementary Table 3.S1 A table showing the percentages of metacyclic like promastigotes present in the stationary phase cultures of wild type *L. donovani*, R231C RagC mutant and the RagC null mutant.

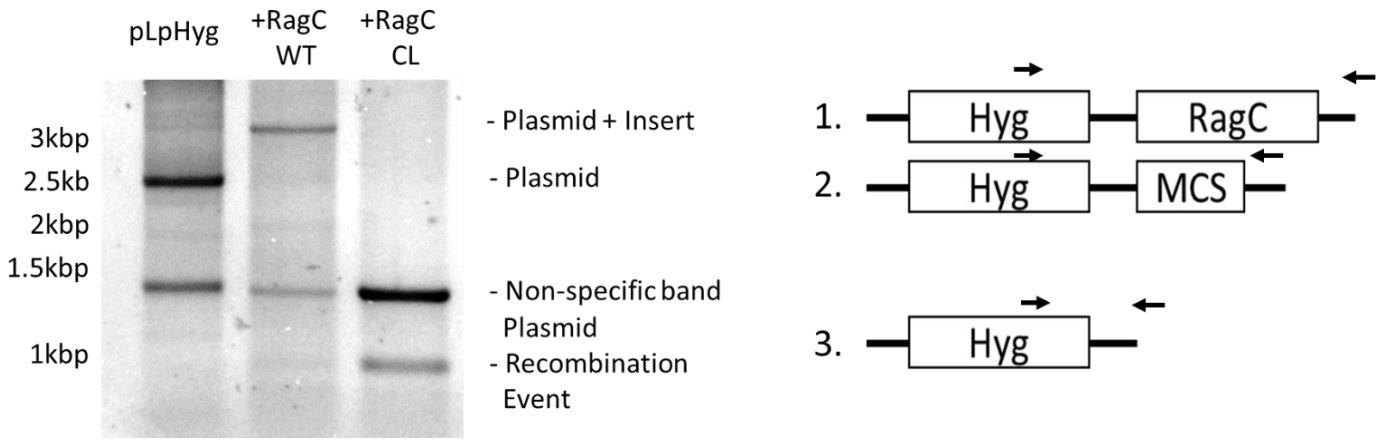
	Cell density in stationary phase	Metacyclic cells from 2 x 10^8 stationary cells	The metacyclic promastigotes rate (%)
WT *L. donovani*	7.925 x 10^7/ml	1.475 x 10^6	0.7375
R231C RagC	6.325 x 10^7/ml	7.4 x 10^6	3.7
RagC null mutant	6.425 x 10^7/ml	6.71 x 10^6	3.355

Supplementary Table 3.S1 Comparison of the proportion of metacyclic cells present in stationary phase cultures.



Supplementary Figure 3.S1 Hypothetical model of the amino acid sensing Rag pathway in *Leishmania*

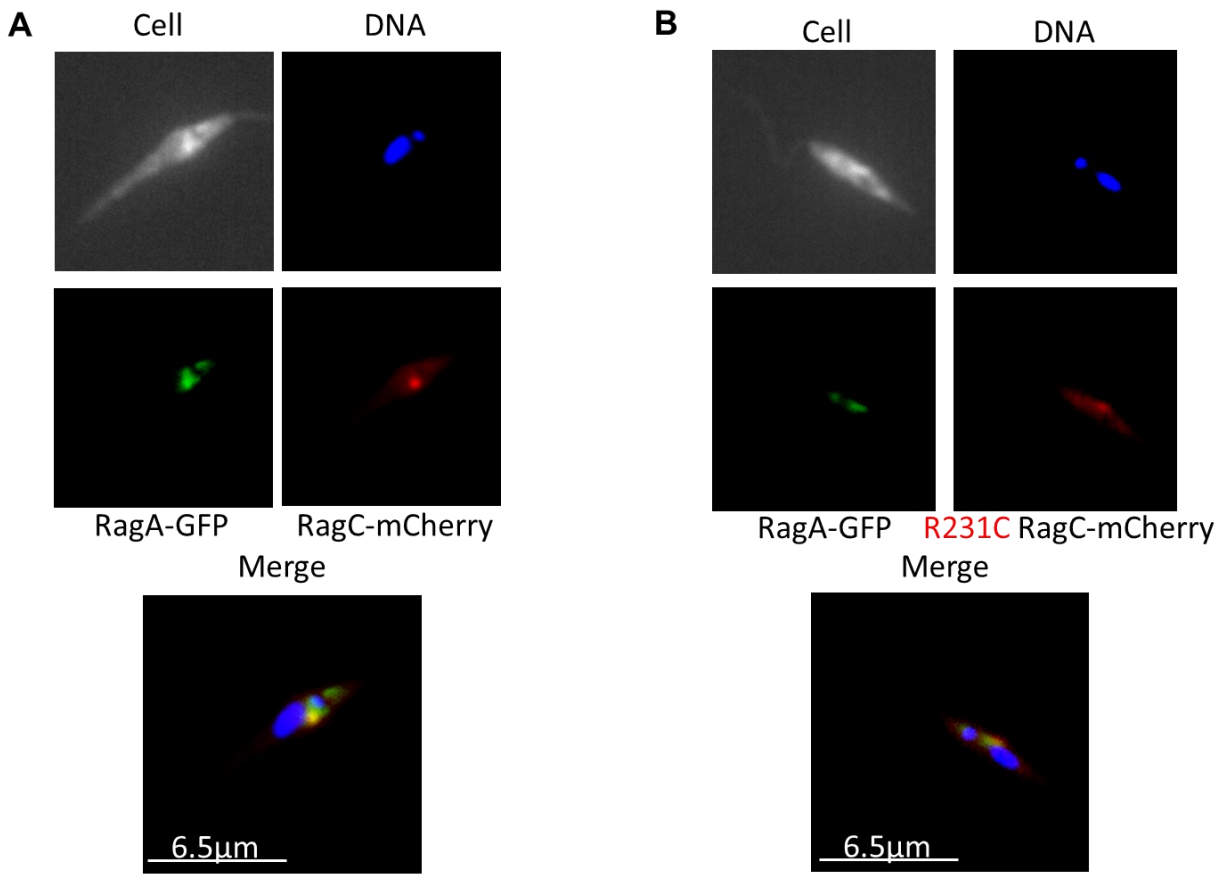
Components of the Rag pathway conserved in *Leishmania* based on sequence homology determined through bioinformatic analysis of the *Leishmania* genome. Note, there are several human components that were not identified in *Leishmania* including RagB and RagD. GATOR1 proteins were identified but with low homology or only partial member homology and are shown in grey. Also indicated are the RagC R231C, and Raptor A969E mutants in red.



Supplementary Figure 3.S2 PCR determination of insert size with the wildtype and mutant RagC protein plasmids

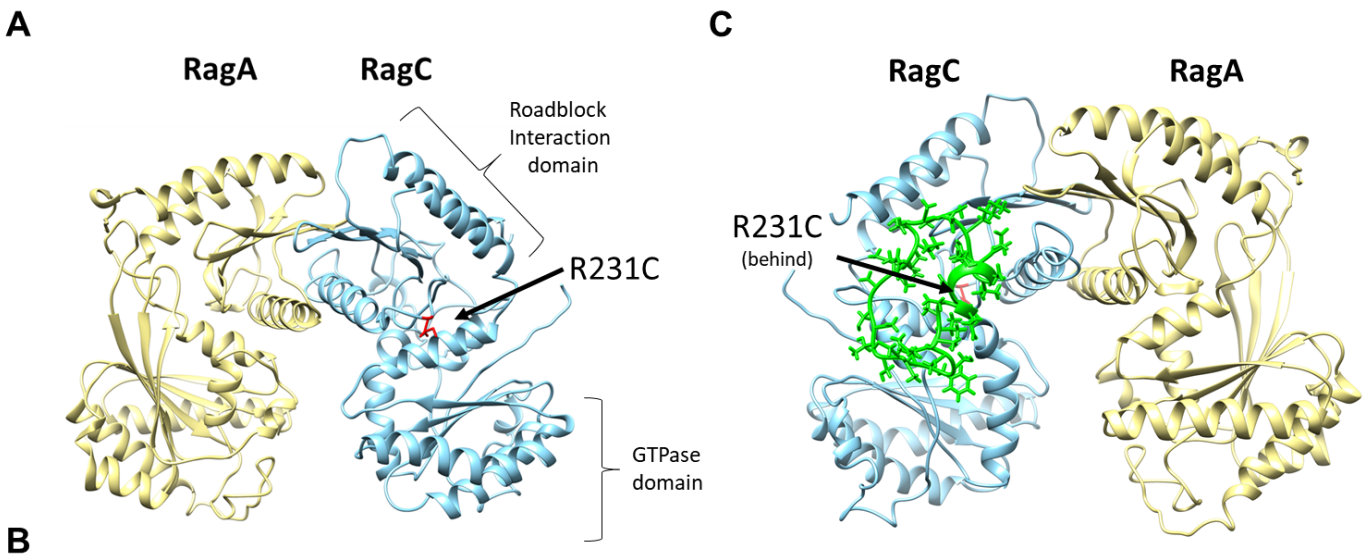
WT promastigotes were transfected with plasmids encoding either the WT or mutant copy of RagC. At 4 weeks post transfection the plasmids were recovered using a plasmid DNA mini-preparation kit.

Primers (arrows) were designed to the C terminal end of the hygromycin resistance cassette (Hyg) and downstream of the protein insertion site (MCS) in order to target episomal copies of *RagC* only. The first lane products originate from an empty plasmid (Band type 2). The second lane shows an increase in band size due to the insertion and retention of the RagC protein on the plasmid (Band type 1). The third lane shows the excision of not only the RagC protein but the flanking UTR as indicated by the large decrease in size of amplicon (Band type 3).



Supplementary Figure 3.S3 Co-localization of RagA with RagC in *L. donovani*

A. Epifluorescence microscopy of RagA and RagC. GFP-tagged RagA and mCherry-tagged RagC form overlapping foci near the kinetoplast and nucleus in transfected promastigotes. **B.** Epifluorescence microscopy of RagA and RagC R231C. GFP-tagged RagA and mCherry-tagged RagC R231C mutant form overlapping foci near the kinetoplast and nucleus in transfected promastigotes.



B

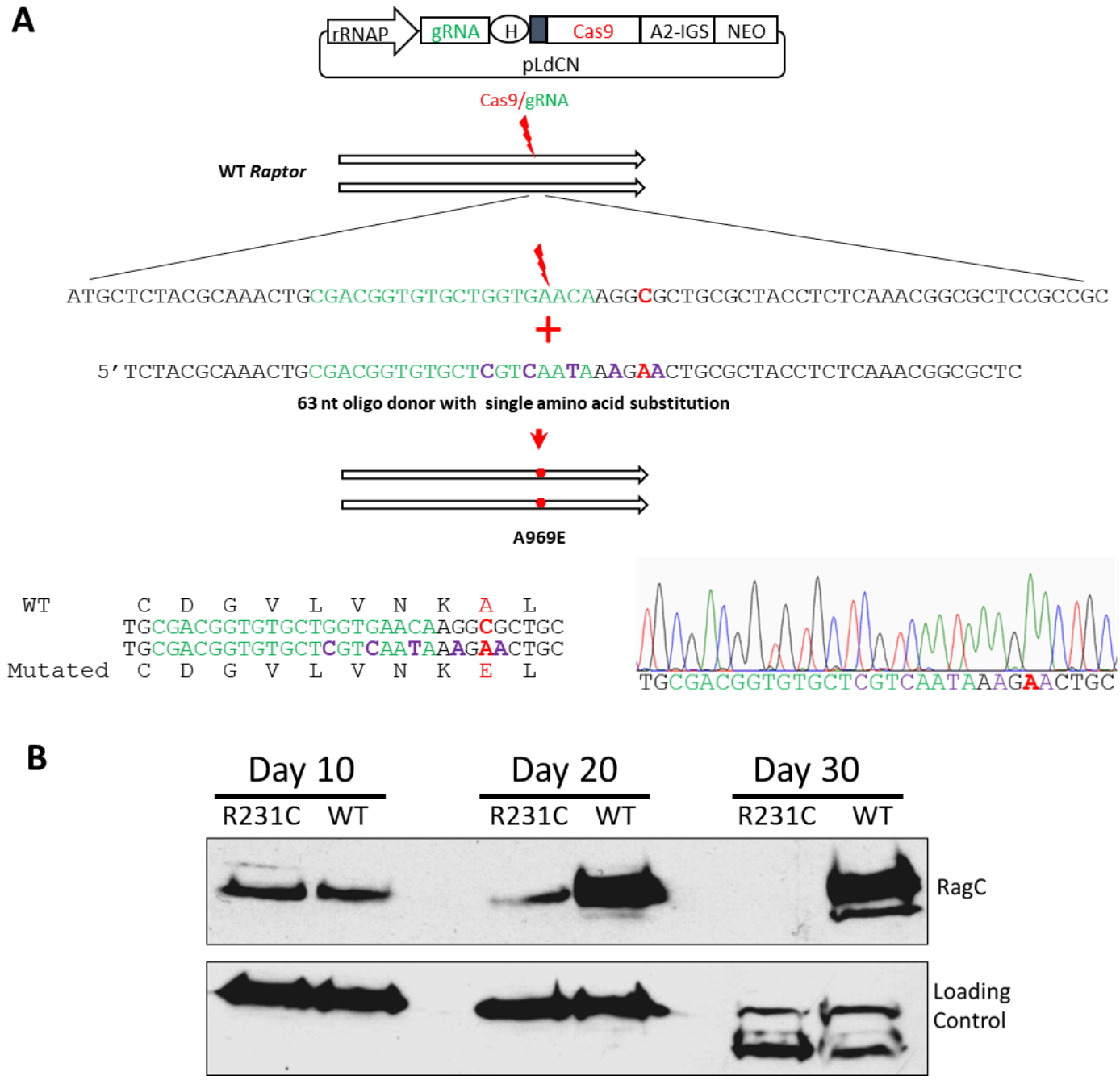
```
Leishmania KLIKSKTPYITELLQMLNSNSNIDLSYLFLSHSKIYVAVDERNRLKSRTYDLCSDAIEVV 247
Human      KLIP-QLPTLENLLNIFISNSGIEKAFLFDVVSKIYIATDS-SPVDMQSYELCCDMIDVV 292
Yeast      KLIP-ELSFLENMLDNLIQHSKIEKAFLFDVNSKIYVSTDS-NPVDIQMYEVCSEFIDVT 238
           ***  :   : ::*: : .:* *: ::**   ****::.*. . :. : *::*.: *:*.

Leishmania VKMSRIYMRQRDGAAGGVKDASADSVHALAASATSCAIFSSLDSADECASDVSELYRGAN 307
Human      IDVSCIYGLKEDGSGSA-------------------------------------YDKESM 315
Yeast      IDLFDLYKAPVLRNSQKSSDK---------------------------DNVINPRNELQ 270
           :.:  :*       .                                      .

Leishmania AVIQLSNEDCIYVRELPSSLTLVLMMKQTHLENR---ALIDRNVDIFYKAAFDIFNTSAS 364
Human      AIIKLNNTTVLYLKEVTKFLALVCILREESFERK---GLIDYNFHCFRKAIHEVFEVGVT 372
Yeast      NVSQLANGVIIYLRQMIRGLALVAIIRPNGTDMESCLTVADYNIDIFKKGLEDIWANARA 330
           : :* *   :*::::   *:** :::    : .    : * *.. * *.  :::  . :
```

Supplementary Figure 3.S4 Location of the RagC R231C mutation

A. Homology modelling of *Leishmania donovani* RagA (yellow) in complex with RagC R231C (blue) shown in ribbon representation. The 231st amino acid position on RagC corresponding to the identified mutation is highlighted in red full atom representation and indicated by arrows. The complex is oriented with the Roadblock interaction domains shown on top and the GTPase domains at the bottom. **B.** Multiple sequence alignment between *Leishmania*, human and yeast RagC homologue sequences with *L. donovani* RagC R231 highlighted. C-terminal portion of the alignment shown with a 37 amino acid long sequence only seen in *Leishmania* highlighted **C.** Homology model of the RagA/RagC complex with the *Leishmania* specific sequence from B. shown in green.



Supplementary Figure 3.S5 Generation of 1S2D *Raptor* mutant parasites does not enable expression of RagC R231C

A. Strategy used to insert the Raptor A969E mutation into the *Raptor* gene of *L. donovani.* A gRNA was designed to target the *Raptor* gene at the site of the desired C to A base change and cloned into the CRISPR vector (pLdCNld251140) expressing this gRNA and Cas9 nuclease. This plasmid was transfected into 1S2D *L. donovani* cells followed by the transfection of a 63 nt donor

oligonucleotide with 25 nt flanking sequences to introduce the desired C to A conversion (red) and 5 silent mutations to prevent further cleavage by Cas9 (purple). The cells were then cloned into 96-well plates and screened by PCR and Sanger sequencing. **B.** Immunoblot following the expression of RagC isoforms in the *Raptor* mutant parasites. The isolated parasites were transfected with plasmids expressing either the WT or R231C isoforms of RagC. At 10 days post-transfection, both isoforms are expressed at comparable low levels. At 20 days, the expression of the WT isoform is stabilized, the R231C isoform appears at slightly reduced levels compared to the 10-day time point. At 30 days post transfection, only the WT isoform is stably expressed.

CHAPTER 4: LEISHMANIA DONOVANI HYBRIDIZATION AND INTROGRESSION IN NATURE: A COMPARATIVE GENOMIC INVESTIGATION

Patrick Lypaczewski and Greg Matlashewski

4.1 Preface

As discussed in **Chapter 2,** several potential virulence factors were identified, possibly contributing to the phenotype seen in Sri Lanka. In **Chapter 3,** evidence was presented that a polymorphism in the *RagC* gene contributes to the attenuated phenotype of the cutaneous *L. donovani* strain in Sri Lanka. However, neither the findings presented in **Chapter 2** or **Chapter 3** were shared in additional strains more recently sequenced and could therefore not explain the phenotype of these new atypical strains from Sri Lanka. We therefore launched a global genomic study of the *L. donovani* species using all available sequencing data made public by researchers around the world to identify the etiological origins of Sri Lankan *L. donovani* samples to explain the atypical disease phenotype of those strains. Our results suggest that inter-species genome hybridization and import may have contributed to the epidemic of cutaneous leishmaniasis in Sri Lanka.

• Adapted from: **Patrick Lypaczewski**, Greg Matlashewski, 2021. *Leishmania donovani* hybridization and introgression in nature: a comparative genomic investigation. *The Lancet Microbe*, (accepted).

4.2 Abstract

Background: Leishmaniasis is a neglected tropical disease transmitted by infected sand flies resulting in diverse human pathologies contingent on the species of *Leishmania*. The *Leishmania donovani* species causes highly virulent fatal visceral leishmaniasis (VL) and the *Leishmania major* and *Leishmania tropica* species cause less virulent cutaneous leishmaniasis (CL) where the infection remains in the skin at the site of the sand fly bite. The aim of this study is to investigate the genetic basis for the emergence of *L. donovani* strains that cause CL instead of VL in Sri Lanka.

Methods: All available sequencing data for *L. donovani* samples from Asia and Africa in the Sequence Read Archive (SRA) were retrieved and filtered to select for genomic material with no region bias, were sequenced on high accuracy Illumina platforms in paired end mode and resulted in coverage of the entire reference genome. These data were used to perform sequence alignments against the reference *L. donovani* genome from Sri Lanka and sequence analysis was used to assess the presence of genomic recombination markers and the presence of foreign genetic sequences in the genomes of *L. donovani* isolates from Sri Lanka associated with CL.

Findings: Here we show that twelve *L. donovani* strains from Sri Lanka form three separate phylogenetic groups. In one group, the density of heterozygous variants is higher than in previously characterized *Leishmania* hybrid strains. BLAST analysis showed this group contains genomes with gene polymorphisms homologous with *L. major* and *L. tropica* genomes in 22% (2160/9757) to 78% (7671/9757) of all genes analysed. Analysis by phylogeny and BLAST generates evidence that the *L. donovani/L. major* and *L. donovani/L. tropica* hybrid strains originated from Africa and are phylogenetically different from the *L. donovani* strains in neighboring India.

Interpretation: Novel *L. donovani* strains may arise in new environments through the integration of genes from another species. Evidence of *L. donovani* hybrid strains generates the hypothesis that hybridization followed by recombination and introgression with genomes from *L. major* and *L. tropica* contributed to the emergence of offspring capable of causing CL in Sri Lanka.

Funding: Canadian Institutes of Health Research and Fonds de recherche du Québec – Santé

4.2.1 Research in context

Evidence before this study: The reason why some *Leishmania* species cause fatal visceral infection while others cause more benign cutaneous infection represents a foremost and unresolved question. Elucidating the evolutionary history of atypical *Leishmania* strains, including genome hybridization, introgression and recombination between different species that cause different pathologies can provide insights into genetic adaptations associated with different epidemiology and disease outcomes. We conducted a literature review on PubMed including the terms Leishmaniasis AND whole genome sequence, Leishmaniasis AND hybridization, *Leishmania* whole genome analysis AND introgression, *Leishmania donovani* AND cutaneous leishmaniasis. The majority of papers concerning genome sequence analysis of atypical *L. donovani* associated with cutaneous leishmaniasis was from studies performed on strains from Sri Lanka, but none of these studies investigated the possibility of *Leishmania* genome hybrids, recombination or introgression involving cutaneous disease-causing *Leishmania* species. These previous studies also limited their comparisons to the native Sri Lankan and Indian *L. donovani* strains due to geographical proximity.

Added value of this study: This study provides evidence that *L. donovani* can form hybrids with both *L. major* and *L. tropica* and following recombination and introgression, the resulting hybrid offspring are viable in nature. These observations advance the hypothesis that *L. donovani* gene polymorphisms derived from *L. major* and *L. tropica* contributed to the ongoing transmission of CL in Sri Lanka.

Implications of all the available evidence: Novel *Leishmania* strains may become indigenous in new locations through the integration of genes from one species to another. Research involving *Leishmania* strains with mixed genomes and atypical human disease phenotypes provides a unique opportunity to identify genetic changes associated with disease pathogenesis and can provide insight into environmental factors that support the propagation of novel *Leishmania* strains associated with newly emerging public health threats.

4.3 Introduction

Leishmaniasis is a neglected tropical disease present throughout the tropics and subtropics and is caused by protozoan parasites from the genus *Leishmania* that are transmitted by infected sand flies (1). There are two major pathologic forms associated with leishmaniasis contingent on the species of Leishmania; cutaneous leishmaniasis (CL) results in skin lesions at the site of the sand fly bite that usually self-heal within several weeks, and the more virulent visceral leishmaniasis (VL) where *Leishmania* egresses the skin and infects the visceral organs and is fatal if not treated. VL is the second most deadly vector-borne parasitic diseases after malaria (1). The highest incidence of VL is in South East Asia and Sub-Saharan Africa where the predominant species is *Leishmania donovani* transmitted by *Phlebotomus spp* sand flies with humans as the

only known reservoir (1). CL is more widespread globally and is caused by numerous *Leishmania* species, all of which have animal reservoirs with the exception of *Leishmania tropica* which has mainly human reservoirs with some animal reservoirs (1, 2).

Genome hybridization between different animal species can provide an influx of genes accompanied by new phenotypic traits. Hybrids between the same or different species of *Leishmania* have been described (3–6). Inter- and intraspecies hybrids have been generated experimentally in the sand fly vector, confirming that the promastigote stage of *Leishmania* can form hybrids and carry out genetic exchange (7–10). Classical chromosome crossing over during meiotic-like recombination and the ability to experimentally conduct backcrosses with F1 progeny has been demonstrated with intraspecies hybrids in sand flies (11). These observations confirm that inter- and intra-species genetic exchange can occur both experimentally and in nature. More common than full hybrids with equal contributions from each genome is the process of introgression where there is an unequal contribution from each genome following backcrossing with a parental strain and recombination of genomes in an incipient hybrid. In the context of this study, the term hybrids include strains that have undergone introgression following an initial hybridization event as introgression in *Leishmania* involves an initial full genome hybridization event. Uncertainty remains however as to what extent interspecies *Leishmania* hybridization and introgression can generate offspring strains with different epidemiology and pathogenesis in nature.

CL has recently become endemic in Sri Lanka with over 15,000 reported cases since 2001 (12, 13). In contrast, there have been only 7 suspected cases of VL in Sri Lanka mostly in individuals with co-morbidities associated with immunosuppression (14) and none of these have been confirmed and reported to the World Health Organization (WHO) (15). CL is atypical in Sri

Lanka because it is caused by *L. donovani* which causes VL in other countries, although CL caused by *L. donovani* has also recently been observed in some regions of India and Nepal (13, 16).

Ongoing research in our laboratory has identified a number of non-synonymous single nucleotide polymorphisms (SNPs) and copy number variations in a *L. donovani* isolate causing CL in Sri Lanka (17, 18). These genetic polymorphisms included genes associated with cell signalling and virulence that may contribute to the atypical CL phenotype of *L. donovani* in Sri Lanka in our original isolate (17, 18). More recently, genome sequences from additional *L. donovani* isolates from Sri Lanka have been reported and deposited in National Center for Bioinformatic Information (NCBI) GenBank (19). Interestingly, these newer isolates did not contain any of the sequence variations that were previously identified (17, 18) suggesting the presence of multiple co-existing *L. donovani* strains in Sri Lanka. The aim of this study was to examine genomic polymorphisms and phylogenetic evolution of *L. donovani* strains to investigate the emergence of atypical CL in Sri Lanka.

4.4 Results

4.4.1 Global distribution of *L. donovani* sequences and the divergence of strains from Sri Lanka

To understand the geographic origins of the *L. donovani* strains circulating in Sri Lanka, we compared their genomes to all available *L. donovani* genomes in GenBank. The initial search yielded 1238 sequencing records of which 554 records were removed due to insufficient quality as outlined in **Figure 4.1A**. The remaining 684 read sets spanning Asia, Africa and Europe used for phylogenetic analysis are listed in **Supplementary Table 4.S1**. Within the 684 records, 12 were of Sri Lankan origin.

A neighbor-joining phylogenetic tree was generated as shown in Figure 1B. This tree is also available in an interactive format at https://itol.embl.de/tree/132216267336879158013475 5, and the accession codes are also listed in **Supplementary Table 4.S1**. The available sequences from the Sri Lankan *L. donovani* isolates formed three distinct groups termed SL1, SL2 and SL3 based on this phylogenetic analysis. The SL1 group comprising 4 isolates was closer to the Indian subcontinent group and included strains originally isolated from Sri Lanka almost 10 years ago (17, 18), as well as sequences from an independent group (NCBI BioProject PRJEB2600). Five isolates clustered in SL2 were much further than any other *L. donovani* cluster from the Indian subcontinent or Africa. Three isolates clustered in the SL3 group and were part of the African *L. donovani* cluster. This demonstrates that the eight Sri Lankan isolates in groups SL2 and SL3, despite being geographically close to India, are quite different in origin from the four SL1 group isolates and that the SL2 group is distinctive compared to any other *L. donovani* strain. All Indian subcontinent genomes cluster closely together including the slightly divergent ISC1 or "Yeti" group consistent with previous phylogenetic analyses (20, 21). Parasites from Africa formed three separate clusters, largely determined by their geographical isolation location, as previously reported (22).

As one of the Sri Lankan groups (SL2, 5 isolates) diverged more than the African to Indian genetic distances based on branch length (**Figure 4.1A**), we manually inspected the alignments to investigate how this may have occurred. It became apparent that the five genomes from the SL2 group were heavily populated with SNPs occurring in the 40-60% frequency range which was not the case for the SL1 (4 isolates) and SL3 (3 isolates) groups. A representative example of these types of SNPs is shown for a 40 bp section of chromosome 1 (**Figure 4.2A**). The frequency of SNPs is also shown for the entire chromosome 1 (**Supplementary Figure 4.S1A**) and across the

entire genome (**Supplementary Figure 4.S1 B, C, Supplementary Table 4.S2**). These data show the different levels of heterozygosity across the entire genome for all groups with SL2 (5 isolates) in the 50% frequency range for diploid chromosomes while group SL3 genomes contain mainly homozygous polymorphisms located on the outer edge of each track.

The high density of heterozygosity in the 50% range could represent regions with equal contributions from homologous chromosomes from different parasite genomes characteristically observed in hybrid parasites. The level of heterozygosity in the Sri Lanka isolates was therefore compared to a group of isolates originating from Ethiopia that were previously reported to be intraspecies hybrids of two distinct *L. donovani* populations and served as a benchmark for hybrid parasites (3, 22, 23). As shown in **Figure 4.2B**, all five isolates from the Sri Lanka SL2 group (red) have a higher ratio of heterozygous SNPs than the hybrid group from Ethiopia (blue). All isolates from the SL1 and SL3 group fall within the normal distribution of heterozygous polymorphisms for *L. donovani.*

4.4.2 Determination of *Leishmania* species contributions to hybrid genomes

Highly heterozygous SNPs (**Figure 4.2**) were of interest because they could be derived from non-*L. donovani* species through hybridization and introgression. To investigate this possibility, nucleotides corresponding to the site of each SNP were altered to correspond to the non-*L. donovani* reference nucleotide to reconstruct the genes contributed by a potential non-*L. donovani* parent as outlined in **Figure 4.3A** and compared using BLAST. This methodology was validated using a previously reported hybrid parasite between *L. major* and *L. infantum* (IMT211 (5), **Figure 4.3B**) confirming that this method of analysis can quantify the sequence contributions from the non-*L. donovani* parent at the whole genome level.

Using this analysis, the four reconstructed genomes from the SL1 group matched almost exclusively to members of the *L. donovani* species complex as expected (17). Analysis of the five reconstructed genomes from the highly heterogenous SL2 group further divided this group into two SL2 A and three SL2 B isolates. This analysis identified SNPs of *L. major* origin were present in 2160 and 3738 of the 9757 SL2A genes (**Figure 4.4A**, & **Supplementary Table 4.S3**). The reconstructed genes from the three SL2B samples matched mostly (7671, 4568, 5977 of 9757) the *L. tropica* reference genome (**Figure 4.4A**, & **Supplementary Table 4.S3**). In comparison, the SL3 group that clustered closer to the African strains (**Figure 4.1B**), contained very few reconstructed gene matches (<325) outside of the *L. donovani* complex (*L. donovani/L. infantum*) (**Figure 4.4A**, & **Supplementary Table 4.S3**). For clarity, only the *Leishmania* species with gene matches are shown from the complete Old World species gene comparison (**Supplementary Figure 4.S2**).

We attempted to further narrow the origins of the hybrid genes present in the isolates from Sri Lanka. The SL1 group (4 isolates) matched mostly with the reference sequences from Sri Lanka and Nepal, while the SL2 A group was more related to *L. major* strain SD75 from Senegal than to *L. major* strain LV39 from Uzbekistan or the reference *L. major* Friedlin strain from Israel (**Figure 4.4B**, & **Supplementary Table 4.S4**). As there is only one reference *L. tropica* strain, the three SL2 B isolates all clustered to this one reference. With respect to the SL3 group, the reconstructed genes were almost exclusively matched to the Ethiopian LdLV9 reference strain (24). To confirm that it was possible to specifically assign the reconstructed genes to either *L. tropica* or *L. major* due to their relatedness, some of the alignments were manually inspected. As shown in the representative alignment in **Figure 4.4C**, the polymorphisms between the *L. tropica* and *L. major* hybrids were frequent enough to be discriminatory.

To verify the possibility that SL2 groups (5 isolates) contained genetic material originating from *L. major* and *L. tropica*, we performed an additional genetic comparison and phylogeny analysis including reference sequences from *L. major* and *L. tropica.* In this tree, all five SL2 isolates are placed at various distances along the same branch as *L. major* and *L. tropica* and this branch originates from the African *L. donovani* lineages (**Supplementary Figure 4.S3**) and the majority of heterozygous SNPs in the SL2 isolates are identical to *L. major* or *L. tropica* alleles (**Supplementary Table 4.S2**).

We further attempted to confirm the validity of the BLAST analysis used in **Figures 4.3 and 4.4** by separating the *L. major* or *L. tropica* haplotypes from *L. donovani* through read-based phasing. While phasing was not accurate across long regions (**Supplementary Figure 4.S4**), analysis of a short segment on chromosome 1 was consistent with the BLAST comparison analysis (**Supplementary Figure 4.S4B, Figure 4.4B**). As expected, the two haplotypes originating from each of the SL2 A and SL2 B isolates clustered on opposite branches between *L. donovani/L. major* and *L. donovani/L. tropica* respectively, demonstrating the read-base phasing analysis agreed with the BLAST analysis.

4.4.3 Chromosomal recombination in *L. donovani* hybrid strains

We next investigated whether there was evidence for recombination and introgression between *L. donovani* and *L. major* homologous chromosomes. As shown in a representative alignment of the two SL2 A *L. donovani/L. major* hybrids for the same section of chromosome 19, blocks of nucleotides consisting of homozygous *L. donovani* sequences and heterozygous *L. donovani/L. major* sequences were evident (**Figure 4.5A**). This pattern of mixed or single parental origin sequences can be seen for the five SL2 isolates throughout all the chromosomes as shown

in the example of chromosome 36 (**Figure 4.5B**) and the whole genome (**Supplementary Figure 4.S5**).

4.4.4 Genetic analysis of *L. donovani* isolates with low heterozygosity (SL3 group)

While the five SL2 isolates were highly heterozygous with a large fraction of SNPs matching the *L. tropica* or *L. major* genomes, the three SL3 isolates contained a low level of heterozygosity (**Figure 4.2B**). Upon close inspection however, these three SL3 isolates also contained some short regions with polymorphisms in common with *L. major.* In contrast, the sequences from the SL1 isolates contained no detectable polymorphisms in common with *L. major.* Although most of the polymorphisms in the SL3 group were no longer present around 50% allele frequency (diploid heterozygous), some of the polymorphisms were in common with the SL2 group as shown in the representative section of chromosome 6 (**Supplementary Figure 4.S6A**). Further, some of the SNPs are retained without a loss of allele frequency to match the *L. major* sequence. Genes with *L. major* non-synonymous polymorphisms retained in all samples of the SL3 group are listed in **Supplementary Table 4.S5**.

We further investigated whether the SL2 and SL3 groups shared other features including conserved aneuploidy across the hybrid subgroups. Notably, the two SL2 A and three SL3 (*L. major/L. donovani*) isolates have a conserved aneuploidy pattern consisting of a decreased copy number of chromosome 2 (red arrow) and an increased copy number of chromosomes 22 and 26 (blue arrows, **Supplementary Figure 4.S6B**). In comparison, the isolate from the SL2 B group (*L. tropica/L. donovani*) does not share this aneuploidy pattern such as for example no decreased copy number of chromosome 2. For clarity, only one of the three SL2 B isolates is plotted in **Supplementary Figure 4.S6B**, but all three *L. tropica/L. donovani* SL2 B isolates are consistent with each other (**Supplementary Figure 4.S6C**). Overall, these observations show that there are

similarities between the SL2 A and SL3 groups (*L. major/L. donovani* introgression groups) with respect to conservation of polymorphisms and pattern of aneuploidy.

4.5 Discussion

This study provides evidence that some atypical *L. donovani* strains in Sri Lanka contain genome polymorphisms homologous with *L. major* and *L. tropica* genomes that likely arose from hybridization followed by recombination and introgression. As *L. major* and *L. tropica* parasites are not present in Sri Lanka, and considering the phylogenetic trees shown in **Figure 4.1B** and **Supplementary Figure 4.S3**, it is likely that the hybrid *L. donovani* strains originated in East Africa and were subsequently imported into Sri Lanka where there appears to be a natural selection for these atypical *L. donovani* parasites. The identification of the *L. donovani* hybrid strains described here advances the hypothesis that gene polymorphisms derived from *L. major* and *L. tropica* could have contributed to the spread of CL in Sri Lanka.

Introgression following hybridization could occur for example through backcrossing with a parental species or through the proposed model of mosaic aneuploidy where *Leishmania* can discard deleterious or retain beneficial alleles (25, 26). Through this process, progeny with a combination of *L. major* or *L. tropica* alleles with *L. donovani* alleles can arise in nature if they acquire a fitness advantage for propagation in a particular environment such as there appears to be in Sri Lanka.

To address the question why diverse atypical *L. donovani* strains have culminated in Sri Lanka, it will be necessary to investigate the vector and potential non-human reservoirs that selects for such atypical *L. donovani* strains. Notably, the probable *L. donovani* vector in Sri Lanka is *Phlebotomus argentipes* subspecies *glaucus* which has a preference for animal rather than human

blood and differs from *P. argentipes* subspecies *sensu lato,* the anthropophilic vector for *L. donovani* in India (27). This Sri Lankan vector is also different from the *Phlebotomus orientalis* and *Phlebotomus alexandri* vectors of *L. donovani* in Africa (22). As inter-species hybridization has previously been shown to confer increased vector competence to hybrid parasites (28), the hybrids described herein could benefit from a wider permissive vector repertoire. Considering that virtually all human CL causing species outside of Sri Lanka have an animal reservoir, it will also be necessary to consider the possibility that a non-human reservoir for atypical *L. donovani* strains may be present in Sri Lanka.

Three parasites with relatively small amounts of hybrid gene polymorphisms in the SL3 group were also identified suggesting that introgression was more ancient in the SL3 isolates than the five SL2 isolates resulting in fewer remaining polymorphic alleles. This is essentially a natural selection experiment where heterozygosity is reduced by the removal of SNPs that are not beneficial and the retention of SNPs that result in more fit parasites (**Supplementary Table 4.S5**). Some of these genes could have been selectively retained during propagation in Sri Lanka and functional analysis of these could identify their role in disease tropism. While the patterns of polymorphisms appear varied across the isolates, the aneuploidy observed in the five SL2 and three SL3 parasites appears to be more conserved (**Supplementary Figure 4.S6**) suggesting that an external pressure is driving those parasites away from a normal diploid genome.

A limitation of the study could result from the relaxed SNP scoring parameters used to preclude removing true SNPs that could have contributed to retaining some false positive SNPs. We have confronted this potential limitation through the judicious choice of relevant controls and statistical analysis to distinguish signal over noise in the data in downstream analyses. Further, as the exact parental strains are unknown and limited reference genomes are available to use as

proxies for BLAST analysis, this will potentially also result in some gene assignment errors (**Supplementary Tables 4.S3, 4.S4**). As our study relies on public data from the SRA, it is susceptible to bias present in the collected data. As such, under-sampled regions could result in missing phylogenetic branches and the small number of whole genome sequences from Sri Lanka (some with debatable pathology as described further below) precludes a definitive association between hybridization and cutaneous disease.

The evidence presented herein for the existence of 4 different populations of *L. donovani* parasites in Sri Lanka (SL1, SL2 A, SL2 B and SL3) could help reconcile differences in lesion morphology (29, 30), spatial distribution (13) and drug susceptibility (19) previously reported across Sri Lanka. Indeed, parasites causing lesions that share features of *L major* or *L. tropica* infection (29) or with variable tolerance to sodium stibogluconate (19) could result from different *L. donovani* hybrids. It would be interesting to compare the pathology caused by SL2 vs SL3 group parasites or the *L. major* vs *L. tropica* (SL2 A vs SL2 B) hybrids. It is also noteworthy that there have been fewer than 7 VL cases in Sri Lanka since 2004 and the majority of these have had co-morbidities (14). For example, the two hybrid isolates labeled as causing VL (SL2 B samples SRR6257366 and SRR6257367) were derived from patients with leukemia and diabetes respectively (19). As there is no active transmission of VL in Sri Lanka, these isolates may have caused this pathology due to the status of their hosts while contributing to the 15000 reported cases of CL in Sri Lanka (12, 13) when transmitted to healthy individuals. Nevertheless, this raises the possibility that some of the hybrid strains could have the potential to cause VL under the right conditions.

Both *L. tropica* and *L. major* are present in Ethiopia in similar animal populations along with their sand fly vectors (31). It has also been suggested that *L. tropica* and *L. donovani* share a

similar animal reservoir in Ethiopia (32). Taken together, it is feasible that *L. donovani* parasites generated hybrid parasites with *L. tropica* and *L. major* in Ethiopia and subsequently entered Sri Lanka as suggested by the SL2 and SL3 grouping in the African cluster (**Supplementary Figure 4.S3**).

Sri Lankan military personnel are routinely deployed in *Leishmania*-endemic countries including Sudan as part of the United Nations' peace-keeping missions, and returning soldiers are routinely screened for malaria upon returning and could also represent a point of entry for *Leishmania* (33). The genetic evidence presented here argues that there have been multiple sources of *L. donovani* entry into Sri Lanka, yet remarkably there is no ongoing VL transmission. Future studies must determine why atypical *L. donovani* parasites have thrived in Sri Lanka instead of VL-causing *L. donovani* which are vastly more widespread in neighboring India and Africa. The ongoing VL elimination program in India must now consider atypical *L. donovani* strains to ensure CL does not become a major public health problem in the Indian subcontinent.

4.6 Methods

4.6.1 Study design and data collection

To understand the origins of the *L. donovani* strains circulating in Sri Lanka, we compared their genome sequences to all available *L. donovani* genome data in the NCBI GenBank/Sequencing Read Archive (SRA). *L. donovani* genome sequences were selected to include high accuracy data as outlined in **Figure 4.1A**. Because of the resulting large sample number (684 whole genome sequences) from regions of Asia and Africa, any potential bias in samples analyzed was not deemed sufficient to affect the conclusions of this study (**Supplementary Method 4.S1**). Ethical approval was not required by institutional review boards for the analysis of publicly available genome sequences in GenBank.

4.6.2 Alignment of all sequenced *L. donovani* isolates to the reference genome

In order to compare all of the genomes and to generate a unified phylogenetic tree, samples were aligned to the same reference sequence and variants identified per sample. Raw data was obtained using the SRA-Toolkit (version 2.9.6), aligned with the Burrows-Wheeler Aligner (version 0.7.17) and analysed using samtools (version 1.10) and VarScan2 (version 2.4.3) to identify SNPs (**Supplementary Method 4.S2**).

4.6.3 *L. donovani* global strains phylogeny

In order to analyse the genetic relationships of the Sri Lankan *L. donovani* samples, a phylogenetic analysis using the global population of *L. donovani* was performed. We combined the SNPs identified above using BCFtools (version 1.10.2), generated a neighbor-joining tree with TASSEL (version 5.0) with no preset cluster number and size for the SL isolates nor the global isolates and IToL (version 5) was used to visualize the tree (**Supplementary Method 4.S3**).

4.6.4 Heterozygosity of *L. donovani* isolates worldwide

As a high density of heterozygous polymorphism is a potential indication of hybridization, we compared the levels of heterozygosity in the Sri Lanka clusters to known *Leishmania* hybrids and the rest of the population (**Supplementary Method 4.S4**) using the alternate SNP frequencies as reported by VarScan2 (version 2.4.3).

4.6.5 Identification of species in the hybrid parasites

To identify the parental species contributing to the hybrid parasites, the reference genome was modified using samtools (version 1.10) and BCFtools (version 1.10.2) to match the polymorphisms and the recreated sequences were analysed by Basic Local Alignment Search Tool

(BLAST v2.7.1) against all Old World *Leishmania spp.* genomes (**Supplementary Method 4.S5**, **Figure 4.3**)

4.6.6 Haplotype phasing

Haplotype phasing was performed to verify the results obtained by BLAST analysis by separating the two putative haplotypes present in samples with high heterozygosity. The samples were phased using samtools (version 1.10) prior to reconstruction analysis as described above (appendix pp 28-29) and phylogenetic analysis using Clustal Omega version 1.2.4 (**Supplementary Method 4.S6**).

4.6.7 Statistical Analysis

Significant variance in the quantitative BLAST analysis to detect species match was determined using 2-way ANOVA in Prism (version 6.01) and p-values reported on **Supplementary Table 4.S3**, **Supplementary Table 4.S4**. Isolates were grouped in their respective clusters (SL2 A, SL2 B, SL3) and the number of species match in each row compared to the matches of the control SL1 cluster.

4.6.8. Role of the funding source

This research was supported by a grant from the Canadian Institutes of Health Research (CIHR) to GM and a doctoral training award from the Fond de Recherche du Quebec en Santé (FRQS) to PL.

The funders of the study played no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and GM had final responsibility for the decision to submit for publication.

4.7 References

1. Burza S, Croft SL, Boelaert M. 2018. Leishmaniasis. Lancet 392:951–970.
2. Scorza BM, Carvalho EM, Wilson ME. 2017. Cutaneous manifestations of human and murine leishmaniasis. Int J Mol Sci 18.
3. Gelanew T, Hailu A, Scho'nian G, Lewis MD, Miles MA, Yeo M. 2014. Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural ethiopian isolates of leishmania donovani. Int J Parasitol 44:751–757.
4. Chargui N, Amro A, Haouas N, Schönian G, Babba H, Schmidt S, Ravel C, Lefebvre M, Bastien P, Chaker E, Aoun K, Zribi M, Kuhls K. 2009. Population structure of Tunisian Leishmania infantum and evidence for the existence of hybrids and gene flow between genetically different populations. Int J Parasitol 39:801–811.
5. Ravel C, Cortes S, Pratlong F, Morio F, Dedet JP, Campino L. 2006. First report of genetic hybrids between two very divergent Leishmania species: Leishmania infantum and Leishmania major. Int J Parasitol 36:1383–1388.
6. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M, Svobodova M, Volf P, Berriman M, Cotton JA, Smith DF. 2014. Genomic Confirmation of Hybridisation and Recent Inbreeding in a Vector-Isolated Leishmania Population. PLoS Genet 10.
7. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, Dobson DE, Beverley SM, Sacks DL. 2009. Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. Science (80-) 324:265–268.
8. Sadlova J, Yeo M, Seblova V, Lewis MD, Mauricio I, Volf P, Miles MA. 2011. Visualisation of leishmania donovani fluorescent hybrids during early stage development in the sand fly vector. PLoS One 6.
9. Inbar E, Akopyants NS, Charmoy M, Romano A, Lawyer P, Elnaiem DEA, Kauffmann F, Barhoumi M, Grigg M, Owens K, Fay M, Dobson DE, Shaik J, Beverley SM, Sacks D. 2013. The Mating Competence of Geographically Diverse Leishmania major Strains in Their Natural and Unnatural Sand Fly Vectors. PLoS Genet 9.
10. Romano A, Inbar E, Debrabant A, Charmoy M, Lawyer P, Ribeiro-Gomes F, Barhoumi M, Grigg M, Shaik J, Dobson D, Beverley SM, Sacks DL. 2014. Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. Proc Natl Acad Sci U S A 111:16808–16813.
11. Inbar E, Shaik J, Iantorno SA, Romano A, Nzelu CO, Owens K, Sanders MJ, Dobson D, Cotton JA, Grigg ME, Beverley SM, Sacks D. 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in leishmania. PLoS Genet 15.
12. WHO. 2018. Status of endemicity of leishmaniasis worldwide, 2018World Health Organisation. Geneva.
13. Karunaweera ND, Ginige S, Senanayake S, Silva H, Manamperi N, Samaranayake N,

Siriwardana Y, Gamage D, Senerath U, Zhou G. 2020. Spatial epidemiologic trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerg Infect Dis 26:1–10.

14. Siriwardana HVYD, Karunanayake P, Goonerathne L, Karunaweera ND. 2017. Emergence of visceral leishmaniasis in Sri Lanka: a newly established health threat. Pathog Glob Health 111:317–326.

15. WHO. 2013. Status of endemicity of visceral and cutaneus Leishmaniasis, worldwide, 2013. Geneva.

16. Thakur L, Singh KK, Shanker V, Negi A, Jain A, Matlashewski G, Jain M. 2018. Atypical leishmaniasis: A global perspective with emphasis on the Indian subcontinent. PLoS Negl Trop Dis 12.

17. Lypaczewski P, Hoshizaki J, Zhang WW, McCall LI, Torcivia-Rodriguez J, Simonyan V, Kaur A, Dewar K, Matlashewski G. 2018. A complete Leishmania donovani reference genome identifies novel genetic variations associated with virulence. Sci Rep 8.

18. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.

19. Samarasinghe SR, Samaranayake N, Kariyawasam UL, Siriwardana YD, Imamura H, Karunaweera ND. 2018. Genomic insights into virulence mechanisms of Leishmania donovani: Evidence from an atypical strain. BMC Genomics 19.

20. Imamura H, Downing T, van den Broeck F, Sanders MJ, Rijal S, Sundar S, Mannaert A, Vanaerschot M, Berg M, de Muylder G, Dumetz F, Cuypers B, Maes I, Domagalska M, Decuypere S, Rai K, Uranw S, Bhattarai NR, Khanal B, Prajapati VK, Sharma S, Stark O, Schönian G, de Koning HP, Settimo L, Vanhollebeke B, Roy S, Ostyn B, Boelaert M, Maes L, Berriman M, Dujardin JC, Cotton JA. 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5.

21. Cuypers B, Berg M, Imamura H, Dumetz F, De Muylder G, Domagalska MA, Rijal S, Bhattarai NR, Maes I, Sanders M, Cotton JA, Meysman P, Laukens K, Dujardin JC. 2018. Integrated genomic and metabolomic profiling of ISC1, an emerging Leishmania donovani population in the Indian subcontinent. Infect Genet Evol 62:170–178.

22. Gelanew T, Kuhls K, Hurissa Z, Weldegebreal T, Hailu W, Kassahun A, Abebe T, Hailu A, Schönian G. 2010. Inference of population structure of leishmania donovani strains isolated from different ethiopian visceral leishmaniasis endemic areas. PLoS Negl Trop Dis 4.

23. Cotton JA, Durrant C, Franssen SU, Gelanew T, Hailu A, Mateus D, Sanders MJ, Berriman M, Volf P, Miles MA, Yeo M. 2020. Genomic analysis of natural intra-specific hybrids among ethiopian isolates of leishmania donovani. PLoS Negl Trop Dis 14:1–26.

24. Camacho E, González-de la Fuente S, Rastrojo A, Peiró-Pastor R, Solana JC, Tabera L, Gamarro F, Carrasco-Ramiro F, Requena JM, Aguado B. 2019. Complete assembly of the Leishmania donovani (HU3 strain) genome and transcriptome annotation. Sci Rep 9.

25. Sterkers Y, Crobu L, Lachaud L, Pagès M, Bastien P. 2014. Parasexuality and mosaic aneuploidy in Leishmania: Alternative genetics. Trends Parasitol 30:429–435.

26. Barja PP, Pescher P, Bussotti G, Dumetz F, Imamura H, Kedra D, Domagalska M, Chaumeau V, Himmelbauer H, Pages M, Sterkers Y, Dujardin JC, Notredame C, Späth GF. 2017. Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. Nat Ecol Evol 1:1961–1969.

27. Senanayake SASC, Abeyewicreme W, Dotson EM, Karunaweera ND. 2015. Characteristics of phlebotomine sand flies in selected areas of Sri Lanka. Southeast Asian J Trop Med Public Health 46:994–1004.

28. Volf P, Benkova I, Myskova J, Sadlova J, Campino L, Ravel C. 2007. Increased transmission potential of Leishmania major/Leishmania infantum hybrids. Int J Parasitol 37:589–593.

29. Siriwardana Y, Deepachandi B, Weliange SDS, Udagedara C, Wickremarathne C, Warnasuriya W, Ranawaka RR, Kahawita I, Chandrawansa PH, Karunaweera ND. 2019. First Evidence for Two Independent and Different Leishmaniasis Transmission Foci in Sri Lanka: Recent Introduction or Long-Term Existence? J Trop Med 2019.

30. Siriwardana Y, Zhou G, Deepachandi B, Akarawita J, Wickremarathne C, Warnasuriya W, Udagedara C, Ranawaka RR, Kahawita I, Ariyawansa D, Sirimanna G, Chandrawansa PH, Karunaweera ND. 2019. Trends in Recently Emerged Leishmania donovani Induced Cutaneous Leishmaniasis, Sri Lanka, for the First 13 Years. Biomed Res Int 2019.

31. Kassahun A, Sadlova J, Benda P, Kostalova T, Warburg A, Hailu A, Baneth G, Volf P, Votypka J. 2015. Natural infection of bats with Leishmania in Ethiopia. Acta Trop 150:166–170.

32. Kassahun A, Sadlova J, Dvorak V, Kostalova T, Rohousova I, Frynta D, Aghova T, Yasur-Landau D, Lemma W, Hailu A, Baneth G, Warburg A, Volf P, Votypka J. 2015. Detection of leishmania donovani and L. tropica in ethiopian wild rodents. Acta Trop 145:39–44.

33. Fernando SD, Dharmawardana P, Semege S, Epasinghe G, Senanayake N, Rodrigo C, Premaratne R. 2016. The risk of imported malaria in security forces personnel returning from overseas missions in the context of prevention of re-introduction of malaria to Sri Lanka. Malar J 15.

34. Leinonen R, Sugawara H, Shumway M. 2011. The sequence read archive. Nucleic Acids Res 39.

35. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM arXiv:1303.

36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

37. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration

discovery in cancer by exome sequencing. Genome Res 22:568–576.

38. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.

39. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res 47.

40. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H. 2009. TriTrypDB: A functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38:D457-62.

41. Danecek P, McCarthy SA. 2017. BCFtools/csq: Haplotype-aware variant consequences. Bioinformatics 33:2037–2039.

42. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645.

43. Madhusoodanan N. 2019. Clustal Omega < Multiple Sequence Alignment < EMBL-EBI. Multisequence Alignment.

4.8 Additional Information

4.8.1 Data availability

Data files: The phylogenetic tree generated in this study is available in an interactive format provided by iToL at: https://itol.embl.de/tree/132216267336879158013475 5

Code availability: All software and methodologies used are described within Methods.

4.8.2 Acknowledgements

We thank Wen Wei Zhang, Kayla Paulini, and Jesse Shapiro for their insightful feedback and suggestions. All DNA sample collection and sequencing for records listed in **Supplementary**

Table 4.S1 were performed by their respective research groups. For complete attribution of each record refer to the corresponding record SRA accession page.

This research was enabled in part by support provided by Calcul Quebec (Beluga cluster, https://www.calculquebec.ca/en/), SciNet (Niagara cluster, https://www.scinethpc.ca/niagara/), WestGrid (Cedar Cluster, https://www.westgrid.ca/) and Compute Canada (www.computecanada.ca).

4.8.3 Author Contributions

PL designed the study, collected and analysed the data and wrote the manuscript. GM helped design the study, wrote and edited the manuscript. Both authors had access and verified the underlying data and have read and approved the final version of the manuscript.

4.8.4 Accession Numbers

All data used in this study were obtained from publicly available sources, all accession numbers used are listed in **Supplementary Table 4.S1**.

4.8.5 Declaration of interests

The authors declare no financial or personal competing interest.

4.9 Figures and Tables

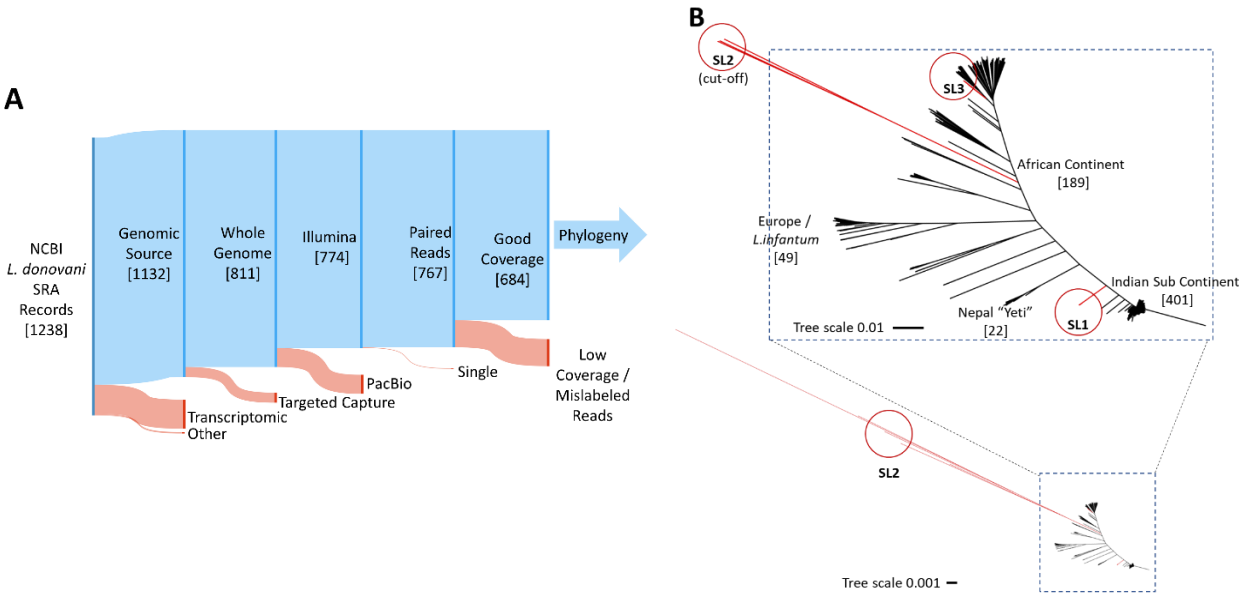


Figure 4.1 Worldwide *L. donovani* phylogenetic tree

A. Selection filters applied to the NCBI 1238 public *L. donovani* Sequence Read Archive (SRA) records to obtain good quality *L. donovani* sequencing data for phylogenetic analysis. Sequencing records were retained if they originated from genomic DNA, had no selection bias for genomic location, were sequenced on high accuracy Illumina platforms in paired sequencing mode and had coverage spanning the entire genome. **B.** Neighbor Joining based tree of all *L. donovani* samples analyzed showing clear geographical groupings. Sri Lankan isolates labeled as **SL1** (Sri Lanka Group 1), **SL2** (Sri Lankan Sri Lanka Group 2), **SL3** (Sri Lanka Group 3)**.**

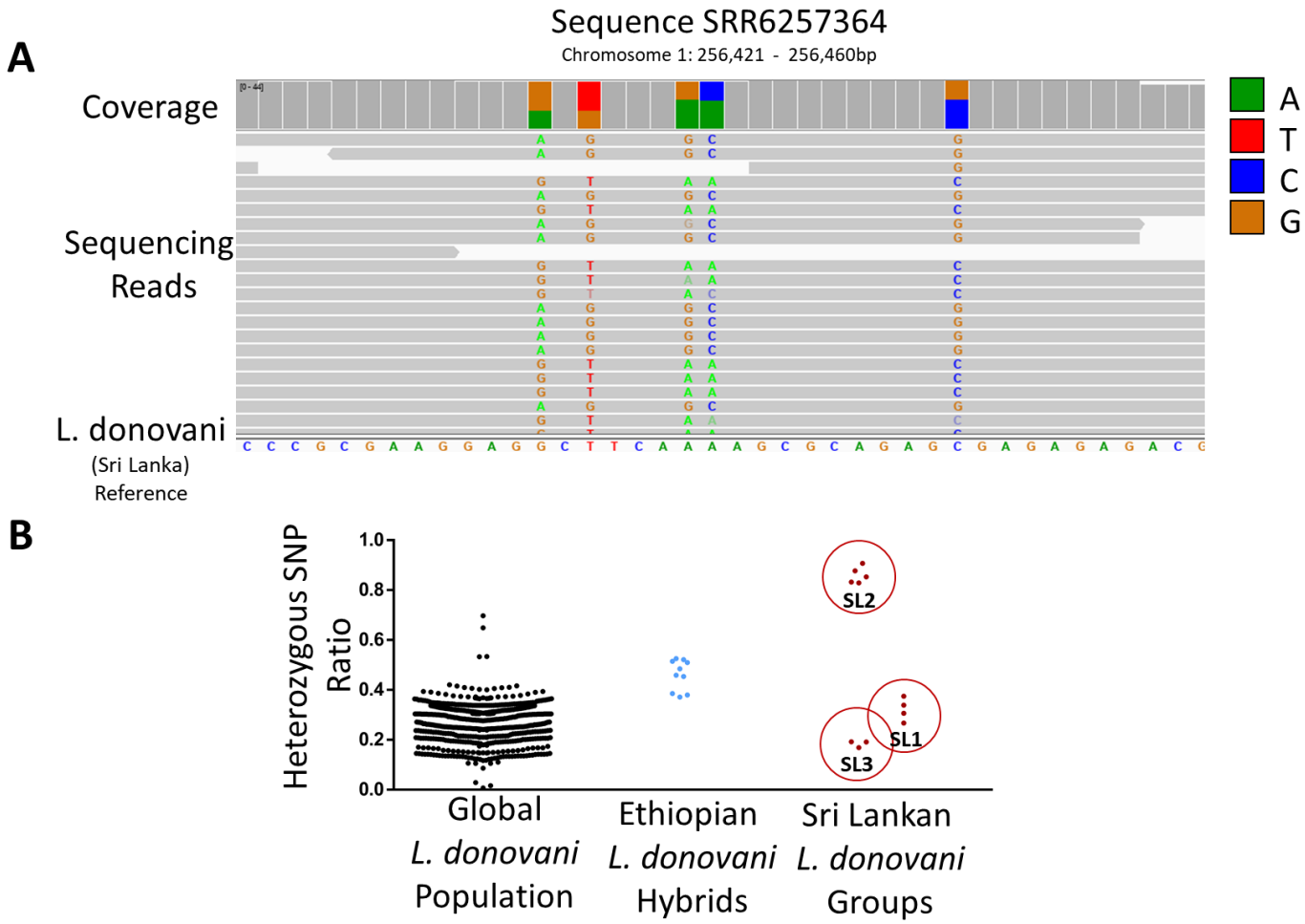


Figure 4.2 Distribution of heterozygous polymorphisms across all samples

A. Representative alignment of SL2 group samples on a 40bp region of *L. donovani* chromosome 1 showing frequent heterozygosity. Reads matching the reference sequence (17) (bottom) with no SNPs are displayed in gray. Loci where reads contain variability are highlighted in colors corresponding to the respective nucleotides. **B.** Comparison of the heterozygous SNP frequency in all *L. donovani* world wide isolates, known Ethiopian hybrid parasites (3, 22) to the Sri Lankan isolates. Each dot represents a single SRA record. Isolates from Ethiopia previously identified as hybrid parasites (3, 22) and their respective frequency shown in blue. Sri Lankan isolates from the SL1, SL2 and SL3 groups are shown in red and their group assignment is highlighted.

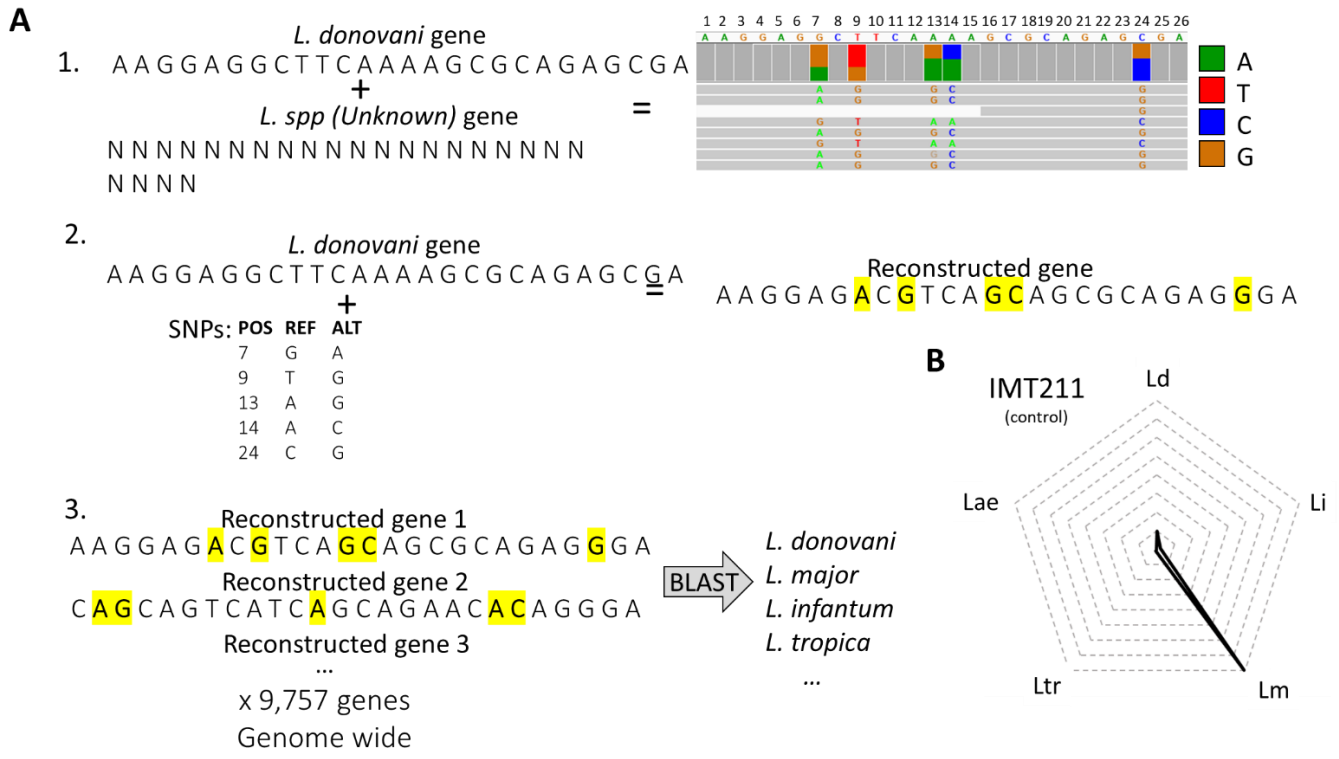
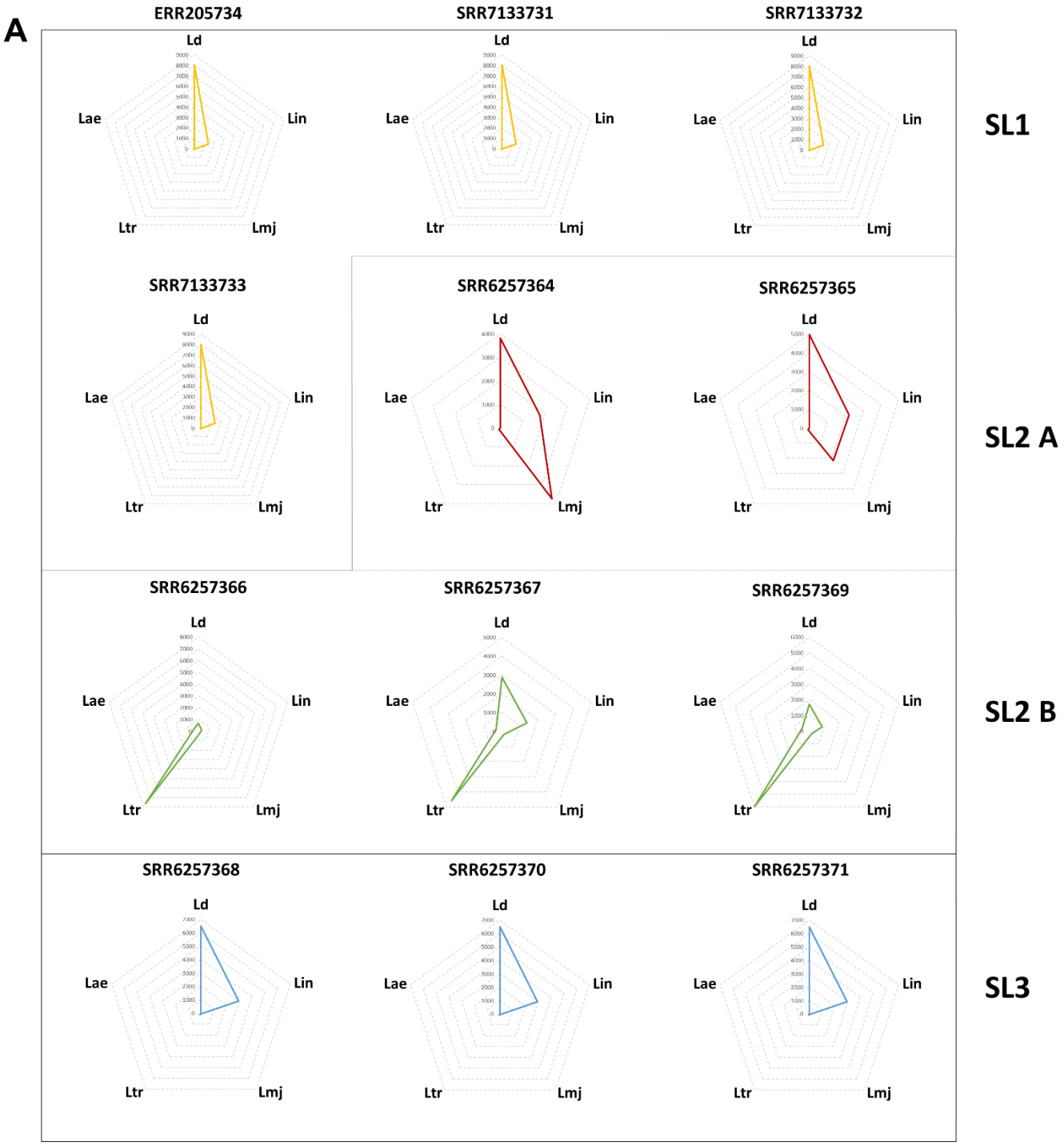


Figure 4.3 Parental strain lineage determination by gene reconstruction

A. Hybrid SNP loci were assumed to have received one allele from *L. donovani* and one allele from an unknown parent resulting in alignments with polymorphisms occurring at +/-50%. Gene sequences from the Sri Lanka reference *L. donovani* (17) (REF) were transformed at the position (POS) of each SNP (ALT) across the entire genome to reconstruct the gene sequences of the unknown parent *Leishmania* species. All reconstructed gene sequences were then compared to a *Leishmania* database containing all Old World *Leishmania* reference genomes by BLAST searches and assigned an originating species and strain. **B.** Control analysis using a known hybrid (IMT211) with *L. major* and *L. infantum* (5) showing reconstructed genes mostly matching *L. major*. Each level in the radar plot corresponds to 1,000 gene matches in the corresponding species (dotted lines).


A
ERR205734
SRR7133731
SRR7133732
SRR7133733
SRR6257364
SRR6257365
SRR6257366
SRR6257367
SRR6257369
SRR6257368
SRR6257370
SRR6257371
Ld
Lin
Lmj
Ltr
Lae
SL1
SL2 A
SL2 B
SL3

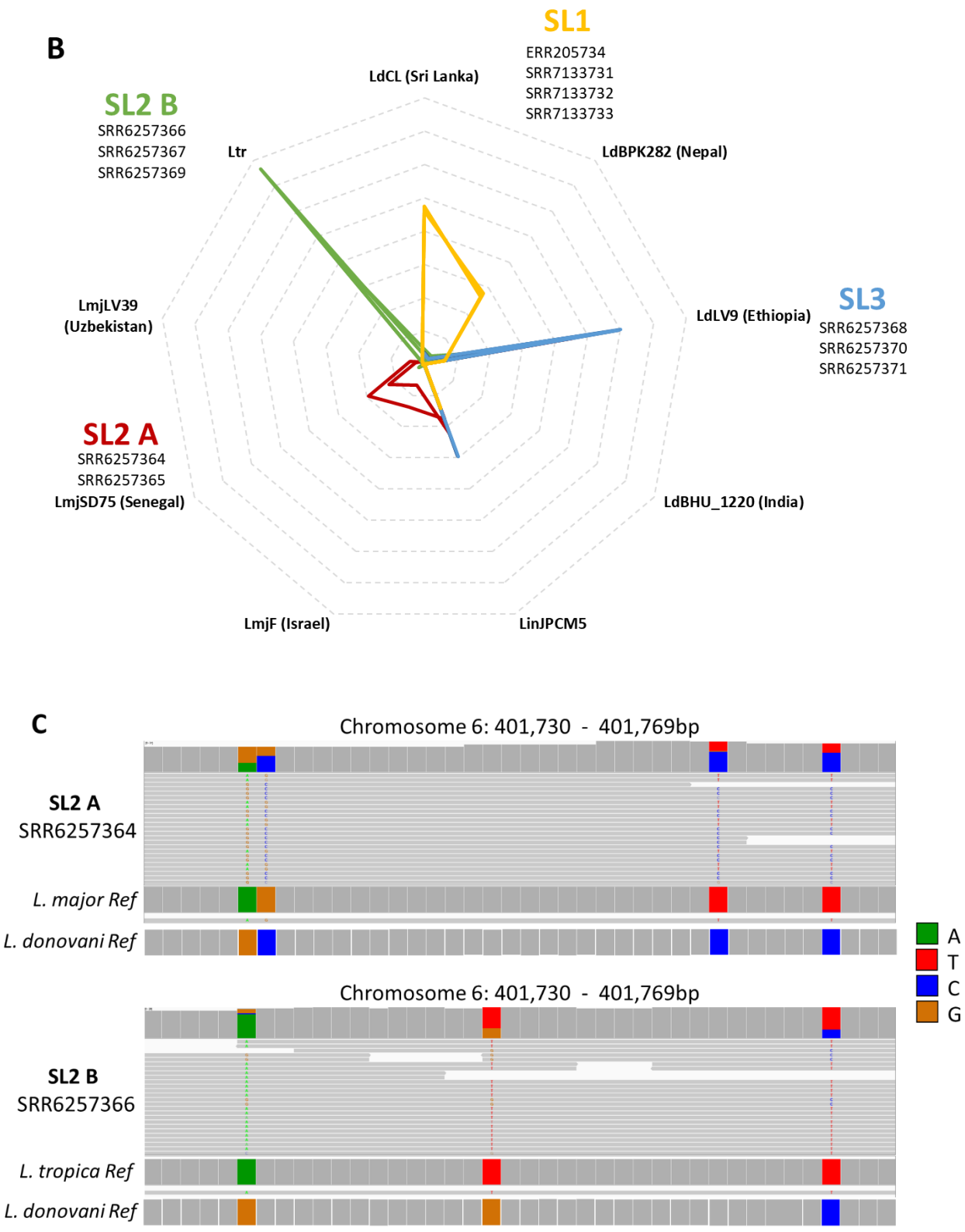

B
SL1
ERR205734
SRR7133731
SRR7133732
SRR7133733
SL2 B
SRR6257366
SRR6257367
SRR6257369
SL3
SRR6257368
SRR6257370
SRR6257371
SL2 A
SRR6257364
SRR6257365
LdCL (Sri Lanka)
LdBPK282 (Nepal)
Ltr
LmjLV39
(Uzbekistan)
LdLV9 (Ethiopia)
LmjSD75 (Senegal)
LdBHU_1220 (India)
LmjF (Israel)
LinJPCM5
C
Chromosome 6: 401,730 - 401,769bp
SL2 A
SRR6257364
L. major Ref
L. donovani Ref
Chromosome 6: 401,730 - 401,769bp
SL2 B
SRR6257366
L. tropica Ref
L. donovani Ref
A
T
C
G

Figure 4.4 Radar plots showing the distribution of the origin of genes at species (A) and strain level (B) and chromosome 6 alignments (C)

A. Distribution of species origin of reconstructed genes as determined by BLAST analysis. Each level in the radar plots corresponds to 1,000 gene matches in the corresponding species (dotted lines). Radar plots are grouped according to isolate groupings with the five SL2 isolates broken down to two SL2A and three SL2B subgroupings. **B.** Distribution of the origin of reconstructed genes to different reference strain genomes as determined by BLAST analysis. Each level in the radar plots corresponds to 1,000 gene matches in the corresponding species (dotted lines). Five SL1 isolates are shown in Orange. The two SL2 A subgroup isolates are shown in red. The three SL2 B subgroup isolates are shown in green. The three SL3 group isolates are shown in blue. **C.** Representative alignment of a *L. major* and *L. tropica* hybrid on the same section of chromosome 6 showing heterozygous polymorphism matching *L. major* and *L. tropica* respectively.

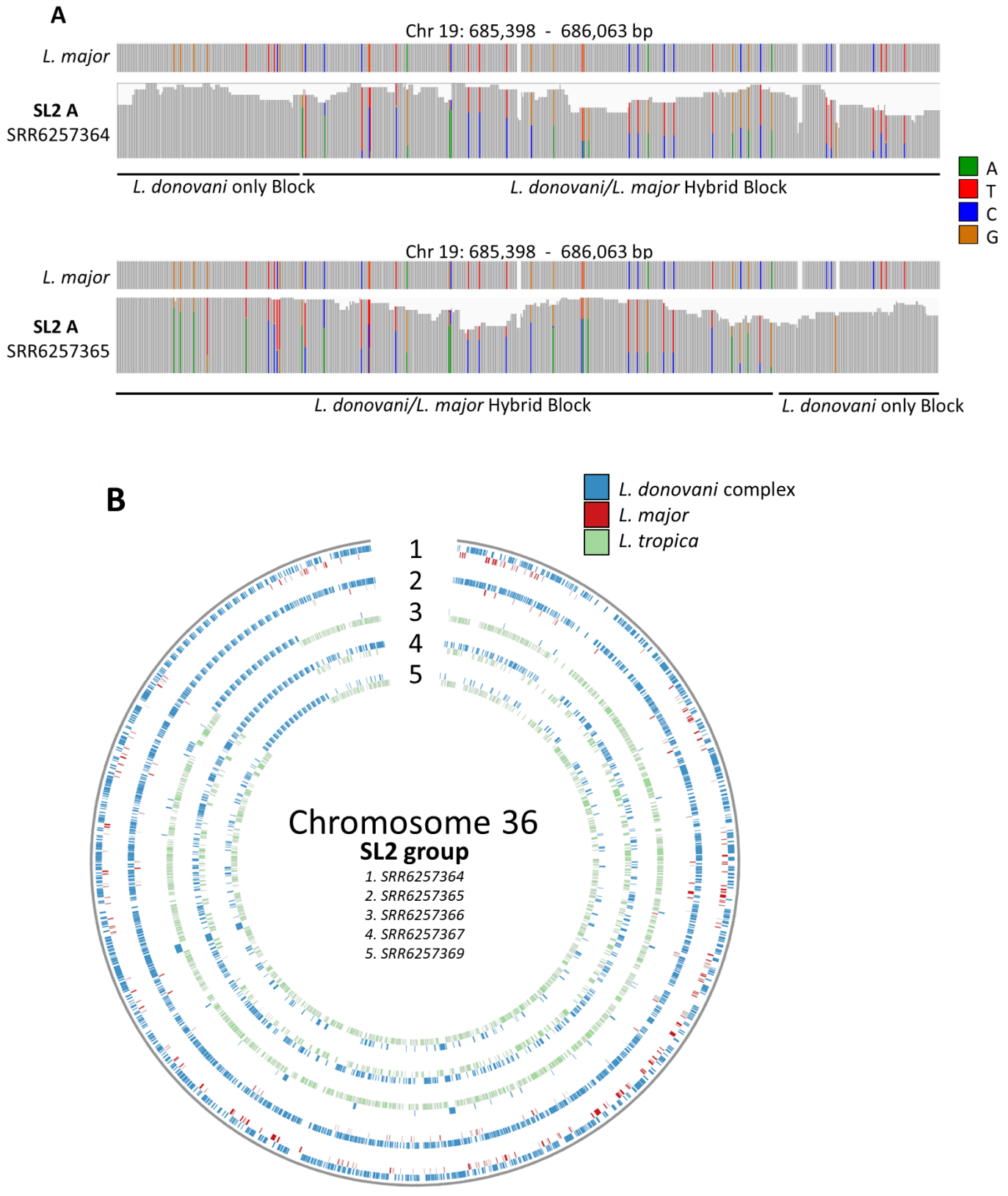

A
Chr 19: 685,398 - 686,063 bp
L. major
SL2 A
SRR6257364
L. donovani only Block
L. donovani/L. major Hybrid Block
A
T
C
G
Chr 19: 685,398 - 686,063 bp
L. major
SL2 A
SRR6257365
L. donovani/L. major Hybrid Block
L. donovani only Block
B
L. donovani complex
L. major
L. tropica
1
2
3
4
5
Chromosome 36
SL2 group
1. SRR6257364
2. SRR6257365
3. SRR6257366
4. SRR6257367
5. SRR6257369

Figure 4.5 Location and origin of genes in SL2 group isolates

A. Representative alignment of genomes from two *L. major* hybrid parasites on the same section of chromosome 19 showing short length blocks of single (*L. donovani* only) or mixed parent ancestry (*L. donovani/ L. major* hybrid). **B.** Representation of chromosome 36 in all isolates in the SL2 group. Each marker represents a single gene. Genes of *L. donovani* species complex origin are marked in blue. Genes with hybrid ancestry (*L. major & L. donovani,* or *L. tropica & L. donovani*) are colored in red and green respectively.

4.10 Supplementary Material

Supplementary Table 4.S1 List of accession numbers used in this study

SRA Accession Numbers
ERR018830 ERR018831 ERR018832 ERR018833 ERR018834 ERR018835 ERR018836
ERR018837 ERR018838 ERR018839 ERR018840 ERR018841 ERR018842 ERR018843
ERR018844 ERR018845 ERR018846 ERR018847 ERR018849 ERR018850 ERR018851
ERR018852 ERR018853 ERR018854 ERR018855 ERR018856 ERR018857 ERR018859
ERR018860 ERR018861 ERR018862 ERR019524 ERR034092 ERR036057 ERR036058
ERR1329883 ERR1329884 ERR205723 ERR205724 ERR205726 ERR205729
ERR205730 ERR205731 ERR205732 ERR205733 ERR205734 ERR205735 ERR205736
ERR205737 ERR205738 ERR205739 ERR205740 ERR205741 ERR205742 ERR205743
ERR205744 ERR205745 ERR205746 ERR205747 ERR205748 ERR205749 ERR205750
ERR205751 ERR205752 ERR205753 ERR205754 ERR205755 ERR205756 ERR205757
ERR205758 ERR205759 ERR205760 ERR205761 ERR205763 ERR205765 ERR205766
ERR205767 ERR205768 ERR205770 ERR205771 ERR205772 ERR205774 ERR205775
ERR205776 ERR205777 ERR205778 ERR205779 ERR205780 ERR205781 ERR205782
ERR205783 ERR205784 ERR205785 ERR205786 ERR205787 ERR205788 ERR205789
ERR205790 ERR205791 ERR205792 ERR205793 ERR205794 ERR205795 ERR205797
ERR205799 ERR205800 ERR205801 ERR205802 ERR205803 ERR205804 ERR205805
ERR205806 ERR205807 ERR205808 ERR205809 ERR205810 ERR205811 ERR205812
ERR205813 ERR205814 ERR205815 ERR205816 ERR205817 ERR205818 ERR205819
ERR205820 ERR206260 ERR206261 ERR206262 ERR206263 ERR206264 ERR206265
ERR206266 ERR206267 ERR206268 ERR206269 ERR206270 ERR206271 ERR206272
ERR206273 ERR206274 ERR206275 ERR206276 ERR206277 ERR206278 ERR206279
ERR206280 ERR206281 ERR206282 ERR206283 ERR206284 ERR206285 ERR206286
ERR206287 ERR206288 ERR206289 ERR206290 ERR206291 ERR206292 ERR206293
ERR206294 ERR206295 ERR206296 ERR206297 ERR206298 ERR206299 ERR206300
ERR206301 ERR206302 ERR206303 ERR206304 ERR206305 ERR206306 ERR206307
ERR206308 ERR206309 ERR206310 ERR206311 ERR206312 ERR206313 ERR206314
ERR206315 ERR206316 ERR206317 ERR206318 ERR206319 ERR206320 ERR206321
ERR206322 ERR206323 ERR206324 ERR206325 ERR206326 ERR206327 ERR206328
ERR206329 ERR206330 ERR206331 ERR206332 ERR206333 ERR206334 ERR206335
ERR206336 ERR206337 ERR206338 ERR206339 ERR206340 ERR206341 ERR206342
ERR206343 ERR206344 ERR206345 ERR206346 ERR206347 ERR206348 ERR206349
ERR206350 ERR206351 ERR206352 ERR206353 ERR206354 ERR206355 ERR206356
ERR206357 ERR206358 ERR206359 ERR206360 ERR206361 ERR206362 ERR206363
ERR206364 ERR206365 ERR206366 ERR206367 ERR206368 ERR206369 ERR206370
ERR206371 ERR206372 ERR206373 ERR206374 ERR206375 ERR206376 ERR206377
ERR206378 ERR206379 ERR206380 ERR206381 ERR206382 ERR206383 ERR206384
ERR206385 ERR206386 ERR206387 ERR206388 ERR206389 ERR206390 ERR206391
ERR206392 ERR206393 ERR206394 ERR206395 ERR206396 ERR206397 ERR206398
ERR206399 ERR206400 ERR206401 ERR206402 ERR206403 ERR206404 ERR206405
ERR206406 ERR206407 ERR206408 ERR206409 ERR206410 ERR206411 ERR206412
ERR206413 ERR206414 ERR206415 ERR206416 ERR206417 ERR206418 ERR206419
ERR206420 ERR206421 ERR206422 ERR206423 ERR206424 ERR206425 ERR206426
ERR206427 ERR206428 ERR206429 ERR206430 ERR206431 ERR206432 ERR206433
ERR206434 ERR206435 ERR206436 ERR206437 ERR206438 ERR206439 ERR206440
ERR206441 ERR206442 ERR206443 ERR206444 ERR206445 ERR206446 ERR206447

ERR206448 ERR206449 ERR206450 ERR206451 ERR206452 ERR206453 ERR206454
ERR206455 ERR206456 ERR206457 ERR206458 ERR206459 ERR206460 ERR206461
ERR206462 ERR206463 ERR206464 ERR206465 ERR206466 ERR206467 ERR206468
ERR206469 ERR206470 ERR206471 ERR206472 ERR206473 ERR206474 ERR206475
ERR206476 ERR206477 ERR206478 ERR206479 ERR206480 ERR206481 ERR206482
ERR206483 ERR206484 ERR206485 ERR206486 ERR206487 ERR206488 ERR206489
ERR206490 ERR206491 ERR206492 ERR206493 ERR206494 ERR206495 ERR206496
ERR206497 ERR206498 ERR206499 ERR206500 ERR206501 ERR206502 ERR206503
ERR206504 ERR206505 ERR206506 ERR206507 ERR206508 ERR206509 ERR206510
ERR206511 ERR206512 ERR206513 ERR206514 ERR206515 ERR206516 ERR206517
ERR206518 ERR206519 ERR206520 ERR206521 ERR206522 ERR206523 ERR206524
ERR206525 ERR206526 ERR206527 ERR206528 ERR206529 ERR206530 ERR206531
ERR206532 ERR206533 ERR206534 ERR206535 ERR206536 ERR206537 ERR206538
ERR206539 ERR206540 ERR206541 ERR206542 ERR206543 ERR206544 ERR206545
ERR206546 ERR206547 ERR206548 ERR206549 ERR206550 ERR206551 ERR206552
ERR206553 ERR206554 ERR206555 ERR206556 ERR206557 ERR206558 ERR206559
ERR206560 ERR206561 ERR206562 ERR206563 ERR206564 ERR206565 ERR206566
ERR206567 ERR206568 ERR206569 ERR206570 ERR206571 ERR206573 ERR206574
ERR206575 ERR206576 ERR206577 ERR206578 ERR206579 ERR206580 ERR206581
ERR206582 ERR206583 ERR206584 ERR206585 ERR206586 ERR206587 ERR206588
ERR206589 ERR206590 ERR206591 ERR206592 ERR206593 ERR206594 ERR206595
ERR206597 ERR206598 ERR206599 ERR206600 ERR206601 ERR206602 ERR206603
ERR206604 ERR206605 ERR206606 ERR206607 ERR206608 ERR206609 ERR206610
ERR206611 ERR206612 ERR206613 ERR206614 ERR206615 ERR206616 ERR206617
ERR206618 ERR207770 ERR207771 ERR207772 ERR207773 ERR207774 ERR207775
ERR207776 ERR207777 ERR207778 ERR207779 ERR207780 ERR2124117
ERR2124119 ERR2124120 ERR2124121 ERR2124122 ERR2124123 ERR2124124
ERR2124125 ERR2124126 ERR2124127 ERR2124128 ERR261858 ERR261859
ERR261860 ERR261861 ERR261862 ERR261863 ERR261864 ERR261865 ERR261866
ERR261867 ERR261868 ERR304754 ERR304755 ERR304756 ERR304757 ERR304758
ERR304759 ERR304760 ERR304761 ERR304762 ERR304763 ERR304764 ERR304765
ERR304766 ERR305913 ERR305914 ERR305915 ERR305916 ERR305917 ERR310823
ERR310824 ERR310825 ERR310826 ERR310827 ERR310828 ERR310829 ERR310830
ERR310831 ERR310832 ERR310833 ERR310834 ERR310835 ERR310836 ERR310837
ERR363217 ERR363218 ERR363219 ERR363220 ERR363221 ERR363222 ERR363223
ERR363224 ERR363225 ERR363226 ERR363227 ERR363228 ERR363229 ERR363230
ERR363231 ERR363232 ERR363233 ERR363234 ERR363235 ERR363236 ERR3723986
ERR3723987 ERR3723988 ERR3723989 ERR376170 ERR376171 ERR376172
ERR376173 ERR376174 ERR376175 ERR376176 ERR376177 ERR376178 ERR376179
ERR376180 ERR376181 ERR376182 ERR376183 ERR376184 ERR376185 ERR376186
ERR412008 ERR438786 ERR438787 ERR438788 ERR438789 ERR438790 ERR438791
ERR438792 ERR438793 ERR438794 ERR438795 ERR438796 ERR539726 ERR539727
ERR539728 ERR539729 ERR539730 ERR539732 ERR539733 ERR539734 ERR539735
ERR539736 ERR539737 ERR539738 ERR539739 ERR541269 ERR541270 ERR541271
ERR541272 ERR541273 ERR541274 ERR541275 ERR541276 ERR568417 ERR568418
ERR568419 ERR600042 ERR600043 ERR600044 ERR600045 ERR600046 ERR600047
ERR600048 ERR600049 ERR600050 ERR600051 ERR600054 ERR600056 ERR754983
ERR754984 ERR862590 ERR862591 ERR862592 ERR862593 ERR862594 ERR862595
ERR862596 ERR862597 ERR925016 ERR925017 ERR925018 ERR925019 ERR925020
ERR925021 ERR925022 ERR925023 ERR925024 ERR925025 ERR925026 SRR1254938

SRR1254940	SRR6257364	SRR6257365	SRR6257366	SRR6257367	SRR6257368
SRR6257369	SRR6257370	SRR6257371	SRR6316153	SRR6316158	SRR6316159
SRR6316161	SRR6369639	SRR6369648	SRR6369649	SRR6369650	SRR6369651
SRR6369652	SRR6369653	SRR6369660	SRR7133731	SRR7133732	SRR7133733

Supplementary Table 4.S2 Summary of polymorphisms across Sri Lankan isolates

Group	Isolate	Homozygous	Heterozygous	Average SNP Quality	Intersection (Het : Lm/Lt)	CL/VL
SL1*	ERR205734	34401	13075	36.21039	211	CL
	SRR7133731	20043	4987	41.78293	114	CL/VL
	SRR7133732	17732	4754	43.5989	129	VL
	SRR7133733	38481	10187	37.66367	120	CL
SL2	SRR6257364	96335	824077	59.55408	767915	CL
	SRR6257365	128317	634899	62.1653	560070	CL
	SRR6257366	268534	1559016	60.00943	1332260	VL #
	SRR6257367	163181	790134	54.89504	672973	VL #
	SRR6257369	110673	1083904	67.31232	931767	CL
SL3	SRR6257368	143212	33995	59.95966	3314	CL
	SRR6257370	141234	28620	53.58575	2784	CL
	SRR6257371	141145	33266	62.02206	3674	CL

*: SL1 group isolates were compared to the LdBPK reference genome to avoid masking SL1 specific homozygous polymorphism as described in Methods.

#: Indicates co-morbidities that can influence pathology

Supplementary Table 4.S3 Old World *Leishmania* species genome matches after alternative allele gene reconstruction

	SL2 A		SL2 B			SL3			SL1			
	SRR6257364	SRR6257365	SRR6257366	SRR6257367	SRR6257369	SRR6257368	SRR6257370	SRR6257371	ERR205734	SRR7133731	SRR7133732	SRR7133733
L. donovani	3821 p<0·0001	4979 p<0·0001	662 p<0·0001	2871 p<0·0001	1711 p<0·0001	6492 p<0·0001	6495 p<0·0001	6475 p<0·0001	8012	8012	8011	8015
L.infantum	1770	2214	296	1404	860	2942 p<0·0001	2940 p<0·0001	2958 p<0·0001	1423	1423	1424	1420
L.major	3738 p<0·0001	2160 p<0·0001	230	203	227	47	46	50	46	46	46	46
L.tropica	93	115	7671 p<0·0001	4568 p<0·0001	5977 p<0·0001	112	112	110	111	111	111	111
L.aethiopica	6	6	412	317	496	0	0	0	0	0	0	0
L.arabica	1	2	8	5	5	0	0	0	0	0	0	0
L.gerbilli	33	21	66	33	69	0	0	0	0	0	0	0
L.enriettii	0	0	0	0	0	0	0	0	0	0	0	0
L.tarentolae	7	6	16	15	16	6	6	6	6	6	6	6
L.turanica	288	254	396	341	396	158	158	158	159	159	159	159

Supplementary Table 4.S3 Cumulative results of the best scoring species matches from genome wide BLAST searches using alternative allege gene reconstruction. In every sample, each gene across the genome was modified to reflect the sample polymorphism. These genes were then compared to the complete reference genomes of all Old World Leishmania species and the highest scoring alignment per gene was counted as one species match for that sample.

Supplementary Table 4.S4 Old World *Leishmania* strains genome matches after alternative allele gene reconstruction

	SL2 A		SL2 B			SL3			SL1			
	SRR6257364	SRR6257365	SRR6257366	SRR6257367	SRR6257369	SRR6257368	SRR6257370	SRR6257371	ERR205734	SRR7133731	SRR7133732	SRR7133733
LdCL	290 p<0·0001	291 p<0·0001	262 p<0·0001	445 p<0·0001	259 p<0·0001	306 p<0·0001	340 p<0·0001	331 p<0·0001	4608	4748	4695	4751
LdBPK282	141 p<0·0001	156 p<0·0001	80 p<0·0001	311 p<0·0001	83 p<0·0001	191 p<0·0001	217 p<0·0001	205 p<0·0001	2772	2646	2695	2646
LdLV9	3389 p<0·0001	4531 p<0·0001	318 p<0·05	2112 p<0·05	1369 p<0·05	5995 p<0·0001	5938 p<0·0001	5938 p<0·0001	629	615	618	615
LdBHU_1220	1	1	2	3	0	0	0	1	3	3	3	3
LinJPCM5	1770	2214	296	1404	860	2942 p<0·0001	2940 p<0·0001	2958 p<0·0001	1423	1423	1424	1420
LmjF	1365 p<0·001	666 p<0·001	23	14	23	0	0	0	0	0	0	0
LmjSD75	1932 p<0·0001	1231 p<0·0001	190	175	190	47	46	50	46	46	46	46
LmjLV39	441	264	17	14	14	0	0	0	0	0	0	0
LtrL590	93	115	7671	4568 p<0·0001	5977 p<0·0001	112	112	110	111	111	111	111

Supplementary Table 4.S4 Cumulative results of the best scoring strain matches from genome wide BLAST searches using alternative allege gene reconstruction. In every sample, each gene across the genome was modified to reflect the sample polymorphism. These genes were then compared to the complete reference genomes of all Old World Leishmania strains and the highest scoring alignment per gene was counted as one strain match for that sample. LdCL = Leishmania donovani strain CL-SL, LdBPK282 = Leishmania donovani strain BPK282A1, LdLv9= Leishmania donovani strain LV9, LinJPCM5 = Leishmania infantum strain JPCM5, LmjF = Leishmania major strain Friedlin, LmjSD75 = Leishmania major strain SD 75.1, LmjLV39 = Leishmania major strain LV39c5, LtrL590 = Leishmania tropica strain L590.

Supplementary Table 4.S5 Non-synonymous polymorphisms identified in the SL3 group in common with *L. major* SNPs

Chr	Position	Gene ID	Protein Name	Effect
LdCL_06	373712	LdCL_060013900	hypothetical protein	Ser2073Thr
LdCL_06	373865			Ala2124Ser
LdCL_06	374067			Thr2191Met
LdCL_06	374691			Leu2399Ser
LdCL_06	375785			Leu2764Met
LdCL_06	379939	LdCL_060014000	dihydrofolate reductase-thymidylate synthase	Thr200Ala
LdCL_06	379942			Thr201Ala
LdCL_06	382826	LdCL_060014100	arginine N-methyltransferase, type III, putative	Glu64Gly
LdCL_06	401369	LdCL_060014600	hypothetical protein	Cys267Tyr
LdCL_06	449436	LdCL_060016100	2OG-Fe(II) oxygenase superfamily, putative	Ala8Val
LdCL_06	483218	LdCL_060017300	protein kinase, putative	His378Arg
LdCL_07	362713	LdCL_070013400	hypothetical protein	Asp33Gly
LdCL_09	135099	LdCL_090009000	DNA photolyase, putative	Arg131Gln
LdCL_09	163075	LdCL_090009800	tyrosine phosphatase, putative	Ala771_Ala774del
LdCL_09	275934	LdCL_090013200	hypothetical protein	Lys504Glu
LdCL_10	169505	LdCL_100009100	pteridine transporter, putative	Thr174Arg
LdCL_11	401388	LdCL_110016000	hypothetical protein	Gln72Pro

LdCL_11	401414			His81Tyr
LdCL_11	490683	LdCL_110018100	ATP-binding cassette subfamily A, member 1, putative	Gly631Asp
LdCL_12	431663	LdCL_120013900	hypothetical protein	Tyr43His
LdCL_12	569653	LdCL_120017400	hypothetical protein	Tyr43His
LdCL_12	569675			Trp50*
LdCL_12	588162	LdCL_120017900	surface antigen protein 2, putative	Gly101Ser
LdCL_12	588178			Thr106Asn
LdCL_12	589089			Asp410Asn
LdCL_12	589891			Ala677Val
LdCL_12	605397	LdCL_120018400	surface antigen protein 2, putative	Glu76Val
LdCL_12	605472			Asp101Gly
LdCL_12	605487			Thr106Asn
LdCL_12	605492			Met108Val
LdCL_12	605503			His111Gln
LdCL_12	605535			Ser122Ile
LdCL_12	605550			Ala127Asp
LdCL_12	605563			Asn131Lys
LdCL_12	605577			Ser136Thr
LdCL_12	605580			Ser137Leu
LdCL_12	605582			Val138Leu
LdCL_13	374514	LdCL_130015300	leucyl-tRNA synthetase, putative	Leu308Phe
LdCL_14	222212	LdCL_140011500	hypothetical protein	Thr687Ala

LdCL_14	346848	LdCL_140014000	hypothetical protein	Gln2642dup
LdCL_14	534346	LdCL_140017800	kinesin K39, putative	Ser1303Asn
LdCL_14	538404			Ala2656Thr
LdCL_14	590582	LdCL_140019500	NADP oxidoreductase coenzyme F420-dependent, putative	Val47Ala
LdCL_14	606148	LdCL_140019900	serine hydroxymethyltransferase (SHMT-L)	Ser24Gly
LdCL_14	659681	LdCL_140021300	hypothetical protein	Ala747Val
LdCL_15	136121	LdCL_150008900	hypothetical protein	Asn394Ser
LdCL_15	162995	LdCL_150010100	tb-292 membrane associated protein-like protein	Glu510Gln
LdCL_15	164150			Glu895Gln
LdCL_16	196816	LdCL_160010600	AAA domain containing protein, putative	Ser85Pro
LdCL_17	88351	LdCL_170007500	receptor-type adenylate cyclase, putative	Ser20Pro
LdCL_17	112477	LdCL_170008200	Kinesin motor domain containing protein, putative	Arg524Cys
LdCL_17	118974	LdCL_170008300	WD domain, G-beta repeat, putative	Ala399Val
LdCL_17	119553			Phe206Ser
LdCL_18	609144	LdCL_180019300	pumilio protein 2, putative	Asn875Thr
LdCL_18	609173			Ser885Gly
LdCL_18	609201			Thr894Met
LdCL_19	478877	LdCL_190016300		Gly3850Asp

LdCL_19	480896		Ankyrin repeats (3 copies)/Ankyrin repeat, putative	Leu4525del
LdCL_22	551683	LdCL_220018200	hypothetical protein	Asn183Ser
LdCL_23	467160	LdCL_230018000	hypothetical protein	Val2766Ala
LdCL_24	154090	LdCL_240009300	Protein of unknown function (DUF3184), putative	Ser362Ala
LdCL_24	161257	LdCL_240009400	Protein of unknown function (DUF3184), putative	Ser357Ala
LdCL_24	324959	LdCL_240014200	hypothetical protein	Ser600Pro
LdCL_24	512669	LdCL_240019700	hypothetical protein	His426Asp
LdCL_24	734342	LdCL_240025400	WD domain, G-beta repeat/PFU (PLAA family ubiquitin binding), putative	Ser402Gly
LdCL_26	800281	LdCL_260026400	hypothetical protein	Ala209Thr
LdCL_26	884803	LdCL_260028200	kynureninase, putative	Ala189Val
LdCL_27	318494	LdCL_270012900	calcium uniporter protein, mitochondrial, putative	Arg37His
LdCL_27	321591	LdCL_270013000	MatE, putative	Ser8Gly
LdCL_27	321790			Arg74His
LdCL_27	334538	LdCL_270013300	Right handed beta helix region/Periplasmic copper-binding protein (NosD), putative	Gln1101del
LdCL_27	581562	LdCL_270019400	protein kinase, putative	Ser1202Pro

LdCL_27	701974	LdCL_270022700	molybdopterin synthase sulfurylase-like protein, putative	Val389Ala
LdCL_27	893453	LdCL_270026900	Transmembrane adaptor Erv26, putative	Arg179Gln
LdCL_28	711944	LdCL_280024500	DNA repair protein-like protein	Asp652Gly
LdCL_28	890779	LdCL_280029500	mitochondrial DNA topoisomerase II	Pro1380Ser
LdCL_28	938512	LdCL_280030800	tRNA methyltransferase complex GCD14 subunit/Protein-L-isoaspartate(D-aspartate) O-methyltransfer...	Thr43Ile
LdCL_29	291220	LdCL_290013000	hypothetical protein	Arg423Leu
LdCL_29	496404	LdCL_290018500	hypothetical protein	Ala883Val
LdCL_29	677115	LdCL_290021200	hypothetical protein	Gly296Ser
LdCL_29	775470	LdCL_290023400	ATP-binding cassette protein subfamily H, member 2, putative	Leu183Ile
LdCL_29	1169010	LdCL_290033900	serine/threonine-protein kinase Nek, putative	Pro364Ser
LdCL_30	516744	LdCL_300020600	p1/s1 nuclease	Ala128Ser
LdCL_30	711383	LdCL_300025400	hypothetical protein	Leu122Arg
LdCL_31	168328	LdCL_310010100	amastin, putative	Ile78Val
LdCL_31	253938	LdCL_310012600	hypothetical protein	Tyr101His
LdCL_31	375755	LdCL_310015700	sodium stibogluconate resistance protein, putative	Glu458Gly

LdCL_31	449919	LdCL_310017200	hypothetical protein	Ala513Val
LdCL_31	466862	LdCL_310017700	hypothetical protein	Leu91Phe
LdCL_31	528422	LdCL_310019500	hypothetical protein	Leu1091Phe
LdCL_31	613221	LdCL_310020500	Cold-shock DNA-binding domain containing protein, putative	Leu31Pro
LdCL_31	619290	LdCL_310020600	hypothetical protein	Asn338Asp
LdCL_31	619307			Asn332Ser
LdCL_31	619502			Arg267Lys
LdCL_31	619821			Thr161Ala
LdCL_31	619989			Met105Val
LdCL_31	620189			Thr38Met
LdCL_31	917260	LdCL_310026100	protein kinase, putative	Lys33Gln
LdCL_31	987807	LdCL_310027400	RING-variant domain containing protein, putative	Ser104Asn
LdCL_31	987835			Leu95Phe
LdCL_31	1084057	LdCL_310029900	Fungal tRNA ligase phosphodiesterase domain containing protein, putative	Arg520Gln
LdCL_31	1151019	LdCL_310030900	hypothetical protein	Asn1902Ser
LdCL_31	1158599	LdCL_310031000	hypothetical protein	Asp359Asn
LdCL_31	1265045	LdCL_310033400	lipase, putative	Met346Ile
LdCL_31	1478843	LdCL_310040100	iron/zinc transporter protein-like protein	Cys269Tyr
LdCL_31	1482804	LdCL_310040200	iron/zinc transporter protein-like protein	Cys269Tyr

LdCL_32	221657	LdCL_320011500	ATP-dependent DEAD/H RNA helicase, putative	Gly19Glu
LdCL_32	844420	LdCL_320027800	hypothetical protein	Arg135His
LdCL_32	1135510	LdCL_320035800	Prefoldin subunit, putative	Thr370Met
LdCL_33	250875	LdCL_330013100	POT family, putative	Ser215Thr
LdCL_33	345811	LdCL_330015500	dnaj chaperone-like protein	Ala533Val
LdCL_33	350414	LdCL_330015800	hypothetical protein	Lys148Arg
LdCL_33	354193	LdCL_330015900	hypothetical protein	Thr59Ser
LdCL_33	562350	LdCL_330021500	protein kinase, putative	Cys1720Tyr
LdCL_34	168645	LdCL_340009900	hypothetical protein	Leu118Phe
LdCL_34	209343	LdCL_340010500	Amastin surface glycoprotein, putative	Ala41Gly
LdCL_34	226078	LdCL_340010900	phosphoglycan beta 1,2 arabinosyltransferase, (SCA like)	Thr81Ile
LdCL_34	372801	LdCL_340013900	serine/threonine-protein phosphatase PP1, putative	His99Gln
LdCL_34	507010	LdCL_340017500	amastin-like surface protein, putative	Leu15Phe
LdCL_34	507037			Gly6Ser
LdCL_34	1194752	LdCL_340033400	methyltransferase-like protein	Asp349del
LdCL_34	1547780	LdCL_340044300	Inositol-pentakisphosphate 2-kinase, putative	Val160Ala
LdCL_34	1777442	LdCL_340051800	sjogren s syndrome nuclear autoantigen 1, putative	Asp59Glu
LdCL_35	163307	LdCL_350010100	proteophosphoglycan ppg3, putative	Val2385Ile

LdCL_35	599964	LdCL_350020200	hypothetical protein	Cys10Tyr
LdCL_35	1412527	LdCL_350042000	hypothetical protein	Gly752Asp
LdCL_35	1716273	LdCL_350050200	hypothetical protein	Ile65Thr
LdCL_35	1716516			Ala146Val
LdCL_35	1900240	LdCL_350055800	hypothetical protein	Leu120Ser
LdCL_36	958054	LdCL_360030700	hypothetical protein	Glu63Gly
LdCL_36	1300075	LdCL_360040600	oxidoreductase, putative	His269Arg
LdCL_36	1309078	LdCL_360040900	Fungal domain of unknown function (DUF1712), putative	Arg64Leu
LdCL_36	1354815	LdCL_360042900	CRAL/TRIO domain containing protein, putative	Thr468Ala
LdCL_36	1354866			Pro451Ser
LdCL_36	1887706	LdCL_360057700	WD domain, G-beta repeat, putative	Gly416Glu
LdCL_36	2255530	LdCL_360068100	mkiaa0324 protein-like protein	Lys86Glu

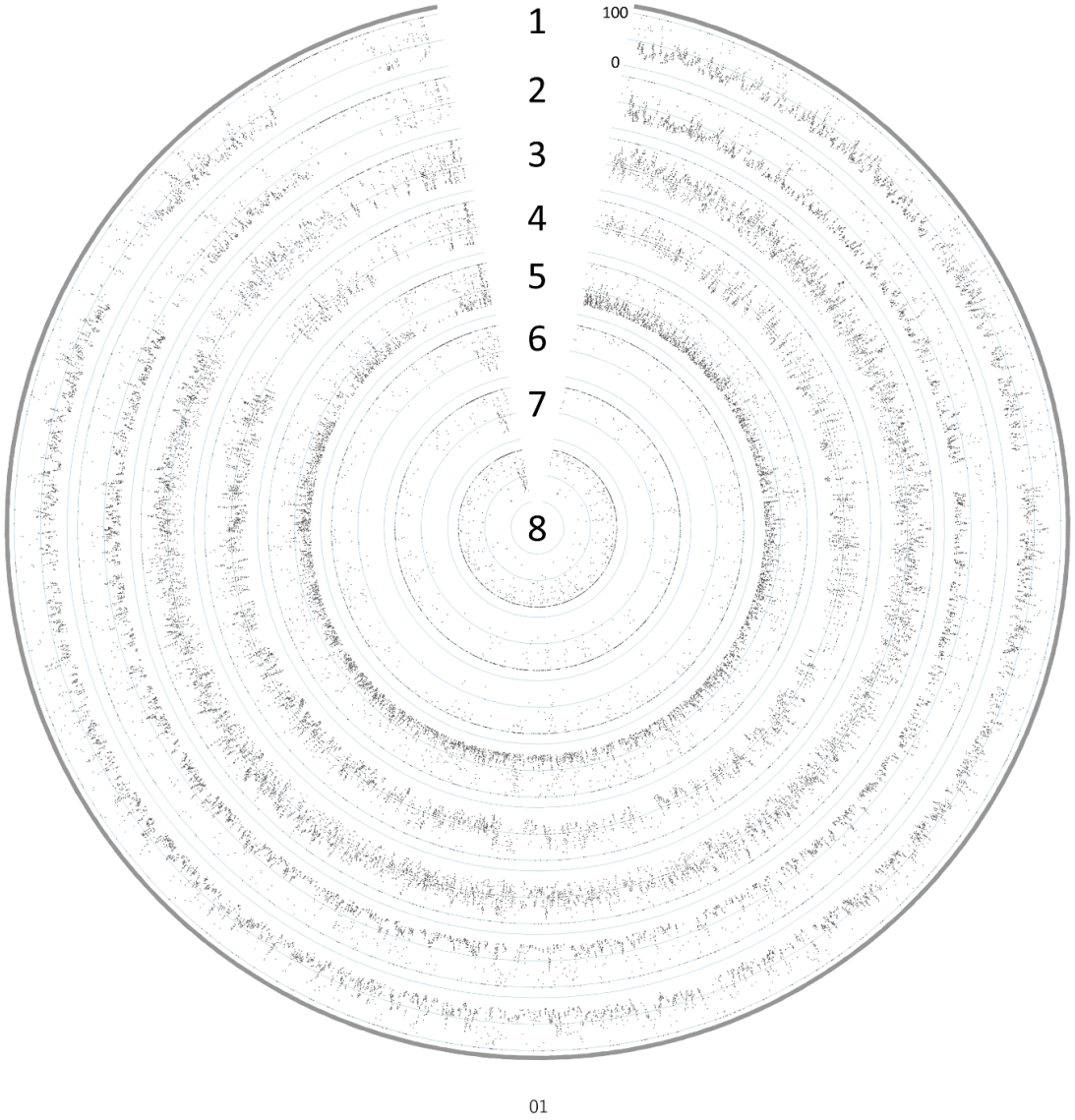
Supplementary Table 4.S2

Polymorphisms identified in the SL3 group were filtered by removing any SNPs in common with visceral *L. donovani* complex species and retained only if present in all three SL3 group isolates and in common with *L. major* alleles.

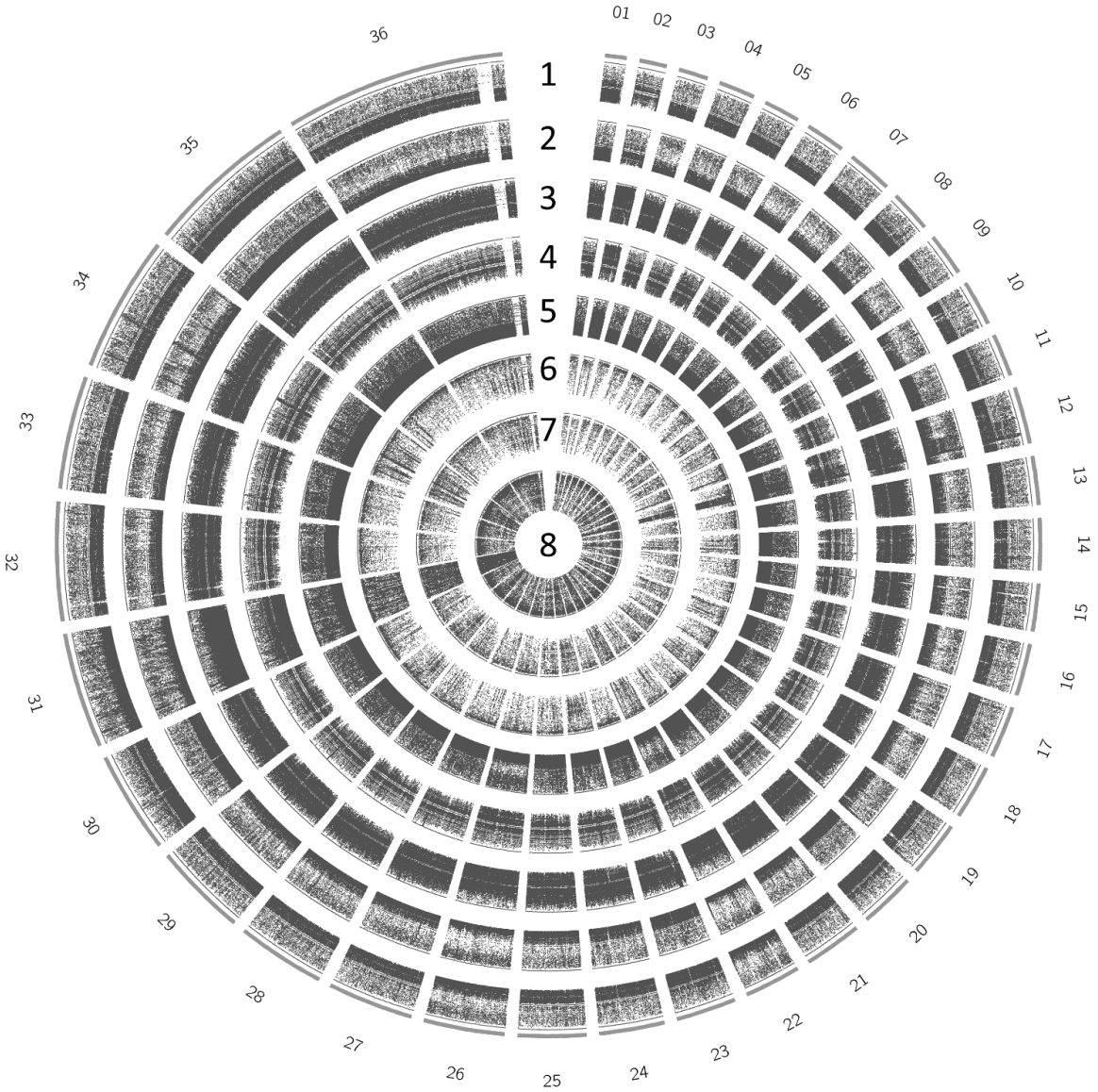
A

SL2
1. SRR6257364
2. SRR6257365
3. SRR6257366
4. SRR6257367
5. SRR6257369

SL3
6. SRR6257368
7. SRR6257370
8. SRR6257371



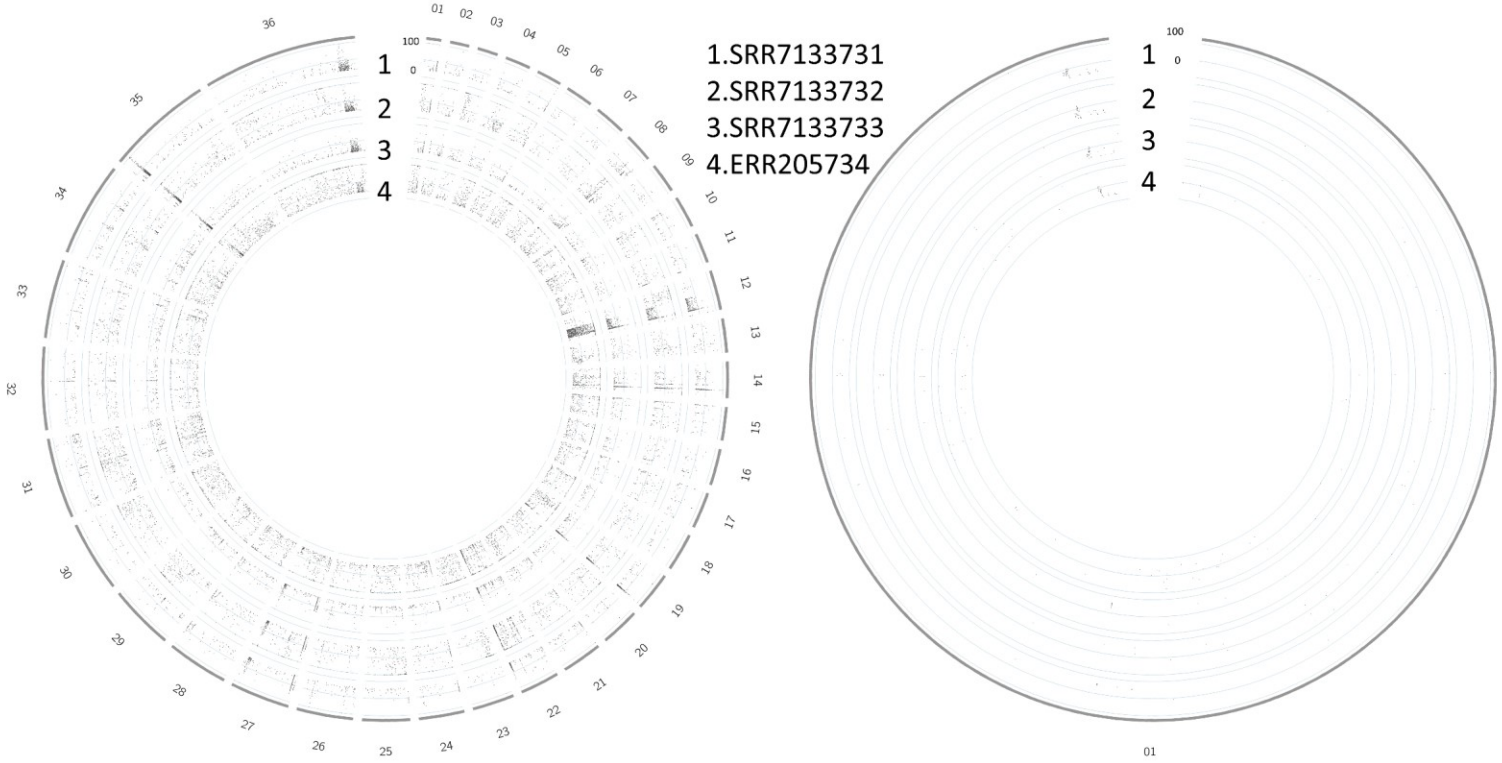
B



SL2
1. SRR6257364
2. SRR6257365
3. SRR6257366
4. SRR6257367
5. SRR6257369

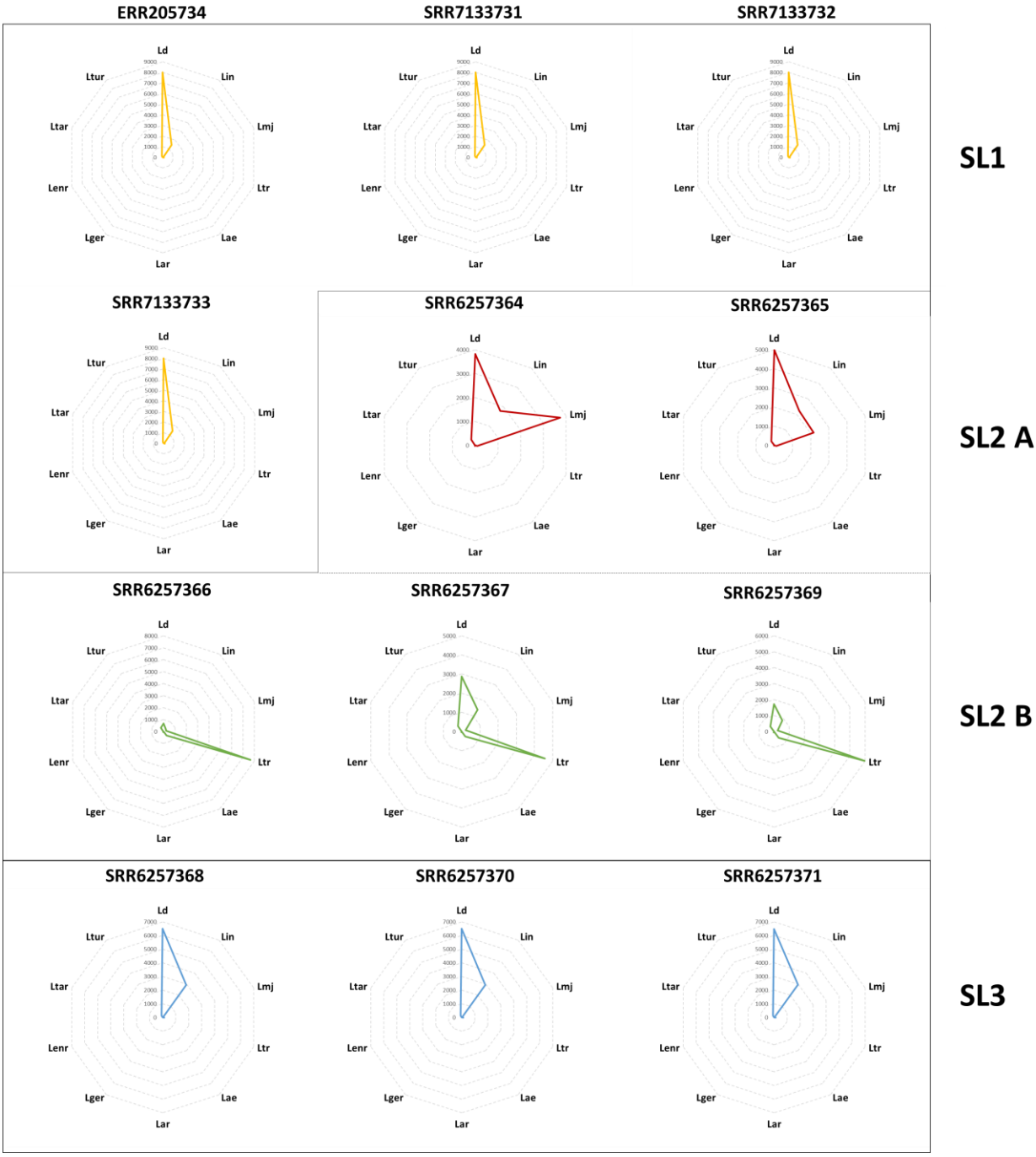
SL3
6. SRR6257368
7. SRR6257370
8. SRR6257371

C



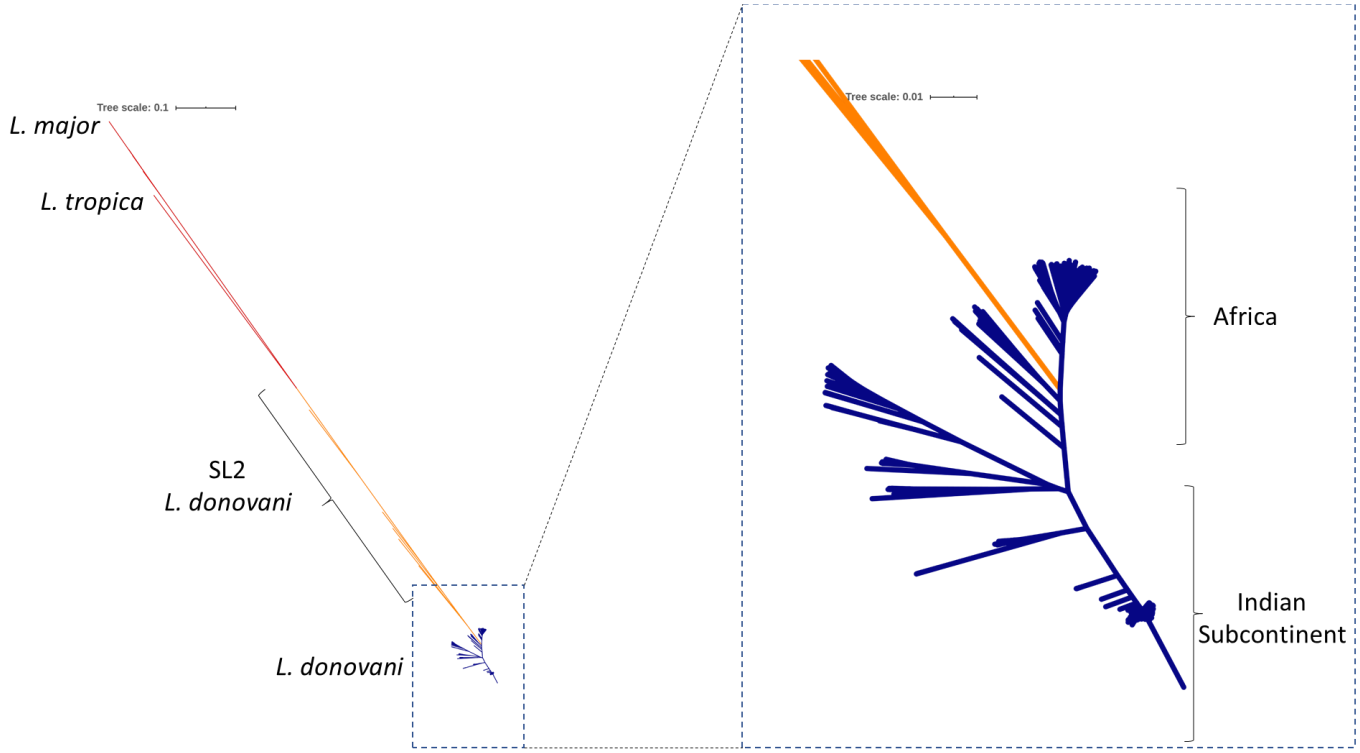
Supplementary Figure 4.S1 Frequency and distribution of polymorphisms across Sri Lankan isolates

The location of every polymorphism is depicted along the X-axis and the underlying read frequency along the Y-axis as a single dot. Points centered around the center of each track represent ~50%, points at the outer edge of each track represent homozygous ~100% polymorphisms. To better distinguish the frequency distribution, the plot is limited to Chromosome 1 (**A**) and observe the distribution of polymorphism across the genome, all 36 chromosomes (**B**) in the SL2 and SL3 groups. **C.** Location and read frequencies of SL1 group isolates across the whole genome or Chromosome 1.



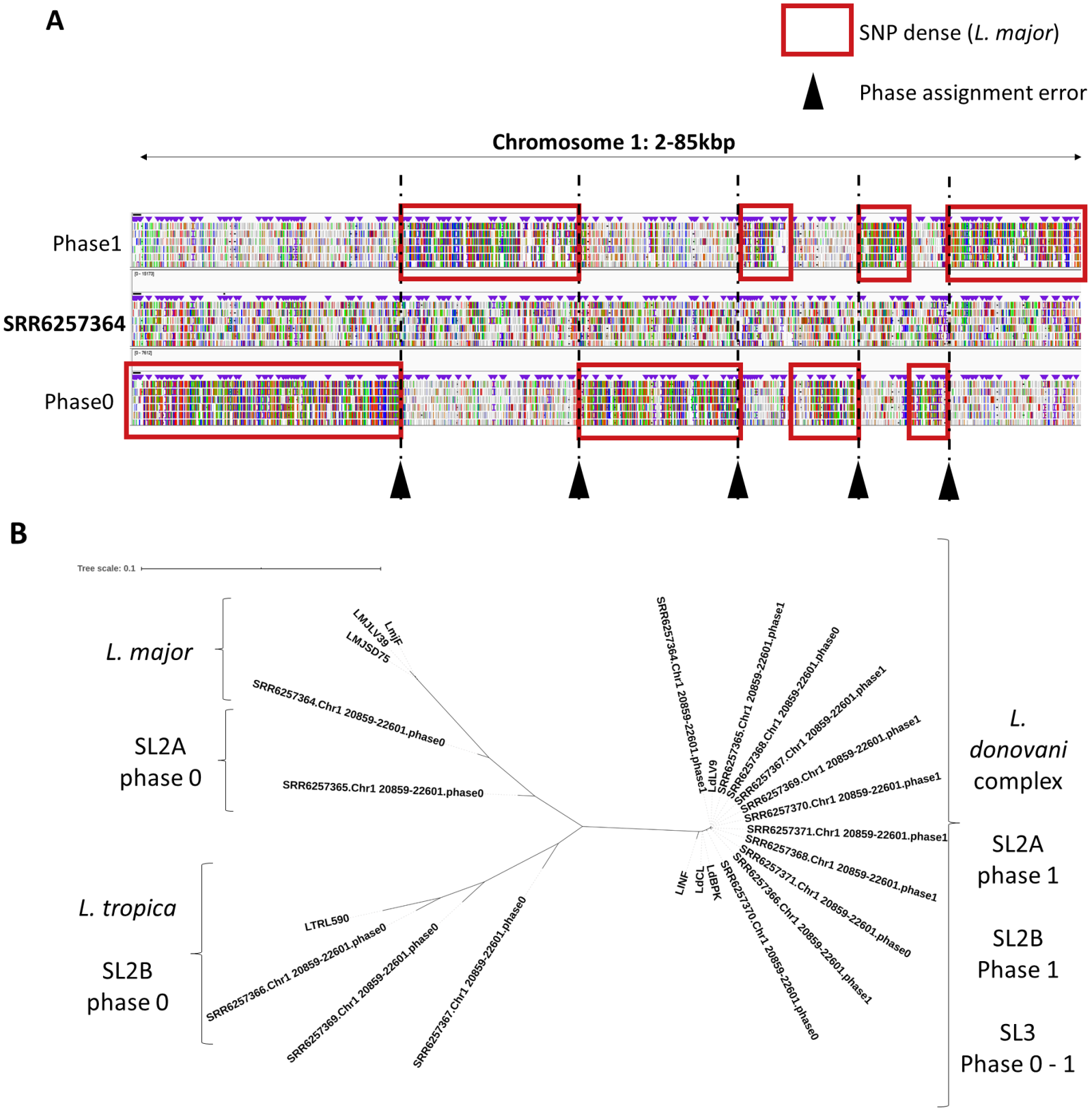
Supplementary Figure 4.S2 Distribution of species origin of reconstructed genes as determined by BLAST analysis

Ld= *Leishmania donovani,*Lin=*Leishmania infantum*, Lmj= *Leishmania major,*Ltr=*Leishmania tropica*, Lae=*Leishmania aethiopica*, Lar=*Leishmania arabica,* Lger=*Leishmania gerbilli*, Lenr=*Leishmania enriettii*, Ltar=*Leishmania tarentolae,* Ltur= *Leishmania turanica.*



Supplementary Figure 4.S3 Phylogenetic analysis of *L. donovani* isolates, *L. major* and *L. tropica*

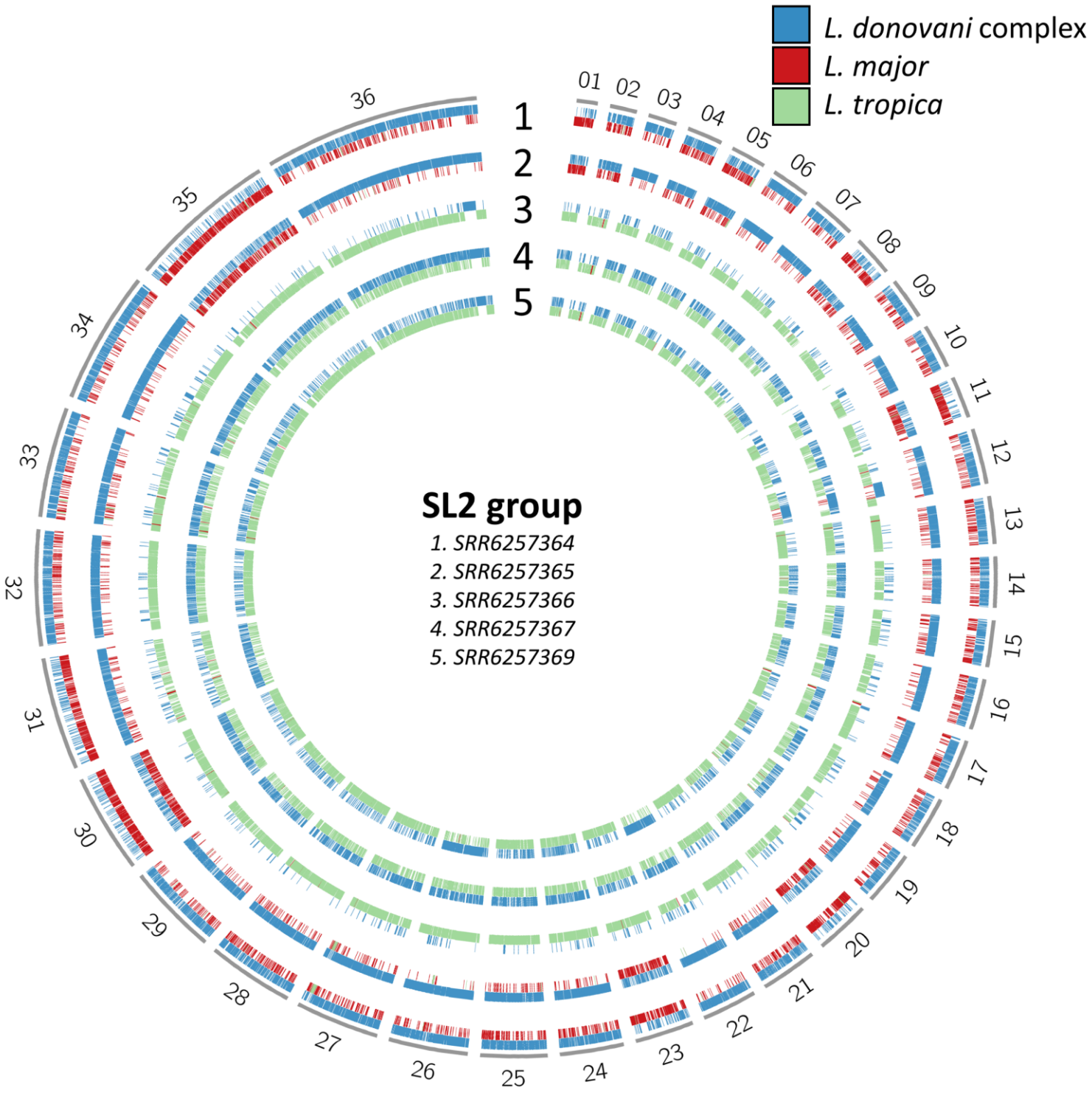
Branches corresponding to *L. donovani* are colored in blue, *L. major* and *L. tropica* in red, and SL2 putative hybrid parasites in orange.



Supplementary Figure 4.S4 Phylogenetic analysis after read-based phasing to separate

haplotypes

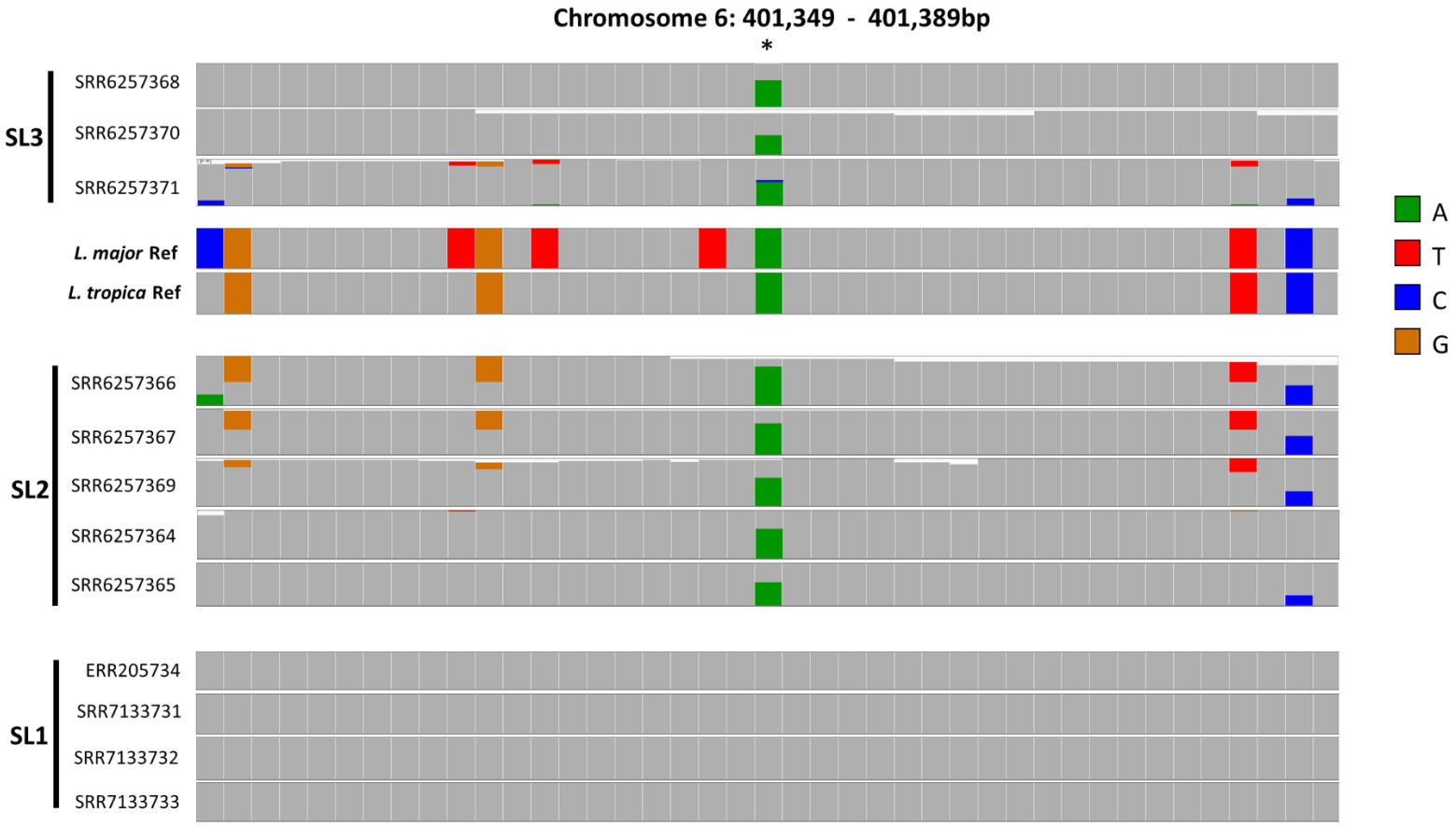
A. Read based phasing results in chimeric haplotypes as read blocks are assigned to the wrong phase sets as shown by the alternating assignment of SNP dense (*L. major*) read blocks in continuous phase sets. **B.** Phylogenetic analysis using phased haplotypes for a section of chromosome 1 showing a different clustering across phases in SL2A *L. major* hybrids and SL2B *L. tropica* hybrids.



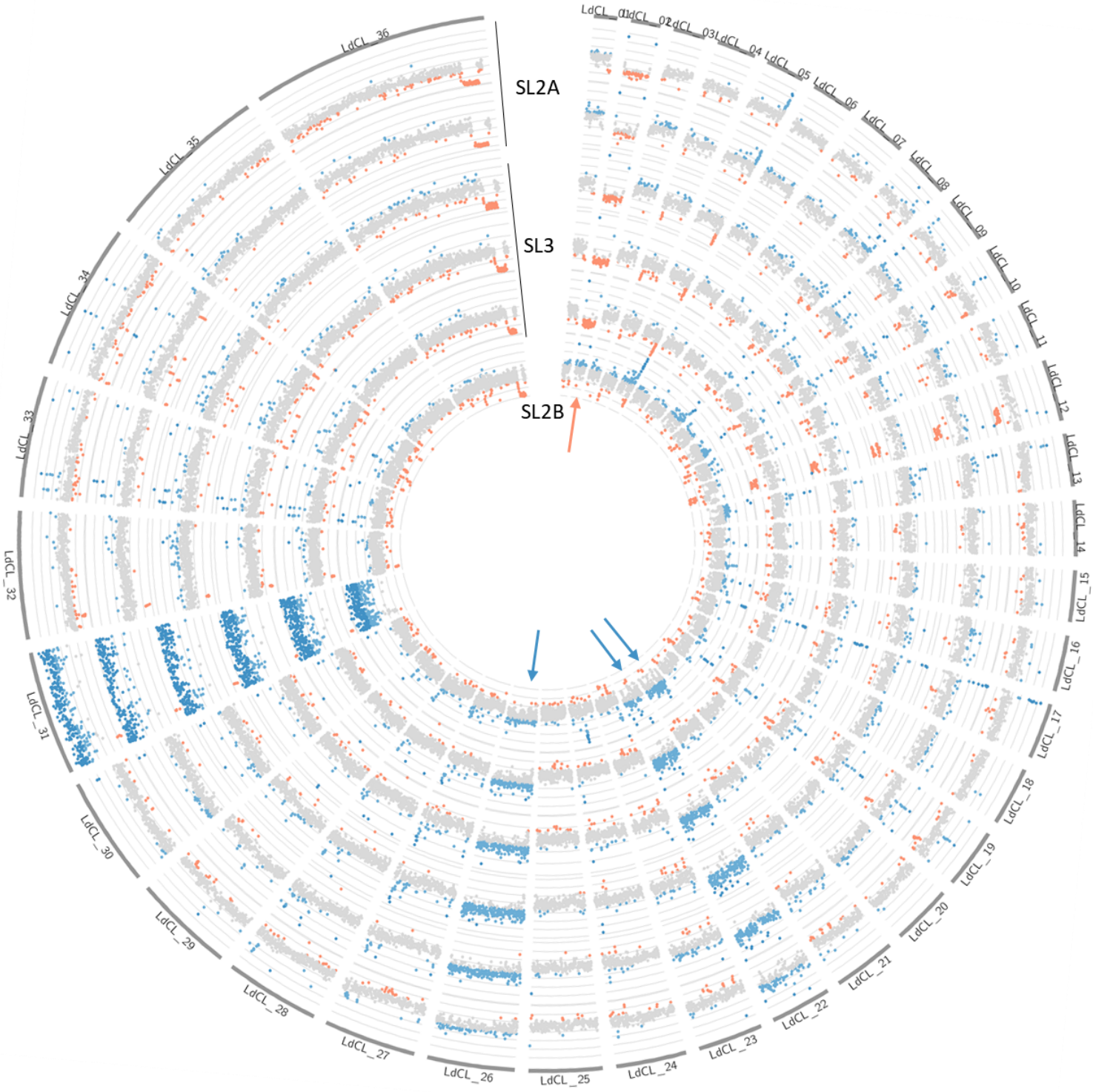
Supplementary Figure 4.S5 Representation of gene ancestry across the entire genome

Genes of *L. donovani* species complex origin are marked in blue. Genes with hybrid ancestry (*L. major & L. donovani,* or *L. tropica & L. donovani*) are colored in red and green respectively.

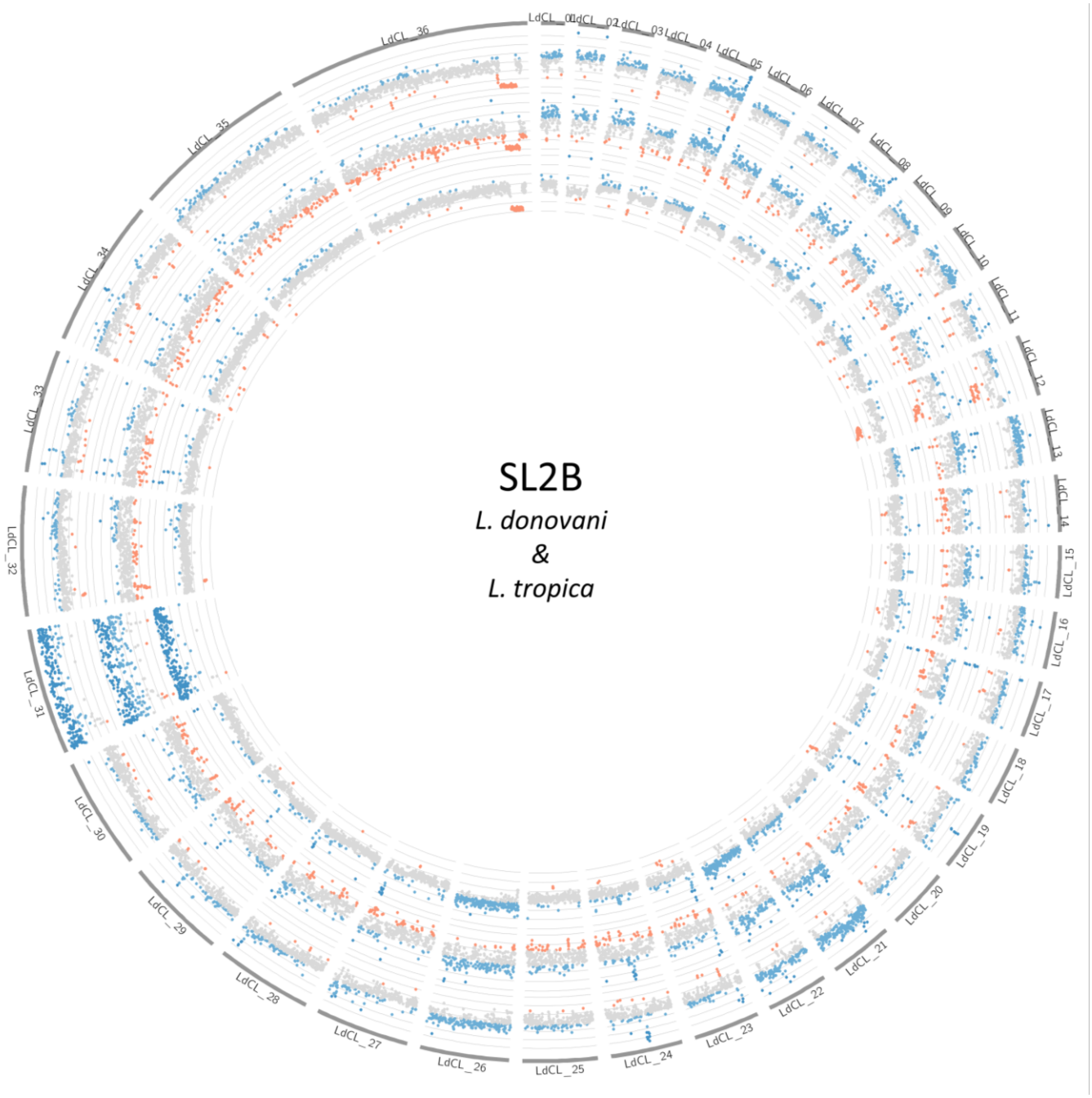
A


Chromosome 6: 401,349 - 401,389bp
*
SL3
SRR6257368
SRR6257370
SRR6257371
L. major Ref
L. tropica Ref
SL2
SRR6257366
SRR6257367
SRR6257369
SRR6257364
SRR6257365
SL1
ERR205734
SRR7133731
SRR7133732
SRR7133733
A
T
C
G

B


SL2A
SL3
SL2B
LdCL_01
LdCL_02
LdCL_03
LdCL_04
LdCL_05
LdCL_06
LdCL_07
LdCL_08
LdCL_09
LdCL_10
LdCL_11
LdCL_12
LdCL_13
LdCL_14
LdCL_15
LdCL_16
LdCL_17
LdCL_18
LdCL_19
LdCL_20
LdCL_21
LdCL_22
LdCL_23
LdCL_24
LdCL_25
LdCL_26
LdCL_27
LdCL_28
LdCL_29
LdCL_30
LdCL_31
LdCL_32
LdCL_33
LdCL_34
LdCL_35
LdCL_36

C



Supplementary Figure 4.S6 Evidence of ancient hybridization in SL3 group samples

A. Prevalence of polymorphisms across all Sri Lanka isolates on a portion of chromosome 6 compared with the reference *L. major* and *L. tropica* sequences. Grey boxes represent Sri Lanka reference *L. donovani* nucleotides. The SL3 group have retained varied levels of *L. major* alleles depending on the sample (upper 3 alignments). The SL2 group retained *L. major* or *L. tropica*

polymorphisms. The SL1 group does not share polymorphisms with either *L. tropica* or *L. major.* The polymorphism highlighted in the center (*) of the alignment and retained in all SL2 and SL3 isolates results in a Cys267Tyr change in the LdCL_060014600 gene and matches the *L. major* allele, shows that some polymorphisms appear to be more stable in the hybrid genomes. **B**. Conservation of aneuploidy patterns across SL3 and SL2A group samples. Chromosomal coverage was determined and colored according to mean coverage (grey), decrease coverage (red) or increase coverage (blue) in sequencing depth. Coverage at each gene location across the entire genome shows the *L. major* hybrid isolates (SL2A) have reduced average copies of chromosome 2, and increased copy number of chromosome 22 and 26. The SL3 isolates show the same aneuploidy pattern as the *L. major* hybrids (SL2A). In comparison, SL2B *L. tropica* hybrid parasites have normal coverage at chromosome 2, a unique increase at chromosome 21, and share the increase across chromosomes 22 and 26. All isolates appear diploid for all other chromosomes with the exception of naturally tetraploid chromosome 31. **C.** Comparison of chromosomal aneuploidy in SL2B (*L. tropica* hybrids) isolates.

Supplementary Method 4.S1 Data collection

A search was performed to obtain all sequencing records for the *Leishmania donovani* organism (NCBI Taxonomy ID 5661). The read sets were filtered to remove any data originating from unknown or transcriptomic cDNA source. The remainder of gDNA reads sets were filtered to remove targeted capture experiments and retained only whole genome random selection libraries to prevent any bias towards particular genomic regions. As long reads are error prone and more suitable for assembly than SNP calling, non-Illumina long read sets were removed. To increase the accuracy of the SNP calls, only reads in paired mode were retained. Other samples with low coverage were removed manually after inspection of the alignments. The retained 684 sequences used for phylogenetic analysis are listed in **Supplementary Table 4.S1**. Included in the sequences were 12 of Sri Lankan origin that included CL and VL cases.

Supplementary Method 4.S2 Alignment of all sequenced *L. donovani* isolates to the reference genome

The filtered 684 sequencing samples were distributed evenly across three Compute Canada clusters, Beluga at Calcul Quebec, Niagara at SciNet, and Cedar at WestGrid. The raw reads for each sample were downloaded using the SRA-Toolkit (version 2.9.6) provided by NCBI (34) using the 'fastq-dump –split-files' command. The alignment of all reads obtained from the SRA was performed as previously described (17). Briefly, Illumina paired reads were aligned to the reference Sri Lanka genome (17) using the Burrows-Wheeler Aligner (35) (version 0.7.17), file formats transformed using samtools (36) (version 1.10), and variant calling was done with VarScan2 (37) (version 2.4.3) to generate VCF files. Per sample candidate SNP were called by VarScan2 with a minimum coverage of 8, a minimum alternate allele frequency of 20% (reads/reads), a minimum average base quality of 15 across the reads and a 90%(CI) significance threshold. This pipeline was automated using in house scripts to process the data in parallel across 684 nodes on the clusters resulting in one variant call form (VCF) file per sample containing a list of all polymorphisms and their respective frequencies.

Supplementary Method 4.S3 Phylogenetic analysis

The resulting 684 VCF files generated at the alignment stage described above were merged using BCFtools (36) (version 1.10.2) with the command 'bcftools merge –missing-to-ref ' resulting in a single VCF file containing all the possible genetic polymorphisms identified in all 684 samples. The merged VCF file was imported into TASSEL (version 5.0) (38) and subjected to a relatedness analysis to generate a distance matrix and a phylogenetic tree using the Neighbor-Joining algorithm. The phylogenetic tree was exported in Newick format and visualized using the Interactive Tree of Life (IToL, version 5) (39) to assign colors to nodes and clades.

After identification of likely inter-species hybridization, additional sequences from *L. major* and *L. tropica* whole genomes (retrieved from TriTrypDb v46 (40)) were aligned to the *L. donovani* LdCL genome (17) to generate a list of polymorphisms between the species and these were added to the global *L. donovani* polymorphism VCF file and processed as above to generate a phylogenetic tree to place the putative hybrid strains.

Supplementary Method 4.S4 Heterozygosity of *L. donovani* isolates worldwide

Variant loci from the VarScan2 annotated VCF files were assigned a HET or HOM value based on the variant allele read frequency for each site for each sample if the software determined the SNP to be heterozygous or homozygous respectively. The frequency of HET and HOM variant annotations across the entire genome was calculated on a per sample basis resulting in a single data point per sample with the Heterozygous ratio defined as (Het/Het+Hom). As the SL1 isolates fall within the reference cluster, resulting in homozygous polymorphisms being masked by the reference which artificially shifts the Heterozygous ratio, these isolates were aligned to the Nepalese reference genome for the purpose of this calculation. The Sri Lankan isolates from BioProject PRJNA413320 (belonging to the SL2 and SL3 groups), our previous studies (17, 18), and one sample from PRJEB2600 (belonging to the SL1 group)were grouped together. 10 samples previously characterized as intra-species hybrids of Ethiopian *L. donovani* (22) were grouped together as a hybrid sample positive control group. The remaining 662 samples were grouped together to represent the natural *L. donovani* distribution.

Supplementary Method 4.S5 Identification of species in the hybrid parasites

The genomic annotation from the *L. donovani* LdCL strain previously generated (17) were used to generate a region list of the genomic coordinates of every gene in the 'chr:start-stop' format. This region list was used as an input to samtools with the 'faidx' command to extract the genomic sequence corresponding to each gene locus from the LdCL reference genome file.

For each sample independently, the VCF file containing the location and nucleotide change of every polymorphism along the genome for that sample was used as an input file for BCFtools (41) with the command 'consensus' to transform the genomic sequences from the reference generated in the previous step into their alternate or reconstructed sequences to reflect the genotype of each sample for 9,757 gene loci.

The complete genomes of all *Leishmania* species reference strains were downloaded from TriTrypDB v46 (40) (*L. major* Friedlin, *L. donovani* BPK282, *L. tropica* L590, *L. tarentolae* Parrot-TarII, *L. turanica* LEM423, *L.gerbilli* LEM452, *L. enriettii* LEM3045, *L. arabica* LEM1108 & *L. aethiopica* L147) in FASTA format. The genomes were concatenated into a single FASTA file and used as an input for NCBI BLAST+ v2.7.1 using the command 'makeblastdb' and '-dbtype nucl' option to create a database of Old World *Leishmania* genomic sequences. For the refined search, all additional non-reference strains genetic sequences were later added (*L. donovani* LV9, *L. donovani* LdCL, *L. donovani* BHU1220, *L. major* SD75, *L. major* LV39).

The alternate or reconstructed sequences for each sample were then used as a list of input queries for this database using the command 'blastn' with the options '-max_target_seqs 1 -max_hsps 1 -outfmt "6 qseqid qcovs pident stitle" to create a report of 9,757 species or strain matches for each queried sample. These reports were then tabulated to generate radar plots depicting the probable

genetic contributions of each species/strain per sample and the genomic coordinates used to generate chromosomal maps of parental blocks using circus (42) to paint the regions on a circular karyotype representation of the *L. donovani* genome .

Supplementary Method 4.S6 Haplotype phasing

For each sample, the BAM alignment file generated by the burrows wheeler aligner (35) was additionally processed prior to polymorphism analysis by VarScan v2 (37) as described above. The alignment was processed using the 'phase' command from samtools (36) with the option '-A' to drop reads with an ambiguous phase resulting in two separate alignment files with segregated reads based on haplotype blocks. The two alignments files containing roughly half of the original reads each were manually inspected for concordance along the coordinates output by the tools as being continuous phase blocks by manual inspection. Alternate gene reconstruction was performed on each phase of each sample independently using BCFtools (41) as described above and aligned with orthologous reference *Leishmania* sequences using Clustal Omega (version 1.2.4) (43) to generate a phylogenetic tree of the phased samples.

Supplementary References

1. Burza S, Croft SL, Boelaert M. 2018. Leishmaniasis. Lancet 392:951–970.

2. Scorza BM, Carvalho EM, Wilson ME. 2017. Cutaneous manifestations of human and murine leishmaniasis. Int J Mol Sci 18.

3. Gelanew T, Hailu A, Scho'nian G, Lewis MD, Miles MA, Yeo M. 2014. Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural ethiopian isolates of leishmania donovani. Int J Parasitol 44:751–757.

4. Chargui N, Amro A, Haouas N, Schönian G, Babba H, Schmidt S, Ravel C, Lefebvre M, Bastien P, Chaker E, Aoun K, Zribi M, Kuhls K. 2009. Population structure of Tunisian Leishmania infantum and evidence for the existence of hybrids and gene flow between genetically different populations. Int J Parasitol 39:801–811.

5. Ravel C, Cortes S, Pratlong F, Morio F, Dedet JP, Campino L. 2006. First report of genetic hybrids between two very divergent Leishmania species: Leishmania infantum and Leishmania major. Int J Parasitol 36:1383–1388.

6. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M, Svobodova M, Volf P, Berriman M, Cotton JA, Smith DF. 2014. Genomic Confirmation of Hybridisation and Recent Inbreeding in a Vector-Isolated Leishmania Population. PLoS Genet 10.

7. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, Dobson DE, Beverley SM, Sacks DL. 2009. Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. Science (80-) 324:265–268.

8. Sadlova J, Yeo M, Seblova V, Lewis MD, Mauricio I, Volf P, Miles MA. 2011. Visualisation of leishmania donovani fluorescent hybrids during early stage development in the sand fly vector. PLoS One 6.

9. Inbar E, Akopyants NS, Charmoy M, Romano A, Lawyer P, Elnaiem DEA, Kauffmann F, Barhoumi M, Grigg M, Owens K, Fay M, Dobson DE, Shaik J, Beverley SM, Sacks D. 2013. The Mating Competence of Geographically Diverse Leishmania major Strains in Their Natural and Unnatural Sand Fly Vectors. PLoS Genet 9.

10. Romano A, Inbar E, Debrabant A, Charmoy M, Lawyer P, Ribeiro-Gomes F, Barhoumi M, Grigg M, Shaik J, Dobson D, Beverley SM, Sacks DL. 2014. Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. Proc Natl Acad Sci U S A 111:16808–16813.

11. Inbar E, Shaik J, Iantorno SA, Romano A, Nzelu CO, Owens K, Sanders MJ, Dobson D, Cotton JA, Grigg ME, Beverley SM, Sacks D. 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in leishmania. PLoS Genet 15.

12. WHO. 2018. Status of endemicity of leishmaniasis worldwide, 2018World Health Organisation. Geneva.

13. Karunaweera ND, Ginige S, Senanayake S, Silva H, Manamperi N, Samaranayake N, Siriwardana Y, Gamage D, Senerath U, Zhou G. 2020. Spatial epidemiologic trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerg Infect Dis 26:1–10.

14. Siriwardana HVYD, Karunanayake P, Goonerathne L, Karunaweera ND. 2017. Emergence of visceral leishmaniasis in Sri Lanka: a newly established health threat. Pathog Glob Health 111:317–326.

15. WHO. 2013. Status of endemicity of visceral and cutaneus Leishmaniasis, worldwide, 2013. Geneva.

16. Thakur L, Singh KK, Shanker V, Negi A, Jain A, Matlashewski G, Jain M. 2018. Atypical leishmaniasis: A global perspective with emphasis on the Indian subcontinent. PLoS Negl Trop Dis 12.

17. Lypaczewski P, Hoshizaki J, Zhang WW, McCall LI, Torcivia-Rodriguez J, Simonyan V, Kaur A, Dewar K, Matlashewski G. 2018. A complete Leishmania donovani reference genome identifies novel genetic variations associated with virulence. Sci Rep 8.

18. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.

19. Samarasinghe SR, Samaranayake N, Kariyawasam UL, Siriwardana YD, Imamura H, Karunaweera ND. 2018. Genomic insights into virulence mechanisms of Leishmania donovani: Evidence from an atypical strain. BMC Genomics 19.

20. Imamura H, Downing T, van den Broeck F, Sanders MJ, Rijal S, Sundar S, Mannaert A, Vanaerschot M, Berg M, de Muylder G, Dumetz F, Cuypers B, Maes I, Domagalska M, Decuypere S, Rai K, Uranw S, Bhattarai NR, Khanal B, Prajapati VK, Sharma S, Stark O, Schönian G, de Koning HP, Settimo L, Vanhollebeke B, Roy S, Ostyn B, Boelaert M, Maes L, Berriman M, Dujardin JC, Cotton JA. 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5.

21. Cuypers B, Berg M, Imamura H, Dumetz F, De Muylder G, Domagalska MA, Rijal S, Bhattarai NR, Maes I, Sanders M, Cotton JA, Meysman P, Laukens K, Dujardin JC. 2018. Integrated genomic and metabolomic profiling of ISC1, an emerging Leishmania donovani population in the Indian subcontinent. Infect Genet Evol 62:170–178.

22. Gelanew T, Kuhls K, Hurissa Z, Weldegebreal T, Hailu W, Kassahun A, Abebe T, Hailu A, Schönian G. 2010. Inference of population structure of leishmania donovani strains isolated from different ethiopian visceral leishmaniasis endemic areas. PLoS Negl Trop Dis 4.

23. Cotton JA, Durrant C, Franssen SU, Gelanew T, Hailu A, Mateus D, Sanders MJ, Berriman M, Volf P, Miles MA, Yeo M. 2020. Genomic analysis of natural intra-specific hybrids among ethiopian isolates of leishmania donovani. PLoS Negl Trop Dis 14:1–26.

24. Camacho E, González-de la Fuente S, Rastrojo A, Peiró-Pastor R, Solana JC, Tabera L, Gamarro F, Carrasco-Ramiro F, Requena JM, Aguado B. 2019. Complete assembly of the Leishmania donovani (HU3 strain) genome and transcriptome annotation. Sci Rep 9.

25. Sterkers Y, Crobu L, Lachaud L, Pagès M, Bastien P. 2014. Parasexuality and mosaic aneuploidy in Leishmania: Alternative genetics. Trends Parasitol 30:429–435.

26. Barja PP, Pescher P, Bussotti G, Dumetz F, Imamura H, Kedra D, Domagalska M, Chaumeau V, Himmelbauer H, Pages M, Sterkers Y, Dujardin JC, Notredame C, Späth GF. 2017. Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. Nat Ecol Evol 1:1961–1969.

27. Senanayake SASC, Abeyewicreme W, Dotson EM, Karunaweera ND. 2015. Characteristics of phlebotomine sand flies in selected areas of Sri Lanka. Southeast Asian J Trop Med Public Health 46:994–1004.

28. Volf P, Benkova I, Myskova J, Sadlova J, Campino L, Ravel C. 2007. Increased transmission potential of Leishmania major/Leishmania infantum hybrids. Int J Parasitol 37:589–593.

29. Siriwardana Y, Deepachandi B, Weliange SDS, Udagedara C, Wickremarathne C, Warnasuriya W, Ranawaka RR, Kahawita I, Chandrawansa PH, Karunaweera ND. 2019. First Evidence for Two Independent and Different Leishmaniasis Transmission Foci in Sri Lanka: Recent Introduction or Long-Term Existence? J Trop Med 2019.

30. Siriwardana Y, Zhou G, Deepachandi B, Akarawita J, Wickremarathne C, Warnasuriya W, Udagedara C, Ranawaka RR, Kahawita I, Ariyawansa D, Sirimanna G, Chandrawansa PH, Karunaweera ND. 2019. Trends in Recently Emerged Leishmania donovani Induced Cutaneous Leishmaniasis, Sri Lanka, for the First 13 Years. Biomed Res Int 2019.

31. Kassahun A, Sadlova J, Benda P, Kostalova T, Warburg A, Hailu A, Baneth G, Volf P, Votypka J. 2015. Natural infection of bats with Leishmania in Ethiopia. Acta Trop 150:166–170.

32. Kassahun A, Sadlova J, Dvorak V, Kostalova T, Rohousova I, Frynta D, Aghova T, Yasur-Landau D, Lemma W, Hailu A, Baneth G, Warburg A, Volf P, Votypka J. 2015. Detection of leishmania donovani and L. tropica in ethiopian wild rodents. Acta Trop 145:39–44.

33. Fernando SD, Dharmawardana P, Semege S, Epasinghe G, Senanayake N, Rodrigo C, Premaratne R. 2016. The risk of imported malaria in security forces personnel returning from overseas missions in the context of prevention of re-introduction of malaria to Sri Lanka. Malar J 15.

34. Leinonen R, Sugawara H, Shumway M. 2011. The sequence read archive. Nucleic Acids Res 39.

35. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM arXiv:1303.

36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

37. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

38. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.

39. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res 47.

40. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H. 2009. TriTrypDB: A functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38:D457-62.

41. Danecek P, McCarthy SA. 2017. BCFtools/csq: Haplotype-aware variant consequences. Bioinformatics 33:2037–2039.

42. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645.

43. Madhusoodanan N. 2019. Clustal Omega < Multiple Sequence Alignment < EMBL-EBI. Multisequence Alignment.

CHAPTER 5: DISCUSSION

5.1 The *Leishmania* reference genome: More approximation than reference

While the integration of genomics research into the *Leishmania* field has enabled ground-breaking discoveries, the use of tools originally designed for the study of human or bacterial genomes comes with its own share of limitations.

Firstly, most sequencing, genome assembly or analysis experiments and tools assume a homogenous sample. In the case of *Leishmania* however, the parasites' genomic instability which provides a rapid evolutionary pathway also results in a mosaic of unique cells even in 'clonal' populations (1–5) that cannot easily be described in a fixed state. Indeed, these parasites can readily amplify or delete genes through circularization of genetic loci or even modulate the copy number of entire chromosomes. Consequently, any genomic analysis originating from a population of cells that can only produce a single output such as one reference genome or one ploidy analysis is only able to detect consensus achieving genetic variation as the whole sample is averaged when pooled data is read. Valuable information contained in the polymorphisms can therefore easily be lost or not considered. While certain analysis pipelines are built to handle such pooled samples (6) the analysis of pooled samples can be complicated by a variety of factors and is not always able to accurately identify variants (7, 8).

As an alternative to address this issue, research is being carried out in an attempt to enable the sequencing and analysis of *Leishmania* genomes at the single cell level (9). While currently lacking in breadth of coverage compared to larger sample analysis, single cell analysis has the potential to uncover genomic variation and evolution that would otherwise be masked. The interest in single-cell genomics is not limited to the study of *Leishmania,* as other biological organisms

and systems such as cancer have been reported to undergo genetic mosaicism (10, 11). Advances in these fields have potential implications in the study of *Leishmania* genetics and should be followed closely.

Secondly, as exemplified in **Chapter 4** and other studies (12–14), *Leishmania* genomes often contain higher than *2n* chromosome copy numbers. In such cases, where some tools are limited to haploid or diploid genomic representation, careful attention or methods specifically designed for the study of non-diploid organisms such as *Leishmania* are needed (7).

5.2 The TOR pathway is a prime target for therapeutics

While the field of *Leishmania* genomics is poised for research opportunities, the contributions of previous research are already measurable. Indeed, the RagC GTPase was previously identified as a potential virulence controlling factor (15). With additional gene annotations resulting from the complete reference genome presented in **Chapter 1**, an additional polymorphism was identified in the gene encoding for Raptor; part of the TORC1 complex downstream of RagC. These concomitant mutations occurring in the same pathway support the importance of the TOR pathway in contributing to tissue tropism.

While the research presented in **Chapter 2**, was unable to elucidate the role of the mutation seen in Raptor, the R231C change in RagC yielded measurable changes in parasite pathogenesis, notably the loss of virulence in visceral organs. It is possible that the Raptor mutation had no effect in the model studied but may have implications in the broader life cycle of *Leishmania* and warrants further investigation. The phenotypic change shown in **Figure 3.3** in combination with the extensive conservation of the pathway as presented in **Table 3.1** and **Figure 3.1** points to a

central role of the TOR pathways in tissue tropism, making the pathway a target for therapeutic intervention. Indeed, soon after the identification of the expanded family of TOR proteins, several mTOR inhibitors were assessed and found to inhibit *Leishmania* and *Trypanosoma* cell proliferation (16).

Further, while the sequence conservation between human and *Leishmania* proteins is not high as shown in **Table 3.1**, the overall structural conservation of proteins appears to be greater as the *Leishmania* RagA-RagC complex illustrated in **Supplementary Figure 3.S4** is remarkably similar to the human homologue protein complex. This overall conservation, if present throughout the pathways, could allow for the repurposing of drugs created against the human TOR pathway while the minor sequence variations could allow for the generation or modification of drugs with a high specificity towards the *Leishmania* proteins.

5.3 Modulation of virulence traits through hybridization

While the mutation of key genes is certainly capable of modulating virulence traits such as the single point mutation in the *RagC* gene show in **Chapter 3**, *Leishmania* appears to prefer genome hybridization. Indeed, as seen in a cutaneous outbreak in Turkey (17) and in multiple cases in Sri Lanka as described in **Chapter 4** where three separate instances of introgression appear to have occurred, genomic hybridization appears to be linked to phenotypic variability. Additionally, laboratory crosses between parasites causing different pathologies have been shown to yield progeny with variable disease phenotype and vector competence (18–20).

Further, our own gain-of-function experiment aimed at reverting a cutaneous parasite into a visceral one as analysed in **Chapter 2** showed very little genetic divergence after several rounds

of selection. Taken together these observations suggest genomic hybridization may play a much larger role in the evolution of the disease by allowing rapid genetic changes which yield different phenotypes. Interestingly, our own unpublished preliminary genomic analysis of a novel outbreak of cutaneous leishmaniasis from Himachal in northern India results in similar findings as those presented in **Chapter 4.** The causative agent in this outbreak was identified as *L. donovani* (21) and is likely the result of hybridization. As shown in **Figure 5.1A,** all 36 chromosomes contained a high density of polymorphism consistent with heterozygosity and ploidy; diploid chromosomes centered around 50%, triploid chromosomes at a 33/66% split and tetraploid chromosomes around a 25/75% bimodal distribution. As shown in **Figure 5.1B**, this range of heterozygosity was most consistent with an interspecies hybridization event and reconstruction of parental haplotypes (**Figure 5.1C**) followed by phylogenetic analysis indicates both parents originate from the ISC1 strain cluster (22). Consistent with hybridization being a key regulator of virulence and an increase in whole genome surveillance, the occurrences of hybrid parasites linked to outbreaks is likely to increase in the future.

5.4 Hybrid parasites as silent virulence carriers: implications for eradication of VL

As CL is not as severe as VL, cutaneous outbreaks such as in Sri Lanka are not always treated as urgent health issues (23). Researchers have previously called for more effective control policies and warned against the possibility of the outbreak becoming more virulent on the island (24). The worries stem from the fact that the main parasite identified as causing CL in Sri Lanka, *L. donovani* MON-37, is not only closely related to parasite strains causing VL in neighboring India, but was also previously reported to cause VL in India, Kenya, Israel and Cyprus (24).

Since then, additional studies identified a closely related *L. donovani* isolate causing VL in Sri Lanka and confirmed the close genetic proximity of these CL and VL causing parasites (15). These same studies and the data presented in **Chapter 3** showed that minute genomic changes can have a large impact on the parasite virulence. Further, as analyzed in **Chapter 2**, a gain-of-function selection experiment to increase virulence (25) did not appear to result in major genome alterations. Taken together, these results support the possibility of phenotypic reversal following genomic changes.

While phenotypic changes are possible following gene coding mutations, the presence of hybrid parasites in Sri Lanka as described in **Chapter 4** raises the possibility of a more easily achievable phenotypic change. As discussed earlier, *Leishmania* readily modulates its chromosome copy number as an evolutionary mechanism (2, 26). Further, experimental crosses of viscerotropic and cutaneous parasites led to the observations that progeny with uneven chromosome ploidy (e.g. *3n*), followed the phenotype of the most represented parent (18).

It would therefore be possible for the hybrid parasites described in **Chapter 4** to preferentially amplify or select the genes and chromosomes of African *L. donovani* origin and revert to a visceral phenotype. Such hybrid parasites could already be contributing to some of the visceral case seen in neighboring countries. Alternatively, these hybrids could provide a latent reservoir of visceral causing parasites that could re-emerge following the relaxation of VL elimination programs or migration of the host to non-endemic areas with amenable sand fly populations (27). This possibility suggests that CL cases need to be considered as part of any country's VL elimination program and screened for the presence of hybrid parasites.

5.5 Hybridization in *Leishmania* parasites, rare or common?

As mentioned in **Chapter 1** and **Chapter 5**, several instances of outbreaks and single isolates have been reported around the globe in which evidence of hybridization both within and across species is seen. While laboratory conditions in which acute co-infections can be generated are hardly reflective of physiological conditions, co-infection of sand flies has been shown to result in hybrid formation in 3 to 40 % of flies analyzed (18, 20, 28, 29) with a possible correlation between genetic similarity and successful hybridization frequencies. In the case of *L. tropica*, the generation of hybrid parasites can even be performed *in vitro* by simply co-culturing two parental lines (30). Similar results could not be obtained using any other cultured *Leishmania* species. It has been previously observed that *L. tropica* hybridizes more frequently both in the laboratory and in nature (31). The molecular mechanism which allows *L. tropica* to form hybrids more easily is unknown but may be related one or more of the 397 genes that appear unique to *L. tropica* through orthology analysis (**Figure 5.2**). Of these genes, the majority bear no annotations or known function and could contribute to this unique phenotype. These results indicate that *Leishmania* promastigotes can readily form hybrids.

Interestingly, as shown in the **SL3** isolates analyzed in **Chapter 4**, parasites that most likely were full genomic hybrids at some point in time can appear mostly homozygous. This loss of the high heterozygosity usually associated with interspecific hybrids is likely the result of haplotype selection and changes in ploidy (2, 5, 26). Further, it has been shown that (at least experimentally) intra-clonal hybrids can exist through the generation of dual-reporter parasites originating from two clonal parents with different reporter genes (32). Without the use of reporter genes, the progeny of parasites with a similar genetic makeup could not be easily classified as genomic

hybrids. Further, as the rate at which hybrid alleles are lost such as in the **SL3** group, uncertainty remains as to how long hybrids can be confidently detected.

Taken together, these observations raise the possibility that hybridization in *Leishmania* occurs more frequently than previously thought. These hard to detect hybridization events may contribute to the maintenance of low levels of heterozygosity in the global *Leishmania* population such as seen for *L. donovani* in **Figure 4.2B**.

This possibility warrants further investigation and calls for more whole genomic analyses to be carried out. Indeed, as exemplified in **Figure 4.5B**, a large number of alleles do not remain heterozygous in hybrid parasites. Therefore, screening of parasites using single alleles might not reveal their hybrid characteristics. Further, analysis of whole genome sequencing results requires careful attention as the hybrid isolates presented in **Chapter 4**, were originally typed as non-hybrid *L. donovani* (33). Additionally, while the analysis conducted in the course of this thesis was focused on *L. donovani* virulence factors, expanding the search of hybridization signals described in **Figure 4.3** to other species in the genus may prove useful.

5.6 Convergence in Sri Lanka

Interestingly, while the additional **SL2** and **SL3** strains analyzed in **Chapter 4** were shown to be very different from the original **SL1** isolates described in **Chapter 2** (and further divergent into **SL2 A** and **SL2 B**), all these strains are part of the same geographically defined epidemic. Indeed, as described in **Chapter 4**, the **SL1** strains likely represent the natural autochthonous population due to its genetic proximity to Indian isolates and gained mutations allowing these strains to cause cutaneous disease in Sri Lanka. Conversely, the **SL2** and **SL3** groups of isolates

appear to have gained their cutaneous phenotypes through hybridization and subsequent import into Sri Lanka, possibly through the return of migrant workers or soldiers from Africa (24). As the precise dynamics of allele loss are not fully understood, the **SL3** group could belong to the same lineage as **SL2 A** and have simply undergone a longer introgression process or they could represent a unique lineage. As shown in **Figure 4.5** however, the **SL2 A** and **SL2 B** groups appear to belong to completely independent lineages. Together, these observations amount to 3 or possibly 4 independent rare genetic events and 2 to 3 parasitic import and dissemination events all converging to Sri Lanka. This epidemic appears to be the culmination of such a complex chain of events that it is hard to believe that it could happen in the absence of selective pressure such as an important animal reservoir. Indeed, as shown in **Figure 4.4**, the RagC R231C needed to modulate tissue tropism appears to be linked to some other polymorphism in the CL strain through epistasis which would have a negative impact on any parasite strain spontaneously acquiring the single mutation. Further, while hybridization of *Leishmania* is more frequent, the probability of a hybrid parasite infecting a host, compounded with the probability that the host will migrate to Sri Lanka, compounded with the probability that a local sand fly will feed on the host and become infected and further spread the hybrid parasites locally is likely an improbable event.

Due to its separation from the main continent, the island of Sri Lanka is home to a variety of unique animals such as the Golden Palm Civet (34), the toque macaque and purple-faced langur (35). Many researchers have speculated on the identity of a possible animal reservoir in Sri Lanka and the ever increasing deforestation and urban expansion has been identified as a potential driver in the epidemic as it could increase human exposure to infected sand flies (36). While a few species of rodents and dogs have been screened, the unique fauna of the island combined with the presence of hybrid parasites may result in an atypical reservoir.

5.7 Overall conclusions

In summary, by combining the benefits of second and third generation sequencing technologies, we were successful in generating a complete reference genome with no gaps for *L. donovani.* The use of this complete genome allowed us to then improve the genomic annotations which in turn lead to improved comparative genomic studies as more polymorphisms were identified between the cutaneous and visceral strains of *L. donovani* in Sri Lanka. This new genome and annotations will continue to contribute to all future genomic studies of *L. donovani.*

Secondly, we complemented the comparative genomics studies performed by characterizing the RagC GTP. We successfully demonstrated the central role for the RagC protein in enabling parasite growth and proliferation in the visceral organs through CRISPR gene editing and identified likely components of the RagC and downstream TOR pathways which represent interesting drug targets.

Finally, we identified the etiological origins of the atypical strains of *L. donovani* circulating in Sri Lanka. Through careful genomic investigation, we were able to demonstrate evidence that multiple inter-species hybridization events occurred between *L. major*, *L. tropica*, and *L. donovani*. This evidence advances the hypothesis that the virulence of *L. donovani* can be modulated through hybridization in nature and identifies the likely geographical origin from which these parasites were imported.

Taken together, this research shines a new light on the leishmaniasis epidemic in Sri Lanka and the roles of both hybrid parasites and the TOR pathway in tissue tropism. The discoveries made throughout this research have far reaching implications for the field of *Leishmania* study ranging from policy and surveillance programs to genomics.

5.8 References

1. Sterkers Y, Crobu L, Lachaud L, Pagès M, Bastien P. 2014. Parasexuality and mosaic aneuploidy in Leishmania: Alternative genetics. Trends Parasitol 30:429–435.

2. Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J, Pescher P, Vanaerschot M, Meehan CJ, Cuypers B, De Muylder G, Späth GF, Bussotti G, Vermeesch JR, Berriman M, Cotton JA, Volf P, Dujardin JC, Domagalska MA. 2017. Modulation of aneuploidy in leishmania donovani during adaptation to different in vitro and in vivo environments and its impact on gene expression. MBio 8.

3. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, De Doncker S, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Schönian G, Stark O, Sundar S, Vanaerschot M, Hertz-Fowler C, Dujardin JC, Berriman M. 2011. Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res 21:2143–2156.

4. Smith M, Bringaud F, Papadopoulou B. 2009. Organization and evolution of two SIDER retroposon subfamilies and their impact on the Leishmania genome. BMC Genomics 10.

5. Lachaud L, Bourgeois N, Kuk N, Morelle C, Crobu L, Merlin G, Bastien P, Pagès M, Sterkers Y. 2014. Constitutive mosaic aneuploidy is a unique genetic feature widespread in the Leishmania genus. Microbes Infect 16:61–66.

6. Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A. 2012. Detecting selective sweeps from pooled next-generation sequencing samples. Mol Biol Evol https://doi.org/10.1093/molbev/mss090.

7. Calarco L, Barratt J, Ellis J. 2020. Detecting sequence variants in clinically important protozoan parasites. Int J Parasitol 50:1–18.

8. Huang HW, Mullikin JC, Hansen NF. 2015. Evaluation of variant detection software for pooled next-generation sequence data. BMC Bioinformatics https://doi.org/10.1186/s12859-015-0624-y.

9. Imamura H, Monsieurs P, Jara M, Sanders M, Maes I, Vanaerschot M, Berriman M, Cotton JA, Dujardin JC, Domagalska MA. 2020. Evaluation of whole genome amplification and bioinformatic methods for the characterization of Leishmania genomes at a single cell level. Sci Rep 10.

10. Alves JM, Posada D. 2018. Sensitivity to sequencing depth in single-cell cancer genomics. Genome Med 10.

11. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: Current state of the science. Nat Rev Genet 17:175–188.

12. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin JC, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. Genome Res 21:2129–2142.

13. Zackay A, Cotton JA, Sanders M, Hailu A, Nasereddin A, Warburg A, Jaffe CL. 2018. Genome wide comparison of Ethiopian Leishmania donovani strains reveals differences potentially related to parasite survival. PLoS Genet 14.

14. Bussotti G, Gouzelou E, Boité MC, Kherachi I, Harrat Z, Eddaikra N, Mottram JC, Antoniou M, Christodoulou V, Bali A, Guerfali FZ, Laouini D, Mukhtar M, Dumetz F, Dujardin JC, Smirlis D, Lechat P, Pescher P, Hamouchi A El, Lemrani M, Chicharro C, Llanes-Acevedo IP, Botana L, Cruz I, Moreno J, Jeddi F, Aoun K, Bouratbine A, Cupolillo E, Späth GF. 2018. Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. MBio 9.

15. Zhang WW, Ramasamy G, McCall LI, Haydock A, Ranasinghe S, Abeygunasekara P, Sirimanna G, Wickremasinghe R, Myler P, Matlashewski G. 2014. Genetic Analysis of Leishmania donovani Tropism Using a Naturally Attenuated Cutaneous Strain. PLoS Pathog 10:e1004244.

16. Diaz-Gonzalez R, Kuhlmann FM, Galan-Rodriguez C, da Silva LM, Saldivia M, Karver CE, Rodriguez A, Beverley SM, Navarro M, Pollastri MP. 2011. The susceptibility of trypanosomatid pathogens to PI3/mTOR kinase inhibitors affords a new opportunity for drug repurposing. PLoS Negl Trop Dis 5.

17. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M, Svobodova M, Volf P, Berriman M, Cotton JA, Smith DF. 2014. Genomic Confirmation of Hybridisation and Recent Inbreeding in a Vector-Isolated Leishmania Population. PLoS Genet 10.

18. Romano A, Inbar E, Debrabant A, Charmoy M, Lawyer P, Ribeiro-Gomes F, Barhoumi M, Grigg M, Shaik J, Dobson D, Beverley SM, Sacks DL. 2014. Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. Proc Natl Acad Sci U S A 111:16808–16813.

19. Seblova V, Myskova J, Hlavacova J, Votypka J, Antoniou M, Volf P. 2015. Natural hybrid of Leishmania infantum/L. donovani: Development in Phlebotomus tobbi, P. perniciosus and Lutzomyia longipalpis and comparison with non-hybrid strains differing in tissue tropism. Parasites and Vectors 8.

20. Inbar E, Akopyants NS, Charmoy M, Romano A, Lawyer P, Elnaiem DEA, Kauffmann F, Barhoumi M, Grigg M, Owens K, Fay M, Dobson DE, Shaik J, Beverley SM, Sacks D. 2013. The Mating Competence of Geographically Diverse Leishmania major Strains in Their Natural and Unnatural Sand Fly Vectors. PLoS Genet 9.

21. Thakur L, Singh KK, Kushwaha HR, Sharma SK, Shankar V, Negi A, Verma G, Kumari S, Jain A, Jain M. 2020. Leishmania donovani infection with atypical cutaneous manifestations, Himachal Pradesh, India, 2014–2018. Emerg Infect Dis 26:1864–1869.

22. Seblova V, Dujardin JC, Rijal S, Domagalska MA, Volf P. 2019. ISC1, a new Leishmania donovani population emerging in the Indian sub-continent: Vector competence of Phlebotomus argentipes. Infect Genet Evol 76.

23. Karunaweera ND, Ginige S, Senanayake S, Silva H, Manamperi N, Samaranayake N,

Siriwardana Y, Gamage D, Senerath U, Zhou G. 2020. Spatial epidemiologic trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerg Infect Dis 26:1–10.

24. Karunaweera ND. 2009. Leishmania donovani causing cutaneous leishmaniasis in Sri Lanka: a wolf in sheep's clothing? Trends Parasitol 25:458–463.

25. McCall LI, Zhang WW, Dejgaard K, Atayde VD, Mazur A, Ranasinghe S, Liu J, Olivier M, Nilsson T, Matlashewski G. 2015. Adaptation of leishmania donovani to cutaneous and visceral environments: In vivo selection and proteomic analysis. J Proteome Res 14:1033–1059.

26. Barja PP, Pescher P, Bussotti G, Dumetz F, Imamura H, Kedra D, Domagalska M, Chaumeau V, Himmelbauer H, Pages M, Sterkers Y, Dujardin JC, Notredame C, Späth GF. 2017. Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. Nat Ecol Evol 1:1961–1969.

27. Gibson W, Lewis MD, Yeo M, Miles MA. 2017. Genetic Exchange in Trypanosomatids and Its Relevance to Epidemiology, p. 459–486. *In* Genetics and Evolution of Infectious Diseases: Second Edition.

28. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, Dobson DE, Beverley SM, Sacks DL. 2009. Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. Science (80-) 324:265–268.

29. Inbar E, Shaik J, Iantorno SA, Romano A, Nzelu CO, Owens K, Sanders MJ, Dobson D, Cotton JA, Grigg ME, Beverley SM, Sacks D. 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in leishmania. PLoS Genet 15.

30. Louradour I, Ferreira TR, Ghosh K, Shaik J, Sacks D. 2020. In Vitro Generation of Leishmania Hybrids. Cell Rep 31.

31. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, Berriman M, Sacks DL, Cotton JA, Grigg ME. 2017. Gene expression in Leishmania is regulated predominantly by gene dosage. MBio 8.

32. Calvo-Álvarez E, Álvarez-Velilla R, Jiménez M, Molina R, Pérez-Pertejo Y, Balaña-Fouce R, Reguera RM. 2014. First Evidence of Intraclonal Genetic Exchange in Trypanosomatids Using Two Leishmania infantum Fluorescent Transgenic Clones. PLoS Negl Trop Dis 8.

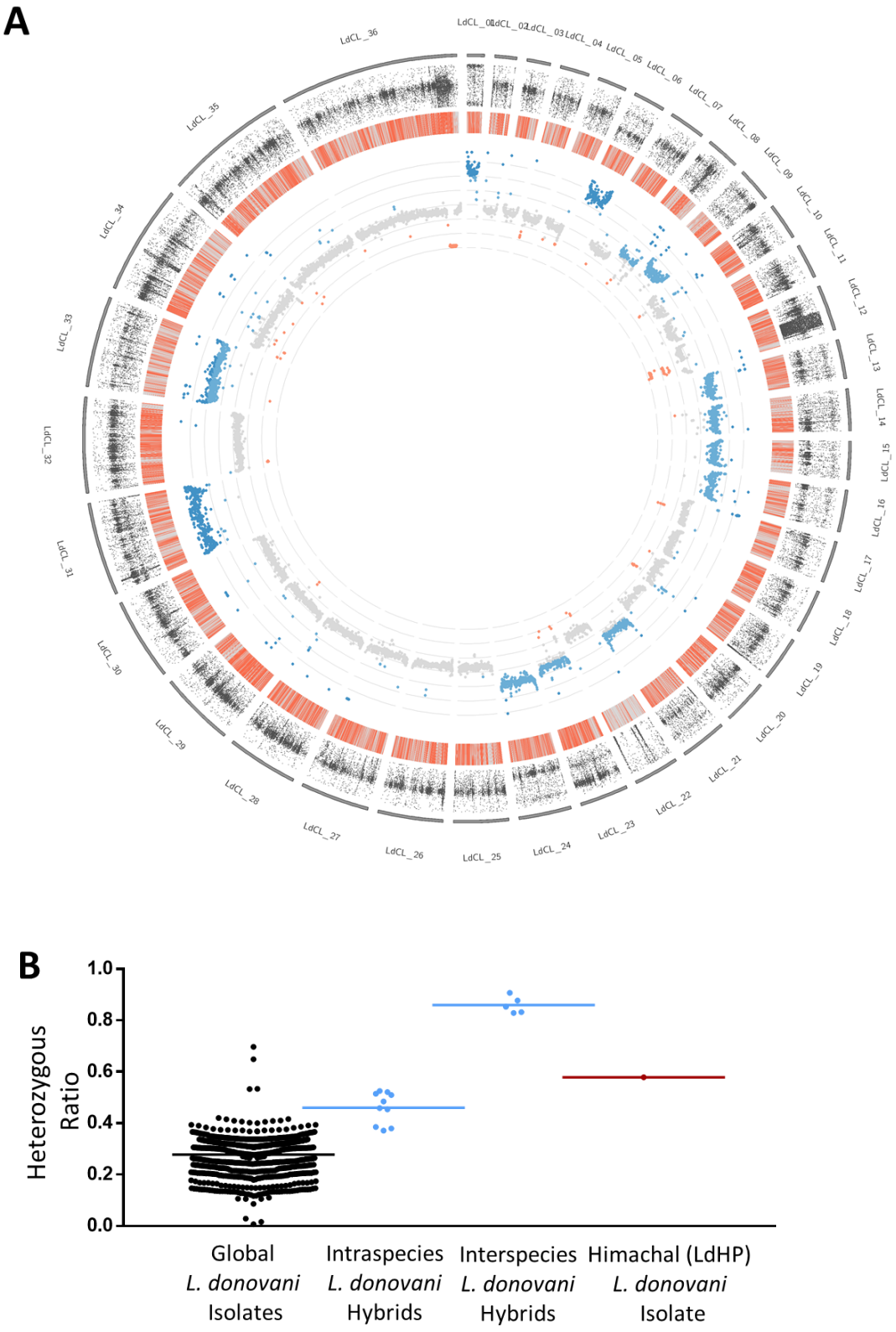
33. Samarasinghe SR, Samaranayake N, Kariyawasam UL, Siriwardana YD, Imamura H, Karunaweera ND. 2018. Genomic insights into virulence mechanisms of Leishmania donovani: Evidence from an atypical strain. BMC Genomics 19.

34. Groves CP, Rajapaksha C, Manemandra-Arachchi K. 2009. The taxonomy of the endemic golden palm civet of Sri Lanka. Zool J Linn Soc 155:238–251.

35. Nahallage CAD, Huffman MA, Kuruppu N, Weerasingha T. 2008. Diurnal Primates in Sri Lanka and People's Perception of Them. Primate Conserv 23:81–87.

36. Wijerathna T, Gunathilaka N, Gunawardana K, Rodrigo W. 2017. Potential Challenges of Controlling Leishmaniasis in Sri Lanka at a Disease Outbreak. Biomed Res Int 2017.

5.9 Figures


A
LdCL_01
LdCL_02
LdCL_03
LdCL_04
LdCL_05
LdCL_06
LdCL_07
LdCL_08
LdCL_09
LdCL_10
LdCL_11
LdCL_12
LdCL_13
LdCL_14
LdCL_15
LdCL_16
LdCL_17
LdCL_18
LdCL_19
LdCL_20
LdCL_21
LdCL_22
LdCL_23
LdCL_24
LdCL_25
LdCL_26
LdCL_27
LdCL_28
LdCL_29
LdCL_30
LdCL_31
LdCL_32
LdCL_33
LdCL_34
LdCL_35
LdCL_36
B
Heterozygous Ratio
1.0
0.8
0.6
0.4
0.2
0.0
Global L. donovani Isolates
Intraspecies L. donovani Hybrids
Interspecies L. donovani Hybrids
Himachal (LdHP) L. donovani Isolate

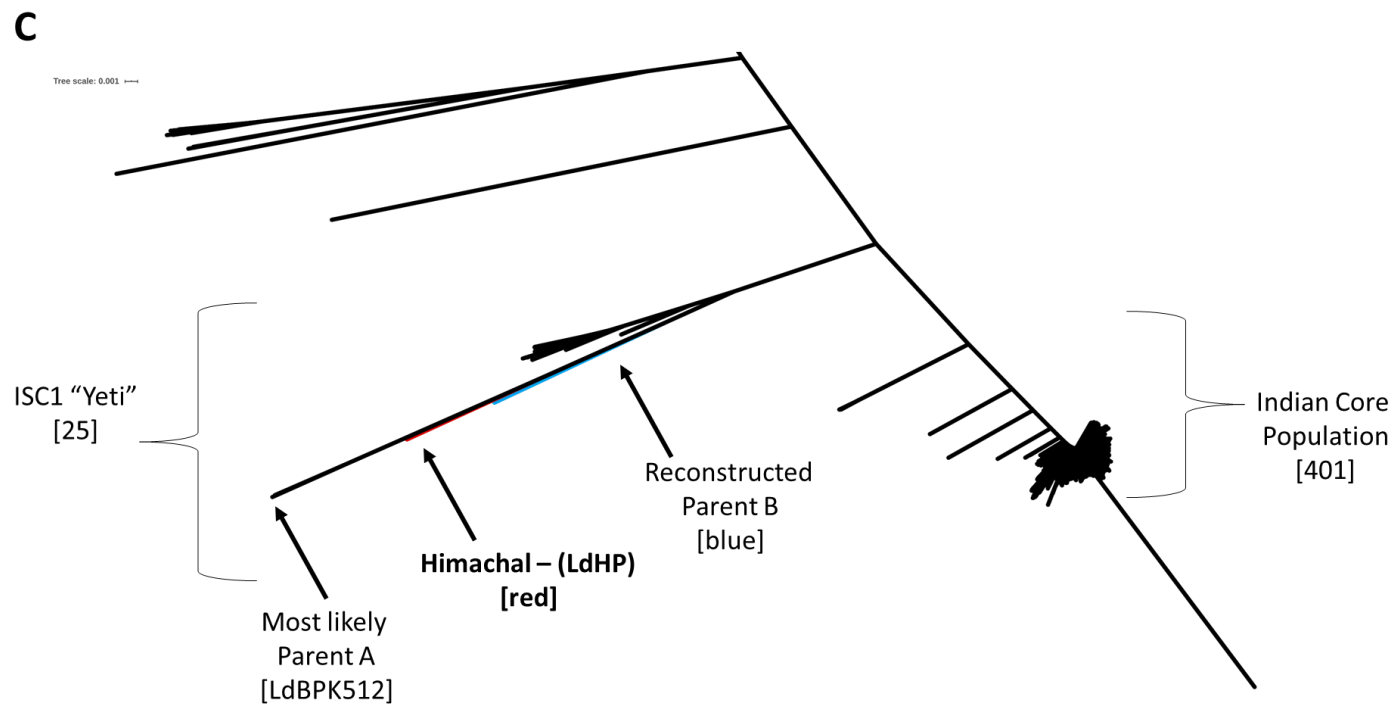


Figure 5.1 Hybrid origins of the cutaneous Himachal isolate

A. SNP density and chromosome ploidy analysis of the Himachal cutaneous isolate. Each dot represents a single SNP and indicates the location along the chromosome (x-axis) and the allele frequency (y-axis, 0-100%). SNPs determined to be heterozygous are marked with a corresponding red line along a grey chromosome bar representation. The inner track is used to plot gene coverage to determine chromosome copy number. Gene coverage consistent with diploid chromosomes is colored in grey, coverage consistent with monoploid chromosomes is colored in red, coverage consistent with triploid and tetraploid chromosomes is colored in light blue and dark blue respectively. **B.** The heterozygous polymorphism ratio of the Himachal cutaneous isolate is consistent with an intraspecies hybrid genotype. The frequency of heterozygous mutations in the cutaneous isolate was compared to those found in the worldwide population of *L. donovani,* and two baseline hybrid population frequencies: an intraspecies hybrid focus from Ethiopia and an interspecies hybrid parasite focus in Sri Lanka. **C.** The likely parents of the hybrid strain both belong the ISC1 *L. donovani* Indian subcontinent parasite group. The phylogeny analysis places the hybrid isolate (red) next to its most likely parent (LdBPK512). The reconstructed genotype for the second parent (blue) also falls with the ISC1 group.

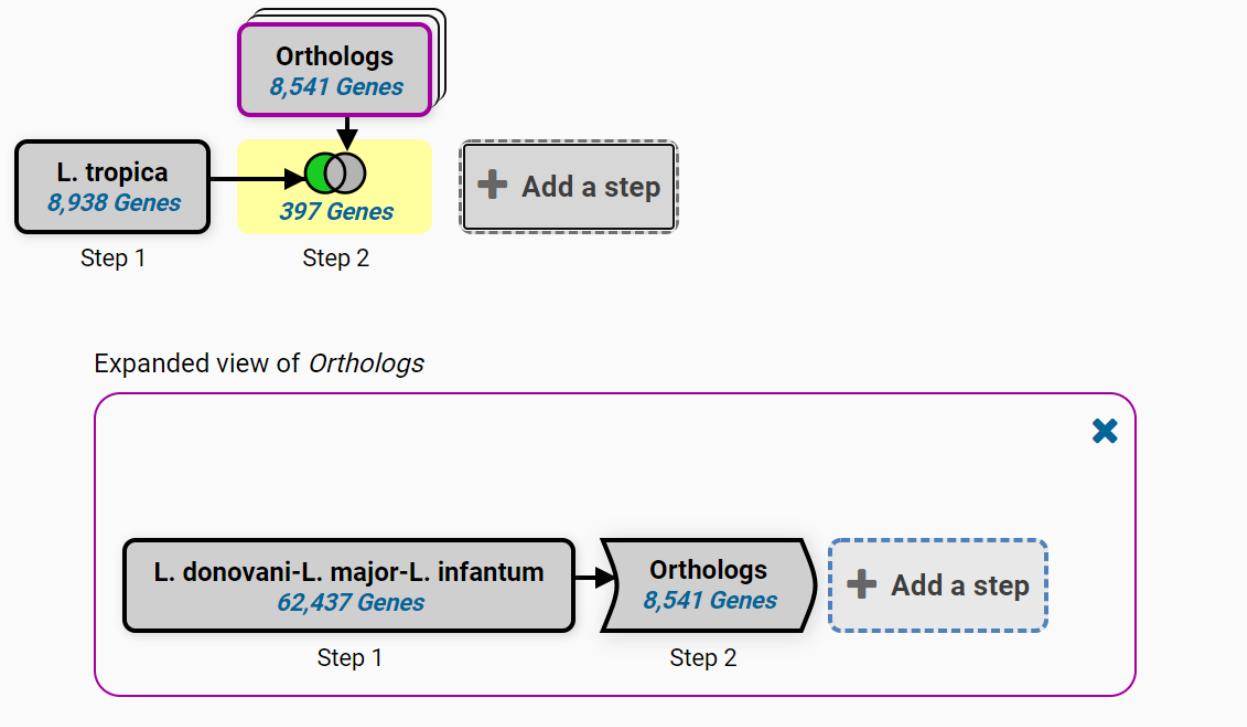


Figure 5.2 Search strategy to identify genes unique to *L. tropica*

An initial search selected all genes from TriTrypDB with the *L. tropica* taxonomy (step 1). A second set of genes containing all *L. donovani, L. major* and *L. infantum* reference strains was transformed into *L. tropica* orthologous genes (step Orthologs). The orthologous genes were then subtracted from the *L. tropica* group to identify 397 genes unique to *L. tropica* (step 2).