Complete Genome Assembly, Annotation and Comparative Analysis of Six *Leishmaniinae* Parasites

PhD Thesis

Hatim Almutairi, BSc, MSc

May 2022

I, Hatim Almutairi, confirm that the work presented in this thesis is my own and has not been submitted in substantially the same form for the award of a higher degree elsewhere. Where information has been derived from other sources, I confirm this has been indicated in the thesis.

Signed

Submitted in part fulfilment of the requirements for the degree of Doctorate of Philosophy

Abstract

Leishmania is a widespread parasite that causes leishmaniasis, a serious but neglected tropical disease reported in nearly 100 countries. Leishmaniasis manifests itself in three main forms: visceral, cutaneous, and mucocutaneous. *Leishmania* is classified into four subgenera: *Leishmania*, *Sauroleishmania*, *Viannia*, and more recently, *Mundinia*, the latter of which accommodates the *L. enriettii* complex as well as other species from a variety of hosts and geographic locations. I detail here sequencing, assembly, and annotation of six *Mundinia* genomes, including those of the Asian species *Leishmania (Mundinia) martiniquensis* and *L. (M.) orientalis*, the American species *L. (M.) enriettii* and *Porcisia hertigi* (formerly *L. hertigi*), and two unnamed African species from Ghana and Namibia, namely *L. (M.)* sp. Ghana and *L. (M.)* sp. Namibia. To maintain chromosome structure while maximising the quality of short read sequencing, genomes were sequenced and assembled using both short and long reads platforms, specifically Illumina and Oxford Nanopore Technologies. They were then annotated using *ab initio* annotation in conjunction with publicly available proteins and transcripts. Each genome contains a complete set of 36 chromosomes and measures between 32.2-35.9 Mega-bases in length with an average N50 of 1,062,685 bases. Each assembly contained an average of 8,126 genes, mRNAs, exons, and protein coding regions. When compared to other *Leishmania* genomes, all were recognisably related to *Mundinia* species, except for *Porcisia hertigi*, which was found to be more closely related to *Endotrypanum monterogeii*, setting it as an outgroup. Phylogenomic analyses revealed that *Mundinia* genomes share a common ancestor with the other three *Leishmania* subgenera, which was estimated to have existed approximately 121 million years ago, during the early Cretaceous period. Selection pressure analysis showed that there are 36 positively selected proteins, four of which may be novel proteins. This work may pave the way for future research on *Leishmania* biology and evolution.

Publications

The following publications have been generated while developing this thesis, and to an extent has guided the thesis into what it has become:

1. Almutairi, H., Urbaniak, M.D., Bates, M.D., Jariyapan, N., Al-Salem, W.S., Dillon, R.J., Bates, P.A. and Gatherer, D., 2021. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) martiniquensis*, Isolate LSCM1, Strain LV760. Microbiology Resource Announcements, 10(24).
2. Almutairi, H., Urbaniak, M., Bates, M., Jariyapan, N., Al-Salem, W., Dillon, R., Bates, P. and Gatherer, D., 2021. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania* (*Mundinia*) *orientalis*, Isolate LSCM4, Strain LV768. Microbiology Resource Announcements, 10(36).
3. Almutairi, H., Urbaniak, M., Bates, M., Thomaz-Soccol, V., Al-Salem, W., Dillon, R., Bates, P. and Gatherer, D., 2021. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania* (*Mundinia*) *enriettii*, Isolate CUR178, Strain LV763. Microbiology Resource Announcements, 10(36).
4. Almutairi, H., Urbaniak, M.D., Bates, M.D., Kwakye-Nuako, G., Al-Salem, W.S., Dillon, R.J., Bates, P.A. and Gatherer, D., 2021. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia)* sp. Ghana, Isolate GH5, Strain LV757. Microbiology Resource Announcements, 10(39).
5. Almutairi, H., Urbaniak, M., Bates, M., Al-Salem, W., Dillon, R., Bates, P. and Gatherer, D., 2021. Chromosome-Scale Assembly of the Complete Genome Sequence of *Porcisia hertigi*, Isolate C119, Strain LV43. Microbiology Resource Announcements, 10(41).
6. Almutairi, H., Urbaniak, M.D., Bates, M.D., Jariyapan, N., Kwakye-Nuako, G., Thomaz-Soccol, V., Al-Salem, W.S., Dillon, R.J., Bates, P.A. and Gatherer, D., 2021. LGAAP: *Leishmaniinae* Genome Assembly and Annotation Pipeline. Microbiology Resource Announcements, 10(29).
7. Almutairi, H., Urbaniak, M., Bates, M., Jariyapan, N., Kwakye-Nuako, G., Thomaz Soccol, V., Al-Salem, W., Dillon, R., Bates, P. and Gatherer, D., 2021. Chromosome-scale genome sequencing, assembly and annotation of six genomes from subfamily *Leishmaniinae*. Scientific Data, 8(1).

List of Abbreviations and Acronyms

BI	Bayesian Inference
CL	Cutaneous Leishmaniasis
DM	Distance Matrix
FDA	Food and Drug Administration
IRS	Indoor Residual Spraying
JD	Jaccard’s Distance
MCL	Mucocutaneous Leishmaniasis
ME	Minimum Evolution
ML	Maximum Likelihood
MP	Maximum Parsimony
NJ	Neighbour Joining
NTDs	Neglected Tropical Disease
SSU rRNA	Small Subunit ribosomal Ribonucleic Acid
TMRCA	Time to Most Recent Common Ancestor
UPGMA	Unweighted Pair Group Method with Arithmetic mean
VL	Visceral Leishmaniasis
WHO	World Health Organization
WMP	Weighted Maximum Parsimony
WP	Weighted Parsimony
L.	*Leishmania genus*
(M.)	*Mundinia* subgenus
(L.)	*Leishmania* subgenus
(S.)	*Sauroleishmania* subgenus
(V.)	*Viannia* subgenus
P.	*Porcisia* species

Table Of Contents

List of Figures

(All Figures are own work, except where specified)

List of Tables

Acknowledgements

First and foremost, I am extremely grateful to my supervisors, Dr Derek Gatherer and Dr Rod Dillon for their invaluable advice, continuous support, and patience during my PhD study. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I would also like to thank Prof. Paul Bates and Dr Mick Urbaniak, for their technical and scientific support on my study. I would like to thank Michelle Bates for helping with wet lab preparations such as parasite culture and DNA extraction for sequencing. I am also deeply grateful to my friend and former boss, Waleed Al Salem, for believing in me and facilitating this amazing opportunity. Finally, I would like to express my gratitude to my mom Sadia, my dad Hamid, my wife Norah, and my children, Mohammed, Jory and Faris. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Chapter 1. Introduction

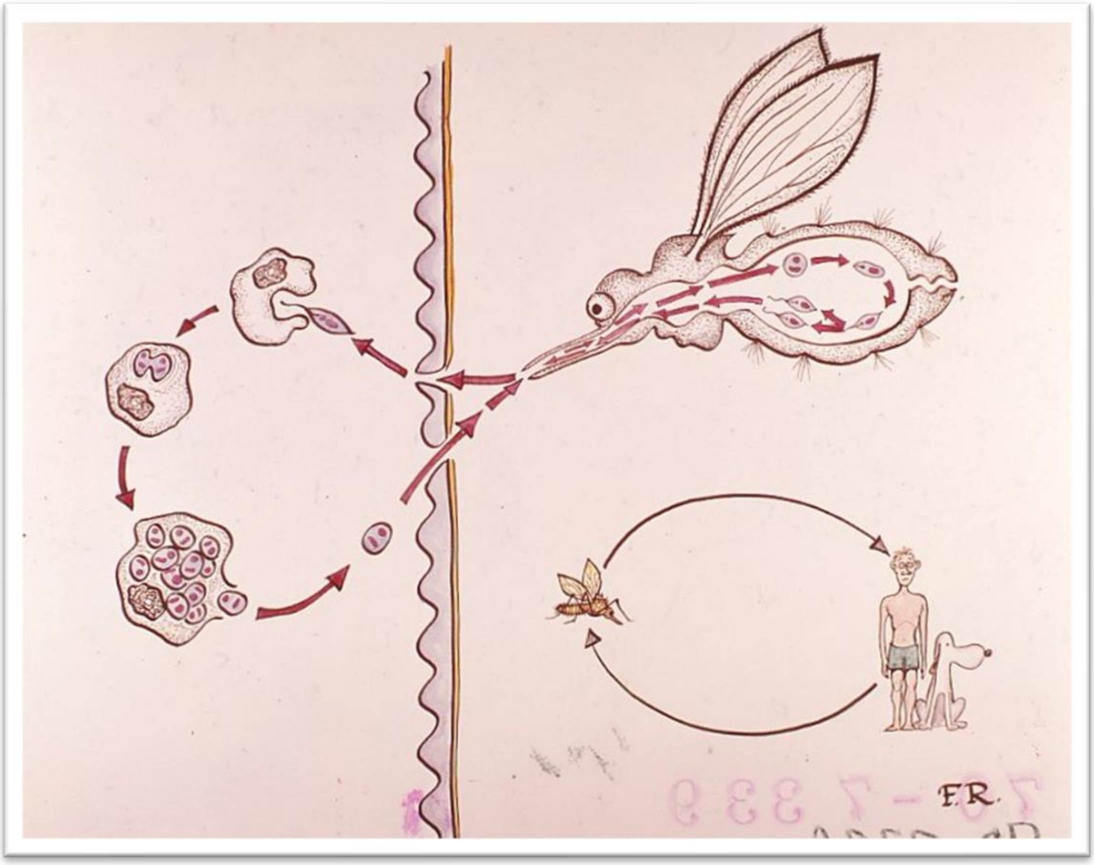


Figure 1.1: Vintage illustration of *Leishmania* life cycle from the National Museum of Health and Medicine, Otis Historical Archives.

Leishmania is a widely distributed genus of parasites that infect a wide variety of hosts, including humans. It is the causative agent of leishmaniasis, can be found on almost every continent, and is one of the Neglected Tropical Disease (NTDs). Leishmaniasis can be manifested mainly in three forms: Visceral Leishmaniasis, the deadliest one if not treated, Cutaneous Leishmaniasis, the most common form, and Mucocutaneous Leishmaniasis, the most stigmatising form. Nearly 100 countries reported cases of leishmaniasis in 2017 (Alvar et al., 2012, Burza et al., 2018).

Previously, a new classification for *Leishmania* was proposed based on integrated molecular data. This classification splits *Leishmania* species into two major evolutionary branches named *Euleishmania* and *Paraleishmania* (Cupolillo et al., 2000). *Paraleishmania* originally comprised the *L. hertigi, L. deanei, L. herreri, L. equatorensis,* and *L. colombiensis,* as well as the formerly known *Endotrypanum* genus, whereas *Euleishmania* is comprised of four subgenera: *Leishmania*, *Viannia*, *Sauroleishmania*, and recently *Mundinia* which previously referred to as the *L. enriettii* complex (Espinosa et al., 2018, Akhoundi et al., 2016). However, there has been ongoing research about the taxonomic classification of *Leishmania Mundinia* subgenus based on the current body of evidence (Akhoundi et al., 2016). Most of that evidence is based on several highly conserved genes and proteins that have been used to estimate Time to Most Recent Common Ancestor (TMRCA) (Noyes, 1998b) or phylogeny reconstruction (Jariyapan et al., 2018b, Butenko et al., 2019b). Therefore, previous studies were either too general, too specific, or too exclusive.

Inferring *Mundinia* taxonomy or reconstructing phylogeny can be challenging based on the current body of evidence and the lack of sufficient high-quality genomes. As a result, assembling *Mundinia* genomes in full entirety could facilitate the creation of more precise taxonomy and phylogeny. Therefore, the goal of this thesis is to *de novo* assemble and annotate six genomes that have never been assembled. Five of them belong to the *Leishmania Mundinia* subgenus and one which was originally classified as *Leishmania* but has been reassigned as *Porcisia hertigi* (Espinosa et al., 2018). The second goal is to compare, using phylogenomic and comparative techniques, the newly assembled ones to other *Leishmania* genomes as well as proteomes with the *Trypanosomatidae* family.

The phylogeny presented in this thesis came from the consilience of two different approaches used to construct phylogeny: genomic phylogeny and proteomic orthology approaches. This thesis presents all evidence produced during the project, much of which has not been published in the papers written during the project (see publications section). As a result, it is regarded as a full description of work of three years on the genomics of *Leishmania.*

First, I shall go through the current state of the *Leishmania* parasite and Leishmaniasis in general followed by a review of *Mundinia* subgenus specifically. Following that, I shall review the literature on *Leishmania* genomics. After that, I shall go over the materials and methods used, including everything from setting up the dry lab or virtual machine to evaluating and selecting the best assembly algorithm, polishing the draft assembly, and identifying chimeric scaffolds, running annotation using evidence and model organisms, and concluding with phylogenomic and comparative analyses. The final chapter shall be a general discussion of the outcomes.

Chapter 2. Literature Review

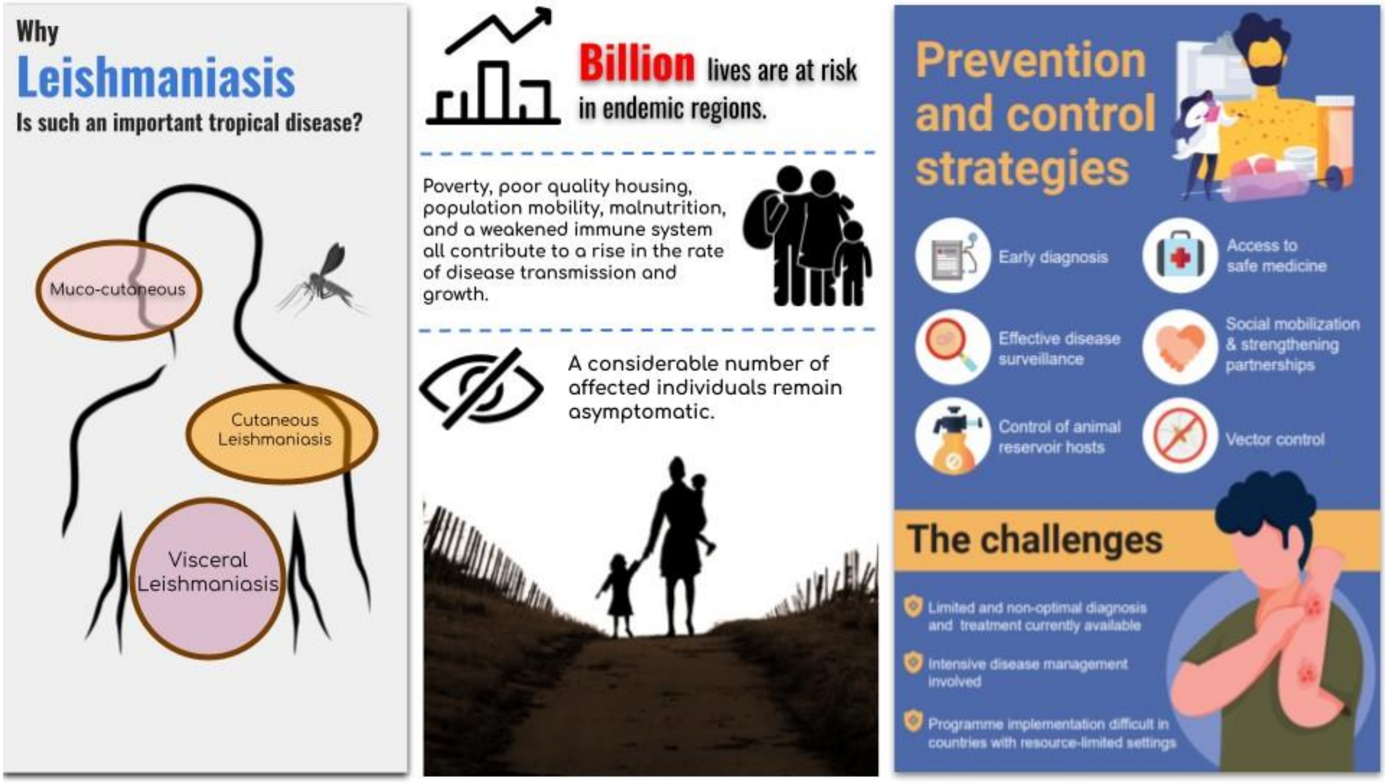

Why
Leishmaniasis
Is such an important tropical disease?
Muco-cutaneous
Cutaneous Leishmaniasis
Visceral Leishmaniasis
Billion lives are at risk in endemic regions.
Poverty, poor quality housing, population mobility, malnutrition, and a weakened immune system all contribute to a rise in the rate of disease transmission and growth.
A considerable number of affected individuals remain asymptomatic.
Prevention and control strategies
Early diagnosis
Access to safe medicine
Effective disease surveillance
Social mobilization & strengthening partnerships
Control of animal reservoir hosts
Vector control
The challenges
Limited and non-optimal diagnosis and treatment currently available
Intensive disease management involved
Programme implementation difficult in countries with resource-limited settings

Figure 2.1: Infographic summary on the disease Leishmaniasis (inspired from WHO Leishmaniasis factsheet).

2.1 The Extent and Impact of Leishmaniasis

Leishmaniasis is a very important tropical disease (Figure 2.1). It has the highest single cause of disease burden among all NTDs (Hotez et al., 2014). The WHO defines Leishmaniasis as a major global health issue with a wide range of clinical symptoms and a potentially lethal outcome (Andrade-Narváez et al., 2001). Nearly 100 countries reported infection of leishmaniasis in 2017 (Burza et al., 2018). It can be found on nearly every continent, with endemic regions in north-eastern Africa, Southern Europe, the Middle East, south-eastern Mexico, and Central and South America (Reithinger et al., 2007). Sandflies, mostly of the genera *Phlebotomus* and *Lutzomyia*, spread the disease in humans (Kashif et al., 2017, Reithinger et al., 2007).

Leishmania parasite has a dixenous life cycle, which means the life cycle is completed in two hosts: mammalian and insect (Maslov et al., 2019). The vector insect is a sandfly belonging to the *Phlebotomus* genus in the case of Old World *Leishmania* and the *Lutzomyia* genus in the case of New World *Leishmania*. During host-vector transition, the parasite undergoes morphological differentiations, mainly amastigotes and promastigotes. The cycle begins with metacyclogenesis — which transforms the parasite from non-pathogenic to pathogenic — which occurs between 7 and 10 days after the sandfly becomes infected with amastigotes taken from a blood meal from an infected host (Sasidharan and Saudagar, 2021). Once inside the sandfly, the parasite undergoes the first differentiation, transforming into procyclic promastigotes, which are flagellated and motile forms of the parasite with a slender body measuring 15-20 µm long and 1.5-3.5 µm wide and with a flagellar measuring approximately 15-28 µm long that assists the parasite in attaching to the sandfly's gut. Then, the procyclic forms divide in the abdominal midgut of the sandfly, resulting in the formation of non-dividing nectomonad forms that migrate from the abdominal to the anterior midgut and then transform into leptomonad promastigotes. The leptomonads then differentiate into metacyclic promastigotes and migrate to the insect's mouth part or proboscis, where they are ready to transmit to a mammalian host via a sandfly bite (Gossage et al., 2003). The bite introduces metacyclic promastigotes into the host along with the sandfly's saliva, and these promastigotes initiate phagocytic activity by adhering to the plasma membrane of macrophages. The promastigotes enter the macrophage unnoticed by the immune system and produce a Parasitophorous Vacuole structure (PV) (Lodge and Descoteaux, 2008), which protects the parasite from the host cell's phagolysosomes, before differentiating into ovoid amastigotes which measuring 2-4 µm in diameter (Herwaldt, 1999). Amastigotes thrive and proliferate inside the PV, surviving both the gut acidity and the macrophages' acidic environment. The macrophage then ruptures, releasing all the mature amastigotes and initiating a chain of infection that ultimately results in one of the clinical manifestations, at which point the amastigotes are ready to be taken in a blood meal by the sandfly, and the life cycle begins again.

Clinical symptoms include a wide range of manifestations with varying degrees of severity, depending on the *Leishmania* species involved and the host's immune response. It can be seen mainly as one of three forms; Visceral Leishmaniasis, which is the most lethal one; Cutaneous Leishmaniasis, which is the most common one; and Mucocutaneous Leishmaniasis, which leads to destruction of tissues that have mucocutaneous structures.

2.1.1 Visceral Leishmaniasis

VL can be manifested clinically as repeated fevers, enlarged organs, anaemia, and considerable weight loss, however some confounding signs may emerge during serology testing because of the resemblance between anti-leishmanial and auto-immune antibodies (Harhay et al., 2011, Burza et al., 2018). However, there is a correlation between high parasite burden and acute malnutrition, especially in young children, but it is unknown whether if this impact is a result of or a cause of malnutrition (Zacarias et al., 2017). Darkening of the skin is most likely a result of increased adrenocorticotropic hormone production triggered by cytokines in cases, hence the Hindi term kala-azar, which loosely translates as black fever (Elkhair, 2014).

VL can be clinically related with HIV co-infection, making its management challenging. HIV was responsible for the recurrence of VL in southern Europe in the late 1990s (Transmissíveis/AIDS, 2004). There are 6-18% VL cases with HIV co-infection rate in endemic areas like Brazil, India and Ethiopia (Yimer et al., 2014). HIV and VL share an immune-pathological route involving macrophages and dendritic cells because of increasing cellular replication, resulting in accelerated disease development. As a result, several countries, such as India, recommend HIV testing for VL patients, as it is recommended to be mandatory in all endemic areas (Mock et al., 2012). Regrettably, half of co-infected patients in India were unaware of their HIV status (Burza et al., 2014), as parasites isolated from the gastrointestinal mucosa, respiratory tract, and liver may easily be found in infrequent disseminated leishmaniasis (Ejara et al., 2010). Additionally, coinfection might result in unusual CL presentations, as was found during a *L. major* related outbreak in Burkina Faso (Guiguemdé et al., 2003).

Seven countries of Brazil, Ethiopia, India, Kenya, Somalia, South Sudan, and Sudan reported more than 90% of all VL cases in 2015. Nonetheless, VL continues to be endemic in over 60 countries (WHO, 2018). Moreover, In east Africa and the Indian subcontinent, devastating epidemics have been reported (Kohn, 2007, Ibrahim, 2002). However, In the last decade, the global incidence of VL has declined significantly, from between 200 - 400 thousand new cases in 2012 to between 50 - 90 thousand new cases in 2017 (Alvar et al., 2012).

Historically, half of the global burden of VL was shared by Asian countries, primarily India, Nepal, and Bangladesh (Burza et al., 2018). They agreed in 2005 to remove VL off the list of public health concerns by 2015, at which point they decided to extend it (WHO, 2005). They did, however, achieve some targets, such as lowering instances to less than one per ten thousand people per year at both the district and sub-district levels. As a result, VL is no longer considered a public health risk in these countries (Burza et al., 2018). VL burden is stable in Africa, with shorter cyclical patterns of 6 - 10 years, with previous epidemics typically linked to conflict-induced forced migration of non-immune individuals into endemic areas (Al-Salem et al., 2016). HIV co-infection may have contributed to increased transmission in Ethiopia with up to 40% of VL hospital patients testing positive for HIV in 2006 (van Griensven et al., 2014). In Latin America, VL cases are also constant, with distribution shifting south-westward with Brazil accounting for the majority of cases (Burza et al., 2018).

Treatment with pentavalent antimonials in two formulations, sodium stibogluconate and meglumine antimoniate, was used for decades to treat VL (Aronson et al., 2017). However, dose recommendations have been adjusted in some countries which reported decreased response, such as India (Sundar and Chakravarty, 2010). Furthermore, the current medication has side effects, including being unpleasant when given intramuscularly for about 28 days, as well as being cardiotoxic and causing arrhythmias (Sundar et al., 2000, Ritmeijer et al., 2001, WHO, 2010). Now, because of resistance, sodium stibogluconate is no longer prescribed in the Indian subcontinent, instead liposomal amphotericin B (LAMB) is the treatment of choice. The FDA and WHO have authorised AmBisome as the sole LAMB formulation for the treatment of VL (Balasegaram et al., 2012). The treatment of choice in east Africa is 20 mg/kg sodium stibogluconate each day, followed by 15 mg/kg intramuscular paromomycin over 17 days (WHO, 2010, Kimutai et al., 2017).

2.1.2 Cutaneous and Mucocutaneous Leishmaniasis

Unlike VL, CL does not cause death, but it can cause significant cosmetic morbidity, social stigmatisation, and psychological repercussions (Yanik et al., 2004, Bennis et al., 2017). CL presents a challenge due to its under-representation as a dynamic disease (Razavinasab et al., 2019). Since CL clinical manifestation may go undiagnosed or unnoticed, it is quite difficult to count all CL cases; consequently, the disease burden may be underestimated when prevalence and Disability-Adjusted Life Years (DALYs) are estimated. For instance, there have been some attempts to broaden the spectrum of CL by dividing it into active and dormant CL (Bailey et al., 2017). Inactive CL cases are frequently excluded from prevalence estimates, which adds to the difficulty of determining the true burden of disease (Gijón-Robles et al., 2018). If inactive cases are factored into the calculations, the estimated burden of CL disease rises by a factor of ten (Bailey et al., 2017). Additionally, because of chronic and sustained conflicts, particularly in the Middle East, updated CL statistics indicate that this disease poses a significant global health and social challenge in endemic countries (Razavinasab et al., 2019).

The lesions, while itchy at times, do not cause the pain that one might expect given their appearance. Multiple lesions are usually associated with different bites, but lymphatic spread is conceivable. Lesions then might grow into nodules, which slowly ulcerate over the next few months (Thomaidou et al., 2015).

CL ulcers are usually painless and heal on their own unless secondary infections occur. However, several factors influence the rate of spontaneous healing, including parasite load and virulence, host immune response, lesion location, and the presence or absence of secondary bacterial infection (Ziaie and Sadeghian, 2008). When scars from *Leishmania* infection ulcerate, they become vulnerable to pathogenic bacteria and yeast colonisation, which can lead to secondary infections (Yehia et al., 2017). *Aerococcus viridans* has been found in the environment and isolated from human skin on occasion (Ruoff, 1995, Kerbaugh and Evans, 1968). Furthermore, *A. viridans*, which makes up 5–10% of the bacterial flora in the air and dust of occupied rooms, was identified in a previous report on the infection of immunocompromised mice (Dagnæs-Hansen et al., 2004, Çetin et al., 2007).

The parasites causing CL are generally categorised geographically into two groups: Old World and New World species. Old World CL species such as *L. major*, *L. tropica*, and *L. aethiopica* are abundant in the Mediterranean Basin, the Middle East, the Horn of Africa, and the Indian subcontinent. New World CL species such as *L amazonensis*, *L mexicana*, *L. braziliensis*, and *L. guyanensis* are endemic in central and South America (de Vries et al., 2015). In Old World CL, lesions may proceed to hyperkeratotic lesions. For example, lesions caused by *L. tropica* and *L. major* heal in a year but leave permanent scars, but lesions caused by *L. aethiopica* take years to heal and can proceed to MCL (Thomaidou et al., 2015). On the other hand, lesions caused by New World CL, such as *L. mexicana*, are often milder and heal faster, whereas ulcerating lesions and MCL are associated with species from the subgenus *Viannia* (Burza et al., 2018).

MCL, on the other hand, is a potentially fatal and very disfiguring condition caused by the late stage loss of the tissues and cartilage around nose and mouth, which can occasionally spread to the larynx and may results in aspiration pneumonia (Burza et al., 2018). Moreover, MCL can cause a severe immunopathological reaction, which can be in a form of severe lesions of the nasal septum, lips, and palate. These lesions usually begin in the nostrils or on the lips, and there is often a history of increased nasal congestion, epistaxis, or discharge (Cincurá et al., 2017). MCL is thought to be more frequent in immunocompromised individuals, and about 90% of the cases exhibit a scar from a previous CL episode, which might have occurred decades earlier (Weina et al., 2004, Hodiamont et al., 2014, Cincurá et al., 2017). As a result, MCL is potentially fatal, can result in severe deformity, and must be recognised and treated promptly (WHO, 2005).

MCL usually manifests itself after a CL infection (Goto and Lauletta Lindoso, 2012). Lesions usually manifest themselves within two years of cutaneous infection, but may take up to 30 years (Samady et al., 1996). Infections can spread through haematogenous or lymphatic routes. Although *L. braziliensis* is responsible for the majority of MCL cases, *L. panamensis, L. guyanensis,* and *L. amazonensis* have also been reported (Strazzulla et al., 2013). Therefore, prevention is critical for disease control. The number of cases of CL with subsequent mucosal involvement is approximately 3-5% in endemic areas but can reach 20% or more in some areas (David et al., 1993).

According to the WHO, 94 % cases occurred in seven countries in 2018: Brazil, India, Kenya, Somalia, South Sudan, Ethiopia, and Sudan. The disease is endemic in 88 nations, 72 of which are developing countries (Desjeux, 2004). VL is more widespread in India, Nepal, Sudan, Brazil, and Bangladesh, whereas CL is concentrated in Afghanistan, Iran, Syria, Brazil, Saudi Arabia, and Peru, accounting for around 90% of the disease. MCL accounts for 90% of cases in Bolivia, Peru, and Brazil (Banuls et al., 2007). The incidence of CL patients has increased significantly as a result of conflict-induced forced migration, with imported cases being common in non-endemic countries (Pavli and Maltezou, 2010, Wall et al., 2012).

CL can either be anthroponotic, as is the case with *L. tropica*, or zoonotic, as is the case with *L. major*, *L. aethiopica*, as well as most New World species (Reithinger et al., 2007). CL is transmitted by sandflies of the species *Phlebotomus* in the Old World or *Lutzomyia* in the New World (Lainson and Shaw, 1987). Transmission by other alternative pathways is unusual. However, skin contact with an active lesion is not infectious, as infection requires the transmission of material from open sores (Araujo et al., 2016).

The Leishmaniasis effects on health is disproportionally distributed by type of disease, gender, and age. Leishmaniasis, for example, is notable for its socioeconomic effects, in which several economic variables serve as proxy for a number of significant global risk factors, such as housing type, malnutrition, livelihood patterns, labour migration, and resource conflicts (Boelaert et al., 2009, Grifferty et al., 2021). A systematic analysis of the socioeconomic risk factors of VL and CL in 2020 indicated that insufficient housing and lack of sanitation contribute to the prevalence of leishmaniasis in impoverished communities (Valero and Uriarte, 2020). Particularly, certain domestic construction materials create ideal circumstances for sandflies, which rest and breed in cracks and crevices in walls and floors. Inadequate sanitation attracts both wild and domestic animals and provides breeding sites for sandflies.

CL is associated with poor housing conditions and the presence of peri-domestic animals (Yadon et al., 2003, Negera et al., 2008). However, closeness to forested regions, sleeping in shelters in crop fields, and domestic animals are also risk factor in the New World species (Davies et al., 1997, Reithinger et al., 2003, de Araújo Pedrosa and de Alencar Ximenes, 2009).

It is worth mentioning that numerous studies have been conducted on domestic animals; traditionally, dogs are considered the primary animal reservoir, while cats and horses have been discovered infected with the parasite in multiple studies (Cardoso et al., 2021). Nonetheless, dogs were found to have a similar or even lower prevalence than wildlife during certain human outbreaks, most likely due to preventative measures that were implemented (Quaresma et al., 2011, Miró et al., 2017). To control future outbreaks and monitor the endemicity of certain regions, it is crucial to investigate the role of wildlife in the infectivity and potential transmission of the parasite.

Leishmania infections have been studied the most in rodents, under both natural and experimental conditions. The presence of *L. braziliens*is and other zoonotic species of the subgenus *Viannia* has been reported in 27 species, including *Rattus rattus, Cerradomys subflavus, Necromys lasiurus, Nectomys squamipe*s, and *Mus musculus*, with the latter being the most frequently investigated species (Azami-Conesa et al., 2021).

In the Americas, CL epidemiology is complex, with numerous *Leishmania* species circulating in the same geographic area, multiple reservoir hosts and sandfly vectors, and varying clinical symptoms and medication responses. Although CL is the most common manifestation, up to 10% of patients infected develop MCL (Burza et al., 2018).

2.1.3 Impact of Leishmaniasis

Leishmaniasis is a disease associated with poverty and conflict, which make its control difficult due to a variety of factors (Alvar et al., 2006, Molyneux et al., 2017). The first factor is biological; leishmaniasis is known to exist in a variety of animal reservoirs, many of which are remote and inaccessible. Along with the expansion of established types of leishmaniasis, recent reports indicate the introduction of new disease manifestations, new vectors, or new potential reservoirs, all of which pose a barrier to the disease's elimination in the near future (Cameron et al., 2016). Second, climate change has been attributed to the introduction of new cases of autochthonous leishmaniasis in countries such as Germany and Austria, as well as outbreaks in new Latin American foci (Dujardin et al., 2008, Carvalho et al., 2015, Obwaller et al., 2016, Seva et al., 2017). The third one is politically driven as is the case in the Middle East, such as in Syria and South Sudan, outbreaks and epidemics of CL and VL have been reported as a result of massive population displacements caused by naive populations being exposed to infected vectors or infected individuals coming in contact with susceptible vectors (Al-Salem et al., 2016, Du et al., 2016b). The last one is economical, leishmaniasis is unlikely to receive the level of funding or stability associated with diseases afflicting wealthy nations, such as cancer, diabetes, or HIV.

In conflict-affected areas of the Middle East, CL from the Old World has resurfaced (Salam et al., 2014), most recently in Syria, as a result of the public health system collapsing and non-immune populations being exposed (Du et al., 2016a). CL incidence was nearly 25,000 in the early 2010s, with the actual situation estimated to be 2–5 times higher than reported numbers (Hayani et al., 2015, mondiale de la Santé and Organization, 2016, Hotez, 2018). Between 2000 and 2012, Lebanon reported six CL incidences, increasing to 1033 in 2013, with 97% of cases occurring among Syrian immigrants (Alawieh et al., 2014). Similar patterns have been reported in Turkey, and an increase in CL cases is projected in the Mediterranean region, where the vector *Phlebotomus sergenti* is prevalent (Khamesipour and Rath, 2016, Koltas et al., 2014).

CL treatment is motivated by a desire to promote cure, reduce scarring, and lower the risk of dissemination or further progression to MCL (Burza et al., 2018). But, the majority of CL lesions will self-heal within two to eighteen months (Pearson and de Queiroz Sousa, 1996, Scott and Novais, 2016). However, immediate therapy should be initiated if there are several lesions, large individual lesions, a duration greater than six months, or if the lesions are located in a sensitive area such as the face or joints (Harhay et al., 2011, Elkhair, 2014).

2.1.4 Prevention and Control

There is currently no approved vaccination to prevent human leishmaniasis. Numerous potential vaccines including a variety of antigens are under pre-clinical development, with several of them still in clinical trials (Moafi et al., 2019). The majority of patients who recover from leishmaniasis are immune to further infection, as demonstrated by the practise of leishmanization, which involves the intradermal inoculation of live *Leishmania* parasites to cause skin lesions in order to provide protective immunity against re-infection following natural healing (Pacheco-Fernandez et al., 2021).

Due to the fact that individuals with untreated leishmaniasis serve as parasite reservoirs, early case detection and management are critical control strategies (Burza et al., 2018). Many countries continue to rely on passive rather than active case detection, implying that many cases will persist within communities for extended periods, particularly when leishmaniasis awareness is low (Burza et al., 2018). However, recent research has shown how critical treatment is in delaying transmission, as these individuals are extremely infectious (Medley et al., 2015). Although mathematical modelling suggested that asymptomatic carriers of *L. donovani* could help maintain VL transmission, actual evidence and current observational research emphasise the critical role of clinical cases in transmission (Das et al., 2016). This demonstrates the critical nature of early detection and treatment for public health reasons, in addition to the clinical benefit to the individuals (Burza et al., 2018).

Control of leishmaniasis is based on three factors: treatment, animal vaccination, and vector control. In Asia, vector control tactics include indoor residual spraying, the use of long-lasting insecticidal nets, and environmental management (Burza et al., 2018). Indoor residual spraying is the primary intervention in the Indian subcontinent's endeavour to eradicate VL (Burza et al., 2014). However, recent reports highlight the emergence of resistance to dichlorodiphenyltrichloroethane (DDT), and India has recently turned to synthetic pyrethroids, which are also used in Bangladesh and Nepal (Coleman et al., 2015). While a long-lasting insecticidal bed net provides some protection against sandfly bites, its efficacy in reducing the incidence of VL at the community level is unknown (Burza et al., 2018). The lack of effect was attributed to transmission occurring outside, near cattle sheds.

In east Africa, it is assumed that the vector mainly bites outdoors and there is no evidence for the effectiveness of insecticide spraying. However, evidence indicates that a long-lasting insecticidal bed net protects approximately 60% of people against VL in south Sudan (Ritmeijer et al., 2007).

Reservoir control plays an important role especially in the case of *L. infantum* and *L. major*, which was used in Brazil and former USSR, respectively (Burza et al., 2018). However, Brazil's attempts to eliminate canine leishmaniasis has been criticized as being inefficient (Quinnell and Courtenay, 2009). Currently, three vaccines have been approved for dogs, two of which (Leishmune and CaniLeish) provide some protection in natural settings. In other countries, such as Iran, giving dogs collars treated with deltamethrin were found to protect children from infection when administered systemically to all dogs (Gavgani et al., 2002).

In 2012, the World Health Organization released an NTDs eradication roadmap, committing to eradicate VL from the Indian subcontinent by 2020 as well as to detect and treat at least 70% of all CL cases in the eastern Mediterranean region (WHO, 2010). The elimination campaign for VL in South Asia, which began in 2005, appears to have had some effect, as evidenced by a steady decline in case numbers. Nevertheless, political and donor attention in VL is likely to die down once the goal is met, there is no guarantee of long-term impact in a disease with cyclical transmission patterns and shifting focal points (Muniaraj, 2014, Adaui et al., 2016). Thus, it is anticipated that research into technical feasibility and treatment safety would be volatile, much more so once these goals are attained, regardless of the other factors indicated above (Burza et al., 2018).

The VL elimination initiative in Asia and the 2012 London declaration on NTDs have improved global awareness of leishmaniasis and significantly increased funding for control (WHO, 2012).

However, limited treatment options, insufficient diagnostic tools, and low community awareness, especially for CL, persist despite this increased focus.

2.1.5 *Leishmania* Evolved to Survive Host Immunity

With multiple hosts, multiple insect vectors, and numerous *Leishmania* species, it is fair to presume that varying host immune responses have played a role in the genus' evolution and diversification. However, there is surprisingly little molecular evidence for this. For instance, various host cells phagocytize the parasites during infection, such as neutrophils, monocytes, monocyte-derived dendritic cells, macrophages, and stromal cells. However, the parasite clearance versus persistence may vary between *Leishmania* species (Kaye and Scott, 2011). Furthermore, the parasite has a variety of methods for manipulating macrophage function, including subverting phagosome biogenesis and maturation control. However, the role of parasite virulence factors, such as lipophosphoglycan, also varies between *Leishmania* species and the host cell type.

Major Surface Protease (MSP), also known as GP63, is one of the most studied mechanisms in *Leishmania* (Castro Neto et al., 2019). By cleaving phosphotyrosine phosphatases like SRC homology 2 domain phosphotyrosine receptor phosphate, GP63 plays an important role in regulating intracellular survival in some host cells (SHP1). After crossing lipid microdomains in the host cell membrane, MSP can access these cytosolic targets (Charmoy et al., 2010). Moreover, Iron is found to be essential for the intracellular survival of Leishmania, and both the host and parasite transporters compete for it (Blackwell et al., 2001, Jacques et al., 2010).

Different models of disease illustrate different aspects of cell-mediated immunity to *Leishmania* infection, such as the importance of CD8+ T cells, in addition to the involvement of monocyte-derived dendritic cells (Ives et al., 2011). Moreover, The presence of interleukin-10 influences both the parasite persistence and the ability to induce good vaccine-induced immunity (Rub et al., 2009). However, T helper 1 (TH1) cells, regulatory T (TReg) cells, B cells, macrophages, and DCs are all sources of interleukin-10, but it is unclear whether they all have the same functional significance (Kaye and Scott, 2011).

To conclude, while much of the research and funding has focused on prevention and control, there is a critical need for more research and funding into parasite and vector biology and transmission. It is critical to address these gaps in our understanding of the parasite's biology and evolution, particularly regarding host interaction and immune system evasion. As new species of the genus *Leishmania* are discovered, we are only beginning to gain an understanding of the real burden of parasitism caused by the parasite throughout the animal kingdom. This can be accomplished by providing additional molecular evidence by utilising the most recent sequencing technologies.

2.2 Theories about the Origins of the Genus *Leishmania*

The genus *Leishmania* is part of the family *Trypanosomatidae*, which are obligatory flagellates that can infect a variety of Insects, leeches, vertebrates, and plants; and they may either have a single host (monoxenous species) or two hosts (dixenous species) during their lifecycle (Maslov et al., 2013, Maslov et al., 2019). The majority of dixenous parasites are included in the genera *Endotrypanum*, *Leishmania*, *Paraleishmania*, *Phytomonas*, and *Trypanosoma*; some of them are medically and economically important (Bruschi and Gradoni, 2018). Most dixenous *Trypanosomatids* are generally believed to have derived from their monoxenous ancestors. Monoxenous genera include *Borovskyia, Crithidia, Leptomonas, Lotmaria, Novymonas,* and *Zelonia*; and dixenous genera include *Endotrypanum, Leishmania,* and *Paraleishmania*, which are classified as members of the subfamily *Leishmaniinae* (Kostygov and Yurchenko, 2017).

Leishmania's origins date all the way back to ancient times. Evidence of fossilised insects found in Burmese amber recorded the presence *Leishmania*-like parasite. The first *Leishmania*-like fossil was discovered in the proboscis and alimentary tract of an extinct sandfly *Palaeomyia burmitis* preserved in 100 million year old Burmese amber (Poinar, 2004, 2004a, 2004b). *P. proterus*, a *Leishmania*-like parasite, was described in a new collective fossil genus *Paleoleishmania* (Poinar and Poinar, 2004b). Furthermore, amastigotes were discovered suggesting that the sandfly acquired the parasite by feeding on the blood of a vertebrate, which indicate that *P. proterus* had a dixenous lifecycle. Following that, the blood cells were classified as reptile derived (Poinar and Poinar, 2004a).

The second *Leishmania*-like fossil, *Paleoleishmania neotropicum*, was discovered in a 20–30 million-year-old Dominican amber in the extinct sand fly *Lutzomyia adiketis* (Poinar, 2004). Promastigotes, paramastigotes, and amastigotes were discovered in the sandfly's gut and proboscis, but no vertebrate blood cells were discovered. Nonetheless, the presence of amastigotes and the absence of monoxenous flagellates in sand flies indicate that *P. neotropicum* had a dixenous life cycle with a vertebrate host. Additionally, this fossil record demonstrates that Neotropical sandflies served as vectors for *Leishmania*-like parasites during the late Oligocene to early Miocene. This evidence supported an early hypothesis about the *Neotropical* origin of *Leishmania* and the evolution of the *Leishmania/Endotrypanum* clade (Noyes, 1998b).

The genus *Leishmania* is divided into four subgenera: *Leishmania*, *Mundinia*, *Sauroleishmania*, and *Viannia*. All subgenera have been extensively investigated except *Mundinia* subgenus. *Mundinia* was newly described to include members formerly classified as part of the *L. enriettii* complex (Espinosa et al., 2018). Mun (Muniz) and din (Medina) inspired the subgenus naming to honour the researchers who discovered this parasite (Paranaiba et al., 2018).

The first recorded case of *Mundinia* was a cutaneous-like leishmaniasis case in 1917 on Martinique. Figure 2.2 depicts a timeline of case reports categorized by continent. However, There was no official description of the causative agent until 2001 (Noyes et al., 2002, Desbois et al., 2014).

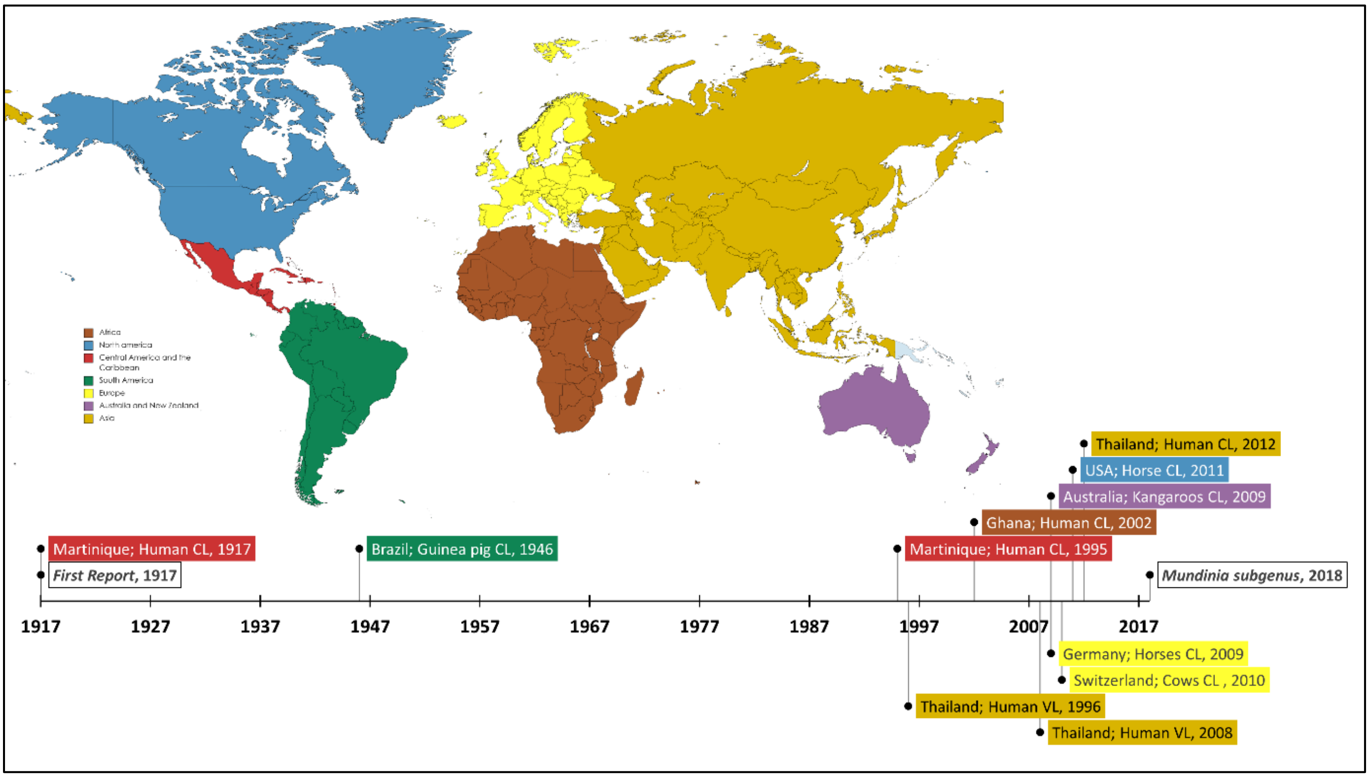


Figure 2.2: Chronological and geographic timeline distribution of the *Leishmania* (*Mundinia*) species (Supplementary Materials for full-scale figure).

Before the official description of *Mundinia*, several CL and VL cases have been reported in animals and humans in geographically sparse areas as a result of non-typical *Leishmania* parasites (Sereno, 2019). Between 2002 and 2003, Ghana reported a total of 8876 possible CL cases, with *L. major* and another *Leishmania* species being identified (Kwakye-Nuako et al., 2015). CL cases affected by *L. major* have been reported in west Africa. A new species was also described in 2008 as a potential causative agent of VL cases in Thailand (Sukmee et al., 2008).

Unexpected CL cases have also been detected in animals, such as horse, cow, and red kangaroo, in north America, Europe, and Australia respectively (Rose et al., 2004, Lobsiger et al., 2010, Reuss et al., 2012). These cases, however, were genetically similar to the newly discovered *L. martiniquensis* in Thailand, which has been associated to a substantial number of CL and VL cases in Thailand since 1999 (Pothirat et al., 2014a). Remarkably, all of these parasites were also genetically related to *L. enriettii*, a parasite previously isolated from a guinea pig in Brazil in 1946 (Pothirat et al., 2014a, Espinosa et al., 2018). In 2018, the subgenus *Mundinia* was formally described (Reuss et al., 2012, Espinosa et al., 2018). It now contains a number of

Leishmania species that cause disease in humans and animals around the world, including *L. martiniquensis* and *L. orientalis*, the most recently described species responsible for CL in Thailand (Jariyapan et al., 2018a, Espinosa et al., 2018).

2.3 The Importance of an Accurate Taxonomy

Taxonomy uses hierarchical grouping to facilitate knowledge of life. Ignoring proper taxonomy means ignoring not only rigorous scientific tradition, but also the similarities and distinctions between living things; it also means ignoring the evolutionary aspects of classification and opting for disorder over order (Bennett and Balick, 2014). Therefore, it is critical to provide the most precise taxonomic classification, particularly in the case of *Leishmania*, to increase our understanding of basic biology and evolution, as well as better treatment, vaccine, and control methods.

The pre-molecular classification scheme for *Leishmania* relied on a limited number of diagnostic characteristics and was mostly based on crude cell morphology and life cycle distinctive features, such as monoxenous vs dixenous mode, as well as host specificity (Hoare and Wallace, 1966, Vickerman and Preston, 1976). Today, performing phylogenetic analyses on nucleotide sequences containing thousands of informative characteristics is a common practise in evolutionary studies (d'Avila-Levy et al., 2015). The first study that used molecular evidence to reconstruct a phylogenetic tree for the purpose of reclassifying *Leishmania* was published in 2004 (Moreira et al., 2004). Since then, many molecular phylogeny studies have enhanced the taxonomy of the genus *Leishmania* (Espinosa et al., 2018). Due to the high conservation of sequences like Small SubUnit ribosomal ribonucleic acid (SSU rRNA), the subfamily *Leishmaniinae* was assigned to a group known as the ’slow evolving’ *Trypanosomatidae* in 2012 (Jirku et al., 2012). This subfamily included mainly dixenous parasites of wild animals that may infect humans accidentally, resulting in diseases collectively referred to as leishmaniasis.

The evidence that parasites other than *Leishmania* species can cause leishmaniasis has been met with scepticism. A variety of genetic evidence for both *Trypanosoma* and *Leishmania* has been revealed through novel discoveries generated from wild animals. Recent studies have shown new degrees of human and animal pathogenetic population diversity, as well as putative *Leishmania* reservoirs (Cupolillo et al., 1998, Asato et al., 2009b, Seblova et al., 2015).

Constructing the phylogenetic tree based on the most conserved gene set is a well-established method for classifying *Leishmania*. However, this approach has limitations due to the amount and length of evidence used, as it may not be an accurate inference, or it may be difficult to estimate the Time to the Most Recent Common Ancestor (TMRCA). This field has room for improvement, as this approach demonstrates how relying on a single set of genes can be slightly inaccurate and occasionally misleading. As the name implies, phylogenomics requires a complete set of genomes as input for reconstructing the tree and inferring taxonomic relationships between species. As a result, additional molecular material and evidence are required for reconstructing a higher-definition tree and, consequently, a more accurate taxonomy.

2.4 Characteristics of *Leishmania* Genomes

Leishmania, as the case with all *Trypanosomatidae* organisms, has a genome structure that is unique among eukaryotes in that it lacks introns and has smaller chromosomes packed with more genes (Kazemi, 2011). The haploid *Leishmania* genome consist of around 32 million base pairs and arranged into 36 chromosomes (Ivens et al., 2005). Each genome typically has approximately 8000 known protein-coding genes, approximately 900 RNA genes, and approximately 40 pseudogenes (Peacock et al., 2007).

McDonagh et al. reported in 2000, during the initial assembly of the *L. major* Friedlin genome, that chromosome 1 has around 79 protein-coding genes, two converted polycistronic clusters of genes, and mRNA transcription is directed to the telomeres (Myler et al., 1999, Myler et al., 2000, McDonagh et al., 2000, Myler et al., 2001). Moreover, genes are structured on one or both DNA strands and are transcribed from unknown promoters as polycistronic transcripts.

Specific expression of *Leishmania* gene products can be categorised into a variety of biological pathways, including structural proteins, transporters, metabolism, amastins, heat shock proteins, and surface proteins (Saxena et al., 2007). However, neither promastigotes nor amastigotes appeared to have a single unified mechanism for surviving in a variety of hosts and environments (Cohen-Freue et al., 2007). Protein expression occurs after replication and during the translation process. However expressions require eukaryotic RNA polymerase II for regulation even though they involve chromatin modification, which makes it different from other regulatory mechanisms in eukaryotes (Ivens et al., 2005).

Leishmania genomes are distinct in comparison to those of other *Trypanosomatidae* species. Synteny findings suggests that the structure of *Leishmania* chromosomes lacks long sub-telomeric regions, which typically carry species-specific genes. This observation was made in one of the seminal studies in the field of *Leishmania* research, in which the authors announced the first assembly of the complete genomes of *L. infantum* and *L. braziliensis*, as well as conducting synteny comparisons with *L. major*, *T. brucei*, and *T. cruzi* (Peacock et al., 2007).

The *Leishmania* genome is divided into two parts: the nucleus, which contains chromosomal DNA, and the kinetoplasts, which contain self-replicating DNA molecules. Additionally, the cytoplasm contains virus-like particles (Molyneux, 1974, Croft and Molyneux, 1979, Tarr et al., 1988). Typically, chromosomes were studied and separated using pulsed-field gel electrophoresis (PFGE), whereas kinetoplasts were separated using ultra-centrifugation. However, the differences between sexual and asexual forms of *Leishmania*, as well as the number of copies of each gene on each chromosome, have not been fully investigated (Lighthall and Giannini, 1992).

The majority of *Leishmania* genomes contain 36 chromosomes. However, earlier research has suggested some *Leishmania* species have evolved with fewer chromosomes because of fission or fusion events. For instance, one paper found that *L. mexicana* has some linkage groups between chromosomes 8 and 29, as well as 30 and 36, suggesting that it has only 34 chromosomes; *L. braziliensis* also has linkage groups between chromosomes 20 and 34, suggesting that it has 35 chromosomes (Britto et al., 1998).

However, this finding was drawn from a study in which the chromosome structure was determined by utilising around 300 loci and PFGE (Levick et al., 1996, Wincker et al., 1996a). They do, however, imply that the rearrangements occurred during the evolution of the genus *Leishmania*, even though the majority of *Leishmania* genomes exhibit a high degree of synteny (Ravel et al., 1995, Myler et al., 1999).

Nonetheless, the publication of these genome assemblies contributed significantly to our understanding about *Leishmania* genomics. Therefore, I shall move on to review how *Leishmania* genomes were assembled and published, as well as why the level of assembly may be critical in defining biological features.

2.5 *Leishmania* Genomes in the Public Domain

Conducting high-confidence genomics, transcriptomics, or proteomics studies requires first and foremost well-assembled genomes. Additionally, there is a high demand for research on *Leishmania* gene expression, pathogenicity, and drug susceptibility. As a result, genome assemblies play a critical role in our fundamental understanding of parasitic infection (Camacho et al., 2019).

Occasionally, the justifications for genome assembly are insufficient, either because the process is laborious and takes an excessive amount of time, or because the genome being assembled has no zoonotic or medical significance (Blake, 2015). However, research is increasingly focusing on more sophisticated areas such as population biology, vaccine development, and molecular diagnostics. Additionally, genomic resources support these areas of development, particularly now that sequencing technology is more feasible and affordable than ever. Therefore, sampling, sequencing, and assembling additional *Leishmania* species increase experimental power and enables a better understanding of the *Leishmania* populations (Cantacessi et al., 2015).

The first *Leishmania* genome to be assembled was announced in 2005 (Ivens et al., 2005). They made the first milestone in genome assembly by sequencing the genome of Friedlin strain of *L. major*. They managed to assemble 32.8 Mbp in 36 chromosomes by Sanger sequencing. The second milestone was when the reference genome of *L. infantum* and *L. braziliensis* were announced two years later (Peacock et al., 2007). They sequenced both parasites by shotgun sequencing and produced five and six-fold of coverage respectively. In 2011, both *L. donovani*, from 16 isolated VL patients from Nepal, and *L. mexicana* were assembled (Rogers et al., 2011, Downing et al., 2011). A year later, the genome of lizard parasite *L. tarentolae* was announced in which it preserves high synteny among other compared *Leishmania* genomes (Raymond et al., 2012).

Since then, more than 58 genomes are available and only 22 of them are set to be reference genomes in the National Centre for Biotechnology Information Assembly database (Kitts et al., 2016). Table 2.1 lists in details all available genomes prior to this project in chronological order. It shows that assembly level is critical because it determines the accuracy of the genome's features and thus facilitates comparative studies by providing a common reference point.

Table 2.1: Chronological list of all publicly representative *Leishmania* genomes. Assembly levels are classified into four categories; complete genomes, which indicate a complete set of chromosomes with no additional unplaced sequences; chromosome level, which is similar to complete genomes level but contains some additional sequences that are not assigned to a particular chromosome; scaffold level assembly, which involves the combining of many contigs to produce a bigger sequence but not a whole set of chromosomes; and finally, contig level assembly, which is the simplest sort of assembly because the assembled contigs have no resemblance to any chromosome.

Date	Organism	Strain	Submitter	Assembly level	Accession
2011	***L. major***	Friedlin	Friedlin Consortium	**Complete Genome**	GCA_000002725.2
2011	***L. braziliensis***	MHOM/BR/75/M2904	The Sanger Institute	**Chromosome**	GCA_000002845.2
2011	***L. infantum***	JPCM5	The Sanger Institute	**Chromosome**	GCA_000002875.2
2011	***L. mexicana***	MHOM/GT/2001/U1103	Wellcome Trust Sanger Institute	**Chromosome**	GCA_000234665.4
2011	***L. donovani***	BPK282A1	Wellcome Trust Sanger Institute	**Chromosome**	GCA_000227135.2
2013	***L. turanica***	LEM423	Kinetoplastid Genomes Consortium	**Scaffold**	GCA_000441995.1
2013	***L. gerbilli***	LEM452	Kinetoplastid Genomes Consortium	**Scaffold**	GCA_000443025.1
2014	***L. panamensis***	MHOM/PA/94/PSC-1	INDICASAT-AIP	**Chromosome**	GCA_000755165.1
2015	***L. sp.***	AIIMS/LM/SS/PKDL/LD-974	All India Institute of Medical Science	**Contig**	GCA_000981925.2
2015	***L. peruviana***	LEM1537 V1	UFMG	**Chromosome**	GCA_001403695.1
2016	***L. sp. MAR***	LEM2494	Washington University School of Medicine	**Chromosome**	GCA_000409445.2
2016	***L. arabica***	LEM1108	Washington University School of Medicine	**Chromosome**	GCA_000410695.2
2018	***L. guyanensis***	204-365	CDC	**Contig**	GCA_003664525.1
2018	***L. lainsoni***	216-34	CDC	**Contig**	GCA_003664395.1
2019	***L. aethiopica***	209-622	CDC	**Contig**	GCA_003992445.1
2019	***L. amazonensis***	UA301	GIMUR	**Complete Genome**	GCA_005317125.1
2019	***L. adleri***	HO174	CBMSO	**Contig**	GCA_902369305.1
2019	***L. tarentolae***	Parrot Tar II	University of Tokyo	**Contig**	GCA_009731335.1
2020	***L. tropica***	CDC216-162	CDC	**Chromosome**	GCA_014139745.1
2020	***L. infantum chagasi***	MCER/BR/1981/M6445/Salvaterra	Instituto Evandro Chagas	**Chromosome**	GCA_014466975.1

Contigs – short for contiguous sequences – are the building blocks of the genome assembly process. A contig is a group of DNA fragments that have been linked together based on their overlapping similarities. As a result, contig assembly can be accomplished by stitching together multiple sequence fragments to form a single long or a few longer contigs. However, the difficulty can sometimes be found in the repeat sequence contigs, which can overlap with multiple contigs, complicating the assembly process. Thereby, the two most critical variables affecting the completeness of the genome assembly are sequencing length and quality.

The first generation of sequencing, Sanger, could produce long sequences of thousands of bases. However, it is limited by the fact that it can only sequence a limited number of short reads per run, making it labour intensive, particularly for large genomes. Therefore, it was insufficient. Then came the second generation of sequencing, primarily Illumina, which had overcome the previous generation's throughput limitations, resulting in millions of sequence-reads but at the cost of shorter read lengths (Mardis, 2008).

That is why genome sequencing will always require a large number of sequences reads, referred to as short reads. However, the difficulty rises significantly when the genome contains many repeat regions, as has been observed in *Leishmania* genomes and was one of the primary reasons for the delay in assembling the first reference genome (McKean et al., 1997, Murray et al., 2005, Karsani, 2006, Rogers et al., 2011).

The third generation of sequencing then entered the market with the goal of bridging that gap through the provision of long read sequencing. There are currently two platforms for long read sequencing: Pacific Biosciences and Oxford Nanopore Technologies (van Dijk et al., 2018). They do, however, make a trade-off between length and base calling accuracy. Using second-generation sequencing technology, it is possible to achieve a base calling with PHRED quality score (Q score) of 40 and an error rate of 1 in 10,000 (Ewing et al., 1998). However, third-generation sequencers can only reach a maximum Q score of 20 and an error rate of 1 in 100 bases. Nonetheless, third-generation sequencing technology has achieved sequencing lengths that were previously unattainable.

Although *Leishmania* genomes have been assembled at a variety of levels as shown in Table 2.1, the most accurate are those that closely resemble the complete set of 36 chromosomes (Wincker et al., 1996b). However, several metrics govern the assembly accuracy of the *Leishmania* genomes.

A good *Leishmania* genome must exhibit the following characteristics:

1. It must be assembled with the fewest possible contigs or scaffolds that match the chromosomes count.
2. It must have the lowest number of gaps and shortest gaps possible. A gap is a sequence of "Ns" referred to as an ambiguous base that is used to connect contigs in order to construct a longer scaffold. (Currie, 1995).
3. It must have a N50 value of around 1 Mb. This value is species-specific, as it may only apply to genomes with a similar size and structure to those of *Leishmania*.

This means that approximately half of the genome's sequence is covered by contigs larger than or equal to the size of the N50 value. In other words, the sum of all contigs with a length of N50 or greater contains at least 50% of the total genome sequence (Miller et al., 2010).

Table 2.2 summarises these metrics for only chromosome-scale genomes that are available prior to our assemblies. The genomes of *Leishmania* have been assembled using nearly all three generations of sequencing. However, no *Leishmania* genome has ever been sequenced using Oxford Nanopore Technologies before the commencement of this work. Nonetheless, assemblies produced highly accurate genomes when sequenced with both long and short reads, such as *L. donovani* (LdCL strain) and *L. tropica* (CDC 216-162 strain). Because chromosome-based assemblies are the closest thing to a truly complete genome, they can enable scientists to conduct more accurate comparative studies.

It is worth noting, however, that we can never be certain that a genome is truly complete and accurate, and that in the wild, polymorphism at the base, repeat or even indel level will always exist even within species. However, end-to-end mapping onto a complete chromosomal scaffold is a strong indicator that the current generation of technology is approaching the optimal effort-to-result ratio.

Table 2.2: Summary of the metrics for all *Leishmania* representative assemblies. Accession numbers are added as hyperlinks to the organism's names.

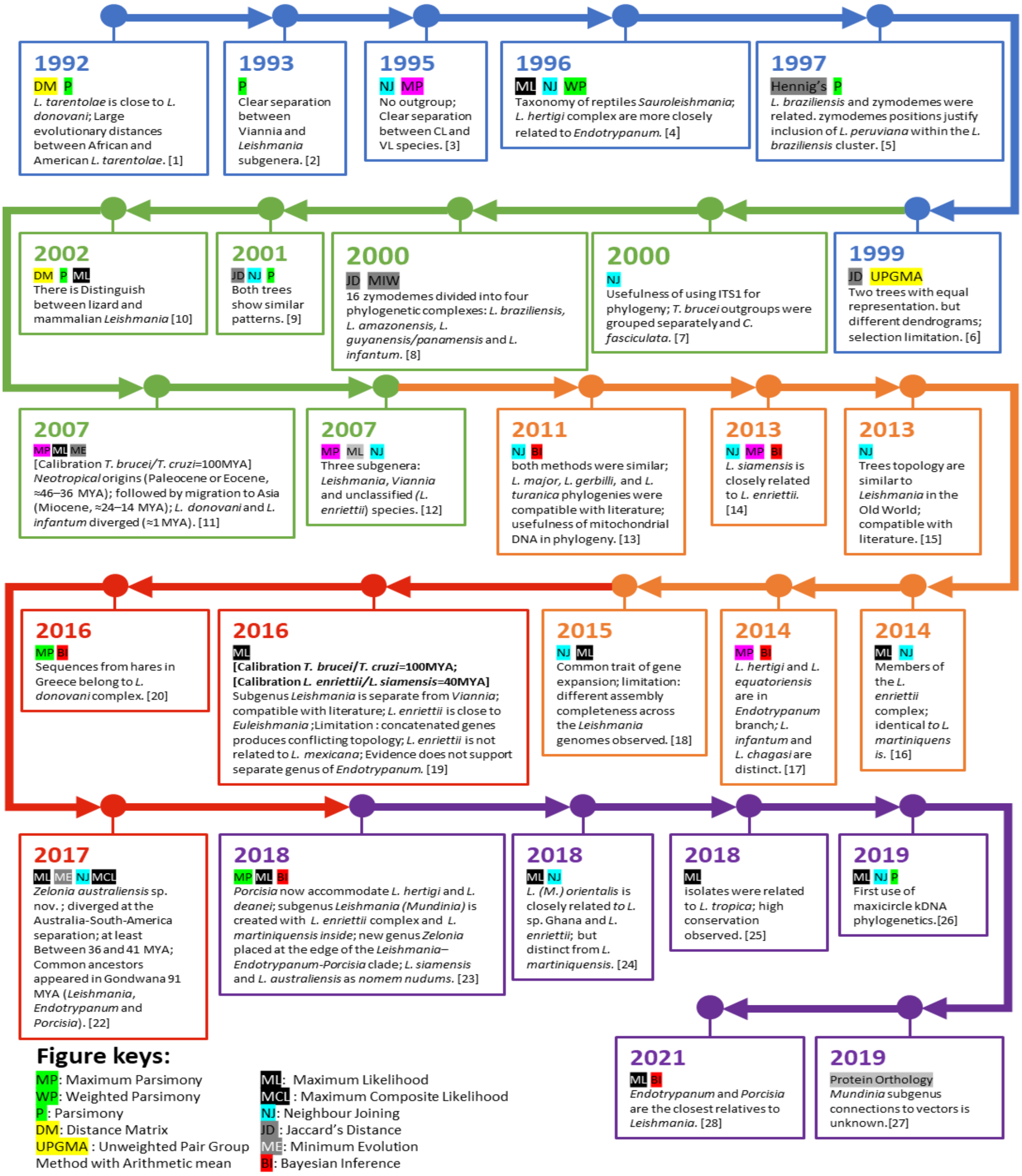
Date	Organism	Sequencing technology	Coverage	Assembly method	Scaffolds	Total length	N50
2011	*L. major*	Sanger			36	32,855,089	
2011	*L. infantum*	Sanger			76	32,122,061	1,043,848
2011	*L. mexicana*	Sanger			588	32,108,741	1,044,075
2011	*L. donovani*	Roche 454; Illumina			36	32,444,968	1,024,085
2011	*L. braziliensis*	Sanger			138	32,068,771	992,961
2013	*L. turanica*	Illumina	108x	AllPaths-LG	336	32,320,007	397,299
2013	*L. gerbilli*	Illumina	140x	AllPaths-LG	492	31,398,648	379,527
2014	*L. panamensis*	Roche 454; Illumina	30x	Newbler; PAGIT	35	30,688,794	1,043,456
2015	*L.* sp.	Illumina HiSeq	110x	A5 assembly pipeline	1,100	27,848,322	61,709
2015	*L. peruviana*		35x		37	33,890,200	1,047,715
2016	*L. arabica*	Illumina	94x	AllPaths-LG	168	31,269,090	1,057,807
2016	*L.* sp. MAR	Illumina	236x	AllPaths-LG	251	30,813,970	873,628
2018	*L. guyanensis*	PacBio RSII	80x	CANU	123	33,816,023	683,170
2018	*L. lainsoni*	PacBio RSII	74x	CANU	137	34,152,029	638,860
2019	*L. aethiopica*	PacBio RSII	74x	CANU	118	33,648,436	763,733
2019	*L. amazonensis*	Illumina HiSeq	99.1x	SMALT	34	32,156,470	N/A
2019	*L. tarentolae*	PacBio RS II	120x	HGAP	179	35,416,496	663,019
2020	*L. tropica*	PacBio RSII; Illumina MiSeq	75x	Flye	43	32,700,668	1,070,514
2020	*L. infantum chagasi*	Illumina MiSeq	150.0x	SOAPdenovo	36	31,924,566	1,043,794

2.6 *Mundinia* Taxonomy: The Knowledge Gap

As previously reviewed, the genus *Leishmania* is divided now into four subgenera: *Leishmania*, *Viannia*, *Sauroleishmania*, and *Mundinia* (Espinosa et al., 2018). *Mundinia* is the most recent and least studied subgenus due to a scarcity of molecular evidence (Lainson, 1997). It accommodates four described species: *L. (M.) enriettii* (Blewett et al., 1971), *L. (M.) macropodum* (Barratt et al., 2017a)*, L. (M.) martiniquensis* (Desbois et al., 2014)*,* and *L. (M.) orientalis* (Jariyapan et al., 2018a)*.*

Species of the subgenus *Mundinia* appears to have a diverse distribution and host range, as well as fewer representative genomes. For instance, they have been reported in a variety of locations across the globe, including Australia (Rose et al., 2004), central Europe (Muller et al., 2009), Ghana (Kwakye-Nuako et al., 2015), Martinique (Desbois et al., 2014) Switzerland (Lobsiger et al., 2010), Thailand (Jariyapan et al., 2018a), and the United States of America (Reuss et al., 2012). Additionally, they have been isolated from a variety of hosts: *L. (M.) orientalis*, *L. (M.) martiniquensis*, and *L. (M.)* sp. Ghana have been isolated from humans; *L. (M.) enriettii* infects guinea pigs; *L. (M.) macropodum* has been isolated from Australian macropods; and some *L. (M.) martiniquensis cases* have been reported in cows and horses. In addition to that, there have been a few instances of infection in immunocompromised patients (Dedet et al., 1995, Chicharro and Alvar, 2003, Bualert et al., 2012).

Numerous studies have been conducted to determine *Leishmania* taxonomy. Inferring taxonomy based on phylogeny had been a well-established practice in *Leishmania* research for nearly three decades (Briones et al., 1992b). The first study to consider molecular phylogenetics to determine taxonomy in trypanosomatids was in 2004 (Moreira et al., 2004). They used 18S rRNA sequences to construct phylogenies. The timeline in Figure 2.3 depicts the main findings of all phylogenetic analyses in chronological sequence done on *Leishmania*. Earlier studies have been published for the subgenera *Leishmania* (Peacock et al., 2007, Cantacessi et al., 2015), *Sauroleishmania* (Raymond et al., 2012, Coughlan et al., 2017), and *Viannia* (Valdivia et al., 2015, Rogers et al., 2011, Llanes et al., 2015), but there have been very few for *Mundinia* (Jariyapan et al., 2018b, Butenko et al., 2019b).



1992
DM P
L. tarentolae is close to *L. donovani*; Large evolutionary distances between African and American *L. tarentolae*. [1]

1993
P
Clear separation between Viannia and *Leishmania* subgenera. [2]

1995
NJ MP
No outgroup; Clear separation between CL and VL species. [3]

1996
ML NJ WP
Taxonomy of reptiles *Sauroleishmania*; *L. hertigi* complex are more closely related to *Endotrypanum*. [4]

1997
Hennig's P
L. braziliensis and zymodemes were related. zymodemes positions justify inclusion of *L. peruviana* within the *L. braziliensis* cluster. [5]

1999
JD UPGMA
Two trees with equal representation. but different dendrograms; selection limitation. [6]

2000
NJ
Usefulness of using ITS1 for phylogeny; *T. brucei* outgroups were grouped separately and *C. fasciculata*. [7]

2000
JD MIW
16 zymodemes divided into four phylogenetic complexes: *L. braziliensis*, *L. amazonensis*, *L. guyanensis/panamensis* and *L. infantum*. [8]

2001
JD NJ P
Both trees show similar patterns. [9]

2002
DM P ML
There is Distinguish between lizard and mammalian *Leishmania* [10]

2007
MP ML ME
[Calibration *T. brucei/T. cruzi*=100MYA] *Neotropical* origins (Paleocene or Eocene, ≈46–36 MYA); followed by migration to Asia (Miocene, ≈24–14 MYA); *L. donovani* and *L. infantum* diverged (≈1 MYA). [11]

2007
MP ML NJ
Three subgenera: *Leishmania, Viannia* and unclassified (*L. enriettii*) species. [12]

2011
NJ BI
both methods were similar; *L. major, L. gerbilli,* and *L. turanica* phylogenies were compatible with literature; usefulness of mitochondrial DNA in phylogeny. [13]

2013
NJ MP BI
L. siamensis is closely related to *L. enriettii.* [14]

2013
NJ
Trees topology are similar to *Leishmania* in the Old World; compatible with literature. [15]

2014
ML NJ
Members of the *L. enriettii* complex; identical to *L. martiniquensis.* [16]

2014
MP BI
L. hertigi and *L. equatoriensis* are in *Endotrypanum* branch; *L. infantum* and *L. chagasi* are distinct. [17]

2015
NJ ML
Common trait of gene expansion; limitation: different assembly completeness across the *Leishmania* genomes observed. [18]

2016
ML
[Calibration *T. brucei/T. cruzi*=100MYA;
[Calibration *L. enriettii/L. siamensis*=40MYA]
Subgenus *Leishmania* is separate from *Viannia*; compatible with literature; *L. enriettii* is close to *Euleishmania* ;Limitation : concatenated genes produces conflicting topology; *L. enriettii* is not related to *L. mexicana*; Evidence does not support separate genus of *Endotrypanum.* [19]

2016
MP BI
Sequences from hares in Greece belong to *L. donovani* complex. [20]

2017
ML ME NJ MCL
Zelonia australiensis sp. nov. ; diverged at the Australia-South-America separation; at least Between 36 and 41 MYA; Common ancestors appeared in Gondwana 91 MYA (*Leishmania, Endotrypanum* and *Porcisia*). [22]

2018
MP ML BI
Porcisia now accommodate *L. hertigi* and *L. deanei*; subgenus *Leishmania (Mundinia)* is created with *L. enriettii* complex and *L. martiniquensis inside*; new genus *Zelonia* placed at the edge of the *Leishmania–Endotrypanum-Porcisia* clade; *L. siamensis* and *L. australiensis* as *nomem nudums.* [23]

2018
ML NJ
L. (M.) orientalis is closely related *to L.* sp. Ghana and *L. enriettii*; but distinct from *L. martiniquensis.* [24]

2018
ML
isolates were related to *L. tropica*; high conservation observed. [25]

2019
ML NJ P
First use of maxicircle kDNA phylogenetics.[26]

2019
Protein Orthology
Mundinia subgenus connections to vectors is unknown.[27]

2021
ML BI
Endotrypanum and *Porcisia* are the closest relatives to *Leishmania.* [28]

Figure keys:

MP: Maximum Parsimony
WP: Weighted Parsimony
P: Parsimony
DM: Distance Matrix
UPGMA : Unweighted Pair Group Method with Arithmetic mean
ML: Maximum Likelihood
MCL : Maximum Composite Likelihood
NJ: Neighbour Joining
JD : Jaccard's Distance
ME: Minimum Evolution
BI: Bayesian Inference

Figure 2.3: A brief timeline history of all *Leishmania* phylogenetic studies done to date. Study keys: [1] (Briones et al., 1992a); [2] (Thomaz-Soccol et al., 1993); [3] (Piarroux et al., 1995); [4] (Croan and Ellis, 1996); [5] (Chouicha et al., 1997); [6] (Banuls et al., 1999); [7] (Dávila and Momen, 2000); [8] (Thomaz-Soccol et al., 2000); [9] (Hide et al., 2001); [10] (Orlando et al., 2002); [11] (Lukes et al., 2007a); [12] (Waki et al., 2007) ;[13] (Cao et al., 2011); [14] (Leelayoova et al., 2013); [15] (Chaouch et al., 2013); [16] (Pothirat et al., 2014a); [17] (Marcili et al., 2014); [18] (Valdivia et al., 2015); [19] (Harkins et al., 2016); [20] (Tsokana et al., 2016); [21] (Zhang et al., 2016); [22] (Barratt et al., 2017b); [23] (Espinosa et al., 2018); [24] (Jariyapan et al., 2018b); [25] (Bamorovat et al., 2018); [26] (Kaufer et al., 2019); [27] (Butenko et al., 2019a); [28] (Albanaz et al., 2021). Full-scale figure can be seen in Supplementary Materials.

Recent comparative genomic study suggested that species from the subgenus *Mundinia* have evolved to survive in the vertebrate host more than the vector (Butenko et al., 2019a). For instance, some proteins that are found in the promastigote stage which are essential in gut interaction in insects, have been observed to be significantly low in *Mundinia*. However, others that are involved in the amastigote stage in which are responsible for surviving inside the host macrophages, have been found to be at the same levels when compared with the subgenera *Leishmania* and *Viannia*.

However, this finding was restricted by under sampling since there were only two proteomes available from the subgenus *Mundinia* at the time of publishing. In addition to under sampling, only three *Mundinia* species were analysed, and phylogenies were constructed using only high conserved genomic characteristics.

As explained earlier, among all phylogenetic reconstructions, *Mundinia* was found to be the most geographically distributed subgenus as well as the deepest branch of *Leishmania*, implying the existence and parasitic circulation of a *proto-Mundinia* organisms prior to the disintegration of the supercontinents (Harkins et al., 2016, Lukes et al., 2018).

2.7 Aims and Objectives

The aims of my thesis are to determine how the subgenus *Mundinia* is taxonomically related to the other *Leishmania* species from the other subgenera; to determine the degree to which *Mundinia* species diverge from the rest of *Leishmania* over time; to determine the genomic structure, chromosome number, and gene content; to examine wither *Mundinia* species are subjected to selection pressure that might affect its role in causing infection or avoiding immunity; and to examine the relationships between *Mundinia* species and other taxa in the family *Trypanosomatidae*.

These aims were accomplished by achieving the following objectives:

1. The essential starting point objective is to *de novo* assemble and annotate multiple *Mundinia* species, with emphasis on chromosome-level assembly.
2. This was achieved by using both short and long sequencing technology to collect high quality data and maintain the integrity of the chromosome backbone structure.
3. Making these genomes publicly available so that they can be compared to other published genomic evidence using a variety of methods based on cutting-edge phylogeny inference technologies.
4. Testing some previous hypotheses about the origin of *Leishmania* and subsequently *Mundinia*.
5. Estimating the time to the most recent common ancestor for *Mundinia* species, which will help to explain how and when subgenus Mundinia evolved.
6. Examining the annotated proteins in these genomes for selection pressure and determining whether any selected protein has been linked to infection or immunity avoidance.

Thus, six genomes have been sequenced, assembled, and annotated; two of them, *L. (M.) martiniquensis* and *L. (M.) enriettii*, already have reference genomes but from separate strains. Another two strains , namely *L. (M.) orientalis* and *L. (M.)* sp. Ghana, were recently added to the subgenus *Mundinia* but do not have representative genomes; one unknown strain was isolated from hyrax in Namibia (Grove and Ledger, 1975); and an outgroup strain, *Porcisia hertigi*, was isolated from the tropical porcupine *Coendou rothschildi* (Herrer, 1971).

This chapter started by explaining the impact of leishmaniasis and the efforts made to better understand the parasites' biology and evolution. As previously stated, the objective is to further our understanding the taxonomic position of the subgenus *Mundinia* in relation to other *Leishmania* species, as well as the genomic structure of its members. Nonetheless, the project exemplifies open science and open data. All data and methods have been made completely public. This commitment, I believe, has been accomplished for this project. Additionally, I believe that everything achieved here is reproducible by others, as described in the following chapters, and that it can be applied to genome projects of similar size.

The following chapters will describe the materials and methods, including how biological samples were sequenced and processed, as well as the computational side, also referred to as the dry lab, and how I maintained the reproducible research aspect throughout the project.

Chapter 3. Materials

To achieve the final output, this project was carried out in two different experimental ecosystems: the wet lab, where the biological sample was collected and prepared for sequencing; and the dry lab, where multiple bioinformatics analyses were carried out in a series to achieve the final output. This chapter will describe all the wet lab materials and methods that were used, as well as how to set up the dry lab and ensure that the results and publications are reproducible. The next chapter will go over the computational methods used in the dry lab in greater detail.

3.1 Sample Selection

Six samples were chosen for the complete genome assembly process: *L. (M.) martiniquensis*, *L. (M.) orientalis*, *L. (M.) enriettii*, *L. (M.)* sp. Ghana, *L. (M.)* sp. Namibia, and *Porcisia hertigi*.

L. (M.) martiniquensis, Chiang Mai 1 (LSCM1) isolate, was initially obtained through bone marrow aspiration from a 52-year-old male who presented with sub-acute fever, huge splenomegaly and pancytopenia from northern Thailand (Pothirat et al., 2014a). Back then, numerous VL cases have been recorded in Thailand since 1996. This isolate, which was given the WHO code MHOM/TH/2012/LSCM1, was identified as a members of the *L. enriettii* complex and appeared to be similar to *L. martiniquensis* previously described from the Caribbean island of Martinique (Desbois et al., 2014).

L. (M.) orientalis, LSCM4 isolate and strain LV768 with the WHO code MHOM/TH/2014/LSCM4, was obtained from a patient diagnosed with CL, an elderly woman who resides in Thailand's Chiang Klang District, Nan Province, and has never travelled beyond her home province. DNA analysis revealed a resemblance to prior HIV-related cases in Thailand (Bualert et al., 2012, Supsrisunjai et al., 2017) as well closely related to *L. (M.) enriettii* and *L. (M.) martiniquensis*.

L. (M.) enriettii; with strain LV763, isolate CUR178, and WHO code MCAV/BR/2001/CUR178;LV763, was one of several isolates from leishmaniasis lesions in guinea pigs in southern Brazil's Curitiba metropolitan area (Thomaz-Soccol et al., 1996). This isolate was obtained from a skin lesion of female guinea pigs (*Cavia porcellus*) and was characterised by isoenzyme electrophoresis to be similar to *L. (M.) enriettii*.

L. (M.) sp. Ghana; Isolate GH5, strain LV757 and WHO code MHOM/GH/2012/GH5;LV757, was described in 2015 as an unnamed parasite of the genus *Leishmania* that was discovered in a case of human CL in Ghana. A PCR-based identification method for active CL was conducted across Ho District of the Volta Region, Ghana. DNA analysis and phylogenetic study revealed it to be part of the subgenus *Mundinia* (Kwakye-Nuako et al., 2015).

L. (M.) sp. Namibia; strain LV425, isolate 253, and WHO code MPRO/NA/1975/253/LV425 was in a cryogenic storage at Liverpool School of Tropical Medicine (LSTM) and then moved to another cryogenic storage facility at Lancaster University (Peters, 1977). Namibia is not commonly considered to have a high number of NTDs, but published reports of over 30 years indicate the possibility of much of the information is buried in historical studies published prior to 1990 (Noden and van der Colf, 2013). The first case was discovered in the 1970s, when the parasite was isolated from rock hyrax (*Procavia capensis*), sandflies, and lesions on the skin of infected humans (Grove and Ledger, 1975, Grove, 1978, Grove, 1989).

Porcisia hertigi, strain LV43, isolate C119, and WHO code MCOE/PA/1965/C119;LV43 was isolated from tropical porcupine (*Coendou rothschildi*) in Panama (Herrer, 1971). This strain was chosen for two reasons: first, *Porcisia* species have been reported to be distinct from *Leishmania* but also to be more closely related than any other species in the family *Trypanosomatidae* (Noyes, 1998a); and second, there was no representative genome for *Porcisia* prior to this project. Furthermore, selecting the best outgroup is a well-established practice among phylogeneticists (Graham et al., 2002), as outgroups can help explain evolutionary conclusions because they share an older ancestor with the in-groups (Nixon and Carpenter, 1993, Barriel and Tassy, 1998, Giribet and Ribera, 1998, De La Torre-bárcena et al., 2009).

3.2 Parasite Culture, Isolation, and DNA Extraction

Parasite culture and isolation were performed using an *in vitro* culture system that was developed previously for *L. (M.) orientalis* and adopted to be used for the other isolates (Chanmol et al., 2019). The isolated parasites were grown initially as promastigotes in Schneider's insect medium (Sigma-Aldrich, St. Louis, MO, USA), supplemented with 20% (v/v) FCS (Life Technologies-Gibco, Grand Island, NY, USA). Parasites then were grown at 26 °C in M199 medium, pH 6.8, supplemented with 10% (v/v) FCS, 2% (v/v) healthy human urine, 1% (v/v) Basal Medium Eagle (BME) vitamins (Sigma-Aldrich, St. Louis, MO, USA), and 25 g/ml gentamicin sulphate (Sigma-Aldrich, St. Louis, MO, USA). Every four days, promastigotes were sub-passaged to fresh medium to maintain parasite growth and viability.

Genomic DNA extraction was done according to the manufacturer's protocol using Qiagen's spin column-based method (Hilden, Germany). The concentrations of extracted DNA were determined using a Qubit fluorometer, a microplate reader, and agarose gel electrophoresis. For further confirmation, PCR, and sequencing for ribosomal protein L23a (RPL23a), for instance from Genbank accession KP006691.1 were done on all isolates using redundant primers 5'-GCGCCAACAAGACTGAGAT-3' and 5'-CGTCACCTTGACGACCTTG-3'. The sequences were then compared to determine whether the extracted DNA was related to the subgenus *Mundinia* or not, using BLAST (Altschul et al., 1990), in order to use them for subsequent *de novo* sequencing.

3.3 Sequencing and Library Preparation

All sequencing libraries were constructed using the same extracted DNA sample to avoid inconsistency. Library construction for sequencing the short reads was contracted to two outsources: the first was BGI (Shenzhen, China), where they used DNBSEQ libraries to produce paired end reads at different insert sizes (170 bp, 270 bp and 500 bp), using the Illumina HiSeq platform. The second was Aberystwyth University (Aberystwyth, UK), where they used TruSeq Nano DNA libraries to produce paired end reads at 300 bp length, using the Illumina MiSeq platform.

The long-read library preparation and sequencing was done using MinION according to the protocol SQK-LSK109 (ONT, UK) on R9 flow cells (FLO-MIN106). Figure 3.1 summarises both reads coverage and file size for both short and long reads.

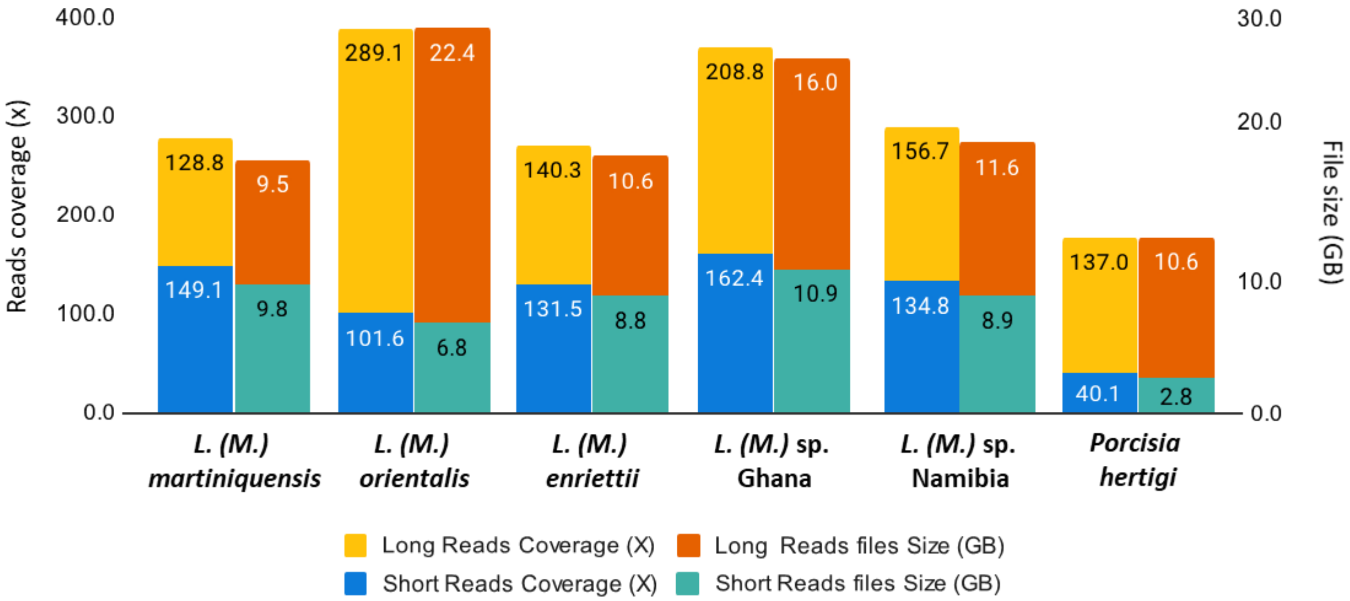


Figure 3.1: Stacked column chart that illustrates the read coverage (left vertical axis) and file size (right vertical axis) for each genome sequence.

3.4 Virtual Machine and Software Management

Most of computational analyses were done on virtual machine with 24 CPUs and 384 Gigabytes of RAM that runs the Ubuntu 18.04 LTS operating system. We used the research file storage system at Lancaster University (LUNA) as the main file storage medium.

Managing software installation compatibility and dependency is a well-known problem in computer science (Jang, 2006), and it can be approached in a variety of ways. At the start of this project, the operating system of choice was Bio-Linux 8, which was based on Ubuntu 14.04 LTS, because it was a stable distribution that came pre-installed with over 250 software packages, many of which were tailored to bioinformatic data analysis (Booth et al., 2012). Additionally, Bio-Linux integrates installation and software dependencies via the Synaptic package manager, which is designed to precisely address these types of issues.

However, the most significant obstacle was that Bio-Linux 8 maintenance was discontinued in 2015, and a lot has happened in software development since then. For instance, a new generation of software management tools, such as Conda, an open-source package management system, and Docker, a platform that utilises virtualization to deliver software in a form of containers, are now becoming widely used in computational biology and bioinformatics. Therefore, they were primarily used in order to achieve reproducible results and because both are well-supported and well-documented (Boettiger, 2015, Gruning et al., 2018).

Bioconda is a Conda package management channel dedicated to bioinformatics (Gruning et al., 2018). It consists of GitHub repositories for recipes, a build mechanism for converting these recipes to Conda packages, and a package repository with over 7000 ready-to-use bioinformatics tools. Over 850 contributors and 570 members contribute, modify, and manage recipes. Conda enables the distribution of packages through repositories, or channels. The defaults channel contains many packages that are frequently used.

Docker is a collection of platform as a service (PaaS) products that deliver software in containers via OS-level virtualization (Boettiger, 2015). Containers are self-contained units that contain their own software, libraries, and configuration files; they communicate via well-defined channels. Containers consume fewer resources than virtual machines because they share the services of a single operating system kernel. Docker Engine is the software that hosts the containers.

3.5 Reproducible Workflow Management

To address the project's ”Protocol Gap” (Weller, 2021), we developed an automated genome assembly and annotation pipeline and successfully applied it to all six genomes by making all methods fully accessible (Almutairi et al., 2021). Following that, the pipeline was created and implemented using the Snakemake workflow management system (Molder et al., 2021). The pipeline is composed of 314 computational steps divided into 21 sequential processes that are divided into two distinct phases (Figure 3.2). Table 3.1 list all the tools used for this project.

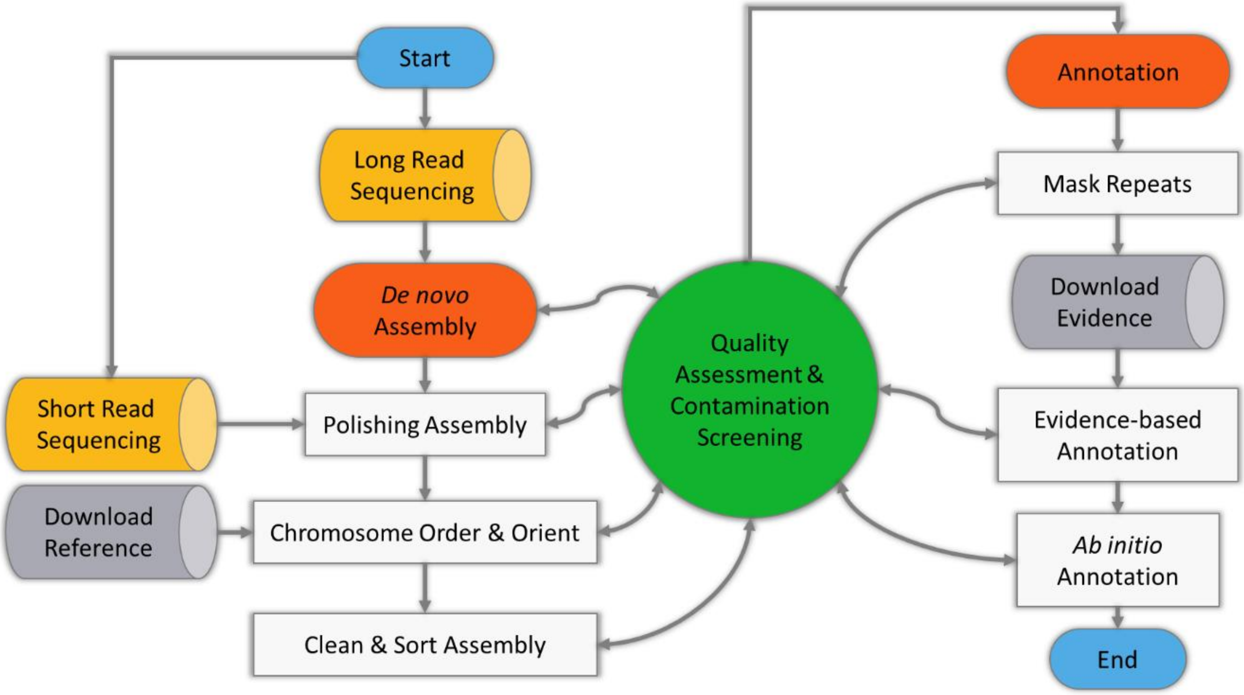


Figure 3.2: A graphical representation of the LGAAP workflow, with the first flow (left) focusing on the assembly process and the second flow (right) containing the annotation steps all the way to the experiment's conclusion, while the green circle (centre) in the centre represents the quality control assessments (Almutairi et al., 2021)

Table 3.1: Tools used in the analysis workflow, as well as their Conda or Docker links.

Tool	Website	Conda or Docker link
AGAT	https://github.com/NBISweden/AGAT	https://anaconda.org/conda-forge/agate
AUGUSTUS	http://bioinf.uni-greifswald.de/webaugustus/about	https://hub.docker.com/r/hatimalmutairi/lmgaap-maker
BCFtools	http://samtools.github.io/bcftools/	https://anaconda.org/bioconda/bcftools
Bedtools	https://bedtools.readthedocs.io/en/latest/	https://anaconda.org/bioconda/bedtools
Blast+	https://blast.ncbi.nlm.nih.gov/Blast.cgi	https://anaconda.org/bioconda/blast
Circa	http://omgenomics.com/circa	
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	https://anaconda.org/bioconda/fastqc
Flye	https://github.com/fenderglass/Flye	https://anaconda.org/bioconda/flye
Funannotate	https://github.com/nextgenusfs/funannotate	https://anaconda.org/bioconda/funannotate
GAAS	https://github.com/NBISweden/GAAS	https://anaconda.org/bioconda/gaas
GeneMark	http://exon.gatech.edu/GeneMark/	https://hub.docker.com/r/hatimalmutairi/lmgaap-maker
Genometools	http://genometools.org/	https://anaconda.org/bioconda/genometools-genometools
Interproscan	https://www.ebi.ac.uk/interpro/search/sequence/	https://hub.docker.com/r/blaxterlab/interproscan
MAKER2	https://www.yandell-lab.org/software/maker.html	https://hub.docker.com/r/hatimalmutairi/lmgaap-maker
Minimap2	https://github.com/lh3/minimap2	https://anaconda.org/bioconda/minimap2
MultiQC	https://multiqc.info/	https://anaconda.org/bioconda/multiqc
MUMmer	http://mummer.sourceforge.net/	https://anaconda.org/bioconda/mummer
Pilon	https://github.com/broadinstitute/pilon/wiki	https://anaconda.org/bioconda/pilon
PycoQC	https://pypi.org/project/pycoQC/	https://anaconda.org/bioconda/pycoqc
RaGOO	https://github.com/malonge/RaGOO	https://anaconda.org/imperial-college-research-computing/ragoo
REPAVER	https://gitlab.com/gringer/bioinfscripts	
RepeatMasker	http://www.repeatmasker.org/	https://hub.docker.com/r/hatimalmutairi/lmgaap-maker
Samtools	https://github.com/samtools/samtools	https://anaconda.org/bioconda/samtools
Snakemake	https://snakemake.readthedocs.io/en/stable/	https://anaconda.org/bioconda/snakemake
TEclass	http://www.compgen.uni-muenster.de/tools/teclass/index.hbi?lang=en	https://hub.docker.com/r/hatimalmutairi/teclass-2.1.3b
Wordcloud		https://anaconda.org/conda-forge/wordcloud

Chapter 4. Methods and Pipeline Analyses

Computational methods and pipeline analyses will be the primary focus of this chapter. Typically, the term pipeline refers to the process of executing multiple computational steps in the direction of a final goal, which involves a series of stages or steps. Because of this, understanding the entire analysis from beginning to end is critical for the pipeline's success (Figure 3.2). I shall describe here the computational methods in the order in which they should be completed, beginning with sequencing, and concluding with comparative analysis.

4.1 Raw Reads Assessments

FASTQ format was used to create both short and long read sequences (Cock et al., 2010) FastQC software was used to evaluate sequences generated on Illumina platforms (Andrews, 2010) while PycoQC was used to evaluate sequences generated on Nanopore technologies (Leger and Leonardi, 2019). Following that, both types of assessments were merged and summarised in a single MultiQC report for each genome (Ewels et al., 2016).

4.2 Assembly Optimisation

Selecting the most efficient assembly tools proved difficult due to the large number of variables affecting the assembly's completion (Dujardin, 2009). As a result, the optimization process was built around the sequencing type. As previously stated, because *Leishmania* genomes contain a high proportion of repeat sequences, these repeats prevent kmer-based *de Bruijn* algorithms from completing the assembly process (Compeau et al., 2011). As a result, three types of algorithms were chosen: short reads assemblers, long reads assemblers, and hybrid assemblers combining long and short reads.

The first strategy was tested only on short read assemblers such as Velvet (Zerbino and Birney, 2008), SPAdes (Bankevich et al., 2012), IDBA (Peng et al., 2010), ABySS (Simpson et al., 2009), Edena (Hernandez et al., 2008), SOAPdenovo (Luo et al., 2012), Ray (Boisvert et al., 2010), and ALLPATHS-LG (Gnerre et al., 2011). The second strategy involved combining long and short reads in a hybrid assembly with Unicycler on all sequenced reads (Wick et al., 2017). The third and final strategy was tested only on long read assembler, for which Flye assembler was used on Nanopore sequences (Flynn et al., 2020). These algorithms were evaluated using a variety of parameters, including number of contigs, GC content, and N50 values.

4.3 Assembly

Prior to running the final assembly pipeline on all six samples, an optimization experiment was performed to determine which assembly strategy would produce the best results. This required the use of sequence reads in three different strategies, as described above: short reads alone, long reads alone, and a combination of short and long reads. The outcomes of the optimisation shall be discussed in subsequent chapters.

Based on that, the assembly pipeline consists of eight sequential processes (Figure 3.2): Long read assembly using version 2.8.2 of Flye assembler (Kolmogorov et al., 2019); followed by mapping the short reads onto assemblies using version 2.17 of Minimap2 (Li, 2016); then a consensus sequence is created using version 1.11 of Samtools (Danecek et al., 2021); polishing of assemblies using version 1.23 of Pilon (Walker et al., 2014); revision of consensus sequences using Samtools; ordering and orientation of the chromosomes and breakage of any chimeric sequences using version 1.1 of RaGOO (Alonge et al., 2019); sorting and removal of any duplicated scaffolds or contigs using version 1.5.3 of Funannotate (Palmer and Nextgenusfs, 2019); and generation of a quality report using version 5.0.2 of QUAST (Gurevich et al., 2013).

4.4 Annotation

Annotation was the pipeline's second major phase. It consists of three steps: scanning for vector contamination, masking repeats, and annotation. Version 2.10.1 of BLAST+ (Camacho et al., 2009) was used to scan assemblies for vector contamination against The UniVec Database (Kitts et al., 2011). Then any contaminants were either masked or deleted using version 2.30 of BEDTools (Quinlan, 2014). The second step starts by using RepeatModeler (Flynn et al., 2020), which was run from version 1.3.1 of Dfam TE Tools Container (Abrusan et al., 2009); followed by transposable elements classification using version 2.1.3b of TEclass running from a docker container (Almutairi, 2021b). Then, any identified complex repeats were masked to allow for more accurate annotation processes.

The annotation process was divided into two rounds; an evidence-based round, which was performed by downloading proteins and transcripts from release 47 of TriTrypDB (Aslett et al., 2010) using version 2.31.10 of MAKER2 (Holt and Yandell, 2011) and running from a docker container (Almutairi, 2021a); and *ab initio* round using version 3.3.2 of Augustus (Hoff and Stanke, 2019). Each round of annotation was completed with annotation quality checking by using both version 1.2.1 of Genometools (Gremme et al., 2013) and version 1.2.0 of GAAS (NBIS, 2021). Both rounds were assessed using Annotation Edit Distance (AED). Then an assignment process for the annotated features was performed using BLAST+ against Uniprot (UniProt,

2021) and version 5.22-61.0 of InterProScan (Jones et al., 2014). The annotation ended with keeping the longest isoforms of each predicted protein, using AGAT (Dainat and Hereñú, 2020).

4.5 Synteny Analysis

MUMmer was used to construct a dot plot against a reference that is believed to be a better depiction of closely related one for all the genomes that were assembled (Kurtz et al., 2004). As a result, three genomes were employed as a reference: *L.* sp. MAR, strain LEM2494, against *L. (M.) martiniquensis*; *L. enriettii*, strain LEM3045, against *L. (M.) orientalis, L. (M.) enriettii, L. (M.) sp.* Ghana and *L. (M.) sp.* Namibia; and *Endotrypanum monterogeii* against *P. hertigi*.

However, all assemblies were cross-referenced against *L. major* Friedlin strain, which is the only available genome without any additional scaffolds and has been used in many karyotyping investigations to determine chromosome representation (Samaras and Spithill, 1987, Bastien et al., 1992, Zhou et al., 2004).

MUMmer was also used to assist in the re-scaffolding of the genomes chosen for the phylogenomic tree construction, which will be detailed in the following section. Genomes assembled at the scaffold or contig level, for example, are difficult to align against chromosomes. As a result, prior to the chromosome alignments, a MUMmer run was performed to ensure a better alignment and, ultimately, a better comparison.

4.6 Phylogenomic Analyses

Since all six genomes were assembled at the chromosomal level, the input for reconstructing the phylogenomic tree must be equivalent. As a result, two datasets were chosen; the first dataset contained 16 chromosome-level assemblies, in addition to the six genomes from this project, which were collected from the 47th release of TriTrypDB (Aslett et al., 2010). They are

L. aethiopica (L147), *L. arabica* (LEM1108), *L. donovani* (BPK282A1), *L. donovani* (CL-SL), *L. donovani* (BHU1220), *L. donovani* (LV9), *L. enriettii* (LEM3045), *L. gerbilli* (LEM452), *L. infantum* (JPCM5), *L. major* (Friedlin), *L. major* (LV39c5), *L. major* (SD-75), *L. sp.* MAR (LEM2494), *L. tarentolae* (Parrot-TarII), *L. tropica* (L590), and *L. turanica* (LEM423).

The second dataset used for phylogenomic analyses contains 60 public assemblies, including non-representative ones, assembled at different levels, as well as three *Porcisia* species used as outgroups. All these assemblies were obtained from the NCBI assembly database and then re-scaffolded using the *L. major* Friedlin strain as a guide, for them to be represented at the chromosome-level and thus included in the comparison (Table 4.1).

Table 4.1: list of all assemblies taken from NCBI assembly database, which used in constructing the phylogenomic trees, and sorted by last update date. (*) represent the genomes that were assembled in this thesis.

Date	Assembly [strain]	Level	Submitted by	Accession
2011	*L. major* [Friedlin]	Complete Genome	Friedlin Consortium	GCA_000002725.2
2012	*L. major* [SD 75.1]	Scaffold	WUGSC	GCA_000250755.2
2012	*L. donovani* [BPK282A1]	Chromosome	Wellcome Trust Sanger Institute	GCA_000227135.2
2012	*L. mexicana* [U1103]	Chromosome	Wellcome Trust Sanger Institute	GCA_000234665.4
2012	*L. braziliensis* [M2904]	Chromosome	The Sanger Institute	GCA_000002845.2
2012	*L. infantum* [JPCM5]	Chromosome	The Sanger Institute	GCA_000002875.2
2013	*L. major* [LV39c5]	Scaffold	WUGSC	GCA_000331345.1
2013	*L. panamensis* [L13]	Scaffold	Kinetoplastid Genomes Consortium	GCA_000340495.1
2013	*L. tropica* [L590]	Scaffold	Kinetoplastid Genomes Consortium	GCA_000410715.1
2013	*L. amazonensis* [LeiAma1.0]	Scaffold	Laboratorio de Genomica e Expressao	GCA_000438535.1
2013	*L. turanica* [LEM423]	Scaffold	Kinetoplastid Genomes Consortium	GCA_000441995.1
2013	*L. gerbilli* [LEM452]	Scaffold	Kinetoplastid Genomes Consortium	GCA_000443025.1
2013	*L. donovani* [BHU1220]	Chromosome	CSIR Central Drug Research Institute Lucknow	GCA_000470725.1
2014	*L. panamensis* [PSC-1]	Chromosome	INDICASAT-AIP	GCA_000755165.1
2015	*L. peruviana* [PAB-4377]	Chromosome	UFMG	GCA_001403675.1
2015	*L. peruviana* [LEM1537]	Chromosome	UFMG	GCA_001403695.1
2016	*L. sp.* MAR [LEM2494]	Chromosome	The Genome Institute - Washington University School of Medicine	GCA_000409445.2
2016	*L. aethiopica* [L147]	Chromosome	Kinetoplastid Genomes Consortium	GCA_000444285.2

Date	Assembly [strain]	Level	Submitted by	Accession
2016	*L. arabica* [LEM1108]	**Chromosome**	The Genome Institute - Washington University School of Medicine	GCA_000410695.2
2016	*L. braziliensis* [M2903]	**Chromosome**	Washington University School of Medicine	GCA_000340355.2
2016	*L. enriettii* [LEM3045]	**Chromosome**	The Genome Institute - Washington University School of Medicine	GCA_000410755.2
2017	*L. donovani* [AG83] [late passage]	**Chromosome**	Indian Institute of Chemical biology	GCA_001989955.1
2017	*L. donovani* [AG83] [early passage]	**Chromosome**	Indian Institute of Chemical biology	GCA_001989975.1
2017	*L. donovani* [1S2D]	**Chromosome**	Centre for Infectious Disease Research	GCA_002243465.1
2018	*L. infantum* [TR01]	**Chromosome**	Public Health General Directorate	GCA_003020905.1
2018	*L. guyanensis* [LgCL085]	**Contig**		dryad.4bm23
2018	*L. naiffi* [LnCL223]	**Contig**		dryad.4bm23
2018	*L. tropica* [2017-IK]	**Scaffold**	Lebanese American University	GCA_003067545.1
2018	*L. braziliensis* [IOC-L3564]	**Chromosome**	Fundação Universidade Federal de Rondônia - UNIR	GCA_003304975.1
2018	*L. tropica* [2015-IK]	**Scaffold**	Lebanese American University	GCA_003352575.1
2018	*L. guyanensis* [204-365]	**Contig**	CDC	GCA_003664525.1
2018	*L. infantum* [HUUFS14]	**Scaffold**	National Institute of Allergy and Infectious Diseases	GCA_003671315.1
2018	*L. lainsoni* [216-34]	**Contig**	CDC	GCA_003664395.1
2018	*L. donovani* [FDAARGOS_360]	**Contig**	US Food and Drug Administration	GCA_003730175.1
2018	*L. donovani* [FDAARGOS_361]	**Contig**	US Food and Drug Administration	GCA_003730215.1
2018	*L. donovani* [LdCL]	**Complete Genome**	McGill University	GCA_003719575.1
2019	*L. aethiopica* [209-622]	**Contig**	CDC	GCA_003992445.1
2019	*L. amazonensis* [210-660]	**Contig**	CDC	GCA_003992505.1
2019	*L. mexicana* [215-49]	**Contig**	CDC	GCA_003992435.1
2019	*L. amazonensis* [UA301]	**Complete Genome**	Grupo de Investigaciones Microbiologicas - UR	GCA_005317125.1
2019	*L. braziliensis* [M2904]	**Complete Genome**	CBMSO	GCA_900537975.1
2019	*L. tarentolae* [Parrot-TarII] [Laval]	**Contig**	*Universite* Laval	GCA_009770625.1
2019	*L. tarentolae* [Parrot-TarII] [Tokyo]	**Contig**	The University of Tokyo	GCA_009731335.1
2020	*L. tropica* [ATCC-50129]	**Contig**	Lebanese American University	GCA_011316065.1
2020	*L. chagasi* [M32502] [IEC]	**Chromosome**	*Instituto Evandro Chagas*	GCA_014466935.1
2020	*L. chagasi* [M6445] [IEC]	**Chromosome**	*Instituto Evandro Chagas*	GCA_014466975.1
2020	*L. tropica* [CDC216-162]	**Chromosome**	CDC	GCA_014139745.1
2021	*L. donovani* [LDHU3]	**Complete Genome**	*Centro Biologia Molecular Severo Ochoa* (CBMSO)	GCA_900635355.2

Date	Assembly [strain]	Level	Submitted by	Accession
2021	*L. infantum* [JPCM5] [CBMSO]	**Complete Genome**	*Centro Biologia Molecular Severo Ochoa* (CBMSO)	GCA_900500625.2
2021	***L. enriettii* [CUR178]***	**Chromosome**	Lancaster University	GCA_017916305.1
2021	***L. martiniquensis* [LSCM1]***	**Chromosome**	Lancaster University	GCA_017916325.1
2021	***L. orientalis* [LSCM4]***	**Chromosome**	Lancaster University	GCA_017916335.1
2021	***L. sp.* Ghana 2012 [LV757]***	**Chromosome**	Lancaster University	GCA_017918215.1
2021	***L. sp.* Namibia [253]***	**Chromosome**	Lancaster University	GCA_017918225.1
2021	***Porcisia hertigi* [C119]***	**Chromosome**	Lancaster University	GCA_017918235.1
2021	*L. chagasi* [M6445] [USP]	**Contig**	University of Sao Paulo - USP	GCA_018291365.1
2021	*L. chagasi* [M32502] [USP]	**Contig**	University of Sao Paulo - USP	GCA_018290745.1
2021	*Porcisia deanei* [TCC-258]	**Scaffold**	University of Ostrava	GCA_018683835.1
2021	*Porcisia hertigi* [TCC-260]	**Scaffold**	University of Ostrava	GCA_019345635.1
2021	*L. major* [Friedlin] [CBMSO]	**Complete Genome**	*Centro Biologia Molecular Severo Ochoa* (CBMSO)	GCA_916722125.1

Then, in addition to the six genomes from this project, all selected genomes in both datasets were split and grouped according to the chromosome number. MAFFT aligner was then used to align each chromosomal set using the default parameters (Yamada et al., 2016). Then, all alignments once performed, were inspected for recombination using SplitsTree4 software (Huson and Bryant, 2005).

A series of parameters required to be added to the configuration files to aid in the reconstruction of each chromosome-scale Bayesian tree. These values were successfully estimated using MEGAX (Kumar et al., 2018) and then fed into BEAUti software (Bouckaert et al., 2014). As a result of this optimization, the following parameters were determined: a General Time Reversible (GTR) substitution model with an estimated five gamma categories for site heterogeneity; 1.2169 as the mean gamma shape parameters (exponential α); a strict clock model (uniform rates across branches); Yule speciation process as tree prior (a process of pure birth) with a crude UPGMA tree as a starting point (Gernhard, 2008); calibrating the pairwise divergent estimate time of 19.6 MYA (14.6 - 24.7 MYA) for *L. infantum* (strain JPCM5) and *L. major* (strain Friedlin), as the estimated times were derived from 2 studies (Lukes et al., 2007b, Harkins et al., 2016) and calculated using TimeTree (Kumar et al., 2017).

The same sitting, which ran ten million states, was applied to all chromosome sets. Using Treeannotator with the default parameters, each converged tree per chromosome set was combined into a single tree. Then, to build a consensus tree for all chromosomes, all 36 chromosomal trees were manually concatenated into a single tree file and run through Treeannotator again (Drummond and Rambaut, 2007). All trees were visualised using Figtree (Rambaut, 2009).

4.7 Orthology Speciation Analysis

The construction of an orthology species tree using OrthoFinder software requires a collection of a proteome dataset (Emms and Kelly, 2015). Therefore, two distinct datasets were collected for consistency reasons. The first dataset was collected to be like the first dataset used for reconstructing the phylogenomic tree, which contained a total of 22 proteomes. The second dataset was more inclusive. It contained all proteomes in the 47th release of TriTrypDB (Aslett et al., 2010). The output from the second dataset was then transformed into a combination of orthology species tree and a heatmap matrix that shows the number of orthologs per proteome.

4.8 Detecting of Selection pressure in *Mundinia* Genomes

I utilized a pipeline that was originally developed for evolution and diversity in human herpes virus genomes HSV-1 and HCMV to detect and quantify natural selection on genes located within chromosome alignments (Szpara et al., 2014). Initially, all *Mundinia* genomes were used for this analysis, where they were reordered into sets of chromosomes, Then, using MAFFT aligner with the default parameters (Yamada et al., 2016), each chromosomal set was aligned and used as an input for the pipeline.

The output of this pipeline is a combination of site-wise likelihood ratio produced by Slr and PAML (Massingham and Goldman, 2005b, Yang, 1997), as well as calculated values of nonsynonymous to synonymous substitution rates (Ω) and transition/transversion rates (K) for all coding sequence regions produced by CodeML (Yang, 2007).

4.9 Assembly Coverage Analysis and End-of-Chromosome Structure Validation

For each genome assembly, all chromosomes were concatenated to form a single continuous sequence separated by a gap of 100 bases of Ns . The number 100 was chosen for two reasons: it is the canonical number of gaps used, and it is less than the length of reads generated by the Illumina platform sequences to detect any chromosome overlap. All raw sequencing data were then mapped onto the concatenated genome using Minimap2 and Samtools. Only the gap areas with 100,000 bases before and after were included for calculating the coverages per bases. Then any reads that overlapped were recorded.

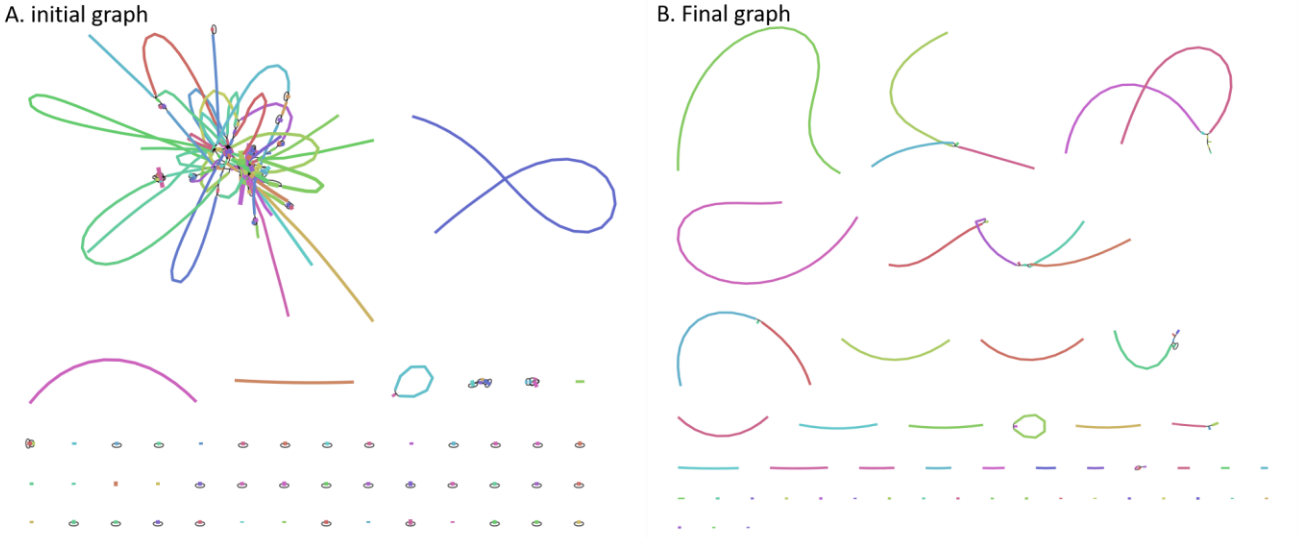
4.10 Detection of Divergent Strand-Switch Regions (dSSRs)

Leishmania genes are known to be clustered and transcribed as large polycistronic transcription units (Puechberty et al., 2007, Chandra et al., 2017). Transcription starts from divergent strand switch regions (dSSRs) (Daniels et al., 2010). To demonstrate dSSRs for the assemblies, the final General Feature Format (GFF) outputs, from the annotation process, were transformed to present the strands. In addition to that, any features contain the term "polymerase" in its description were selected to be visualised as well using a circular plot created with Circa software.

4.11 Chromosome and Karyogram visualisation

Multiple visualisations were used to validate assembly, annotation, and chromosome structure. The R package chromPlot was used to show entire genome data in a non-circular fashion across all chromosomes (Oróstica and Verdugo, 2016). REPAVER code, a R script for visualising DNA sequence repeats, was also used to view repeats, particularly centromere repeats. In addition to that, multiple circos plots were created with Circa software (Table 3.1).

Chapter 5. Assembly Results



Graph size

Node count: 170
Edge count: 253
Total length (no overlaps): 31,796,222

Graph connectivity

Dead ends: 47
Percentage dead ends: 13.82%
Connected components: 50
Largest component: 26,011,554 bp (81.81%)
Total length orphaned nodes: 4,992,685 bp (15.70%)

Node sizes

N50: 858,119 bp
Shortest node: 502 bp
Lower quartile node: 1,614 bp
Median node: 3,993 bp
Upper quartile node: 89,367 bp
Longest node: 2,965,686 bp

Figure 5.1: *De Bruijn* graph and statistics illustrate how repeat sequences entangled the assembly (much like entangled cotton threads), where the nodes represent the similar repeat sequences found in the majority of contigs, preventing the assembly from generating larger chromosome-like contigs.

Choosing the best possible assembly algorithm proved challenging. The three-strategy approach significantly aided in the assembly's completion. SPAdes, IDBA, and Ray assemblers produced the best results with the short-reads-only strategy. However, none of the assemblers used in this strategy is as effective as those used in the other two strategies. The hybrid strategy, on the other hand, was better than using only short reads. The long-read assemblies, on the other hand, were the best. They preserved chromosomal structures despite the low-quality scores generated by Nanopore's long reads, which were significantly lower than those generated by Illumina. Figures 5.1 and 5.2 demonstrate the clear superiority of the long reads assembler over the other two. The assessment in this case was made using Nx values, cumulative length, and GC content (Table 5.1).

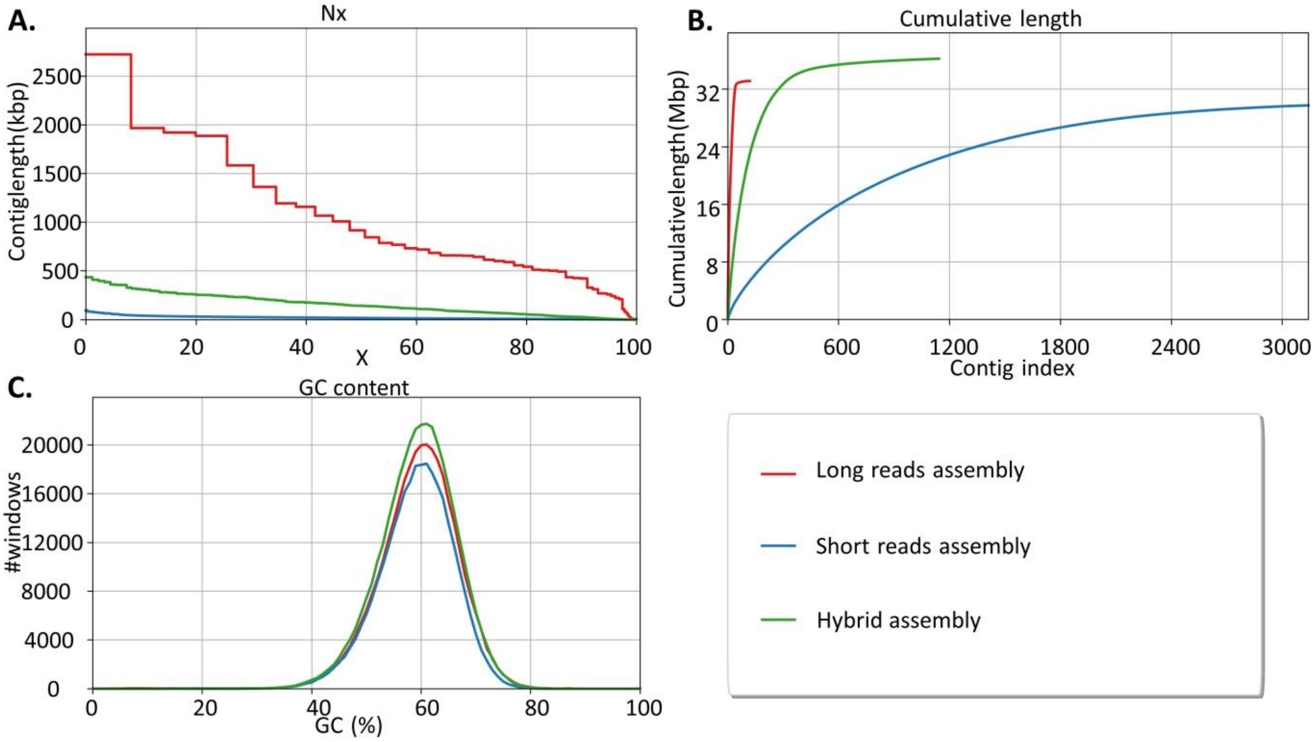


Figure 5.2: Plots comparing the three assembly strategies used in *L. (M.)* sp. Namibia as a test case during the optimisation. Section A plots the Nx values, which represent the length of the contig that accounts for at least x% of the assembly's bases, ranging from 0 to 100%. Section B illustrates the cumulative contig lengths. Section C shows the GC content.

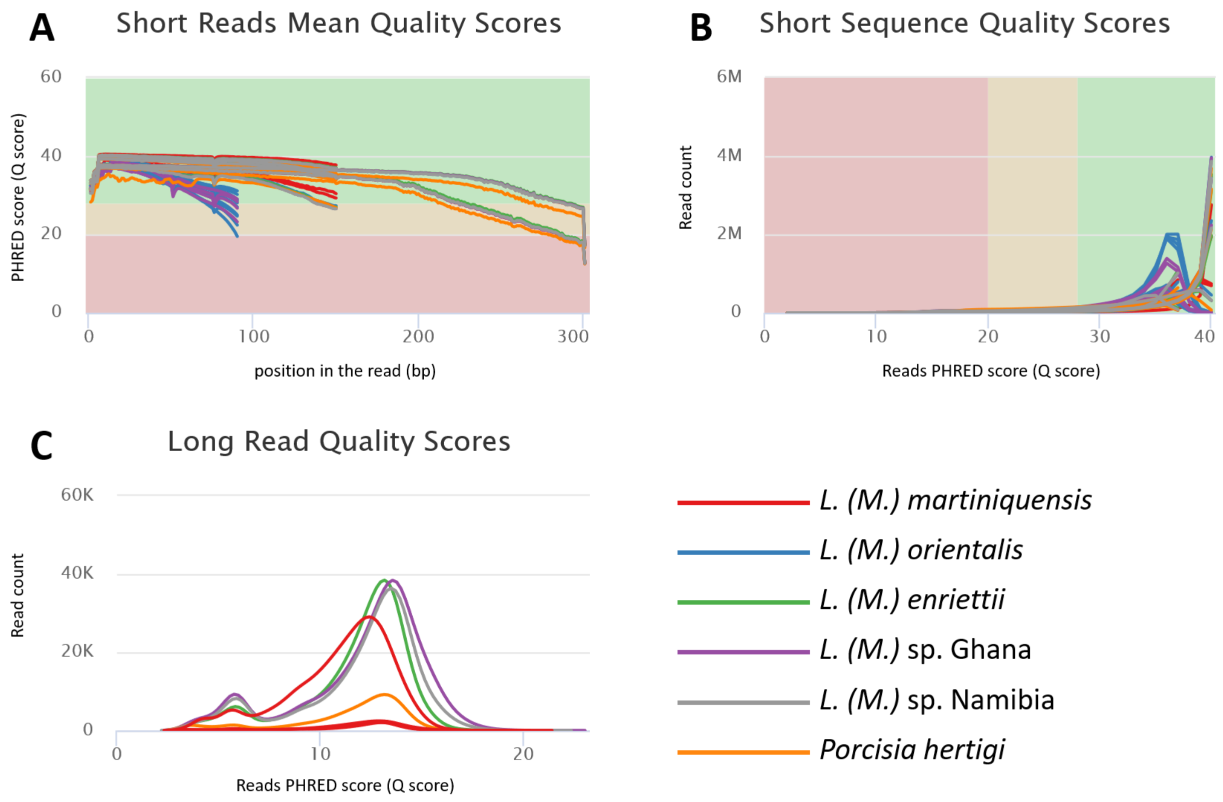


Figure 5.3: Sequence quality scores. A displays the average quality scores across short reads, with the x-axis representing the read's position and the y-axis representing the Q score. B is also connected to the short reads. It plots the Q scores against the total number of reads in each library. C, on the other hand, demonstrates the same as B except for the long reads (more details in Supplementary Materials).

Table 5.1: Statistics for each of the three assembly strategies we developed using *L. (M.)* sp. Namibia as a case study.

Assembly	Long reads only (Flye)	Short reads only (SPAdes)	Hybrid assembly (Unicycler)
Number of contigs	120	3141	1143
Largest contig	2,721,116	92,010	433,539
Total length	33,097,966	29,748,773	36,205,818
GC (%)	59.56	59.14	59.40
N50	916,499	16,965	142,159
N75	601,329	9,052	68,033
L50	11	536	80
L75	23	1138	171
Number of total reads	24,283,788	24,972,674	24,330,561
Mapped (%)	93.94	90.79	93.71
Properly paired (%)	75.57	67.78	74.75
Avg. coverage depth	266	260	233

5.1 Genome Assemblies

After sequencing, the data sizes and total yield per sample are summarised in Figure 5.3. The combined file size of all samples was 139.327 Gigabytes, resulting in 58.698 Giga-Bases and 23.708 Giga-Reads (Table 5.3). All sequences were assigned BioSample and BioProject accession numbers in NCBI database, as shown in table 5.2.

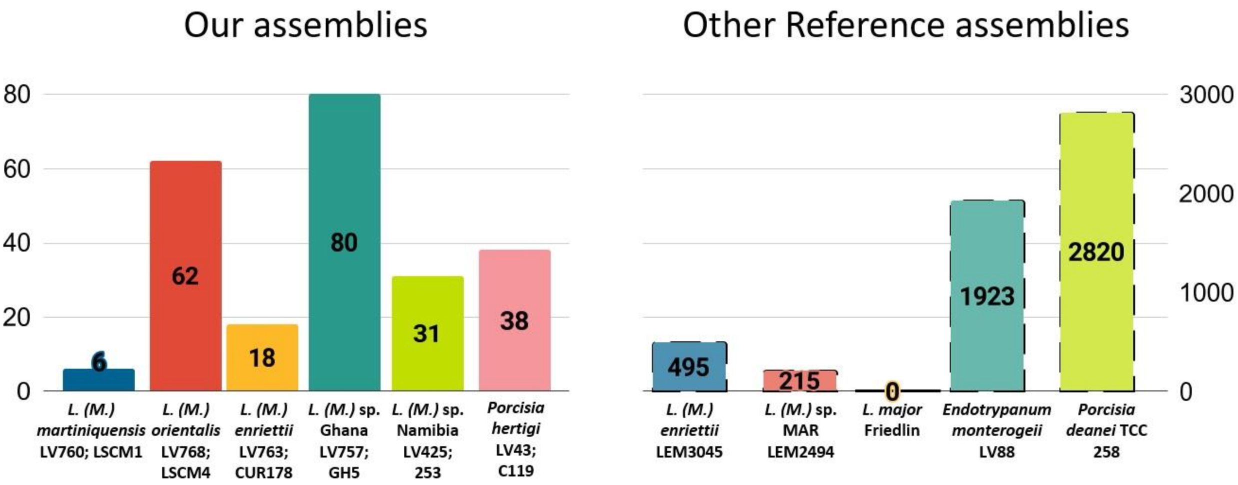


Figure 5.4: Comparison plot of the unplaced contigs that were not assigned to any chromosome between our assemblies and those of other public assemblies. Vertical axes on both sides were created separately to accommodate the range of values for this project's genomes (left) and a few other representative genomes (right).

Table 5.2: List of sample assembly descriptions.

Sample	Strain	Isolate	BioSample accession	BioProject
L. (M.) martiniquensis	LV760	LSCM1	SAMN17294109	PRJNA691531
L. (M.) orientalis	LV768	LSCM4	SAMN17294111	PRJNA691532
L. (M.) enriettii	LV763	CUR178	SAMN17294112	PRJNA691534
***L. (M.)* sp. Ghana**	LV757	GH5	SAMN17294115	PRJNA691536
***L. (M.)* sp. Namibia**	LV425	253	SAMN17294129	PRJNA689706
Porcisia hertigi	LV43	C119	SAMN17294121	PRJNA691541

Table 5.3: Descriptions of the sequencing stage as well as information about the reads, bases, and file sizes.

Species	Platform	Accessions	Reads (GigaReads)	Bases (GigaBases)	Size (Gigabyte)
L. (M.) martiniquensis	Illumina	SRR13558784 SRR13558792 SRR13558785	2.318	4.153	10.135
	Nanopore	SRR13558786 SRR13558788 SRR13558790 SRR13558793	0.086	4.809	9.684
L. (M.) orientalis	Illumina	SRR13558774 SRR13558775 SRR13558776 SRR13558777 SRR13558778 SRR13558779 SRR13558780 SRR13558781	7.996	9.553	27.760
	Nanopore	SRR13558782	0.054	3.357	6.756
L. (M.) enriettii	Illumina	SRR13558795 SRR13558796 SRR13558797	2.600	4.656	11.365
	Nanopore	SRR13558798	0.072	4.365	8.786
***L. (M.)* sp. Ghana**	Illumina	SRR13558800 SRR13558801 SRR13558802 SRR13558803 SRR13558804	4.844	6.932	18.563
	Nanopore	SRR13558805	0.077	5.390	10.840
***L. (M.)* sp. Namibia**	Illumina	SRR13558764 SRR13558765 SRR13558766	2.858	5.087	12.434
	Nanopore	SRR13558767	0.068	4.377	8.807
Porcisia hertigi	Illumina	SRR13558754 SRR13558755 SRR13558756	2.717	4.654	11.455
	Nanopore	SRR13558757	0.019	1.364	2.742

5.2 Polishing Assemblies

After implementing the optimisation process, chromosome-scale genomes were assembled successfully. This is critical, as it became clear when comparing previous assemblies to the ones created for this project that having a complete set of chromosomes assisted in the annotation process. The total number of contigs (including chromosomes), number of contigs, N50 values, and other assembly statistics can be seen in table 5.4.

When all six genomes were benchmarked using BUSCO – a Benchmark for Universal Single-Copy Orthologs – they all had a BUSCO content of 98 % or higher (Figure 5.4). The reference lineage dataset used, however, contained only 130 BUSCOs. They were collected and curated to be universal across all the Phylum *Euglenozoa* species. The missing BUSCOs across all six assemblies were rapamycin binding domain, GTP-binding protein, ATP-dependent zinc metallopeptidase, metallo-peptidase, and 3-oxo-5-alpha-steroid 4-dehydrogenase.

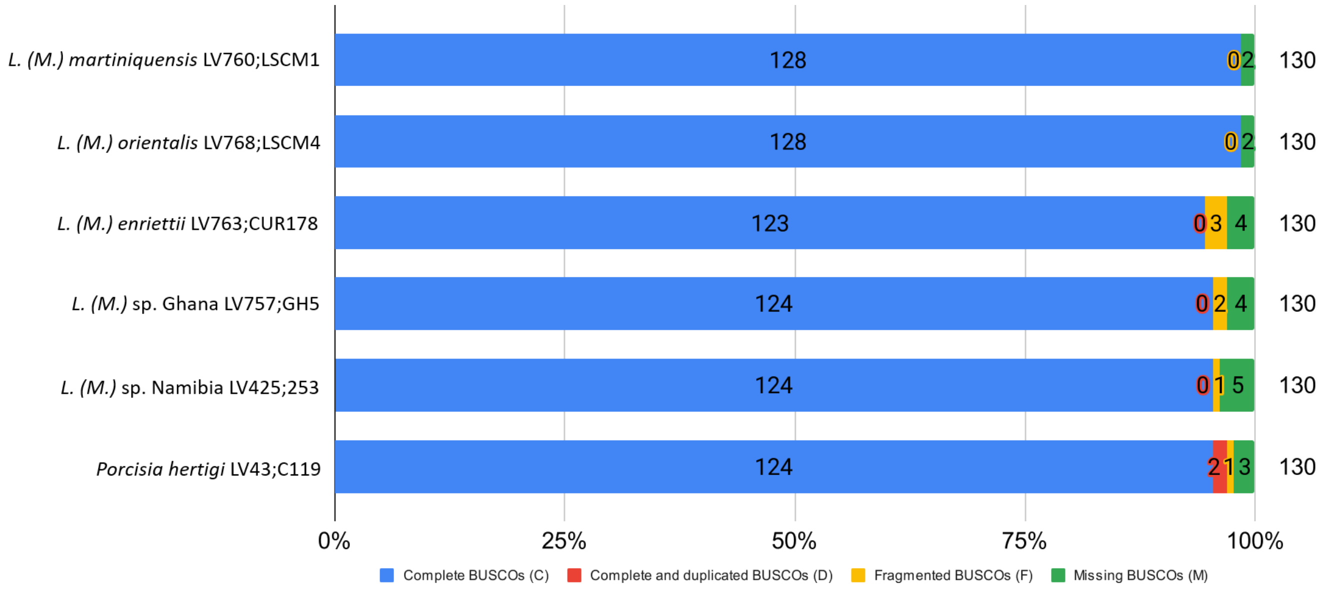


Figure 5.5: Summary of the benchmarking for all six assemblies in BUSCO notation. *Euglenozoa* is the reference lineage dataset (31 species: 130 BUSCOs).

Table 5.4: Summary statistics for all assemblies.

Features	*L. (M.) martiniquensis*	*L. (M.) orientalis*	*L. (M.) enriettii*	***L. (M.)* sp. Ghana**	***L. (M.)* sp. Namibia**	***Porcisia hertigi***
Total number of reads	24,128,044	80,540,904	26,789,424	49,308,106	29,347,348	27,383,632
Number of yield bases (Gbp)	19.24	29.20	19.41	26.93	20.51	13.41
Genome coverage (x)	277.9x	390.7x	271.8x	371.2x	291.5x	177.1x
Total number of scaffolds	42	98	54	116	67	74
Genome size in bases	32,413,670	34,194,276	33,318,864	35,953,538	34,118,624	34,958,538
N50	1,046,741	1,120,138	1,075,649	1,100,365	1,066,046	967,170
GC-content	59.90%	59.70%	59.60%	59.70%	59.50%	56.00%
Number of Ns (% of genome)	50 (0.0002%)	1707 (0.005%)	380 (0.001%)	481 (0.001%)	530 (0.002%)	320 (0.001%)

Although the majority of the contigs were assembled initially chromosome-sized, we used RaGOO — a reference-guided scaffolding tool — to order and align all contigs into chromosome-length scaffolds using Minimap2 alignments between contigs and a reference assembly. It generated 36 pseudomolecules (corresponding to 36 chromosomes) and small unplaced scaffolds for all assemblies (further details in table 5.5).

Table 5.5: Order and length of assemblies prior to and following the use of RaGOO.

Species	**Contigs length Before (Mb)**	**Contigs length after (Mb)**	**% length change**	**Unplaced length in bp**
L. (M.) martiniquensis	72 (32.46)	42 (32.41)	99.84%	70,152
L. (M.) orientalis	171 (34.39)	98 (34.19)	99.44%	257,579
L. (M.) enriettii	100 (33.39)	54 (33.32)	99.80%	76,607
***L. (M.)* sp. Ghana**	158 (36.10)	116 (35.95)	99.59%	1,077,537
***L. (M.)* sp. Namibia**	126 (34.20)	67 (34.12)	99.77%	248,213
Porcisia hertigi	168 (35.22)	74 (34.96)	99.26%	1,892,991

To ensure that the correctness of *Leishmania* genomes, I created a BLAST+ database containing all *Leishmania* species genomic sequences extracted from TriTrypDB, then I queried our assemblies against that database, returning no more than one hit per query and only of Highest Scoring Pairs (HSP). The output was then processed and visualised as a word cloud, with the most frequently occurring word displayed in the largest size (Figure 5.5).

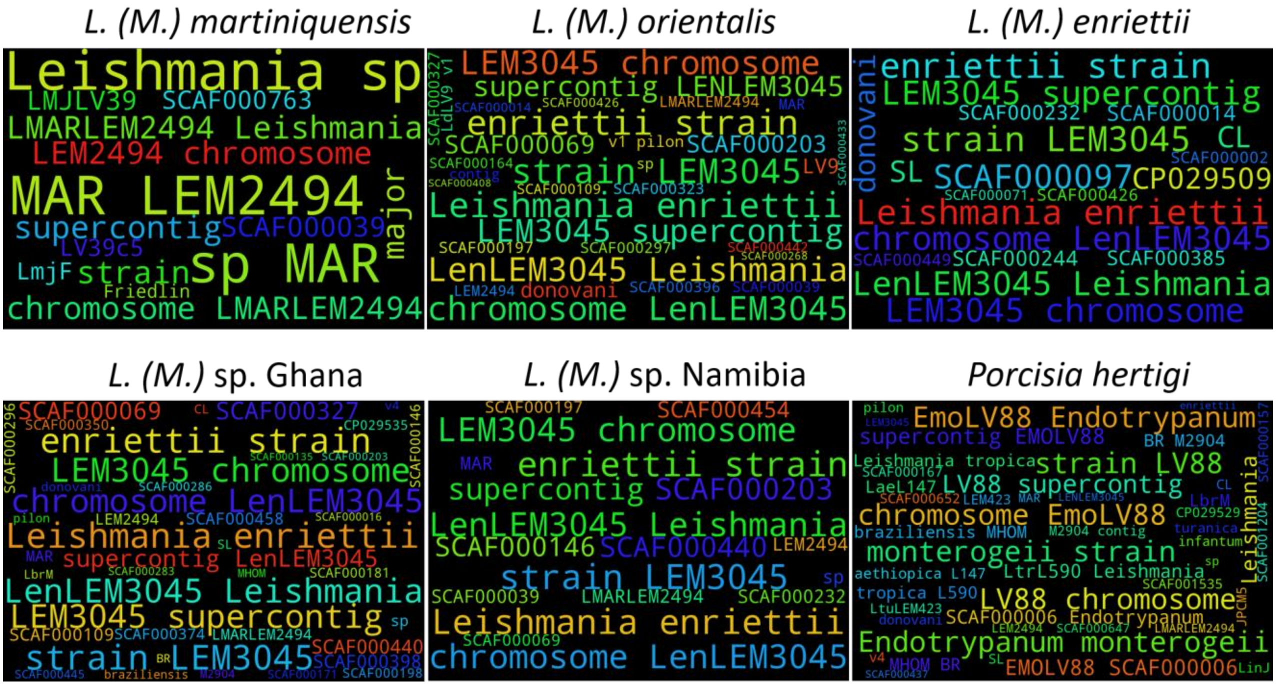


Figure 5.6: A word cloud is used to represent the best hits from BLAST+, with the largest word being the most frequently mentioned.

To determine the assembly's similarity to the closest genome, syntenic dot plots were generated against two public genomes: *L. (M.) enriettii* LEM3045 strain against all assemblies except *P. hertigi*, which was compared to *Endotrypanum monterogeii* strain LV88 (Figure 5.6) as described before, syntenic dot plot is a type of scatterplot in which each axis represents an end-to-end sequence of the genome. Each point on the scatterplot represents a possible homologous match between these two genomes. Syntenic dot plots are a highly useful tool for determining synteny between genomes relating to different taxa. The results indicate that there is a high degree of synteny between all contigs and subject scaffolds. However, when plotted against our assemblies except *P. hertigi*, chromosome 31 from *L. (M.) enriettii* LEM3045 exhibited a typical deletion signal (Figure 5.7 and 5.8). That deletion occurred because of a misassembly artefact in the *L. (M.) enriettii* LEM3045 genome, as the majority of 31 chromosome alignments were matched and located at the unplaced scaffolds, but completely matched the new genome's chromosome 31.

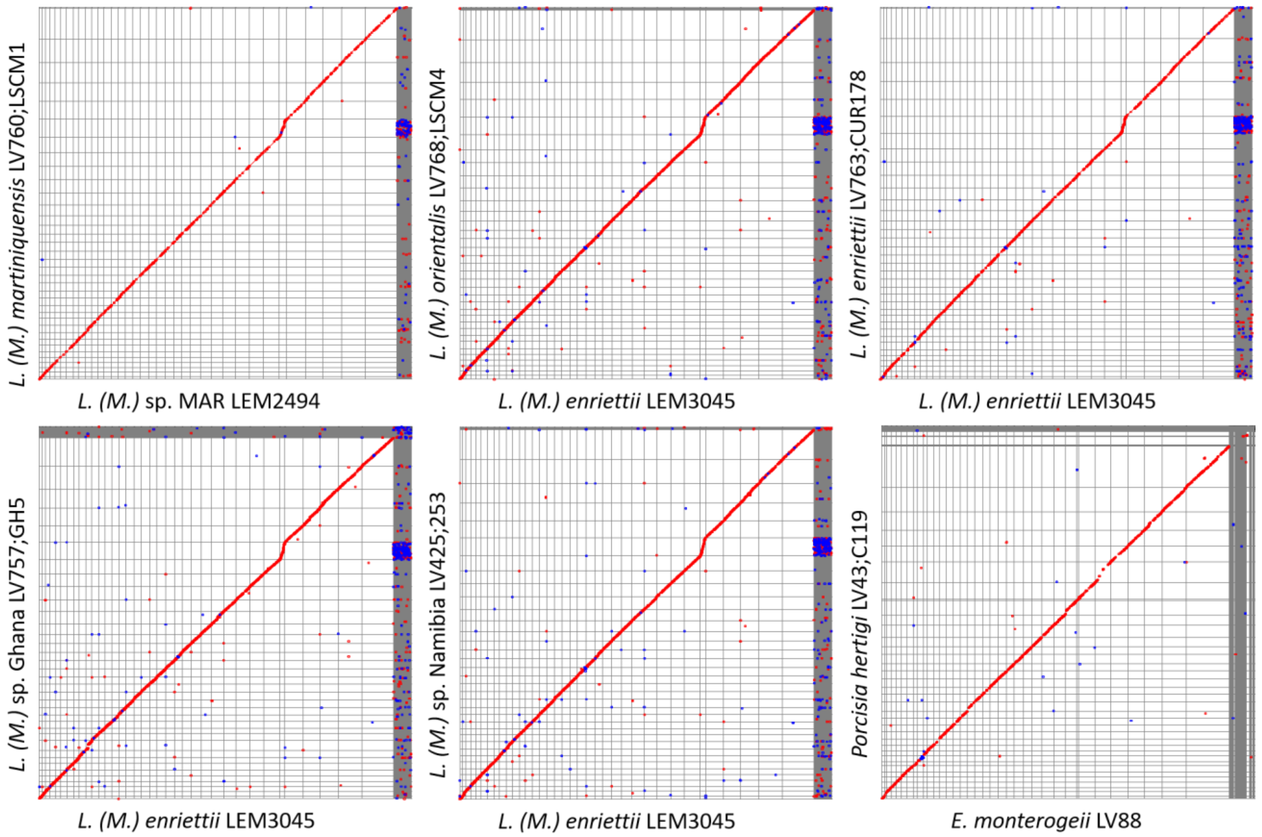


Figure 5.7: dot plot graph shows the similarity between the new assemblies (vertical axis) against three publicly available genome (horizontal axis); *L. (M.) martiniquensis* strain LEM2494*, L. (M.) enriettii* strain LEM3045 and *Endotrypanum monterogeii* strain LV88 (see Supplementary Materials for full-scale figure).

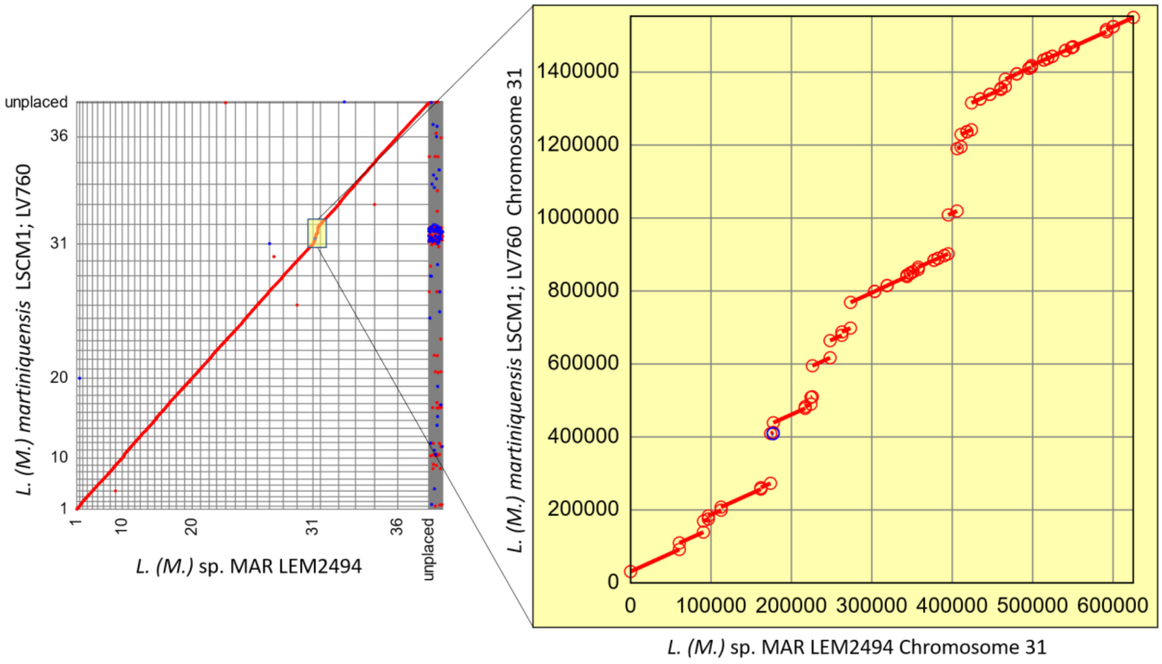


Figure 5.8: comparing our assembly of *L. (M.) martiniquensis* strain LV760 (vertical axis) to *L. (M.) martiniquensis* strain LEM2494 (horizontal axis) with the focus on Chromosome 31 in the yellow square. The chromosomes are ordered by their size rather than their numerical order (see Supplementary Materials for full-scale figure).

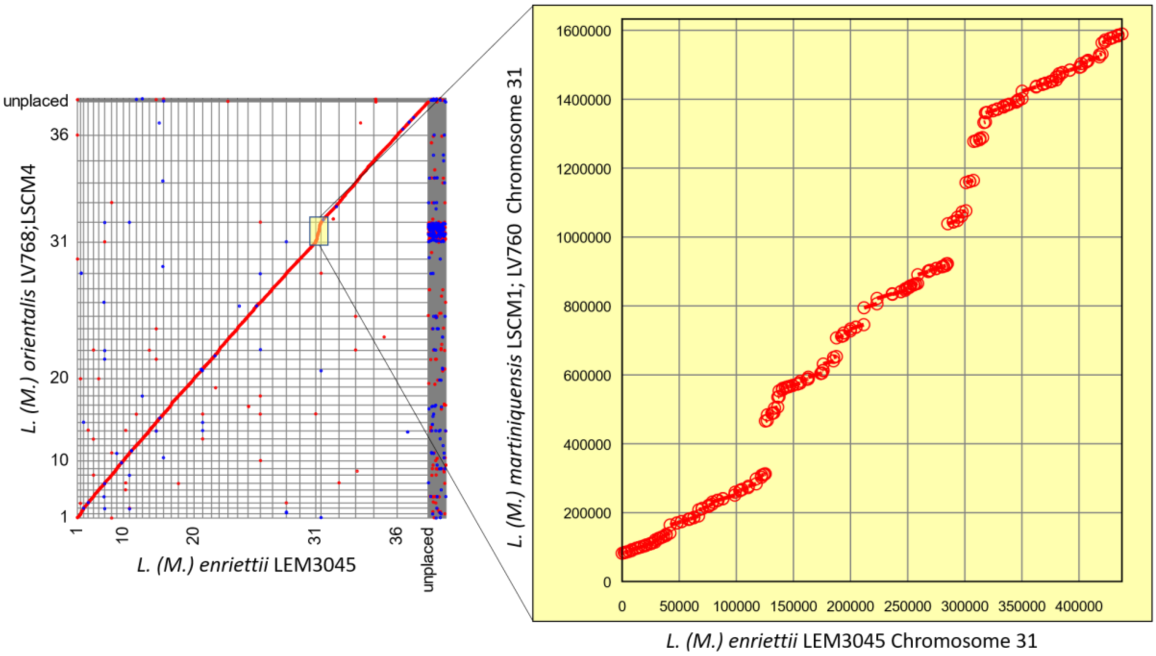
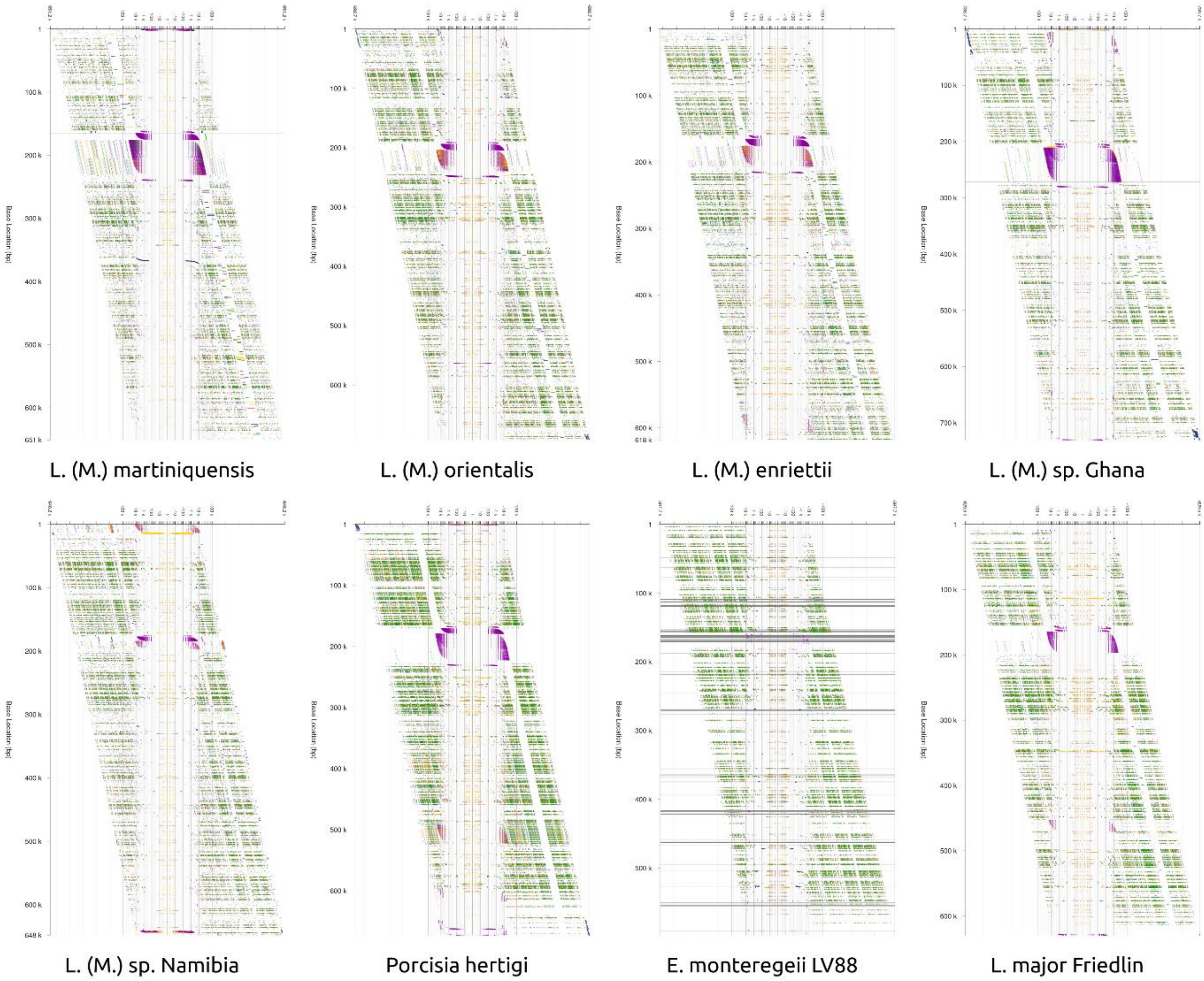


Figure 5.9: comparing the new assembly of *L. (M.) orientalis* strain LV768 (vertical axis) to *L. (M.) enriettii* strain LEM3045 (horizontal axis) with the focus on Chromosome 31 in the yellow square. The chromosomes are ordered by their size rather than their numerical order (see Supplementary Materials for full-scale figure).

At no point during the assembly process were chimeric sequences detected. The only sequences that observed unusual were those found at either ends of the chromosomes. This small number of vector-derived contaminants, such as sequence adaptors, has been removed. Moreover, contaminants were identified through the UniVec Database (Kitts et al., 2011) in which BLAST+ algorithm was used.

5.3 Chromosomal Inspection and Karyogram Visualisation

To further explore syntenic similarity, we visualised repetitive patterns across the chromosomes using the REPAVER script. The repeat patterns revealed some similarities between regions on the same chromosome and between *Leishmania* species. For example, repetitive patterns resembling centromeres were observed in chromosome 15 across all six assemblies and the *L. major* Friedlin strain (Figure 5.9). Additionally, comparable patterns were observed in different chromosomes. The same pattern, however, was mostly absent in the *Endotrypanum monterogeii* strain LV88 genome due to large gaps located exactly in the same corresponding positions. This can be explained by the low assembly quality and the large number of unplaced contigs, the majority of which were repeats that were difficult to assemble.



L. (M.) martiniquensis

L. (M.) orientalis

L. (M.) enriettii

L. (M.) sp. Ghana

L. (M.) sp. Namibia

Porcisia hertigi

E. monteregeii LV88

L. major Friedlin

Figure 5.10: repetitive pattern visualisation of chromosome 15 for 8 species. The vertical axis represents the chromosome length while the horizontal axis represents the repeat distance from the start of the pattern (see Supplementary Materials for full-scale figure).

Additionally, a complementary inspection was performed on all assemblies to ensure that no two chromosomes were unintentionally fused to generate a chimeric chromosome. This was accomplished mostly by using sequence reads, which were remapped to both ends of chromosomes with an intentional gap between them. Across all six assemblies, none of the chromosomes exhibit any read coverage in the gap between the chromosomes. As an example, figure 5.10 showed the reads coverage between chromosomes 22 and 23 including the 100 Ns gap.

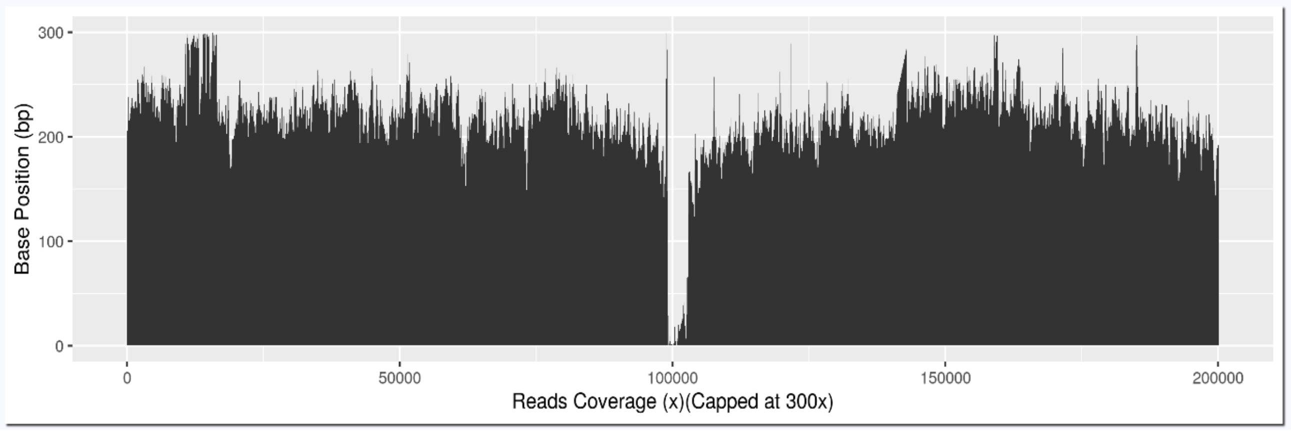


Figure 5.11: read coverage capped at 300x of 10,000 bases at both the end of chromosome 22 and the start of chromosome 23, separated by 100 Ns, which resulted in no overlapping reads over the gap and the coverage dropping to zero (see Supplementary Materials for full-scale figures).

5.4 Assemblies Annotation

5.4.1 Repeat Analysis

At this point, all assemblies have been thoroughly scrutinised and inspected for artefacts or foreign sequences from contamination or vector origin, as well as ordered and oriented using the best reference genome as a guide. As a result, the annotation process began by detecting and classifying repeat regions to initiate the annotation process (Abrusan et al., 2009). Most annotations have been performed using one of two types: evidence-based annotations and *ab initio* gene predictions.

After identifying all repeats with RepeatModeler, the number of repeats representing non-coding Transposable Elements (TEs) varied between 202 and 380 across all assemblies. The majority of these TEs were DNA transposons, followed by long terminal repeats, long and short interspersed nuclear elements, and uncleared repeats, which were labelled as unidentified repeats (Figure 5.12).

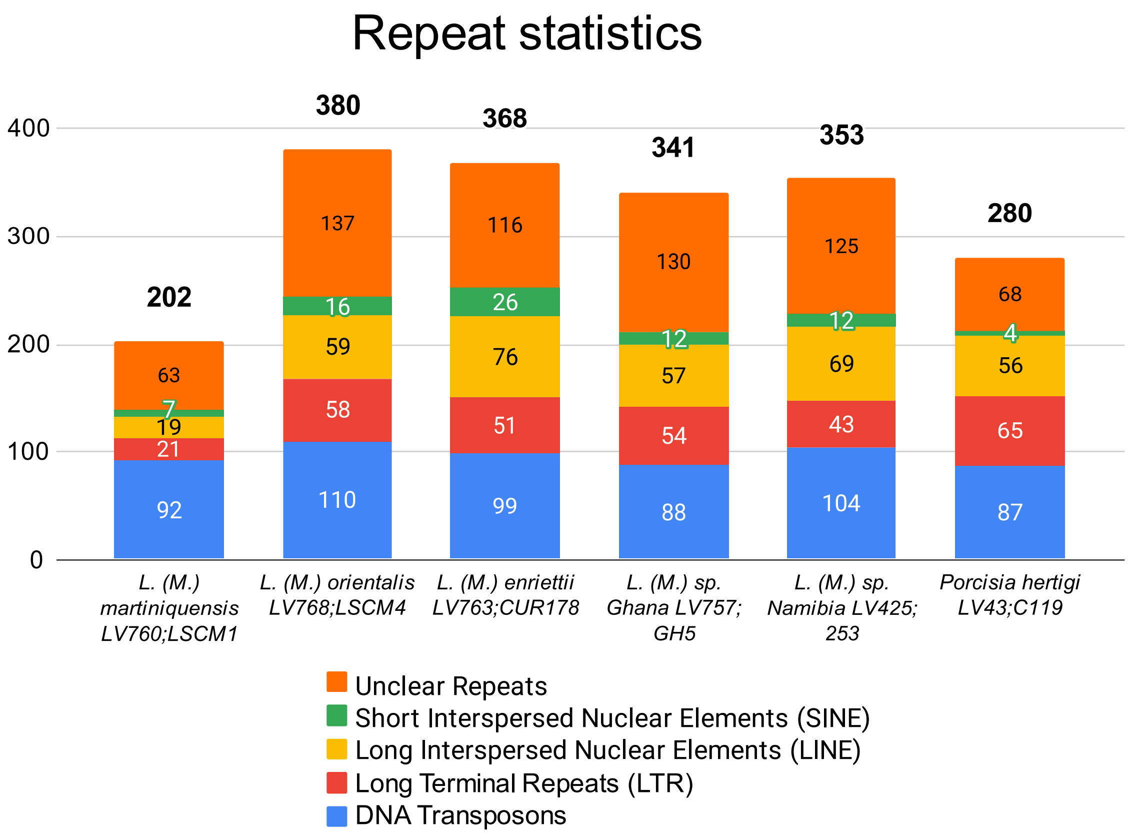


Figure 5.12: Summary graph of the repeats identified in each assembly according to the non-coding transposable elements' type (TEs).

Initially, the repeats were identified and classified for a reason; masking repeats is critical for a smooth annotation process and, more broadly, for assembly. As a result, identifying complex repeats and masking them for annotation must be done, while leaving the simple repeats unmasked because they are part of the coding regions and thus useful for detecting genes and proteins (Figure 5.13).

The interspersed repeat landscape then was estimated using the Kimura 2-Parameter (K2P) model (Kimura, 1980), which is included in the RepeatMasker package as a ready-to-use utility script (Figure 5.14). The interspersed repeat landscape is useful for revealing copy-divergence between different TE classes. The landscape's vertical axis depicts the percentages of various TEs in the genome, which are sorted according to their Kimura values on the horizontal axis. The older copies are located on the right side of the graph, whereas more recent copies are on the left.

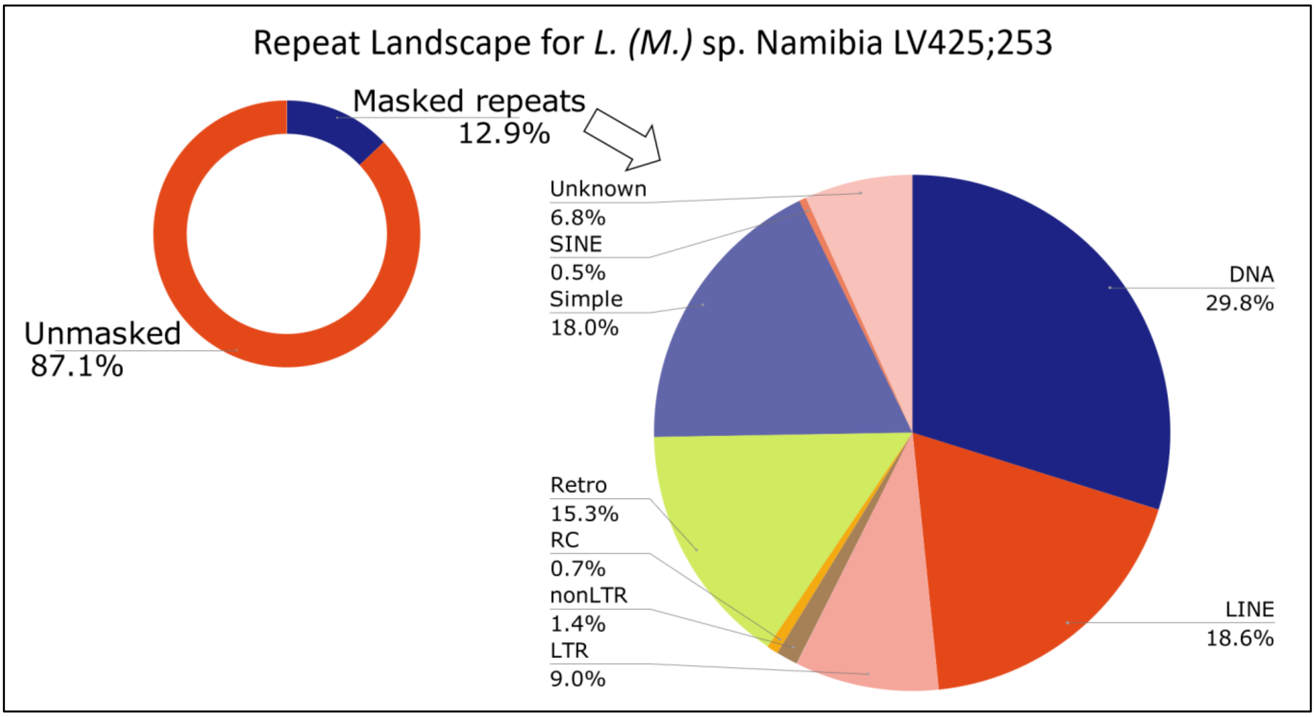


Figure 5.13: percentages chart of repeats in *L. (M.)* sp. Namibia LV425;253 as an example. SINE: Short Interspersed Nuclear Elements; LTR: Long Terminal Repeat Elements; LINE: Long Interspersed Nuclear Elements; and DNA: DNA Repeat Elements (see Supplementary Materials for other genomes).

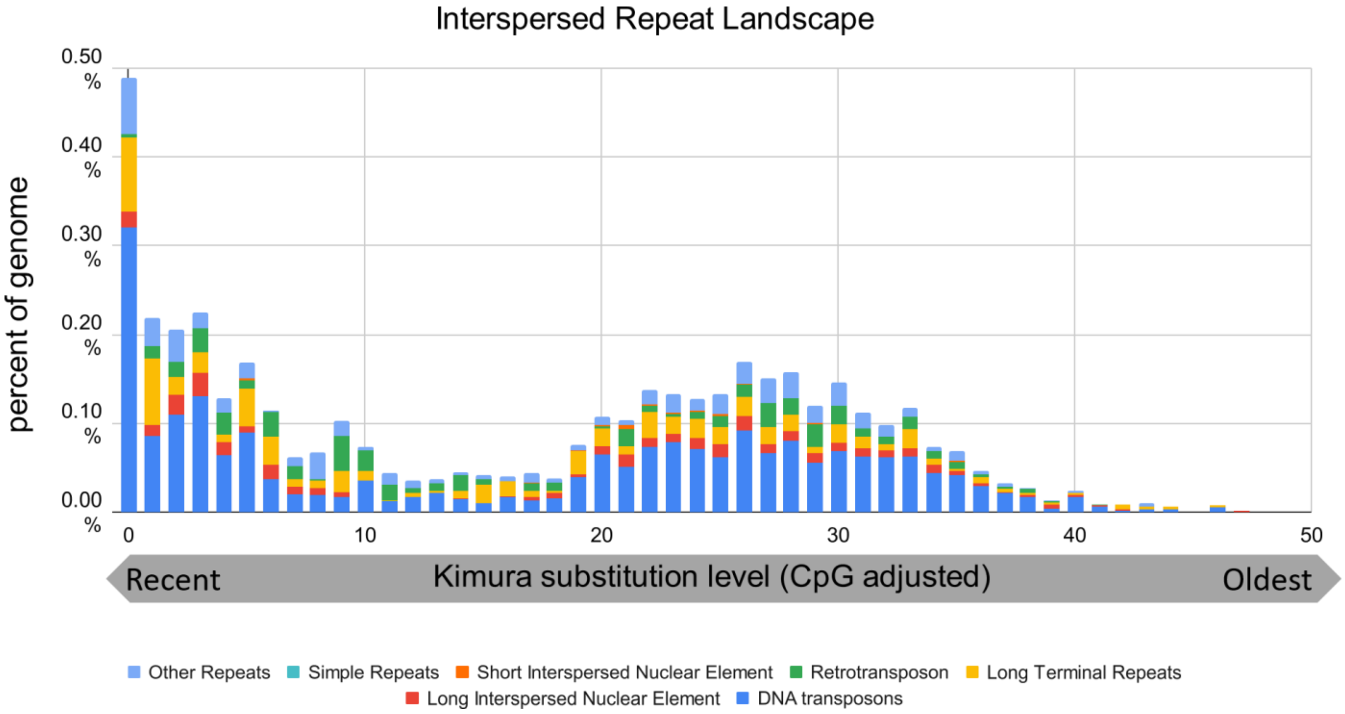


Figure 5.14: Histogram below shows Kimura distance values (X-axis; from 0 to 50; pairwise substitutions/site) done for *L. (M.)* sp. Namibia as an example, for each TE class in relation to the number of copies in the genome (Y-axis; % of genome). Peaks represent insertion waves of elements into the genome. Older TEs insertion waves are shown on the right side of the graph, while newer insertions are on the left side. Different colours show distinct TEs classes, as described at the legend (see Supplementary Materials for other genomes).

5.4.2 Evidence-based and *ab initio* Annotations

The process of evidence-based annotation began by collecting Annotated Proteins and Transcripts, as well as General Feature Format files (GFF) corresponding to the genomic features and its coordinates. All evidence were collected from 14 *Leishmania* species; *L. aethiopica*, *L. amazonensis*, *L. arabica*, *L. braziliensis*, *L. donovani*, *L. enriettii*, *L. gerbilli*, *L. infantum*, *L. major*, *L. mexicana*, *L. panamensis*, *L. tarentolae*, *L. tropica*, and *L. turanica*. Moreover, they were in a total of 68 files and were obtained from the 47th release of TriTrypDB.

As mentioned previously, to determine the quality of each annotation round, Annotation Edit Distance (AED) was used (Figure 5.15). The AED model, developed for Sequence Ontology, is the best model for supported gene annotation discrepancies (Eilbeck et al., 2009). AED was initially created to quantify the changes made to individual annotations between releases. This performance metric was widely used in the field of gene prediction. However, the best feature is its ability to address previously unaddressed sensitivity and specificity measures, such as alternative splicing (Burset and Guigo, 1996). AED quantifies the extent to which an annotation undergoes structural changes, such as agreement between intron-exon annotated structures.

In this case, multiple rounds of annotations were considered as releases and, therefore, AED was used for each round of genomes annotation. AED is used to determine the degree to which an annotation corresponds to the evidence that supports it. It is a numeric value between 0 and 1, with one indicating perfect concordance with available evidence and zero indicating complete lack of support for the annotated gene model (Figure 5.15). In other words, the AED score indicates the degree to which each annotated transcript corresponds to the supporting evidence.

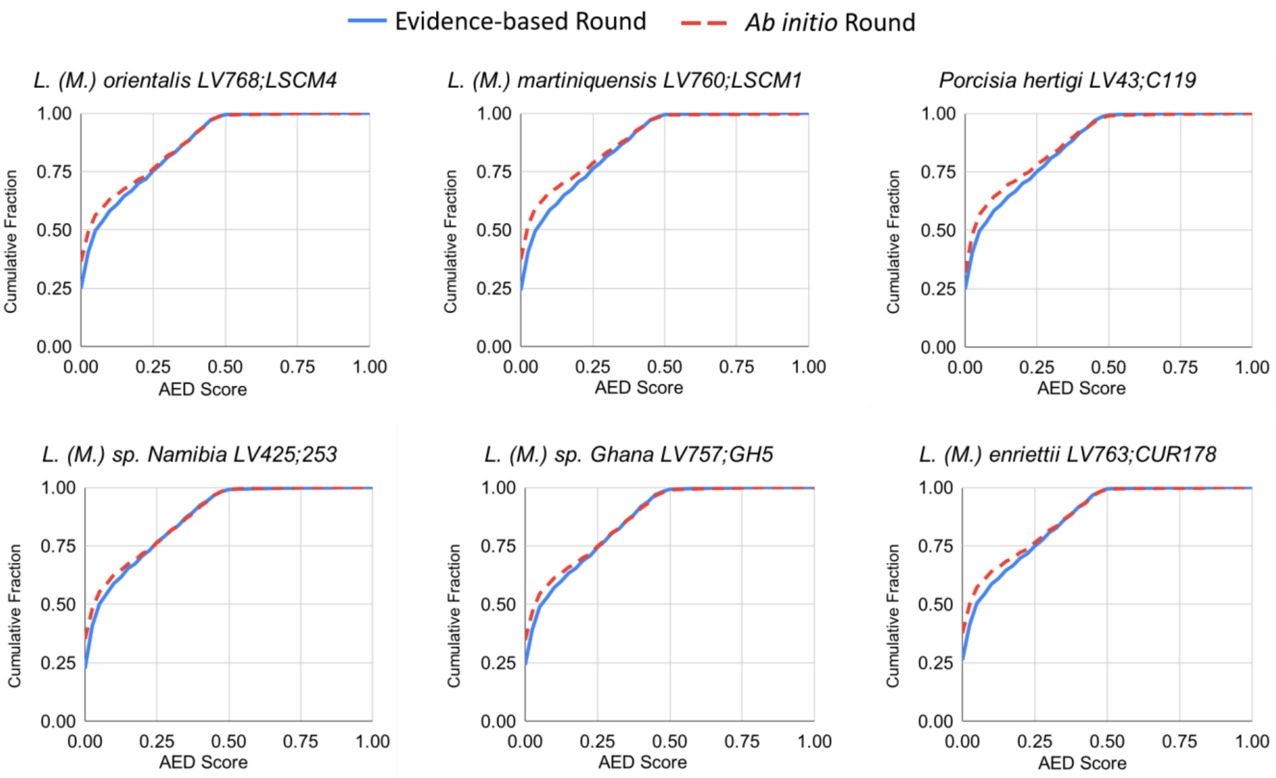


Figure 5.15: Annotation Edit Distance (AED) Scores for both evidence-based and *ab initio* rounds for all assemblies, visualised.

In all annotations, AED scores for *ab initio* rounds were slightly higher than those for evidence-based rounds (Figure 5.15). This can be explained using pre-built training set for *L. tarentolae* with AUGUSTUS predictor software, as it can predict the 5'UTR and 3'UTR including introns more accurately in its trained sets.

After the two rounds are complete, I ended up with a range of 7891 to 8535 annotated genes at a density of 225.7 to 250.7 genes/Mbp (Figure 5.16 and Table 5.6). Additionally, multiple files containing proteins and transcripts in FASTA format were generated, as well as GFF files containing genomic coordinates for each annotated genome. All annotations were visualised globally in relation to chromosomes in accordance with their feature type (Figure 5.17 of *L. (M.) martiniquensis*, LSCM1; LV760 as an example).

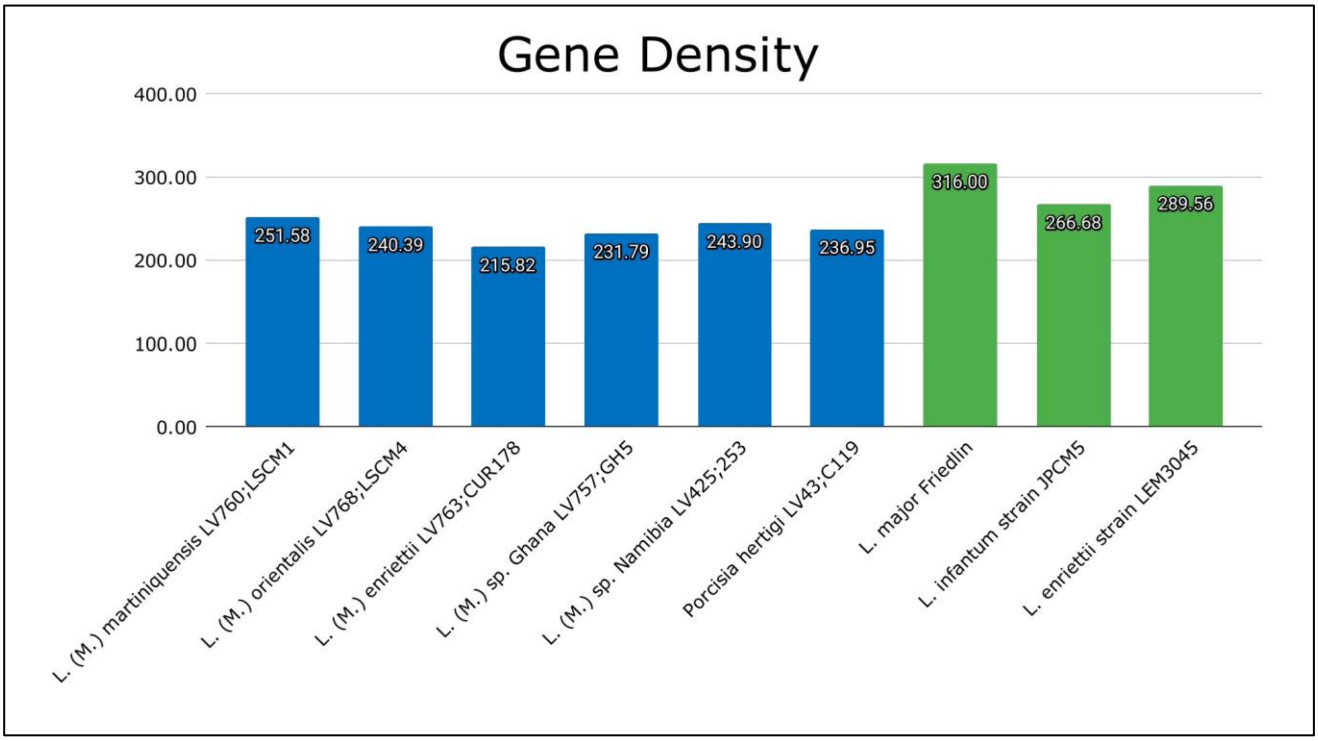


Figure 5.16: Total Gene Density (Genes/Mbp) of our assemblies (blue) in comparison to other publicly available *Leishmania* genomes (green).

Table 5.6: Annotation statistics for each assembly after Evidence-based and *ab initio* rounds.

Features	***L. (M.) martiniquensis***	***L. (M.) orientalis***	***L. (M.) enriettii***	***L. (M.)* sp. Ghana**	***L. (M.)* sp. Namibia**	***Porcisia hertigi***
Number of genes	7,967	8,158	8,353	8,119	8,266	7,891
Gene density in (genes/Mb)	245.8	238.6	250.7	225.8	242.3	225.7
Number of exons	7,969	8,488	8,584	8,119	8,529	8,270
Mean gene length	1,857	1,938	1,897	1,838	1,919	1,908
CDSs total length (Mb) (genome%)	14.80 (45.66%)	15.40 (45.05%)	15.46 (46.40%)	14.92 (41.51%)	15.48 (45.37%)	14.70 (42.06%)

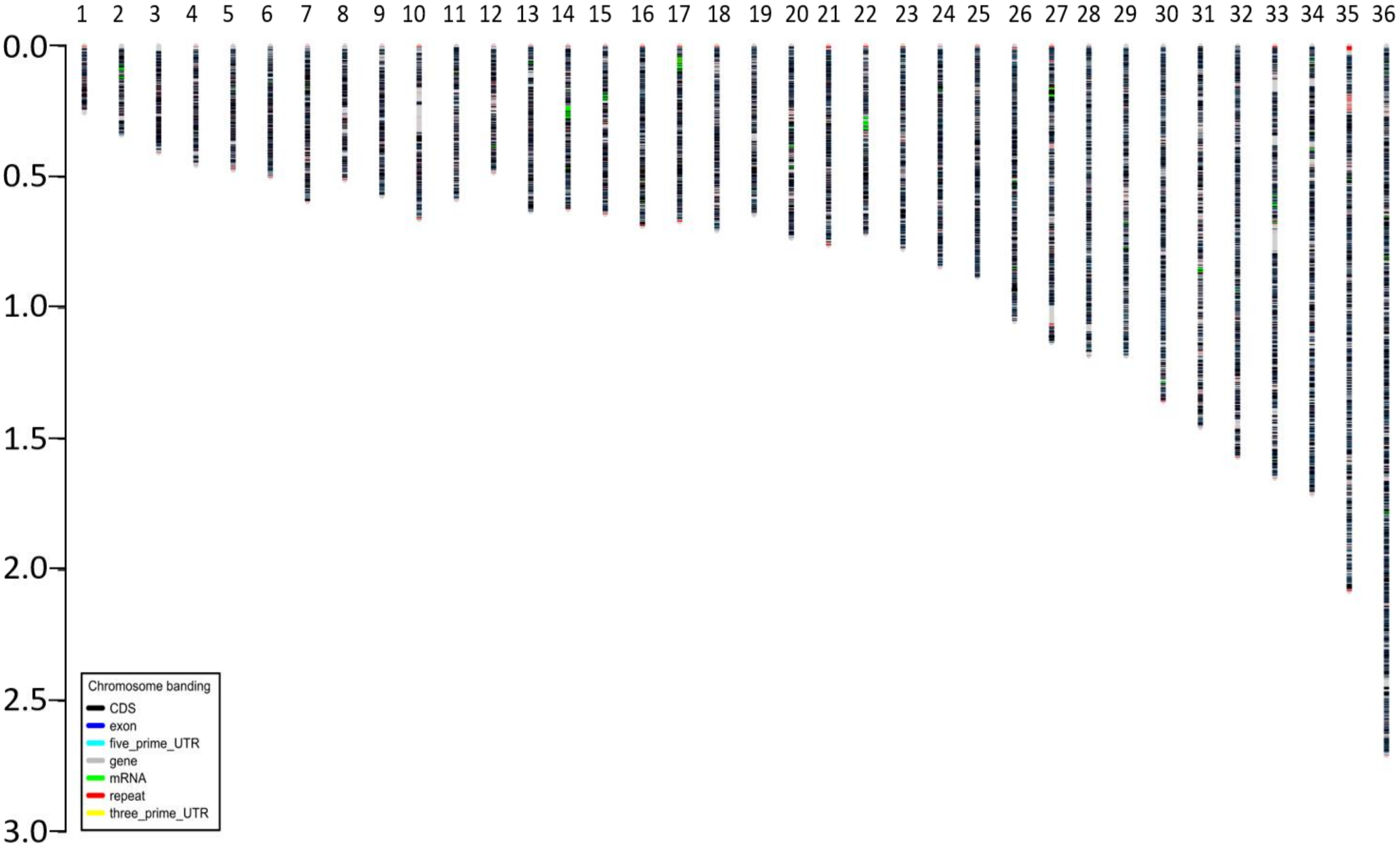


Figure 5.17: Visualisation of chromosomal annotations of *L. (M.) martiniquensis* (LSCM1), showing how and where they were located within each chromosome and coloured according to their type as indicated in legends. (Higher resolution figure for each genome can be seen in Supplementary Materials).

5.4.3 Functional Annotation Assignment

I used the InterProScan software on both the proteins and transcripts outputs to assist in assigning putative functions derived from the Pfam database collection of protein family domains and the Gene Ontology (GO) Term Enrichment database; as well as the UniProtKB/Swiss-Prot database, a high-quality manually annotated and nonredundant protein sequence database; Then I used the output of that to produce a word cloud (Figure 5.18).

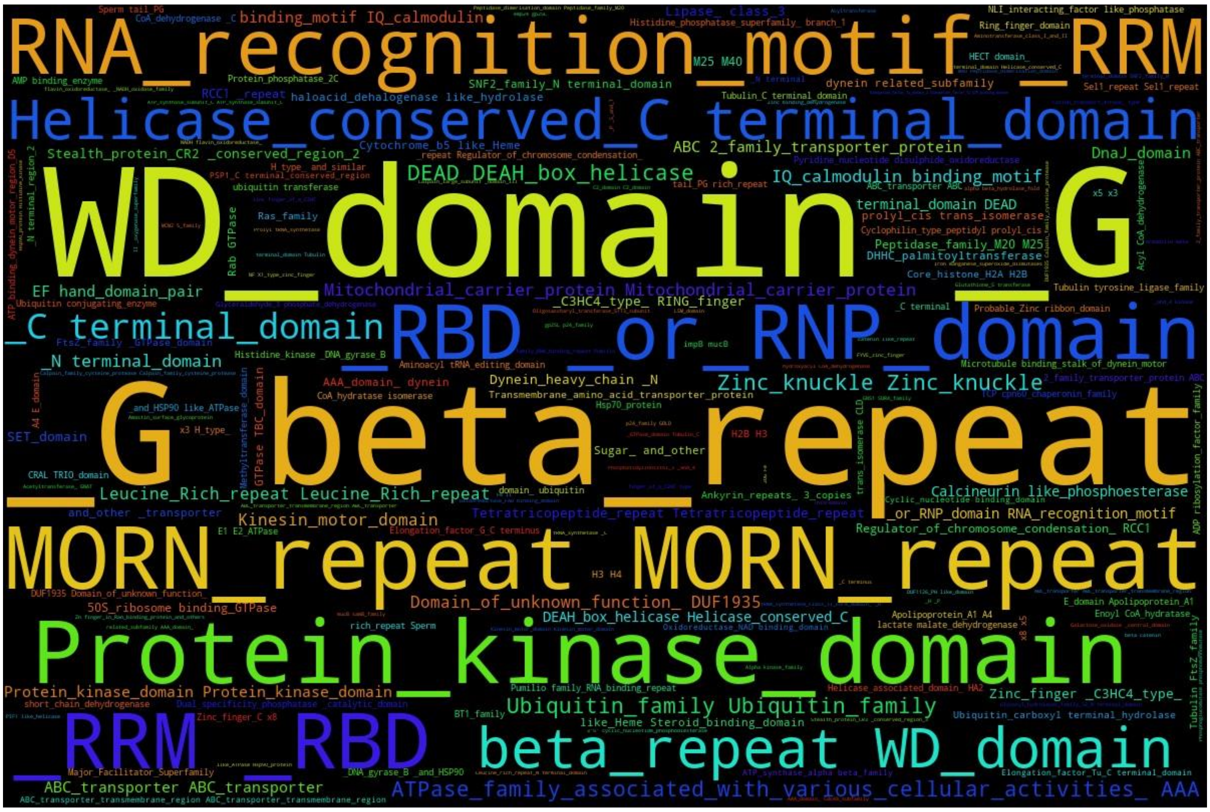


Figure 5.18: Word cloud of annotations that represents the assignment of functional annotations to all annotated features across all assemblies.

Unfortunately, none of the annotation assignments were used in the final published genomes since they, by definition, violated protein nomenclature guidelines because they were done based on similarity rather than experimental confirmation (Uhlen et al., 2010). Instead, they were all labelled as “hypothetical proteins” because no proof of confirmation has yet been reported.

Chapter 6. Comparative Analyses Results

6.1 Phylogenomic Trees

At this point, having completed the annotation process, the genomes were ready for a detailed comparison with other *Leishmania* species. To accomplish this, two approaches were chosen. The first is to reconstruct a chromosome-scale Bayesian phylogenomic tree to assess and estimate the time to the most recent common ancestor (TMRCA), as well as to establish the topology order for the new assemblies within the *Leishmaniinae* family tree. The second approach, which will be detailed in the next section, is to take advantage of the new annotated proteins and combine them with other publicly available proteomes to perform a deeper analysis on how these proteins are grouped in terms of similarity, and then infer a phylogenetic orthology species tree.

Chromosome-scale construction phylogeny has several benefits and drawbacks. For a project of this scale, the disadvantages of using Bayesian phylogenomic analysis are that the alignment process is computationally intensive because each set of chromosomes must be aligned individually, as some chromosomes exceed 2.7 Mb in length. Therefore, it necessitates a significant amount of computational power. The best benefit, however, is that it provides a high-resolution phylogeny. Nonetheless, some pre-start optimisations are required.

Although phylogenomic analysis was done on two different datasets, the optimisation parameters were similar due to the high degree of synteny between all genomes (Raymond et al., 2012). As previously indicated, each genome assembly, including the new six genomes, was divided into individual chromosomes, and then combined with the corresponding chromosomes of the other *Leishmania* species for alignment. All optimizations were carried out on the shortest chromosome, and once they are satisfactory, the analyses were expanded to include all chromosome sets and eventually produced a consensus tree using all 36 chromosomes.

Assessing multiple substitution models was done to rank all possible model combinations based on their BIC scores (Bayesian Information Criterion) as well as *lnL* (Log Likelihood value). This was all done using the MEGAX software (Kumar et al., 2018) (Table 6.1). The best substitution model with the lowest BIC and *lnL* scores was the General Time Reversible model with discrete Gamma distribution (GTR+G), which had the lowest BIC and *lnL* value (Table 6.1).

Table 6.1: Maximum Likelihood fits of top 5 out of 24 nucleotide substitution models sorted based on the lowest BIC scores. Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter.

Model	**Parameters**	**BIC**	**AIC**	***lnL***
GTR+G	40	4146179.336	4145645.182	-2072782.591
GTR+G+I	41	4146194.69	4145647.182	-2072782.591
HKY+G	36	4148192.742	4147712.004	-2073820.002
TN93+G	37	4148230.392	4147736.3	-2073831.15
T92+G	34	4148254.574	4147800.543	-2073866.271

Following that, the maximum likelihood substitution matrix was estimated using the GTR+G model, which uses a discrete Gamma distribution to explain evolutionary rate variation across sites (5 categories, [+G], alpha parameter = 1.2649). The maximum *lnL* for this computation was -2,072,622.483. This analysis involved 17 nucleotide sequences. There was a total of 604,328 positions in the optimisation dataset. However, no significant differences were found when the same estimation was done on the second dataset that contained 60 species.

After that, using the SplitsTree4 software, the alignments output were examined for recombination (Huson and Bryant, 2005). No signal of recombination was found in any of the 36 sets of chromosome alignments. Additionally, the temporal interval between *L. major* (Friedlin) *vs L. infantum* (JPCM5), which was estimated to be 19.6 MYA (14.6 - 24.7 MYA) using TimeTree (Lukes et al., 2007b, Harkins et al., 2016, Kumar et al., 2018), was used as a prior in constructing the phylogenomic tree.

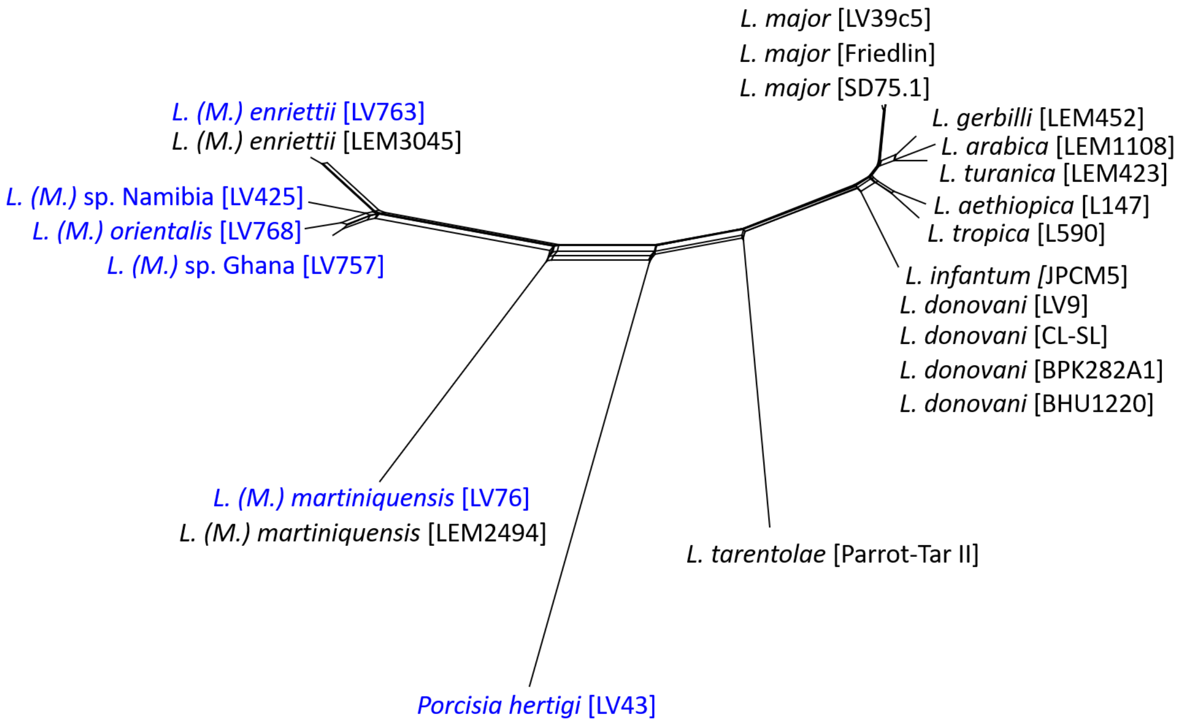
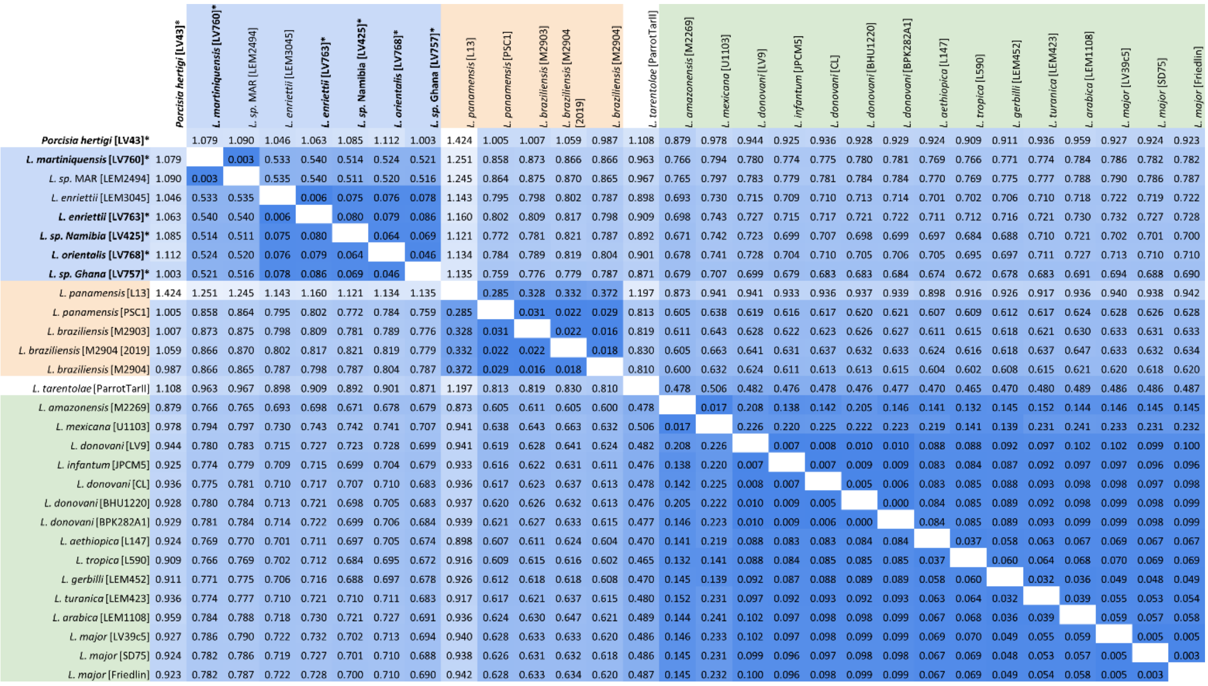


Figure 6.1: Phylogenetic network of chromosome 5 for all 22 species of *Leishmania*. The blue labelled species are the newly assembled genomes (see *Supplementary Materials* for full-scale figure per chromosome).



	Porcisia hertigi* [LV43]	***L. martiniquensis* [LV760]***	*L. sp.* MAR [LEM2494]	*L. enriettii* [LEM3045]	***L. enriettii* [LV763]***	***L. sp.* Namibia [LV425]***	***L. orientalis* [LV768]***	***L. sp.* Ghana [LV757]***	*L. panamensis* [L13]	*L. panamensis* [PSC1]	*L. braziliensis* [M2903]	*L. braziliensis* [M2904 [2019]	*L. braziliensis* [M2904]	*L. tarentolae* [ParrotTarII]	*L. amazonensis* [M2269]	*L. mexicana* [U1103]	*L. donovani* [LV9]	*L. infantum* [JPCM5]	*L. donovani* [CL]	*L. donovani* [BHU1220]	*L. donovani* [BPK282A1]	*L. aethiopica* [L147]	*L. tropica* [L590]	*L. gerbilli* [LEM452]	*L. turanica* [LEM423]	*L. arabica* [LEM1108]	*L. major* [LV39c5]	*L. major* [SD75]	*L. major* [Friedlin]
Porcisia hertigi* [LV43]		1.079	1.090	1.046	1.063	1.085	1.112	1.003	1.424	1.005	1.007	1.059	0.987	1.108	0.879	0.978	0.944	0.925	0.936	0.928	0.929	0.924	0.909	0.911	0.936	0.959	0.927	0.924	0.923
L. martiniquensis* [LV760]	1.079		0.003	0.533	0.540	0.514	0.524	0.521	1.251	0.858	0.873	0.866	0.866	0.963	0.766	0.794	0.780	0.774	0.775	0.780	0.781	0.769	0.766	0.771	0.774	0.784	0.786	0.782	0.782
L. sp. MAR [LEM2494]	1.090	0.003		0.535	0.540	0.511	0.520	0.516	1.245	0.864	0.875	0.870	0.865	0.967	0.765	0.797	0.783	0.779	0.781	0.784	0.784	0.770	0.769	0.775	0.777	0.788	0.790	0.786	0.787
L. enriettii [LEM3045]	1.046	0.533	0.535		0.006	0.075	0.076	0.078	1.143	0.795	0.798	0.802	0.787	0.898	0.693	0.730	0.715	0.709	0.710	0.713	0.714	0.701	0.702	0.706	0.710	0.718	0.722	0.719	0.722
L. enriettii* [LV763]	1.063	0.540	0.540	0.006		0.080	0.079	0.086	1.160	0.802	0.809	0.817	0.798	0.909	0.698	0.743	0.727	0.715	0.717	0.721	0.722	0.711	0.712	0.716	0.721	0.730	0.732	0.727	0.728
L. sp. Namibia* [LV425]	1.085	0.514	0.511	0.075	0.080		0.064	0.069	1.121	0.772	0.781	0.821	0.787	0.892	0.671	0.742	0.723	0.699	0.707	0.698	0.699	0.697	0.684	0.688	0.710	0.721	0.702	0.701	0.700
L. orientalis* [LV768]	1.112	0.524	0.520	0.076	0.079	0.064		0.046	1.134	0.784	0.789	0.819	0.804	0.901	0.678	0.741	0.728	0.704	0.710	0.705	0.706	0.705	0.695	0.697	0.711	0.727	0.713	0.710	0.710
L. sp. Ghana* [LV757]	1.003	0.521	0.516	0.078	0.086	0.069	0.046		1.135	0.759	0.776	0.779	0.787	0.871	0.679	0.707	0.699	0.679	0.683	0.683	0.684	0.674	0.672	0.678	0.683	0.691	0.694	0.688	0.690
L. panamensis [L13]	1.424	1.251	1.245	1.143	1.160	1.121	1.134	1.135		0.285	0.328	0.332	0.372	1.197	0.873	0.941	0.941	0.933	0.936	0.937	0.939	0.898	0.916	0.926	0.917	0.936	0.940	0.938	0.942
L. panamensis [PSC1]	1.005	0.858	0.864	0.795	0.802	0.772	0.784	0.759	0.285		0.031	0.022	0.029	0.813	0.605	0.638	0.619	0.616	0.617	0.620	0.621	0.607	0.609	0.612	0.617	0.624	0.628	0.626	0.628
L. braziliensis [M2903]	1.007	0.873	0.875	0.798	0.809	0.781	0.789	0.776	0.328	0.031		0.022	0.016	0.819	0.611	0.643	0.628	0.622	0.623	0.626	0.627	0.611	0.615	0.618	0.621	0.630	0.633	0.631	0.633
L. braziliensis [M2904 [2019]	1.059	0.866	0.870	0.802	0.817	0.821	0.819	0.779	0.332	0.022	0.022		0.018	0.830	0.605	0.663	0.641	0.631	0.637	0.632	0.633	0.624	0.616	0.618	0.637	0.647	0.633	0.632	0.634
L. braziliensis [M2904]	0.987	0.866	0.865	0.787	0.798	0.787	0.804	0.787	0.372	0.029	0.016	0.018		0.810	0.600	0.632	0.624	0.611	0.613	0.613	0.615	0.604	0.602	0.608	0.615	0.621	0.620	0.618	0.620
L. tarentolae [ParrotTarII]	1.108	0.963	0.967	0.898	0.909	0.892	0.901	0.871	1.197	0.813	0.819	0.830	0.810		0.478	0.506	0.482	0.476	0.478	0.476	0.477	0.470	0.465	0.470	0.480	0.489	0.486	0.486	0.487
L. amazonensis [M2269]	0.879	0.766	0.765	0.693	0.698	0.671	0.678	0.679	0.873	0.605	0.611	0.605	0.600	0.478		0.017	0.208	0.138	0.142	0.205	0.146	0.141	0.132	0.145	0.152	0.144	0.146	0.145	0.145
L. mexicana [U1103]	0.978	0.794	0.797	0.730	0.743	0.742	0.741	0.707	0.941	0.638	0.643	0.663	0.632	0.506	0.017		0.226	0.220	0.225	0.222	0.223	0.219	0.141	0.139	0.231	0.241	0.233	0.231	0.232
L. donovani [LV9]	0.944	0.780	0.783	0.715	0.727	0.723	0.728	0.699	0.941	0.619	0.628	0.641	0.624	0.482	0.208	0.226		0.007	0.008	0.010	0.010	0.088	0.088	0.092	0.097	0.102	0.102	0.099	0.100
L. infantum [JPCM5]	0.925	0.774	0.779	0.709	0.715	0.699	0.704	0.679	0.933	0.616	0.622	0.631	0.611	0.476	0.138	0.220	0.007		0.007	0.009	0.009	0.083	0.084	0.087	0.092	0.097	0.097	0.096	0.096
L. donovani [CL]	0.936	0.775	0.781	0.710	0.717	0.707	0.710	0.683	0.936	0.617	0.623	0.637	0.613	0.478	0.142	0.225	0.008	0.007		0.005	0.006	0.083	0.085	0.088	0.093	0.098	0.098	0.097	0.098
L. donovani [BHU1220]	0.928	0.780	0.784	0.713	0.721	0.698	0.705	0.683	0.937	0.620	0.626	0.632	0.613	0.476	0.205	0.222	0.010	0.009	0.005		0.000	0.084	0.085	0.089	0.092	0.098	0.099	0.098	0.099
L. donovani [BPK282A1]	0.929	0.781	0.784	0.714	0.722	0.699	0.706	0.684	0.939	0.621	0.627	0.633	0.615	0.477	0.146	0.223	0.010	0.009	0.006	0.000		0.084	0.085	0.089	0.093	0.099	0.099	0.098	0.099
L. aethiopica [L147]	0.924	0.769	0.770	0.701	0.711	0.697	0.705	0.674	0.898	0.607	0.611	0.624	0.604	0.470	0.141	0.219	0.088	0.083	0.083	0.084	0.084		0.037	0.058	0.063	0.067	0.069	0.067	0.067
L. tropica [L590]	0.909	0.766	0.769	0.702	0.712	0.684	0.695	0.672	0.916	0.609	0.615	0.616	0.602	0.465	0.132	0.141	0.088	0.084	0.085	0.085	0.085	0.037		0.060	0.064	0.068	0.070	0.069	0.069
L. gerbilli [LEM452]	0.911	0.771	0.775	0.706	0.716	0.688	0.697	0.678	0.926	0.612	0.618	0.618	0.608	0.470	0.145	0.139	0.092	0.087	0.088	0.089	0.089	0.058	0.060		0.032	0.036	0.049	0.048	0.049
L. turanica [LEM423]	0.936	0.774	0.777	0.710	0.721	0.710	0.711	0.683	0.917	0.617	0.621	0.637	0.615	0.480	0.152	0.231	0.097	0.092	0.093	0.092	0.093	0.063	0.064	0.032		0.039	0.055	0.053	0.054
L. arabica [LEM1108]	0.959	0.784	0.788	0.718	0.730	0.721	0.727	0.691	0.936	0.624	0.630	0.647	0.621	0.489	0.144	0.241	0.102	0.097	0.098	0.098	0.099	0.067	0.068	0.036	0.039		0.059	0.057	0.058
L. major [LV39c5]	0.927	0.786	0.790	0.722	0.732	0.702	0.713	0.694	0.940	0.628	0.633	0.633	0.620	0.486	0.146	0.233	0.102	0.097	0.098	0.099	0.099	0.069	0.070	0.049	0.055	0.059		0.005	0.005
L. major [SD75]	0.924	0.782	0.786	0.719	0.727	0.701	0.710	0.688	0.938	0.626	0.631	0.632	0.618	0.486	0.145	0.231	0.099	0.096	0.097	0.098	0.098	0.067	0.069	0.048	0.053	0.057	0.005		0.003
L. major [Friedlin]	0.923	0.782	0.787	0.722	0.728	0.700	0.710	0.690	0.942	0.628	0.633	0.634	0.620	0.487	0.145	0.232	0.100	0.096	0.098	0.099	0.099	0.067	0.069	0.049	0.054	0.058	0.005	0.003	

Figure 6.2: Estimates of Evolutionary Divergence between Sequences. The average number of base substitutions per site from between sequences are shown. Analyses were conducted using the Maximum Composite Likelihood model (Tamura et al., 2004). This analysis involved 29 sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 781282 positions in this dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018) (see Supplementary Materials for more details).

Along with the optimal parameters, multiple tree construction variations were configured using the BEAST software package to achieve convergence by experimenting with different molecular clock speeds and tree priors. Strict molecular clock and Yule process speciation was chosen for tree priors (Gernhard, 2008, Bouckaert et al., 2014). Then, all chromosome sets were subjected to the same parameters, and ran for ten million states. Each converged tree for each chromosome set was merged into a single tree using the default parameters of TreeAnnotator. The phylogeny of *Leishmania* parasites' genomes was investigated using a consensus tree constructed from 36 sets of chromosome sequences from our six assemblies and the other chromosome-scale genomes.

The consensus phylogenomic tree for the first dataset, which includes only chromosome-scale and complete genome assemblies from TriTrypDB, and for all 36 chromosomal sets is depicted in Figure 6.3 and 6.4. The tree placed the new six assemblies outside the *Leishmania* subgenus, and the time to their most recent common ancestor (TMRCA) was estimated to be

105.11 Million Years Ago (MYA) with the 95% Highest Posterior Density (HPD) ranged between 84.62 and 138.48 MYA; and between *L. (M.) martiniquensis* strains (LSCM1 and LEM2494) and the rest of *Mundinia* species estimated to be 62.25 MYA (95% HPD: 48.40 - 86.78 MYA). *Porcisia hertigi* was identified as an outgroup from the genus *Leishmania* as expected. The TMRCA between *P. hertigi* and the genus *Leishmania* is estimated to be 131.52 MYA (95% HPD: 100.10 - 182.58 MYA). Moreover, the TMRCA for the subgenus *Sauroleishmania* and *Leishmania* is estimated to be 57.73 MYA (95% HPD: 46.12 - 75.3 MYA), suggesting that *L. tarentolae* Parrot-TarII distinctly belongs to the subgenus *Sauroleishmania.* In this dataset, no genome from the subgenus *Viannia* was shown to be representative due to the difference in chromosome numbers.

The geologic time scale places the split between the genus of ancestors *Leishmania* from *P. hertigi* and later *Mundinia* from the other *Leishmania* subgenera in the early Cretaceous period (Figure 6.3). *L. (M.) martiniquensis* strains diverged significantly from the rest of the *Mundinia* species during the Paleogene. Simultaneously, *L. tarentolae* separated from the subgenus *Leishmania*. The rest of splits observed in this tree occurred during the Neogene and Quaternary periods, which span the last 25 MYA.

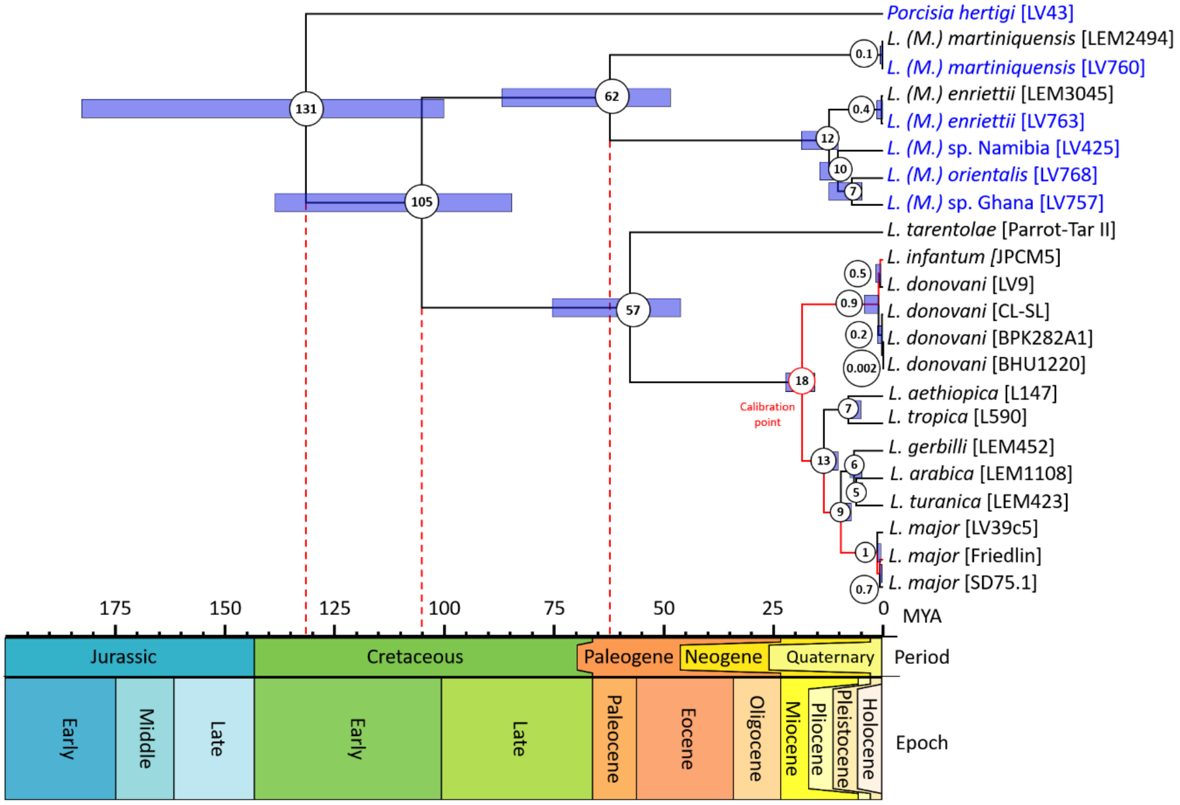


Figure 6.3: Phylogenomic tree construction of all 36 chromosomes for 22 species mentioned in the method section. The blue coloured species are the new assemblies. The horizontal (bottom) shows the time scale in reverse order (MYA) and Geologic Time Scale (GTS) (Ogg, 2020) (see *Supplementary Materials* for details).

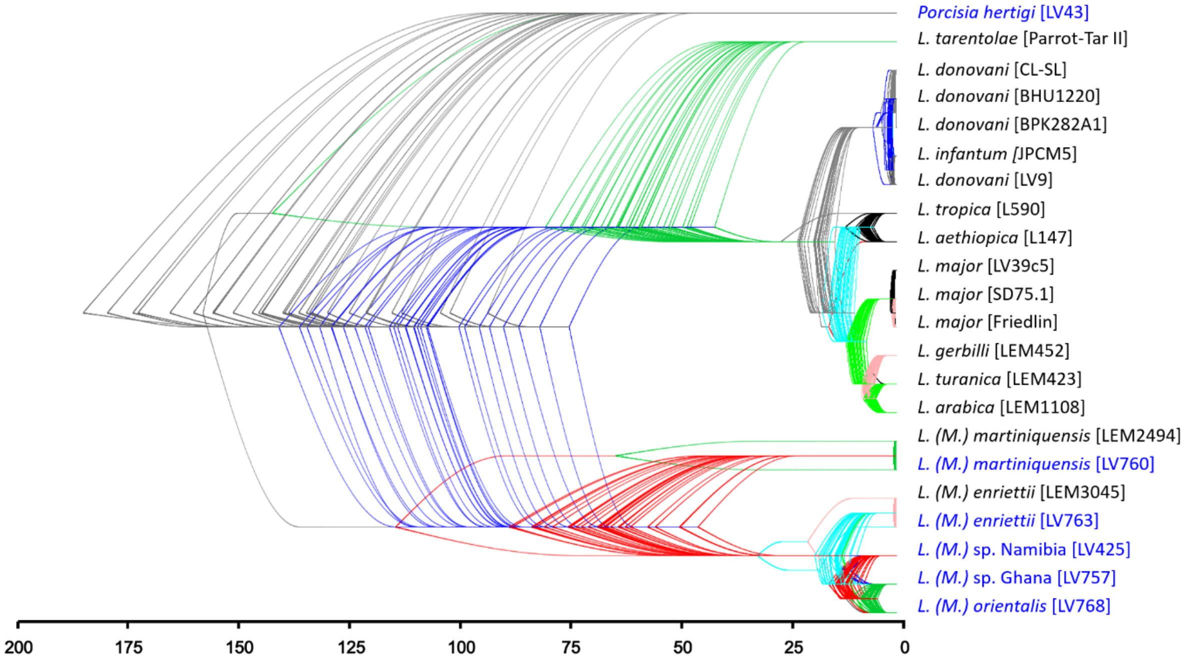


Figure 6.4: All 36 chromosome trees added together to create a single consensus tree using DensiTree software. The colours were generated automatically to differentiate each clade (see *Supplementary Materials*).

The phylogenomic tree for the second dataset verified what occurred in the first one but with considerably extra details. As stated before, this dataset includes 60 species. As shown in figure 6.5 and 6.6, all four subgenera of *Leishmania* were clearly denoted by branches on the tree. This tree also confirms that the new assemblies are, apart from *P. hertigi*, all in the subgenus *Mundinia*. However, the time estimation in this tree were slightly different than the first one.

The TMRCA between the outgroup, which are *Porcisia* species, and the rest of *Leishmania* assemblies is estimated to be around 152.54 MYA (95% HPD of 110.6 – 170.37 MYA). TMRCA between the subgenus *Mundinia* assemblies and the other four subgenera is estimated to be around 121.15 MYA (95% HPD of 91.13 – 135.93 MYA). The same distinction between *L. (M.) martiniquensis* strains (LSCM1 and LEM2494) and the rest of *Mundinia* species was also estimated to be around 74.3 MYA (95% HPD of 53.7 – 84.09 MYA) (Figure 6.5). Moreover, the TMRCAs between the three subgenera were also estimated. The TMRCA between the subgenus *Viannia* and both *Sauroleishmania* and *Leishmania* subgenera is estimated to be 105.51 MYA (95% HPD of 74.6 – 118.34 MYA). The TMRCA between the subgenera *Sauroleishmania* and *Leishmania* is estimated to be 63.79 MYA (95% HPD of 48.94 – 74.52 MYA) (Figure 6.5).

Other observations were also made with the other genomes. For instance, *L. (L.) tropica* species (strain CDC216-162), that was isolated from an American traveller in Afghanistan (Unoarumhi et al., 2021), seemed to be more closely related to the *L. infantum-chagasi and L. donovani* clade than the other *L. tropica* species. In *Viannia* clade, *L. (V.) braziliensis, L. (V.) guyanensis* and *L. (V.) panamensis* appeared to be difficult to be differentiated since they appear to have different topologies in each chromosome tree (Figure 6.6).

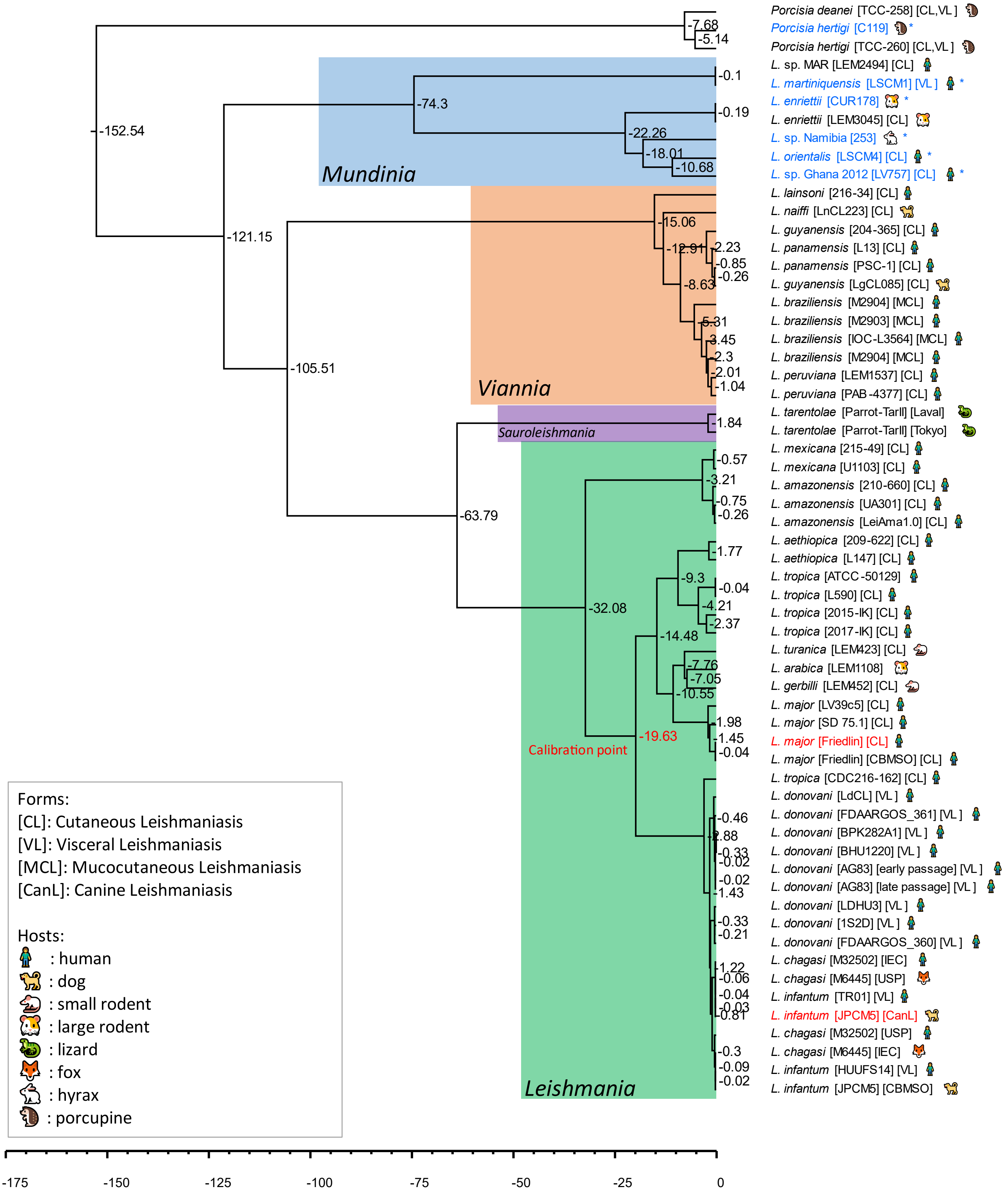

Porcisia deanei [TCC-258] [CL,VL]
Porcisia hertigi [C119] *
Porcisia hertigi [TCC-260] [CL,VL]
L. sp. MAR [LEM2494] [CL]
L. martiniquensis [LSCM1] [VL] *
L. enriettii [CUR178] *
L. enriettii [LEM3045] [CL]
L. sp. Namibia [253] *
L. orientalis [LSCM4] [CL] *
L. sp. Ghana 2012 [LV757] [CL] *
L. lainsoni [216-34] [CL]
L. naiffi [LnCL223] [CL]
L. guyanensis [204-365] [CL]
L. panamensis [L13] [CL]
L. panamensis [PSC-1] [CL]
L. guyanensis [LgCL085] [CL]
L. braziliensis [M2904] [MCL]
L. braziliensis [M2903] [MCL]
L. braziliensis [IOC-L3564] [MCL]
L. braziliensis [M2904] [MCL]
L. peruviana [LEM1537] [CL]
L. peruviana [PAB-4377] [CL]
L. tarentolae [Parrot-TarII] [Laval]
L. tarentolae [Parrot-TarII] [Tokyo]
L. mexicana [215-49] [CL]
L. mexicana [U1103] [CL]
L. amazonensis [210-660] [CL]
L. amazonensis [UA301] [CL]
L. amazonensis [LeiAma1.0] [CL]
L. aethiopica [209-622] [CL]
L. aethiopica [L147] [CL]
L. tropica [ATCC-50129]
L. tropica [L590] [CL]
L. tropica [2015-IK] [CL]
L. tropica [2017-IK] [CL]
L. turanica [LEM423] [CL]
L. arabica [LEM1108]
L. gerbilli [LEM452] [CL]
L. major [LV39c5] [CL]
L. major [SD 75.1] [CL]
L. major [Friedlin] [CL]
L. major [Friedlin] [CBMSO] [CL]
L. tropica [CDC216-162] [CL]
L. donovani [LdCL] [VL]
L. donovani [FDAARGOS_361] [VL]
L. donovani [BPK282A1] [VL]
L. donovani [BHU1220] [VL]
L. donovani [AG83] [early passage] [VL]
L. donovani [AG83] [late passage] [VL]
L. donovani [LDHU3] [VL]
L. donovani [1S2D] [VL]
L. donovani [FDAARGOS_360] [VL]
L. chagasi [M32502] [IEC]
L. chagasi [M6445] [USP]
L. infantum [TR01] [VL]
L. infantum [JPCM5] [CanL]
L. chagasi [M32502] [USP]
L. chagasi [M6445] [IEC]
L. infantum [HUUFS14] [VL]
L. infantum [JPCM5] [CBMSO]
-7.68
-5.14
-152.54
-0.1
-74.3
-0.19
-22.26
-18.01
-10.68
Mundinia
-121.15
-15.06
-12.91
-2.23
-0.85
-0.26
-8.63
-5.31
-3.45
-2.3
-2.01
-1.04
-105.51
Viannia
-1.84
Sauroleishmania
-0.57
-3.21
-0.75
-0.26
-63.79
-1.77
-9.3
-0.04
-4.21
-2.37
-32.08
-14.48
-7.76
-7.05
-10.55
-1.98
-1.45
-0.04
-19.63
Calibration point
-0.46
-2.88
-0.33
-0.02
-0.02
-1.43
-0.33
-0.21
-1.22
-0.06
-0.04
-0.03
-0.81
-0.3
-0.09
-0.02
Leishmania
Forms:
[CL]: Cutaneous Leishmaniasis
[VL]: Visceral Leishmaniasis
[MCL]: Mucocutaneous Leishmaniasis
[CanL]: Canine Leishmaniasis
Hosts:
: human
: dog
: small rodent
: large rodent
: lizard
: fox
: hyrax
: porcupine
-175
-150
-125
-100
-75
-50
-25
0

Figure 6.5: Phylogenomic tree construction of all 36 chromosomes for 60 species from the second dataset. The blue coloured species are the new assemblies. The horizontal (bottom) shows the time scale in reverse order (MYA). The red coloured species are the ones used for TMRCA calibration (see *Supplementary Materials* for full-scale figure).

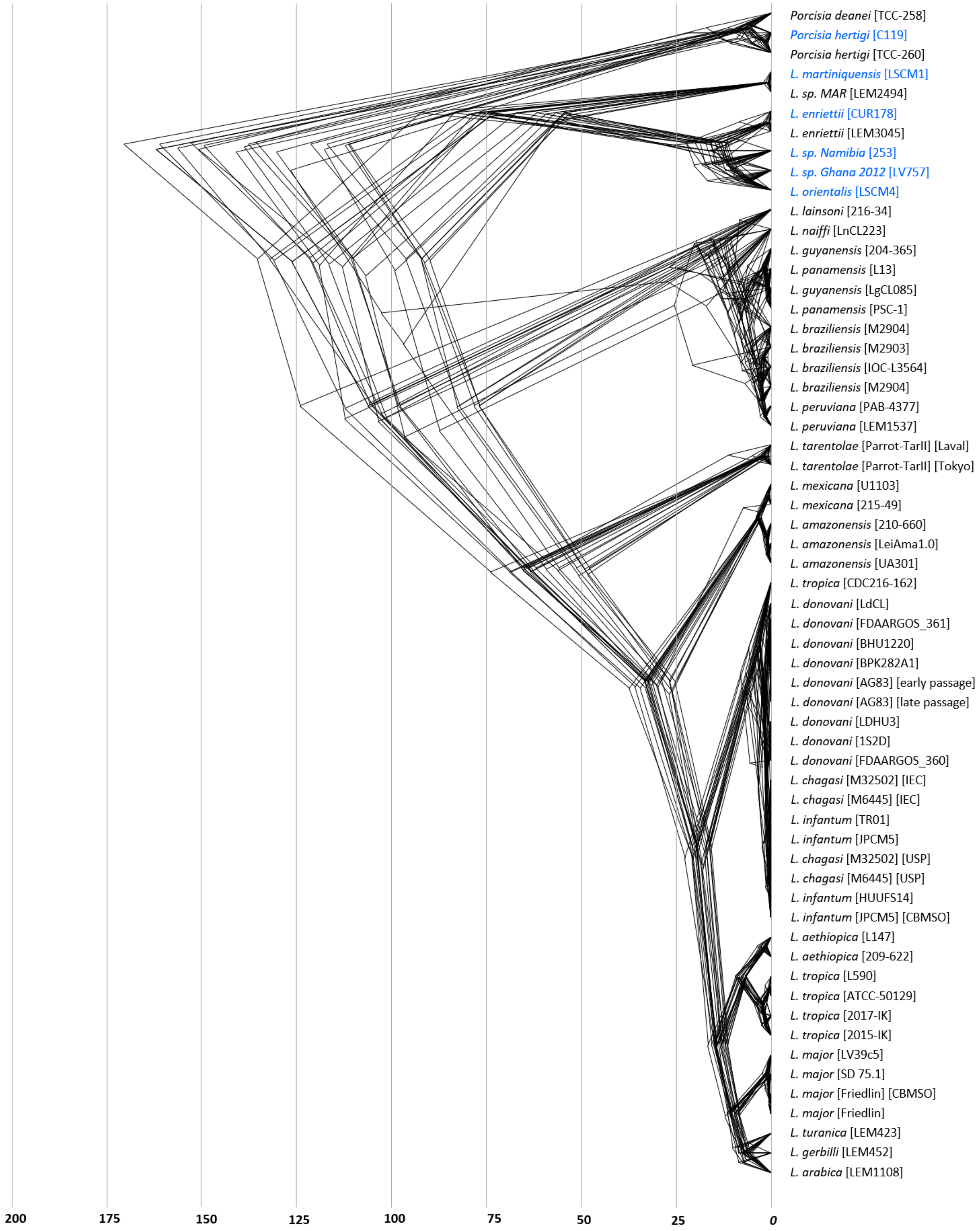


Figure 6.6: All 36 chromosome trees added together to create a single consensus tree using DensiTree software. The bottom axis represents the time scale in reverse order in million years (see *Supplementary Materials* for full-scale figure).

6.2 Orthology-based Tree

Given that the total number of annotated CDSs for all six assemblies is 48,405 (details in Table 5.6), an orthology-based species tree was created utilising 55 proteomes from TriTrypDB. The tree topology of the species tree created by the OrthoFinder algorithm and the two Bayesian phylogenomics trees showed a high degree of agreement. The species tree is shown in Figure 6.7 with a heatmap matrix representing the number of one-to-one orthology similarities. All four *Leishmania* subgenera were clustered in the same way as the phylogenomic tree. Additionally, as expected, *P. hertigi* were an outgroup for the four *Leishmania* subgenera and were closely related to the proteome of *Endotrypanum monterogeii* strain LV88. The heatmap revealed a high degree of orthologue similarity between all *Leishmania* species, including *Porcisia* and *Endotrypanum* species, ranging between 6,540 and 8,268 orthologs, whereas the *Mundinia* clade had a range of 7,076 to 8,250 orthologs.

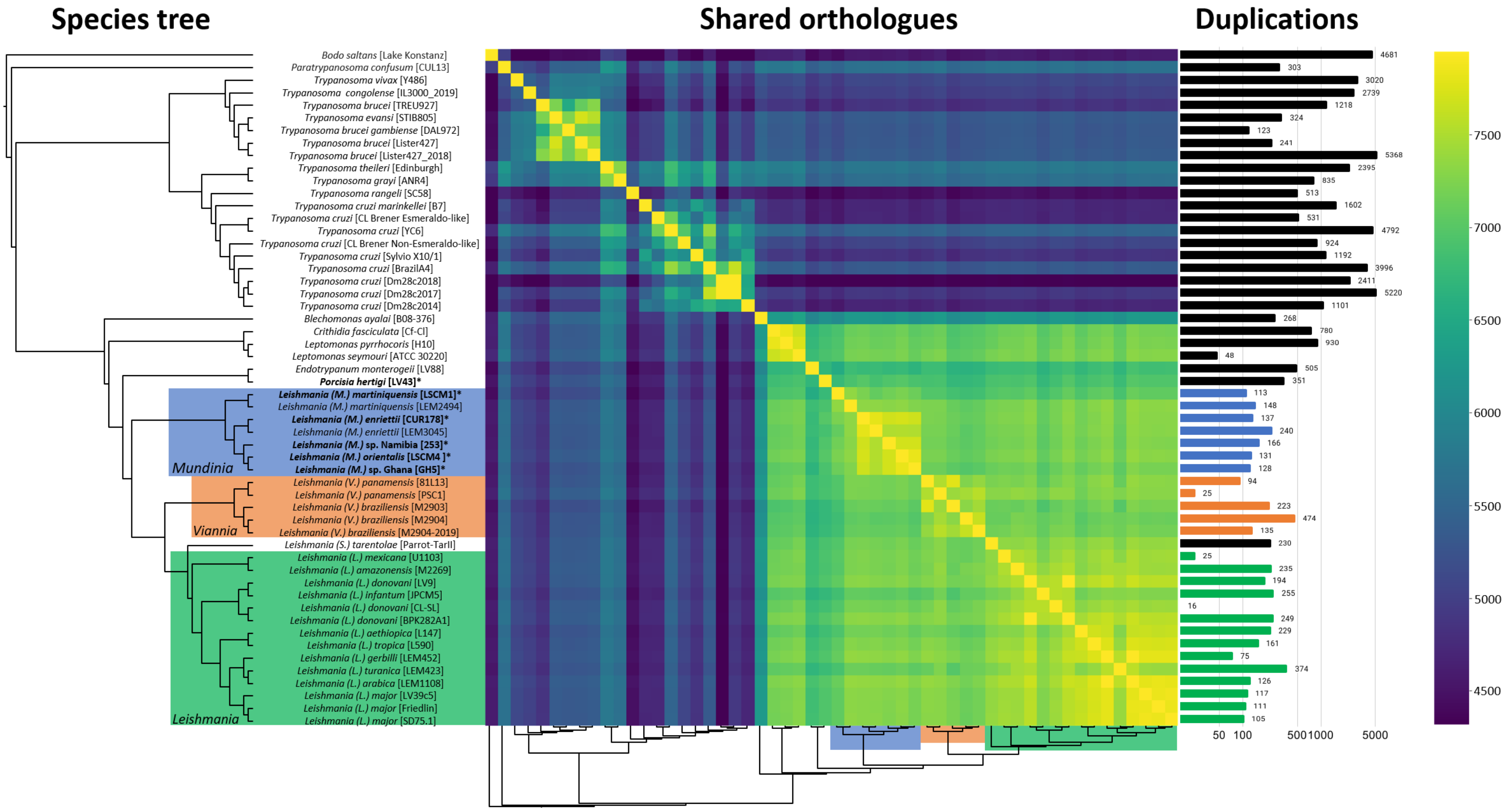


Figure 6.7: Species tree, shared orthologues matrix and duplications generated by OrthoFinder. The heatmap is coloured according to the number of orthologs in the matrix. Names of the new genomes are followed by asterisk (see Supplementary Materials for details).

6.3 Identification of Polycistronic Transcription Units (PTUs)

Genes in *Leishmania* are known to be arranged in clusters and transcribed as long Polycistronic Transcription Units (PTUs) (Pannunzio and Lieber, 2016, Chandra et al., 2017). Transcription starts from divergent Strand Switch Regions (dSSRs) (Daniels et al., 2010).

The GFF outputs generated during the annotating process were repurposed. They were transformed into coordinate tables to demonstrate both upstream and downstream transcription PTUs, and then shown in circular form. Additionally, a layer has been added to indicate the position of any feature that has been fully or partially labelled as "polymerase" as a potential polymerase binding site in SSRs. This was done to determine whether there is any correlation between the beginning of PTUs and the relative position of any feature labelled as polymerase throughout all chromosomes, as previously described (Martínez-Calvillo et al., 2003).

The polymerase location and strand direction combination confirm that the new genome assemblies shared a high number of polymerase locations with the reference. Additionally, most of them can be found at either end of the PTUs (Figure 6.8).

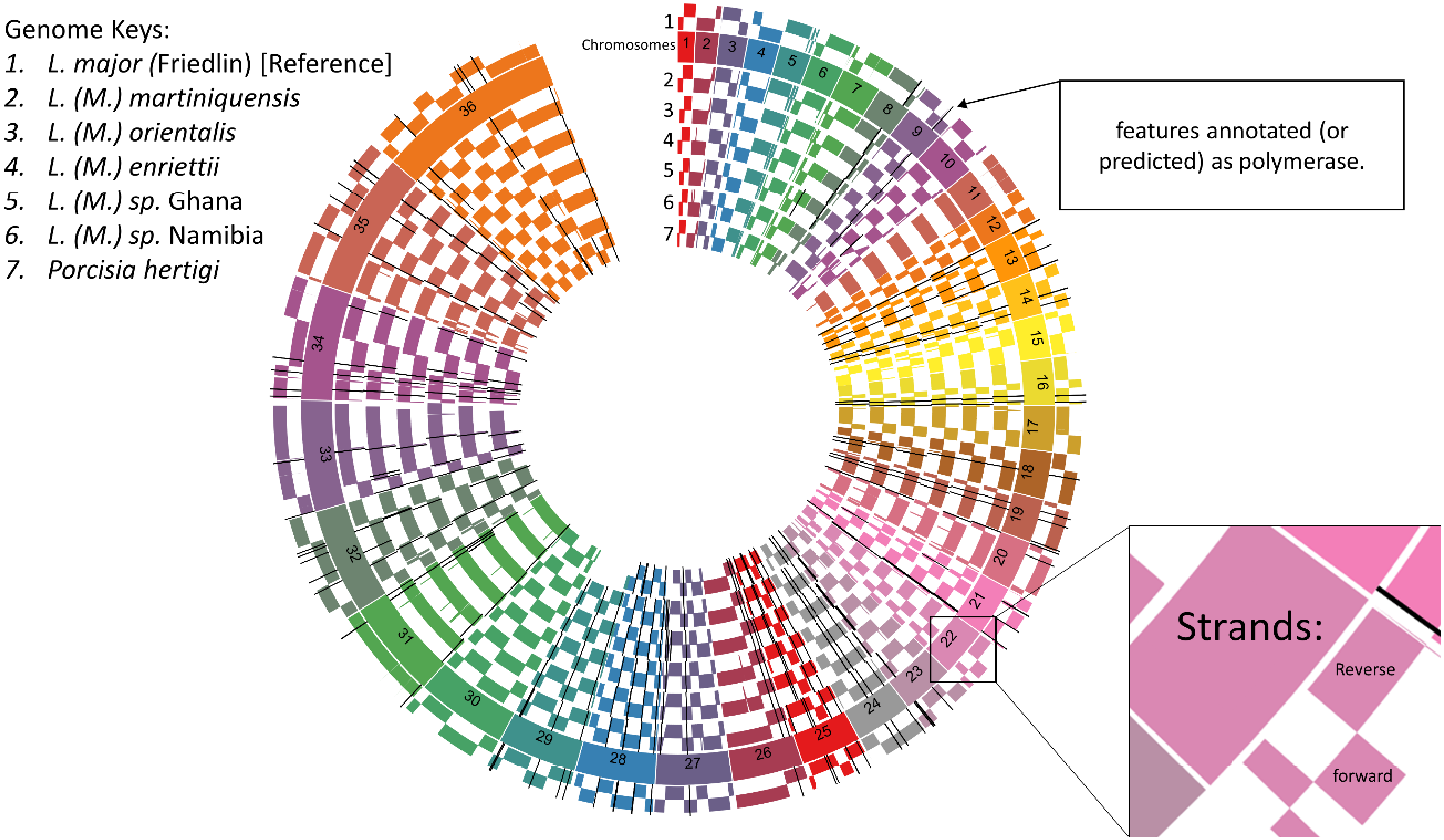


Figure 6.8: Strand switching regions features across all assembled genomes in comparison to *L. major* Friedlin strain as a reference. The black coloured lines represent features annotated (or predicted) as "polymerase". Some of the polymerase ribbons are coordinated with strand switch locations. The topology of stranding across chromosomes shows similarity between *Mundinia* subgenus as well as *P. hertigi* and when compared to the genome of *L. major* Friedlin strain (see Supplementary Materials for full-scale).

6.4 Detection of Selection Pressure

The detection of selection pressure was focused to *Mundinia* species and genes for which the gene slicer script was able to generate sufficiently high-quality gene alignments from the original chromosomal alignments (Szpara et al., 2014). Due to the alignment quality threshold imposed by the gene slicing script, only 69% (5,519) of all CDS were tested for selection pressure. A total of 36 (0.65%) sites were detected to have positive selective pressure (Table 6.2).

The proteins encoding these 36 sites were subjected to a protein family search using both the profile hidden Markov Models HMMSCAN (Potter et al., 2018) and Position-Specific Iterated BLAST methods (Altschul et al., 1997). Among those sites, we find domains annotated as: "protein of unknown function", C2 domain-containing protein, Poly(A)-specific ribonuclease, Deoxyhypusine hydroxylase, Protein seedling plastid development, Transmembrane 9 superfamily member, Patatin-like phospholipase domain-containing protein, and MTOR-associated protein MEAK7. Only 4 proteins did not have any match with both HMMSCAN and PSI-BLAST which are in chromosomes 3, 6, 33 and 34 (see table 6.3). These four proteins may be inferred to be novel.

Table 6.2: statistics of CDS subjected to selection pressure analysis (see *Supplementary Materials* for details).

Chromosome	Number of CDS	CDS included	included for Selection (%)	Slr hits	Slr% among included CDS	Slr% among All
1	78	64	82%	1	1.56%	1.28%
2	70	40	57%	0	0.00%	0.00%
3	91	64	70%	2	3.13%	2.20%
4	124	86	69%	0	0.00%	0.00%
5	125	90	72%	0	0.00%	0.00%
6	125	96	77%	2	2.08%	1.60%
7	125	64	51%	1	1.56%	0.80%
8	110	77	70%	0	0.00%	0.00%
9	153	115	75%	1	0.87%	0.65%
10	132	95	72%	1	1.05%	0.76%
11	123	92	75%	0	0.00%	0.00%
12	99	65	66%	0	0.00%	0.00%
13	173	111	64%	1	0.90%	0.58%
14	138	104	75%	0	0.00%	0.00%
15	178	102	57%	1	0.98%	0.56%
16	171	118	69%	0	0.00%	0.00%
17	153	93	61%	3	3.23%	1.96%
18	159	134	84%	1	0.75%	0.63%
19	156	107	69%	0	0.00%	0.00%
20	159	96	60%	3	3.13%	1.89%
21	213	153	72%	1	0.65%	0.47%
22	152	126	83%	0	0.00%	0.00%
23	194	132	68%	1	0.76%	0.52%
24	237	168	71%	0	0.00%	0.00%
25	261	208	80%	3	1.44%	1.15%
26	273	189	69%	3	1.59%	1.10%
27	258	196	76%	0	0.00%	0.00%
28	327	244	75%	1	0.41%	0.31%
29	278	228	82%	1	0.44%	0.36%
30	379	290	77%	1	0.34%	0.26%
31	315	54	17%	0	0.00%	0.00%
32	415	160	39%	0	0.00%	0.00%
33	354	261	74%	3	1.15%	0.85%
34	405	309	76%	3	0.97%	0.74%
35	528	427	81%	0	0.00%	0.00%
36	732	561	77%	2	0.36%	0.27%
Grand Total	**7967**	**5519**	**69%**	**36**	**0.65%**	**0.45%**

Table 6.3: Protein Families detected using both HMMSCAN and NCBI BLAST search. The families are ◎: NCBI superfamily, 🔴: NCBI conserved domains, 🟠: COG, 🟡: SMART/cl, 🟢: Supfam, 🔵: TIGR, 🟣: TigrFAM, ⚫: Gene3D, ⚪: TreeFAM, 🟤: PIRSF, 🔺: Pfam, 🔻: PSI-BLAST, and 🔶:PDB. The rows coloured in light red do not have any protein family detected while the ones coloured in light green have a parasite-host interaction function.

Accession	Chr	Sites	Protein Families	Names	Functions
KAG5488126.1	1	1	⚫ 🔻		
KAG5487731.1	3	1	🔻		
KAG5487718.1	3	1	🔻		
KAG5486723.1	6	1	🔻		
KAG5486712.1	6	1	⚫ 🔻		
KAG5486216.1	7	1	◎ 🟡 🟢 ⚫ 🔺 🔻 🔶	• Anaphase-promoting complex subunit 4 WD40 domain • DNA polymerase III subunits gamma and tau • YVTN repeat-like/Quinoprotein amine dehydrogenase • beta-propeller	• Mediate ubiquitin-dependent proteasomal degradation in cell cycle • Transport proteins
KAG5485498.1	9	1	🔺 🔻	• Unstructured region C-term to UIM in Ataxin3	• Ubiquitin binding and ubiquitylation promoter
KAG5485011.1	10	1	◎ 🟠 🟡 ⚫ 🔻		
KAG5484037.1	13	1	◎ 🟠 🟡 🟢 ⚫ ⚪ 🔺 🔻 🔶	• Protein kinase C conserved region 2 • Ca^{2+}-dependent lipid-binding protein C2- domain Calcium/lipid binding domain	• Regulate signal transduction processes at the membrane surface
KAG5482904.1	15	1	🟡 🟢 ⚫ 🔺 🔻 🔶	• DNA clamp	• Promoting factor in DNA replication
KAG5481582.1	17	1	◎ ⚪ 🔺 🔻	• Oxidored-nitro domain-containing protein isoform 1	
KAG5481538.1	17	1	⚫ 🔺 🔻 🔶	• mt-LAF8	• Part of kinetoplastids mitoribosomes
KAG5481498.1	17	1	◎ 🔴 🟡 🔻	• Transcriptional regulator ICP4 • Adventurous gliding motility protein GltJ	• Transcriptional regulatory protein • Gliding motility mechanism

Accession	Chr	Sites	Protein Families											Names	Functions
KAG5481068.1	18	1					🟢		⚫		🔺	🔻	🔶	• Mitochondrial carrier	• Amino acids transportation
KAG5480045.1	20	2	◎				🟢		⚫		🔺	🔻	🔶	• Golgi alpha-mannosidase II • Smp-1-like	• N-glycan synthesis
KAG5480006.1	20	2	◎	🔴	🟠	🟡	🟢	🔵	⚫		🔺	🔻	🔶	• tRNA pseudouridine synthase D	• Synthesis of pseudouridine from uracil-13 in transfer RNAs.
KAG5479963.1	20	2					🟢		⚫			🔻		• Metallo-dependent phosphatases	• Metabolic damage-control (housecleaning)
KAG5479324.1	21	1					🟢		⚫		🔺	🔻	🔶	• BAR/IMD domain-like • Ciliary rootlet component, centrosome cohesion	• Cell signalling and sensory perception
KAG5478657.1	23	1	◎			🟡	🟢		⚫	⚪	🔺	🔻	🔶	• CAF1 family ribonuclease • Rnase H	• Remove RNA primers during DNA replication
KAG5477112.1	25	1										🔻			
KAG5477073.1	25	1										🔻			
KAG5476895.1	25	1	◎			🟡			⚫		🔺	🔻	🔶	• BRCT domain • BRCA1 C-terminus domain	• Cell cycle checkpoint functions and responsive to DNA damage
KAG5476530.1	26	1	◎		🟠			🔵	⚫			🔻		• Chromosome segregation ATPase	• Cell cycle control • Cell division • Chromosome partitioning
KAG5476493.1	26	1	◎	🔴	🟠	🟡	🟢		⚫	⚪	🔺	🔻	🔶	• ARM • Leucine-rich Repeat Variant • phycocyanin lyase	• Structural function • Increases oxygen production
KAG5476398.1	26	1	◎	🔴		🟡	🟢		⚫		🔺	🔻	🔶	• Phosphotyrosine-binding domain (PTB) • Meiotic cell cortex C-terminal pleckstrin homology • Pathogenicity factor • large tegument protein UL36 • Plant-like Phospholipase C (PLC) • pleckstrin homology (PH) domain	• Signal transduction • DNA replication and evasion of the immune response • Focal-adhesion molecule to plasma membrane
KAG5474511.1	28	1	◎	🔴		🟡	🟢		⚫		🔺	🔻	🔶	• calmodulin binding domain • C-terminal middle region of Androglobins (Adgbs) • P-loop containing nucleoside triphosphate hydrolases	• Mediates Ca^{2+} signalling • Spermatogenesis protein folding

Accession	Chr	Sites	Protein Families	Names	Functions
KAG5473777.1	29	1	◉ 🟡 🔺 🔻	• PSP1 C-terminal conserved region	
KAG5472426.1	30	3	◉ 🔴 🟠 🟡 🟢 🔵 🔺 🔻 🔶	• Sporulation stage III, protein AA • P-loop containing nucleoside triphosphate hydrolases	• sporulation to survive harsh environment
KAG5469588.1	33	1	🔻		
KAG5469580.1	33	1	◉ 🔴 🟢 ⚫ 🔻 🔶	• P-loop containing nucleoside triphosphate hydrolases • Kinesin • large tegument protein UL36 • Med15 subunit of Mediator complex	• Generate force and displacement along microtubules • Host interaction • General transcriptional cofactor
KAG5469529.1	33	1	◉ 🔴 🟡 🔻		
KAG5468413.1	34	1	◉ ○ 🔺 🔻	• Endomembrane protein 70	• Permits various cell functions to be compartmentalized
KAG5468408.1	34	1	◉ 🔴 🟠 🟡 🟢 ⚫ 🔺 🔻 🔶	• Triacylglycerol lipase 3 • Predicted acylesterase/phospholipase RssA • FabD/lysophospholipase-like	• Metabolic function • Lipid mediator production
KAG5468130.1	34	1	🔻		
KAG5464423.1	36	1	🟡 🔺 🔻	• Rab3 GTPase-activating protein regulatory subunit N-terminus • beta-propeller	• Intracellular regulations and vesicle traffic
KAG5464022.1	36	1	◉ 🔴 🟠 🟡 🔺 🔻 🔶	• TLD	• Cell death and DNA replication

Chapter 7. Discussion

Recent years have seen a substantial increase in our understanding of genomes and their function in health and disease. Two decades earlier, researchers were conducting early studies of the first reference human genome sequences, which cost more than $1 billion to assemble (Venter et al., 2001, Lander et al., 2001). Thousands of genomes have been sequenced since then. Advances in sequencing technologies have accelerated the pace of research. It was estimated ten years ago, that a human genome can be sequenced completely in days for less than a thousand dollars, with costs expected to continue to decline in the coming years (Mardis, 2011).

Making sense of genomic data demands parallel advancement of both sequencing and computational technologies (Stein, 2010). Both continue to advance, enabling an ever-increasing capacity for accurate disease detection and the creation of effective and focused treatment. They also provide opportunity to further analyse in far greater detail than previously done, especially in the case of NTDs, which may result in more suitable approach for control, treatment, and prevention.

7.1 Computational Challenges

Working with large amounts of data proves challenging. To process and analyse all the data, for example, a large amount of storage space and enough powerful computing units are required. Cloud computing, rather than relying on traditional computers, may be the answer.

Cloud computing is the product of the combining a group of traditional computing units in such a way that it can make efficient use of any accessible resource pools for compute, storage, and memory (Iosup et al., 2011, Prasad and Rao, 2014, Li et al., 2015). This resource pools can also be made available to other users over the Internet. One of the most essential resources in Cloud Computer technology is the virtual machine (VM). Some VMs are dependent on resource scheduling methods that prioritise quality of service over the profitability. Currently, resource scheduling and management are the most significant and challenging tasks in cloud computing technology (Eswaraprasad and Raja, 2017).

Scheduled-based VM is usually centralised computation service designed to support users who require high-performance and high-throughput computing, including workloads that exceed the capabilities of the Interactive Unix Service (IUS) or a desktop PC. It contains high number central processing units (CPUs), terabytes of aggregate random-access memory (RAM), and high-speed file storage.

However, because it is based on a job scheduling system, most bioinformatics analyses require continuous interaction with the machine, which requires adding more jobs and thus decreasing the priority of running the job, as scheduling systems use this as a penalty parameter for prioritizing and thus extending the analysis time.

On the other hand, some virtual machines can be considered dedicated units equipped with the Interactive Unix Service (IUS) and a limited number of CPUs and RAMs but with far better processing capability and storage than a desktop PC. This type of virtual machine is well-suited for interactive analysis because the entire machine is often managed by a single user.

7.2 *Mundinia* Lineage and Evolution

The results demonstrated that high-quality genomes, with most of their features annotated and predicted, can be important. They also showed that they can be used to create accurate findings about how *Leishmania* parasites evolved and how they may be taxonomically classified.

The findings suggest that the subgenus *Mundinia* was the first subgenus lineage to branch off from the *Palaeoleishmania* hypothetical common ancestor. The early origin of *Mundinia* explains two things: 1) its geographical diversity – as it travelled along with the splitting continents, rather than having to cross oceans, and 2) its genetic diversity, since a lot of time has elapsed since the *Mundinia* MRCA. Additionally, the fact that the most recent ancestor is estimated to have lived in the early Cretaceous period suggests that branching happened prior to the supercontinent split (Dixon et al., 2001). According to some fossil evidence, the early Cretaceous period is thought to be the period of the origin of mammals in general (Marlowe, 2005). However, other evidence suggests that it could have occurred as early as the late Jurassic, as the early Cretaceous is the time of the radiation of eutherian mammals, such as the TMRCA of *Afrotheria* (elephants, hyraxes), and the lineage leading to all other mammals is around 105 MYA (Rook and Hunter, 2014).

The phylogenomic trees revealed evolution of the ancestors of modern genus *Leishmania* across 3 distinct periods. The first one was between the late Jurassic and early Cretaceous periods. It included the splits of *Euleishmania* from *Paraleishmania*, averaged around 138 MYA, as well as the splits of the subgenus *Viannia* from both *Sauroleishmania* and *Leishmania*, averaged around 103 MYA. The second period encompasses both early and late Cretaceous epochs. *Mundinia* was split from the other three subgenera, *Viannia*, *Sauroleishmania*, and *Leishmania*, averaged around 113 MYA. The third one, averaged around 60 MYA, occurs during the late Cretaceous and Paleogene periods. It comprises two main splits: one between *L. (M.) martiniquensis* strains and the rest of *Mundinia*, and another one between the subgenera *Sauroleishmania* and *Leishmania*.

As a result, the question arose as to what had occurred to either mammalian hosts or insect vectors during these time periods. Furthermore, were there any major events that coincided with the findings?

The primary event that represents the time coordinate of the root of the trees, around 175 MYA, coincides with the start of supercontinent breakup, between 201 and 174 MYA (Early Jurassic), which resulted in the formation of the modern continents and the Atlantic and Indian oceans (Wilson, 1963). However, due to its separation from other landmasses such as Baltica, Laurentia, and Siberia, Gondwana is not considered to be a supercontinent (Bradley, 2015).

Another significant time period in the evolution of mammals was the Paleocene-Eocene Thermal Maximum (PETM) (McInerney and Wing, 2011), which occurred approximately 56 million years ago. There are fossil traces of large mammalian groups migrating from Asia to north America during that time period (Bowen et al., 2002). Additionally, other studies have suggested that the expansion of primates begins at this stage with fossils found in Europe, North America, and Asia (Beard, 2008, Smith et al., 2006).

According to earlier theory, the divergence of *Paraleishmania* from *Euleishmania*, as depicted in Figure 6.5 by the genus *Porcisia*, occurred approximately 140 MYA, as opposed to the 26 MYA estimated previously. This major discrepancy in values could be explained by the fact that when these theories were developed, they relied on limited molecular evidence and outdated mathematical models (Noyes et al., 1997, Noyes et al., 2000). Moreover, it was also assumed that *L. hertigi*, as it was then called, would be more closely related to *Leishmania*, so earlier divergence dates remained plausible, until more molecular phylogenetic evidence appeared.

In general, previous analyses supported my conclusion that *P. hertigi* is closely linked to *P. deanei*, as indicated in the results, as well as *Porcisia* being closely related to *Endotrypanum*, as I have indicated in the orthology species tree (Croan et al., 1997, Cupolillo et al., 2000, Noyes et al., 2002, Asato et al., 2009a, Pothirat et al., 2014b). Nonetheless, the phylogenomic analysis revealed that TMRCA of sub-family *Leishmaniinae* is estimated to be older than that previously hypothesised.

7.3 Implications of Accurately Assembling Genomes

The new genomes, particularly *L. (M.) martiniquensis* and *L. (M.) enriettii*, were assembled more accurately than the previous ones. The syntenic dotplot for chromosome 31 revealed a typical deletion signal (Figures 5.7 and 5.8 in chapter 5). However, this deletion signal was matched with some unplaced contigs and scaffolds that were located the end of both previous assemblies. This artefact was confirmed when the same analysis was performed on *L. major* Friedlin strain, where the deletion signals were absent. Additional verifications were also performed by comparing all assemblies to other reference genomes, and no deletions were found.

Several of the features that were studied in earlier *Leishmania* genomes were found to be present in all six assemblies, providing additional evidence that the annotation process was thorough, such as the strand switching region (Figure 6.8 in chapter 6). Other features were, however, not studied before due to the absence of any publicly available data of this subgenus. For instance, when selection pressure was analysed in all *Mundinia* protein coding sequences, the majority of CDS were negatively selected while the CDS with positively selected residues – all still have an overall $\Omega < 1$ – were only thirty-six coding regions, all of which were assigned as hypothetical proteins. Although efforts have been made here to characterise some of these residues, the findings necessitate more examination into their possible involvement in evading immunity in both vectors and hosts, as seen in previous studies (Neto et al., 2019).

Generally, the results presented here are a consilience of evidence. It demonstrates my strategy of inference to the best explanation, as I used a range of methods and tools reaching the conclusion that all points lead towards the same conclusion, although each alone is incomplete proof when considered in isolation. However, the cumulative effect of estimating the taxonomy of subgenus *Mundinia* in a range of ways makes the acceptance of my conclusion irresistible. Indeed, rejection of the conclusion would be unreasonable, given the weight of independent evidence from different angles of enquiry. This implies that the inference is correct, as described above.

Looking back at the assembly's foundations, combining both long and short reads for assembly proves to be the best way to overcome the inherited difficulty of assembling repeat regions with conventional short reads using *de Bruijn* graph-based assemblers (Dujardin, 2009). Long reads sequences with Nanopore technology, on the other hand, have low quality scores, resulting in a high rate of incorrect base calling. Additionally, the use of alternative long read platforms, such as long reads on PacBio Hi-C platforms, is advocated. As a result, high-quality long reads can help in accurate assemblies.

7.4 Annotation Challenges and Opportunities

The annotation process resulted in a substantial number of annotated features that are compatible with other annotated *Leishmania* genomes (Table 5.6). However, a serious challenge is that on average of 2.08% (Figure 7.1) of the annotated genes in our submitted assemblies have multiple exons (Steven L Salzberg, personal communication, 08 December 2021). Even though the majority of *Leishmania* genes lack introns, these results indicate that one of the annotation rounds predicted the presence of introns (Fong and Lee, 1988, Kazemi, 2011).

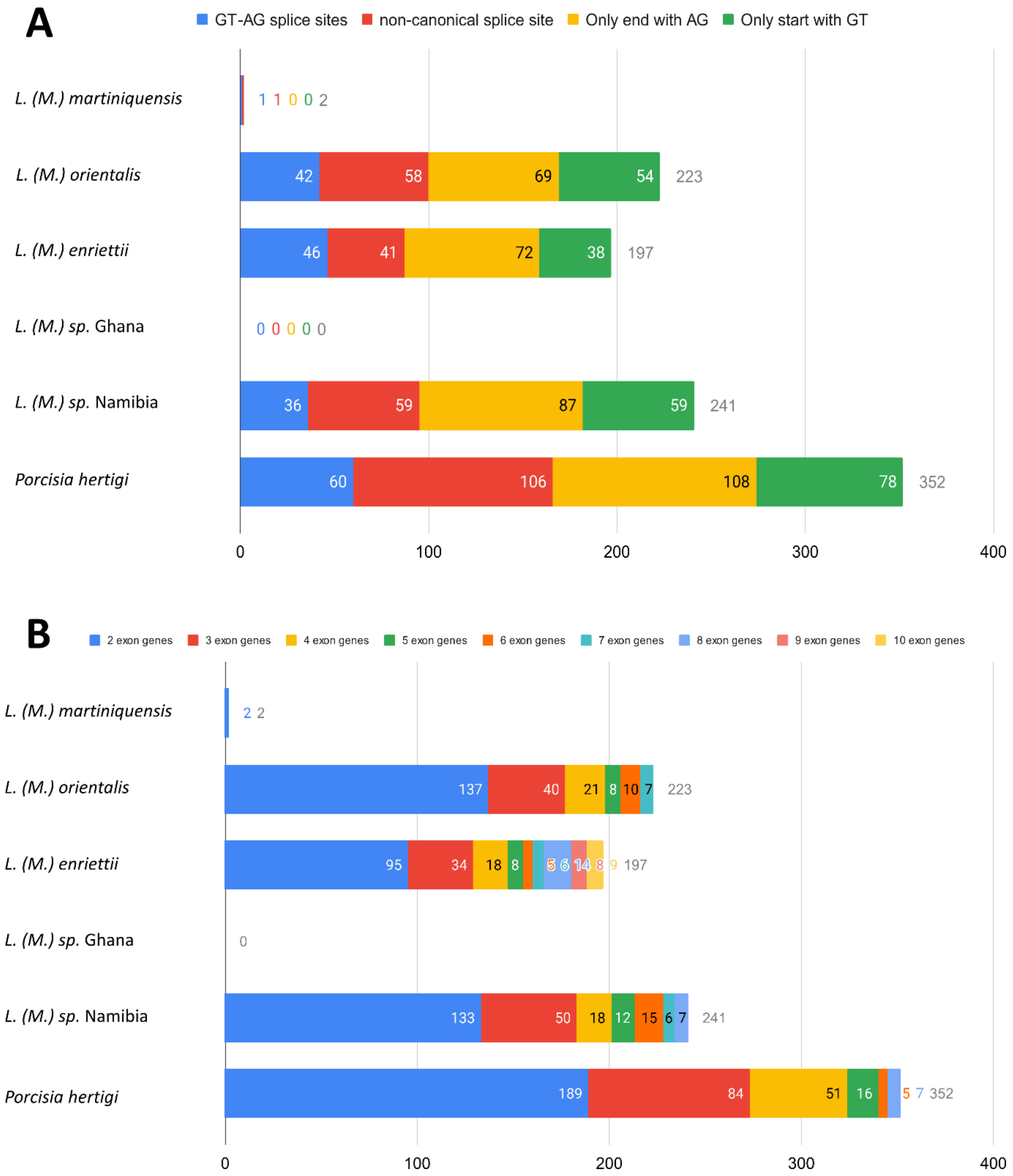


Figure 7.1: A: Number and type of splice sites per genome. B: Number of introns across annotated genes.

This could indicate that the evidence used for annotation already had genes with multiple exons, or that the prediction organism model, *L. tarentolae*, predicted genes with multiple exons incorrectly. The other possibility is that, as part of the process of finalising the annotation, the longest isoform of the annotated feature was chosen, as multiple isoforms are not acceptable for submission of genome assemblies to Genbank.

This constraint, on the other hand, can be viewed as a future opportunity to improve the annotation process. One strategy that might be used in this case is to use native RNA sequencing and transcriptome assembly via Oxford Nanopore Technologies as the primary evidence for the annotation process, rather than relying on external evidence from other species (Soneson et al., 2019). However, not all transcripts will be captured since some may be present in a form of the parasite that was not cultivated in cells and is only detected in the vector, such as the promastigote form.

Another approach that may help and has been used previously for multiple *Leishmania* parasites, even prior to the assembly of their genomes (Mann and Pandey, 2001), is to search against all six-frame translated nucleotide sequences from a genome or expressed sequence tag using an experimental proteome from tandem mass spectrometry (MS/MS) experiment (Choudhary et al., 2001, Pawar et al., 2012, Sanchiz et al., 2020). This technique, when used in conjunction with conventional genome annotation, has the potential to overcome this challenge.

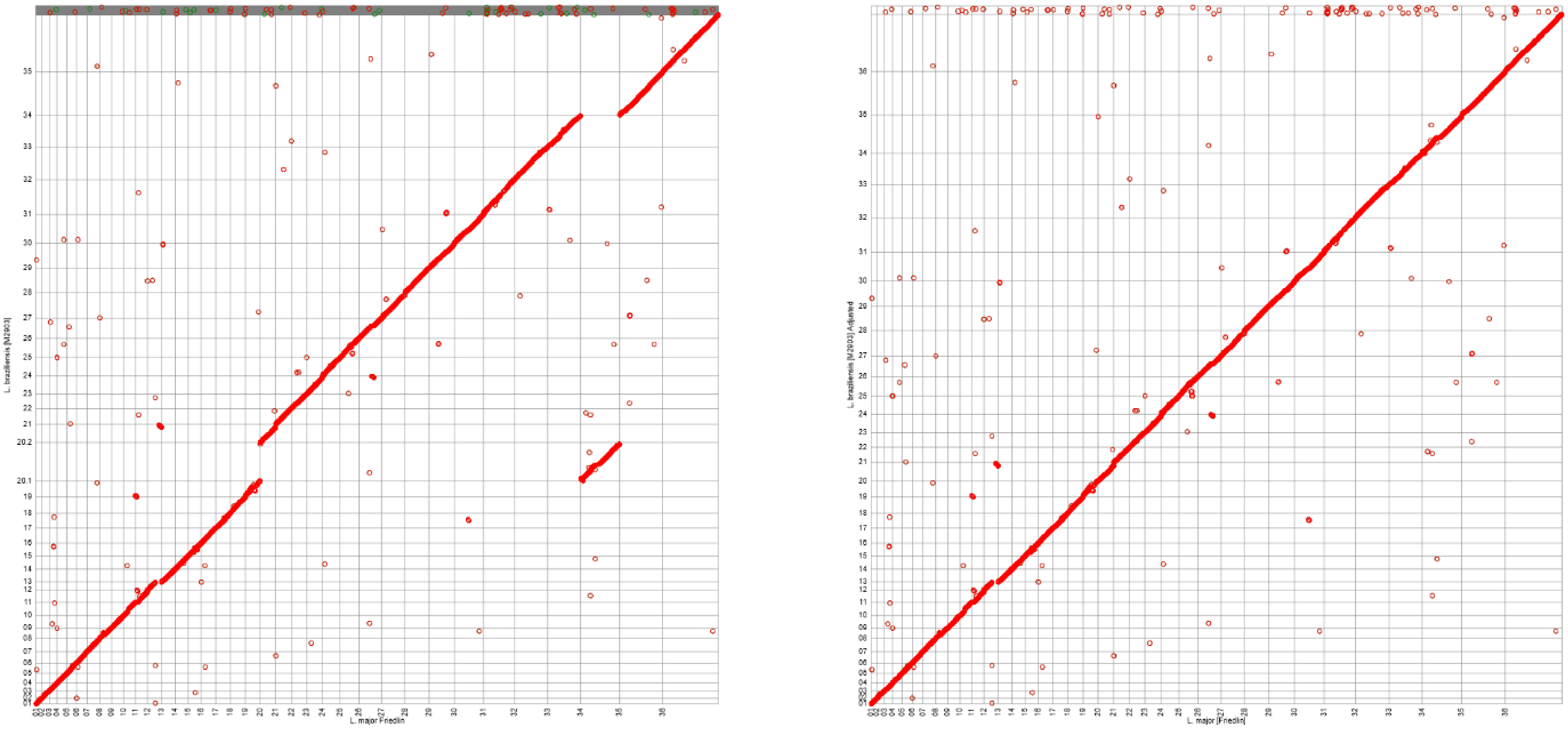
In terms of the optimization, results demonstrated that, at least in the case of *Leishmania* genome assembly, using hybrid assembly algorithms appears to have the opposite effect, which is not to be confused with the approach taken in this project, which uses long reads for assembly and short reads for polishing. Hybrid assembly generates a final assembly by using both short and long reads. This clearly demonstrated that it could produce not only shorter contigs but also a longer genome in total, which in this case resulted in a genome of 36 Mbps, more than 4 Mbps larger than the target (Table 5.1).

7.5 Implications and limitations

Most findings have added to the wealth of evidence indicating that *Mundinia* is a distinct sub-genus of genus *Leishmania* within the subfamily *Leishmaniinae* (Kostygov and Yurchenko, 2017, Jariyapan et al., 2018a, Butenko et al., 2019a). However, considering the clear separation of branches between *L. (M.) martiniquensis* and other *Mundinia* species in phylogenomic trees, *L. (M.) martiniquensis* may be considered as a representative species within a possible new subgenus as they diverge around 60 - 70 MYA (Figure 6.3 and 6.5). Therefore, this distinction should be formally made, as around the same period, a similar distinction was made between the *Leishmania* and *Sauroleishmania* subgenera.

When species were initially chosen to be included in reconstructing the phylogenomic trees, genomes assembled at the chromosome level were chosen solely for consistency. This, however, resulted in a less informative tree (Figure 6.3). For instance, species with known fission and fusion events in their chromosomes, primarily genomes from the *Viannia* subgenus with smaller chromosome numbers, had to be excluded. However, because the adjustment step used previously in the LGAAP pipeline demonstrated its validity, it was also used to adjust and reorient the genomes to fit them into chromosome-scale assemblies in a way that allows for alignment and subsequent tree reconstruction to be insightful (Figure 6.5). As a result, the *L. major* Friedlin genome was used once more as a guided reference genome to order and orient the chromosomes for those that were not assembled at that level, as well as to fit the genomes with inconsistent chromosome numbers, as it is the most reliable reference genome for *Leishmania* species so far.

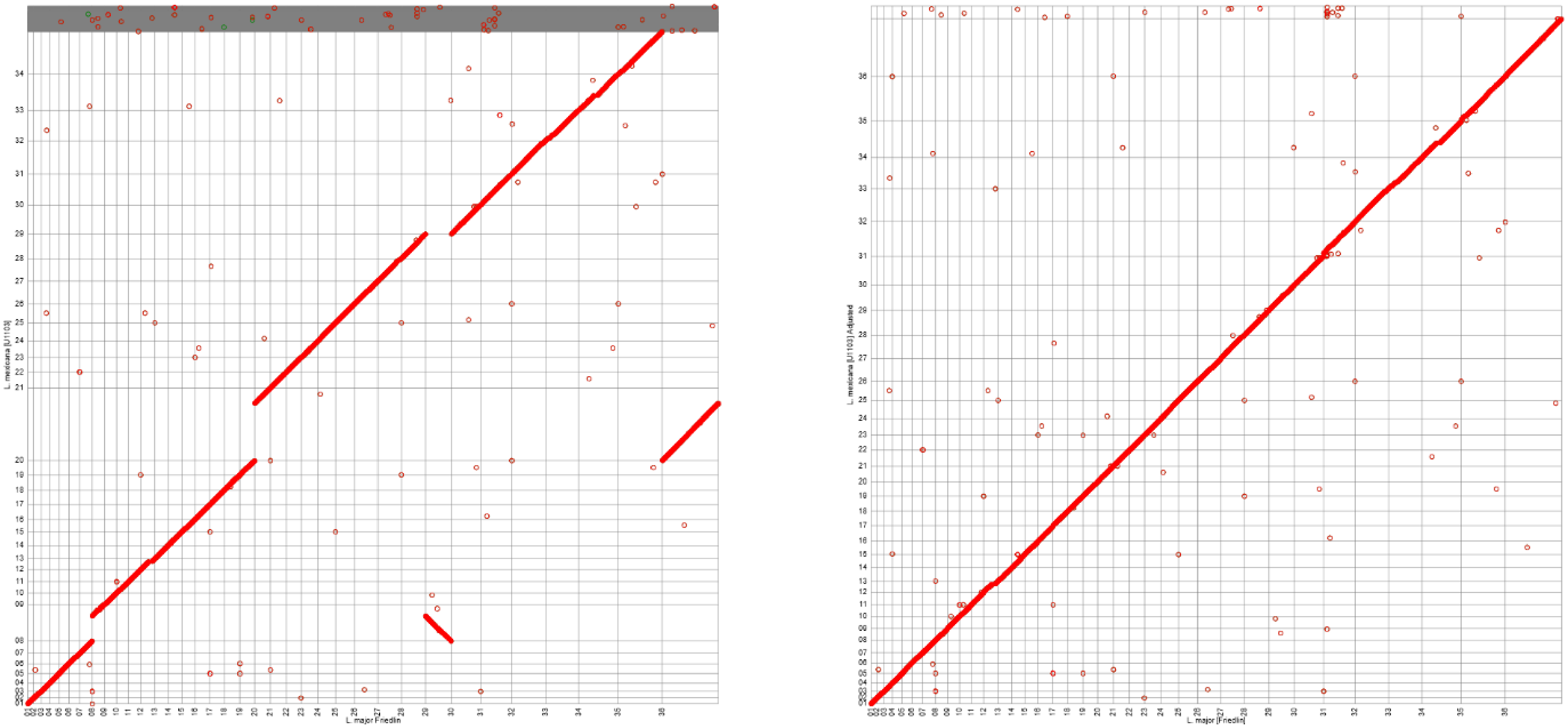
However, when *L. major* Friedlin genome was used to adjust for previously known species with fewer chromosomes, such as *L. (V.) braziliensis* and *L. (L.) mexicana*, the results revealed unexpectedly perfect alignments with different chromosomes (Figure 7.1 and 7.2). Given the strong alignments and the fact that the different chromosome numbers were reported based on using around 300 markers 20 years ago, it is possible that lower chromosome numbers might be wrong. This, of course, emphasises the importance of rectifying this issue with additional genome assemblies, particularly those with fewer chromosomes, using long read sequencing technologies and a pipeline roughly equivalent to the LGAAP pipeline.



Original

Adjusted

Figure 7.2 Dotplot comparison between the original genome of *L. (V.) braziliensis* M2903 (vertical axis) against *L. major* Friedlin stain (horizontal axis) before and after the adjustment. The figure shows that chromosomes 20.1 and 34 are matching completely (see *Supplementary Materials* for full-scale figures).



Original

Adjusted

Figure 7.3: Dotplot comparison between the original genome of *L. (L.) Mexicana* U1103 (vertical axis) subjected against *L. major* Friedlin stain (horizontal axis) before and after the adjustment. The figure shows that chromosome 8 and 29 as well as chromosomes 20 and 36 are matching with clear separation (see *Supplementary Materials* for full-scale figures).

Despite our encouraging results, a critical examination of current and previous assemblies is necessary to facilitate the development of new next generation assemblies for *Leishmania* species that incorporate both long and short reads. This will contribute to the development of more precise and accurate genomes. In addition to that, the results contribute a clearer understanding of the basic biology of *Mundinia* genomics and evolution, and eventually will aid in drug discovery, vector control, and treatment.

As previously stated, the results are limited in their generalizability due to the pipeline being tailored to assemble and annotate only six genomes. However, because the methodology is publicly available and open sourced, this provides an opportunity for others to obtain the pipeline code and adapt it for other genomes, not just *Leishmania*, but for other eukaryotic single cell organisms as well.

I attempted to be as reproducible as possible; however, because it is novel, it required many trials and errors to produce. This influences the time required to complete the task. However, once it has been thoroughly tested and retested, its application is quick, simple, and reliable.

7.6 Selection Pressure and Host Interaction

Positive selection analysis tool Slr was chosen over other selection algorithms in order to perform a statistical likelihood-ratio test with a particular emphasis on the strength of selection at each site. The strength of selection at each site is calculated by comparing the rate of nonsynonymous substitutions, which alter amino acids, to the rate of synonymous substitutions, which are considered to be a silent mutations and thus evolve in a strictly neutral manner (Massingham and Goldman, 2005a).

Only four of the 36 positively selected CDS are completely novel since they have no homologs in any of the protein families searched (Table 6.3). Among the remaining 32 CDS, two were found to be associated with evasion of the immune system, adhesion to plasma cells, and pathogenesis: one was found on chromosome 26 and contains Phosphotyrosine-binding domain (PTB), Meiotic cell cortex C-terminal pleckstrin homology, Pathogenicity factor, large tegument protein UL36, and Plant-like Phospholipase C (PLC) pleckstrin homology (PH) domain (Accession numbers: KAG5476398.1, KAG5502146.1, KAG5476121.1, KAG5476671.1, KAG5499538.1, and KAG5476398.1).

The phosphotyrosine-binding domain (PTB) is located at the C-terminus of the tensin protein. Tensin is a multi-domain protein that binds to actin filaments and serves as a focal-adhesion molecule, which are plasma membrane regions through which cells attach to the extracellular matrix (CHEN et al., 2000). During *Leishmania* infection, PTB has been linked to phagolysosome biogenesis and phagosome function during *L. major* infection. Dok proteins are expressed in macrophages and are involved in the negative regulation of signalling in response to lipopolysaccharide and various cytokines and growth factors (Boulais et al., 2010, de Celis et al., 2015).

The second CDS associated to immune system evasion, plasma cell adhesion, and pathogenesis was found on chromosome 33. It contains P-loop nucleoside triphosphate hydrolases, kinesin, large tegument protein UL36, and Med15 subunit of Mediator complex (Accession numbers: KAG5469580.1, KAG5494493.1, KAG5469225.1, KAG5469935.1, and KAG5493729.1).

Kinesins, a molecular motor superfamily, use microtubules as tracks to transport a variety of cellular cargoes (Lawrence et al., 2004). The 350 amino acid motor domain of all kinesins is highly conserved. *Leishmania* has a much larger kinesin repertoire than the amoeboid parasite, with many of them appearing to have evolved through multiple gene duplications (Richardson et al., 2006). The reason for *Leishmania* to have many kinesins is an intriguing possibility to consider. It is unclear how many of these kinesins are functional. It is currently unknown whether these kinesins play a role in facilitating *Leishmania*-host-cell interaction.

It is critical to note that the selection pressure analyses were conducted exclusively on *Mundinia* for two reasons: first, because those species are closely related, they have more similar CDS sequences and thus can be easily tested for selection; second, because some other

Leishmania species from other subgenera have not been annotated, which made them difficult to include in this analysis.

7.7 Future Opportunities

This project, in my opinion, has not yet been completed in its entirety. For example, preliminary findings indicate that kDNA was assembled within the unplaced contigs. These sequences are identical in length to the kDNA found in several *Leishmania* genomes. Additionally, When the kDNA sequence from unplaced contigs of *L. (M.) martiniquensis* (LSCM1) was examined as a pilot study using a mitochondrial annotation tool such as MITOS(Bernt et al., 2013), many mitochondrial genes and proteins were located and annotated in the unplaced contig number 11. The annotation contains ATP synthase subunits, cytochrome b, cytochrome c oxidase subunits, and NADH dehydrogenase subunits. Due to time constraints, this analysis can be considered for future work because it requires additional steps to validate the results and expand to other assemblies.

Future additional genomes need to be assembled, particularly those claimed to belong to *Mundinia but* with unclear taxonomy or do not have representative genomes, such as: the recently described *L. macropodum* (Barratt et al., 2017a); additional *L. (M.) martiniquensis* strains should be assembled, as there have been reports of different reservoirs and host in horses in Central Europe and the United States, as well as bovines in Switzerland, and therefore more diverse genomes can improve taxonomic phylogeny (Lobsiger et al., 2010, Muller et al., 2009); genomes from the subgenus *Viannia*, such as *L. (V.) braziliensis, L. (V.) panamensis* and *L. (V.) peruviana ;* genomes from the subgenus *Leishmania, such as L. (L.) mexicana*, as there have been no assemblies using short and long reads in the same way as this project, which makes it compelling to assemble given the chromosome number has been investigated here; additional genomes from the *Sauroleishmania* subgenus, as there are only two genomes from the same species; genomes from the genus *Zelonia*, including *Z. costaricensis* and *Z.*

australiensis (Barratt et al., 2017a); and genomes from the genera *Porcisia* and *Endotrypanum*, including *E. colombiensis*, *E. equatorensis*, and *E. herreri* (Espinosa et al., 2018, Kostygov and Yurchenko, 2017).

For more accurate and reliable annotation, a dedicated study on whole transcriptome assembly should be conducted. Furthermore, the unplaced contigs should be investigated to understand more about ploidy and kinetoplastids DNA. Moreover, on the bioinformatics side, additional technical research is required to determine whether any of the reference genomes can be used as a guide for assembling only short reads.

Chapter 8. Supplementary Materials

Due to the nature of the evidence presented in this thesis, appendices are not feasible. As a result, all data, including some tables and some full-scale figures, can be viewed at the link (https://doi.org/10.17635/lancaster/researchdata/509), which are indexed according to the chapter numbers.

Chapter 9. References

ABRUSAN, G., GRUNDMANN, N., DEMESTER, L. & MAKALOWSKI, W. 2009. TEclass-a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics,* 25**,** 1329-1330.

ADAUI, V., LYE, L.-F., AKOPYANTS, N. S., ZIMIC, M., LLANOS-CUENTAS, A., GARCIA, L., MAES, I., DE DONCKER, S., DOBSON, D. E. & AREVALO, J. 2016. Association of the endobiont double-stranded RNA virus LRV1 with treatment failure for human leishmaniasis caused by Leishmania braziliensis in Peru and Bolivia. *The Journal of infectious diseases,* 213**,** 112-121.

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 Sandflies. *PLOS Neglected Tropical Diseases,* 10**,** e0004349.

AL-SALEM, W., HERRICKS, J. R. & HOTEZ, P. J. 2016. A review of visceral leishmaniasis during the conflict in South Sudan and the consequences for East African countries. *Parasites & vectors,* 9**,** 1-11.

ALAWIEH, A., MUSHARRAFIEH, U., JABER, A., BERRY, A., GHOSN, N. & BIZRI, A. R. 2014. Revisiting leishmaniasis in the time of war: the Syrian conflict and the Lebanese outbreak. *International Journal of Infectious Diseases,* 29**,** 115-119.

ALBANAZ, A. T. S., GERASIMOV, E. S., SHAW, J. J., SADLOVA, J., LUKES, J., VOLF, P., OPPERDOES, F. R., KOSTYGOV, A. Y., BUTENKO, A. & YURCHENKO, V. 2021. Genome Analysis of Endotrypanum and Porcisia spp., Closest Phylogenetic Relatives of Leishmania, Highlights the Role of Amastins in Shaping Pathogenicity. *Genes,* 12.

ALMUTAIRI, H. 2021a. *hatimalmutairi/lmgaap-maker* [Online]. Available: https://hub.docker.com/r/hatimalmutairi/lmgaap-maker [Accessed].

ALMUTAIRI, H. 2021b. *hatimalmutairi/teclass: v2.1.3b* [Online]. University of Muenster. Available: https://hub.docker.com/r/hatimalmutairi/teclass-2.1.3b [Accessed].

ALMUTAIRI, H., URBANIAK, M. D., BATES, M. D., JARIYAPAN, N., KWAKYE-NUAKO, G., THOMAZ-SOCCOL, V., AL-SALEM, W. S., DILLON, R., BATES, P. A. & GATHERER, D. 2021. LGAAP: Leishmaniinae Genome Assembly and Annotation Pipeline. *Microbiology Resource Announcements,* 10.

ALONGE, M., SOYK, S., RAMAKRISHNAN, S., WANG, X. G., GOODWIN, S., SEDLAZECK, F. J., LIPPMAN, Z. B. & SCHATZ, M. C. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology,* 20.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research,* 25**,** 3389-3402.

ALVAR, J., VÉLEZ, I. D., BERN, C., HERRERO, M., DESJEUX, P., CANO, J., JANNIN, J., BOER, M. D. & TEAM, W. L. C. 2012. Leishmaniasis worldwide and global estimates of its incidence. *PloS one,* 7**,** e35671.

ALVAR, J., YACTAYO, S. & BERN, C. 2006. Leishmaniasis and poverty. *Trends in parasitology,* 22**,** 552-557.

ANDRADE-NARVÁEZ, F. J., VARGAS-GONZÁLEZ, A., CANTO-LARA, S. B. & DAMIÁN-CENTENO, A. G. 2001. Clinical picture of cutaneous leishmaniases due to Leishmania (Leishmania) mexicana in the Yucatan peninsula, Mexico. *Memórias do Instituto Oswaldo Cruz,* 96**,** 163-167.

ANDREWS, S. 2010. Babraham bioinformatics-FastQC a quality control tool for high throughput sequence data. *URL: https://www. bioinformatics. babraham. ac. uk/projects/fastqc*.

ARAUJO, A. R. D., PORTELA, N. C., FEITOSA, A. P. S., SILVA, O. A. D., XIMENES, R. A. A., ALVES, L. C. & BRAYNER, F. A. 2016. Risk factors associated with American cutaneous leishmaniasis in an endemic area of Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo,* 58.

ARONSON, N., HERWALDT, B. L., LIBMAN, M., PEARSON, R., LOPEZ-VELEZ, R., WEINA, P., CARVALHO, E., EPHROS, M., JERONIMO, S. & MAGILL, A. 2017. Diagnosis and Treatment of Leishmaniasis: Clinical Practice Guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH). *The American Journal of Tropical Medicine and Hygiene,* 96**,** 24-45.

ASATO, Y., OSHIRO, M., MYINT, C. K., YAMAMOTO, Y.-I., KATO, H., MARCO, J. D., MIMORI, T., GOMEZ, E. A., HASHIGUCHI, Y. & UEZATO, H. 2009a. Phylogenic analysis of the genus Leishmania by cytochrome b gene sequencing. *Experimental parasitology,* 121**,** 352-361.

ASATO, Y., OSHIRO, M., MYINT, C. K., YAMAMOTO, Y., KATO, H., MARCO, J. D., MIMORI, T., GOMEZ, E. A. L., HASHIGUCHI, Y. & UEZATO, H. 2009b. Phylogenic analysis of the genus Leishmania by cytochrome b gene sequencing. *Experimental Parasitology,* 121**,** 352-361.

ASLETT, M., AURRECOECHEA, C., BERRIMAN, M., BRESTELLI, J., BRUNK, B. P., CARRINGTON, M., DEPLEDGE, D. P., FISCHER, S., GAJRIA, B., GAO, X., GARDNER, M. J., GINGLE, A., GRANT, G., HARB, O. S., HEIGES, M., HERTZ-FOWLER, C., HOUSTON, R., INNAMORATO, F., IODICE, J., KISSINGER, J. C., KRAEMER, E., LI, W., LOGAN, F. J., MILLER, J. A., MITRA, S., MYLER, P. J., NAYAK, V., PENNINGTON, C., PHAN, I., PINNEY, D. F., RAMASAMY, G., ROGERS, M. B., ROOS, D. S., ROSS, C., SIVAM, D., SMITH, D. F., SRINIVASAMOORTHY, G., STOECKERT, C. J., JR., SUBRAMANIAN, S., THIBODEAU, R., TIVEY, A., TREATMAN, C., VELARDE, G. & WANG, H. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res,* 38**,** D457-62.

AZAMI-CONESA, I., GÓMEZ-MUÑOZ, M. T. & MARTÍNEZ-DÍAZ, R. A. 2021. A systematic review (1990–2021) of wild animals infected with zoonotic leishmania. *Microorganisms,* 9**,** 1101.

BAILEY, F., MONDRAGON-SHEM, K., HOTEZ, P., RUIZ-POSTIGO, J. A., AL-SALEM, W., ACOSTA-SERRANO, A. & MOLYNEUX, D. H. 2017. A new perspective on cutaneous leishmaniasis—Implications for global prevalence and burden of disease estimates. *PLoS neglected tropical diseases,* 11**,** e0005739.

BALASEGARAM, M., RITMEIJER, K., LIMA, M. A., BURZA, S., ORTIZ GENOVESE, G., MILANI, B., GASPANI, S., POTET, J. & CHAPPUIS, F. 2012. Liposomal amphotericin B as a treatment for human leishmaniasis. *Expert opinion on emerging drugs,* 17**,** 493-510.

BAMOROVAT, M., SHARIFI, I., MOHAMMADI, M. A., EYBPOOSH, S., NASIBI, S., AFLATOONIAN, M. R. & KHOSRAVI, A. 2018. Leishmania tropica isolates from non-healed and healed patients in Iran: A molecular typing and phylogenetic analysis. *Microbial Pathogenesis,* 116**,** 124-129.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A. A., DVORKIN, M., KULIKOV, A. S., LESIN, V. M., NIKOLENKO, S. I., PHAM, S., PRJIBELSKI, A. D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology,* 19**,** 455-477.

BANULS, A.-L., HIDE, M. & PRUGNOLLE, F. 2007. Leishmania and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in parasitology,* 64**,** 1-458.

BANULS, A. L., BRISSE, S., SIDIBE, I., NOEL, S. & TIBAYRENC, M. 1999. A phylogenetic analysis by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting of the Leishmania genome project Friedlin reference strain. *Folia Parasitol (Praha),* 46**,** 10-4.

BARRATT, J., KAUFER, A., PETERS, B., CRAIG, D., LAWRENCE, A., ROBERTS, T., LEE, R., MCAULIFFE, G., STARK, D. & ELLIS, J. 2017a. Isolation of Novel Trypanosomatid, Zelonia australiensis sp nov (Kinetoplastida: Trypanosomatidae) Provides Support for a Gondwanan Origin of Dixenous Parasitism in the Leishmaniinae. *Plos Neglected Tropical Diseases,* 11.

BARRATT, J., KAUFER, A., PETERS, B., CRAIG, D., LAWRENCE, A., ROBERTS, T., LEE, R., MCAULIFFE, G., STARK, D. & ELLIS, J. 2017b. Isolation of Novel Trypanosomatid, Zelonia australiensis sp. nov. (Kinetoplastida: Trypanosomatidae) Provides Support for a Gondwanan Origin of Dixenous Parasitism in the Leishmaniinae. *PLoS Negl Trop Dis,* 11**,** e0005215.

BARRIEL, V. & TASSY, P. 1998. Rooting with multiple outgroups: consensus versus parsimony. *Cladistics,* 14**,** 193-200.

BASTIEN, P., BLAINEAU, C. & PAGES, M. 1992. Molecular karyotype analysis in Leishmania. *Subcell Biochem,* 18**,** 131-87.

BEARD, K. C. 2008. The oldest North American primate and mammalian biogeography during the Paleocene–Eocene Thermal Maximum. *Proceedings of the National Academy of Sciences,* 105**,** 3815-3818.

BENNETT, B. C. & BALICK, M. J. 2014. Does the name really matter? The importance of botanical nomenclature and plant taxonomy in biomedical research. *Journal of ethnopharmacology,* 152**,** 387-392.

BENNIS, I., THYS, S., FILALI, H., DE BROUWERE, V., SAHIBI, H. & BOELAERT, M. 2017. Psychosocial impact of scars due to cutaneous leishmaniasis on high school students in Errachidia province, Morocco. *Infectious diseases of poverty,* 6**,** 1-8.

BERNT, M., DONATH, A., JÜHLING, F., EXTERNBRINK, F., FLORENTZ, C., FRITZSCH, G., PÜTZ, J., MIDDENDORF, M. & STADLER, P. F. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution,* 69**,** 313-319.

BLACKWELL, J. M., GOSWAMI, T., EVANS, C. A., SIBTHORPE, D., PAPO, N., WHITE, J. K., SEARLE, S., MILLER, E. N., PEACOCK, C. S. & MOHAMMED, H. 2001. SLC11A1 (formerly NRAMP1) and disease resistance. *Cellular microbiology,* 3**,** 773.

BLAKE, D. P. 2015. Eimeria genomics: where are we now and where are we going? *Veterinary Parasitology,* 212**,** 68-74.

BLEWETT, T. M., KADIVAR, D. M. & SOULSBY, E. J. 1971. Cutaneous leishmaniasis in the guinea pig. Delayed-type hypersensitivity, lymphocyte stimulation, and inhibition of macrophage migration. *Am J Trop Med Hyg,* 20**,** 546-51.

BOELAERT, M., MEHEUS, F., SANCHEZ, A., SINGH, S., VANLERBERGHE, V., PICADO, A., MEESSEN, B. & SUNDAR, S. 2009. The poorest of the poor: a poverty appraisal of households affected by visceral leishmaniasis in Bihar, India. *Tropical medicine & international health,* 14**,** 639-644.

BOETTIGER, C. 2015. An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review,* 49**,** 71-79.

BOISVERT, S., LAVIOLETTE, F. & CORBEIL, J. 2010. Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies. *Journal of Computational Biology,* 17**,** 1519-1533.

BOOTH, T., BICAK, M., GWEON, H. S., FIELD, D. & AFGAN, E. Bio-Linux as a tool for bioinformatics training. 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), 2012. IEEE, 578-582.

BOUCKAERT, R., HELED, J., KUHNERT, D., VAUGHAN, T., WU, C. H., XIE, D., SUCHARD, M. A., RAMBAUT, A. & DRUMMOND, A. J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol,* 10**,** e1003537.

BOULAIS, J., TROST, M., LANDRY, C. R., DIECKMANN, R., LEVY, E. D., SOLDATI, T., MICHNICK, S. W., THIBAULT, P. & DESJARDINS, M. 2010. Molecular characterization of the evolution of phagosomes. *Molecular systems biology,* 6**,** 423.

BOWEN, G. J., CLYDE, W. C., KOCH, P. L., TING, S., ALROY, J., TSUBAMOTO, T., WANG, Y. & WANG, Y. 2002. Mammalian dispersal at the Paleocene/Eocene boundary. *Science,* 295**,** 2062-2065.

BRADLEY, D. C. 2015. Mineral evolution and Earth history. *American Mineralogist,* 100**,** 4-5.

BRIONES, M. R., NELSON, K., BEVERLEY, S. M., AFFONSO, H. T., CAMARGO, E. P. & FLOETER-WINTER, L. M. 1992a. Leishmania tarentolae taxonomic relatedness inferred from phylogenetic analysis of the small subunit ribosomal RNA gene. *Mol Biochem Parasitol,* 53**,** 121-7.

BRIONES, M. R. S., NELSON, K., BEVERLEY, S. M., AFFONSO, H. T., CAMARGO, E. P. & FLOETERWINTER, L. M. 1992b. Leishmania-Tarentolae Taxonomic Relatedness Inferred from Phylogenetic Analysis of the Small Subunit Ribosomal-Rna Gene. *Molecular and Biochemical Parasitology,* 53**,** 121-128.

BRITTO, C., RAVEL, C., BASTIEN, P., BLAINEAU, C., PAGÈS, M., DEDET, J.-P. & WINCKER, P. 1998. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World Leishmania genomes. *Gene,* 222**,** 107-117.

BRUSCHI, F. & GRADONI, L. 2018. *The leishmaniases: old neglected tropical diseases*, Springer.

BUALERT, L., CHARUNGKIATTIKUL, W., THONGSUKSAI, P., MUNGTHIN, M., SIRIPATTANAPIPONG, S., KHOSITNITHIKUL, R., NAAGLOR, T., RAVEL, C., EL BAIDOURI, F. & LEELAYOOVA, S. 2012. Autochthonous disseminated dermal and visceral leishmaniasis in an AIDS patient, southern Thailand, caused by Leishmania siamensis. *Am J Trop Med Hyg,* 86**,** 821-4.

BURSET, M. & GUIGO, R. 1996. Evaluation of gene structure prediction programs. *genomics,* 34**,** 353-367.

BURZA, S., CROFT, S. L. & BOELAERT, M. 2018. Leishmaniasis. *The Lancet,* 392**,** 951-970.

BURZA, S., MAHAJAN, R., SANZ, M. G., SUNYOTO, T., KUMAR, R., MITRA, G. & LIMA, M. A. 2014. HIV and visceral leishmaniasis coinfection in Bihar, India: an underrecognized and underdiagnosed threat against elimination. *Clinical infectious diseases,* 59**,** 552-555.

BUTENKO, A., KOSTYGOV, A. Y., SADLOVA, J., KLESCHENKO, Y., BECVAR, T., PODESVOVA, L., MACEDO, D. H., ZIHALA, D., LUKES, J., BATES, P. A., VOLF, P., OPPERDOES, F. R. & YURCHENKO, V. 2019a. Comparative genomics of Leishmania (Mundinia). *Bmc Genomics,* 20.

BUTENKO, A., KOSTYGOV, A. Y., SÁDLOVÁ, J., KLESCHENKO, Y., BEČVÁŘ, T., PODEŠVOVÁ, L., MACEDO, D. H., ŽIHALA, D., LUKEŠ, J., BATES, P. A., VOLF, P., OPPERDOES, F. R. & YURCHENKO, V. 2019b. Comparative genomics of Leishmania (Mundinia). *BMC Genomics,* 20.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST plus : architecture and applications. *Bmc Bioinformatics,* 10.

CAMACHO, E., GONZALEZ-DE LA FUENT, S., RASTROJO, A., PEIRO-PASTOR, R., SOLANA, J. C., TABERA, L., GAMARRO, F., CARRASCO-RAMIRO, F., REQUEN, J. M. & AGUADO, B. 2019. Complete assembly of the Leishmania donovani (HU3 strain) genome and transcriptome annotation. *Scientific Reports,* 9.

CAMERON, M. M., ACOSTA-SERRANO, A., BERN, C., BOELAERT, M., DEN BOER, M., BURZA, S., CHAPMAN, L. A., CHASKOPOULOU, A., COLEMAN, M. & COURTENAY, O. 2016. Understanding the transmission dynamics of Leishmania donovani to provide robust evidence for interventions to eliminate visceral leishmaniasis in Bihar, India. *Parasites & vectors,* 9**,** 1-9.

CANTACESSI, C., DANTAS-TORRES, F., NOLAN, M. J. & OTRANTO, D. 2015. The past, present, and future of Leishmania genomics and transcriptomics. *Trends in Parasitology,* 31**,** 100-108.

CAO, D.-P., GUO, X.-G., CHEN, D.-L. & CHEN, J.-P. 2011. Species delimitation and phylogenetic relationships of Chinese Leishmania isolates reexamined using kinetoplast cytochrome oxidase II gene sequences. *Parasitology research,* 109**,** 163-173.

CARDOSO, L., SCHALLIG, H., PERSICHETTI, M. F. & PENNISI, M. G. 2021. New epidemiological aspects of animal leishmaniosis in Europe: The role of vertebrate hosts other than dogs. *Pathogens,* 10**,** 307.

CARVALHO, B. M., RANGEL, E. F., READY, P. D. & VALE, M. M. 2015. Ecological niche modelling predicts southward expansion of Lutzomyia (Nyssomyia) flaviscutellata (Diptera: Psychodidae: Phlebotominae), vector of Leishmania (Leishmania) amazonensis in South America, under climate change. *PLoS One,* 10**,** e0143282.

CASTRO NETO, A. L., BRITO, A. N., REZENDE, A. M., MAGALHÃES, F. B. & DE MELO NETO, O. P. 2019. In silico characterization of multiple genes encoding the GP63 virulence protein from Leishmania braziliensis: identification of sources of variation and putative roles in immune evasion. *BMC genomics,* 20**,** 1-17.

ÇETIN, M., OCAK, S. & ERTUNÇ, D. 2007. An unusual case of urinary tract infection caused by Aerococcus viridans.

CHANDRA, U., YADAV, A., KUMAR, D. & SAHA, S. 2017. Cell cycle stage-specific transcriptional activation of cyclins mediated by HAT2-dependent H4K10 acetylation of promoters in Leishmania donovani. *PLoS Pathog,* 13**,** e1006615.

CHANMOL, W., JARIYAPAN, N., SOMBOON, P., BATES, M. D. & BATES, P. A. 2019. Axenic amastigote cultivation and in vitro development of Leishmania orientalis. *Parasitol Res,* 118**,** 1885-1897.

CHAOUCH, M., FATHALLAH-MILI, A., DRISS, M., LAHMADI, R., AYARI, C., GUIZANI, I., BEN SAID, M. & BENABDERRAZAK, S. 2013. Identification of Tunisian Leishmania spp. by PCR amplification of cysteine proteinase B (cpb) genes and phylogenetic analysis. *Acta Trop,* 125**,** 357-65.

CHARMOY, M., BRUNNER-AGTEN, S., AEBISCHER, D., AUDERSET, F., LAUNOIS, P., MILON, G., PROUDFOOT, A. E. & TACCHINI-COTTIER, F. 2010. Neutrophil-derived CCL3 is essential for the rapid recruitment of dendritic cells to the site of Leishmania major inoculation in resistant mice. *PLoS pathogens,* 6**,** e1000755.

CHEN, H., ISHII, A., WONG, W.-K., CHEN, L. B. & LO, S. H. 2000. Molecular characterization of human tensin. *Biochemical Journal,* 351**,** 403-411.

CHICHARRO, C. & ALVAR, J. 2003. Lower trypanosomatids in HIV/AIDS patients. *Ann Trop Med Parasitol,* 97 Suppl 1**,** 75-8.

CHOUDHARY, J. S., BLACKSTOCK, W. P., CREASY, D. M. & COTTRELL, J. S. 2001. Matching peptide mass spectra to EST and genomic DNA databases. *TRENDS in Biotechnology,* 19**,** 17-22.

CHOUICHA, N., LANOTTE, G., PRATLONG, F., CUBA CUBA, C. A., VELEZ, I. D. & DEDET, J. P. 1997. Phylogenetic taxonomy of Leishmania (Viannia) braziliensis based on isoenzymatic study of 137 isolates. *Parasitology,* 115 (Pt 4)**,** 343-8.

CINCURÁ, C., DE LIMA, C. M. F., MACHADO, P. R., OLIVEIRA-FILHO, J., GLESBY, M. J., LESSA, M. M. & CARVALHO, E. M. 2017. Mucosal leishmaniasis: a retrospective study of 327 cases from an endemic area of Leishmania (Viannia) braziliensis. *The American journal of tropical medicine and hygiene,* 97**,** 761.

COCK, P. J. A., FIELDS, C. J., GOTO, N., HEUER, M. L. & RICE, P. M. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research,* 38**,** 1767-1771.

COHEN-FREUE, G., HOLZER, T. R., FORNEY, J. D. & MCMASTER, W. R. 2007. Global gene expression in Leishmania. *International journal for parasitology,* 37**,** 1077-1086.

COLEMAN, M., FOSTER, G. M., DEB, R., SINGH, R. P., ISMAIL, H. M., SHIVAM, P., GHOSH, A. K., DUNKLEY, S., KUMAR, V. & COLEMAN, M. 2015. DDT-based indoor residual spraying suboptimal for visceral leishmaniasis elimination in India. *Proceedings of the National Academy of Sciences,* 112**,** 8573-8578.

COMPEAU, P. E., PEVZNER, P. A. & TESLER, G. 2011. Why are de Bruijn graphs useful for genome assembly? *Nature biotechnology,* 29**,** 987.

COUGHLAN, S., MULHAIR, P., SANDERS, M., SCHONIAN, G., COTTON, J. A. & DOWNING, T. 2017. The genome of Leishmania adleri from a mammalian host highlights chromosome fission in Sauroleishmania. *Sci Rep,* 7**,** 43747.

CROAN, D. & ELLIS, J. 1996. Phylogenetic relationships between Leishmania, Viannia and Sauroleishmania inferred from comparison of a variable domain within the RNA polymerase II largest subunit gene. *Molecular and biochemical parasitology,* 79**,** 97-102.

CROAN, D. G., MORRISON, D. A. & ELLIS, J. T. 1997. Evolution of the genus Leishmania revealed by comparison of DNA and RNA polymerase gene sequences. *Molecular and biochemical parasitology,* 89**,** 149-159.

CROFT, S. & MOLYNEUX, D. 1979. Studies on the ultrastructure, virus-like particles and infectivity of Leishmania hertigi. *Annals of Tropical Medicine & Parasitology,* 73**,** 213-226.

CUPOLILLO, E., MEDINA-ACOSTA, E., NOYES, H., MOMEN, H. & GRIMALDI, G. 2000. A revised classification for Leishmania and Endotrypanum. *Parasitology today,* 16**,** 142-144.

CUPOLILLO, E., PEREIRA, L. O. R., FERNANDES, O., CATANHO, M. P., PEREIRA, J. C., MEDINA-ACOSTA, E. & GRIMALDI, G. 1998. Genetic data showing evolutionary links between Leishmania and Endotrypanum. *Memorias Do Instituto Oswaldo Cruz,* 93**,** 677-683.

CURRIE, L. A. 1995. Nomenclature in Evaluation of Analytical Methods Including Detection and Quantification Capabilities (Iupac Recommendations 1995). *Pure and Applied Chemistry,* 67**,** 1699-1723.

D'AVILA-LEVY, C. M., BOUCINHA, C., KOSTYGOV, A., SANTOS, H. L. C., MORELLI, K. A., GRYBCHUK-IEREMENKO, A., DUVAL, L., VOTYPKA, J., YURCHENKO, V., GRELLIER, P. & LUKES, J. 2015. Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Memorias Do Instituto Oswaldo Cruz,* 110**,** 956-965.

DAGNÆS-HANSEN, F., KILIAN, M. & FUURSTED, K. 2004. Septicaemia associated with an Aerococcus viridans infection in immunodeficient mice. *Laboratory animals,* 38**,** 321-325.

DAINAT, J. & HEREÑÚ, D. 2020. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format.(Version v0. 4.0). Zenodo.

DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M. & LI, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience,* 10.

DANIELS, J. P., GULL, K. & WICKSTEAD, B. 2010. Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev,* 74**,** 552-69.

DAS, V. N. R., PANDEY, R. N., SIDDIQUI, N. A., CHAPMAN, L. A., 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 neglected tropical diseases,* 10**,** e0005196.

DAVID, C., DIMIER-DAVID, L., VARGAS, F., TORREZ, M. & DEDET, J. P. 1993. Fifteen years of cutaneous and mucocutaneous leishmaniasis in Bolivia: a retrospective study. *Trans R Soc Trop Med Hyg,* 87**,** 7-9.

DAVIES, C. R., LLANOS-CUENTAS, E. A., SHARP, S. J., CANALES, J., LEON, E., ALVAREZ, E., RONCAL, N. & DYE, C. 1997. Cutaneous leishmaniasis in the Peruvian Andes: factors associated with variability in clinical symptoms, response to treatment, and parasite isolation rate. *Clinical infectious diseases,* 25**,** 302-310.

DÁVILA, A. & MOMEN, H. 2000. Internal-transcribed-spacer (ITS) sequences used to explore phylogenetic relationships within Leishmania. *Annals of Tropical Medicine & Parasitology,* 94**,** 651-654.

DE ARAÚJO PEDROSA, F. & DE ALENCAR XIMENES, R. A. 2009. Sociodemographic and environmental risk factors for American cutaneous leishmaniasis (ACL) in the State of Alagoas, Brazil. *The American journal of tropical medicine and hygiene,* 81**,** 195-201.

DE CELIS, H. Á., GOMEZ, C. P., DESCOTEAUX, A. & DUPLAY, P. 2015. Dok proteins are recruited to the phagosome and degraded in a GP63-dependent manner during Leishmania major infection. *Microbes and infection,* 17**,** 285-294.

DE LA TORRE-BÁRCENA, J. E., KOLOKOTRONIS, S.-O., LEE, E. K., STEVENSON, D. W., BRENNER, E. D., KATARI, M. S., CORUZZI, G. M. & DESALLE, R. 2009. The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS One,* 4**,** e5764.

DE VRIES, H. J., REEDIJK, S. H. & SCHALLIG, H. D. 2015. Cutaneous leishmaniasis: recent developments in diagnosis and management. *American journal of clinical dermatology,* 16**,** 99-109.

DEDET, J. P., ROCHE, B., PRATLONG, F., CALES-QUIST, D., JOUANNELLE, J., BENICHOU, J. C. & HUERRE, M. 1995. Diffuse cutaneous infection caused by a presumed monoxenous trypanosomatid in a patient infected with HIV. *Trans R Soc Trop Med Hyg,* 89**,** 644-6.

DESBOIS, N., PRATLONG, F., QUIST, D. & DEDET, J. P. 2014. Leishmania (Leishmania) martiniquensis n. sp. (Kinetoplastida: Trypanosomatidae), description of the parasite responsible for cutaneous leishmaniasis in Martinique Island (French West Indies). *Parasite,* 21**,** 12.

DESJEUX, P. 2004. Leishmaniasis: current situation and new perspectives. *Comparative immunology, microbiology and infectious diseases,* 27**,** 305-318.

DIXON, D., BENTON, M. J., KINGSLEY, A. & BAKER, J. 2001. *Atlas of Life on Earth*, Barnes & Noble.

DOWNING, T., IMAMURA, H., DECUYPERE, S., CLARK, T. G., COOMBS, G. H., COTTON, J. A., HILLEY, J. D., DE DONCKER, S., MAES, I., MOTTRAM, J. C., QUAIL, M. A., RIJAL, S., SANDERS, M., SCHONIAN, G., STARK, O., SUNDAR, S., VANAERSCHOT, M., HERTZ-FOWLER, C., DUJARDIN, J. C. & BERRIMAN, M. 2011. Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Research,* 21**,** 2143-2156.

DRUMMOND, A. J. & RAMBAUT, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol,* 7**,** 214.

DU, R., HOTEZ, P. J., AL-SALEM, W. S. & ACOSTA-SERRANO, A. 2016a. Old world cutaneous leishmaniasis and refugee crises in the Middle East and North Africa. Public Library of Science San Francisco, CA USA.

DU, R., HOTEZ, P. J., AL-SALEM, W. S. & ACOSTA-SERRANO, A. 2016b. Old world cutaneous leishmaniasis and refugee crises in the Middle East and North Africa. Public Library of Science San Francisco, CA USA.

DUJARDIN, J.-C., CAMPINO, L., CAÑAVATE, C., DEDET, J.-P., GRADONI, L., SOTERIADOU, K., MAZERIS, A., OZBEL, Y. & BOELAERT, M. 2008. Spread of vector-borne diseases and neglect of Leishmaniasis, Europe. *Emerging infectious diseases,* 14**,** 1013.

DUJARDIN, J. C. 2009. Structure, dynamics and function of Leishmania genome: resolving the puzzle of infection, genetics and evolution? *Infect Genet Evol,* 9**,** 290-7.

EILBECK, K., MOORE, B., HOLT, C. & YANDELL, M. 2009. Quantitative measures for the management and comparison of annotated genomes. *Bmc Bioinformatics,* 10.

EJARA, E. D., LYNEN, L., BOELAERT, M. & VAN GRIENSVEN, J. 2010. Challenges in HIV and visceral Leishmania co-infection: future research directions.

ELKHAIR, E. B. 2014. Elevated cortisol level due to visceral leishmaniasis and skin hyper-pigmentation are causally related. *Int J Sci Commer Humanit,* 2**,** 7.

EMMS, D. M. & KELLY, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology,* 16.

ESPINOSA, O. A., SERRANO, M. G., CAMARGO, E. P., TEIXEIRA, M. M. G. & SHAW, J. J. 2018. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as Leishmania and Endotrypanum. *Parasitology,* 145**,** 430-442.

ESWARAPRASAD, R. & RAJA, L. 2017. A review of virtual machine (VM) resource scheduling algorithms in cloud computing environment. *Journal of Statistics and Management Systems,* 20**,** 703-711.

EWELS, P., MAGNUSSON, M., LUNDIN, S. & KALLER, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics,* 32**,** 3047-3048.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome research,* 8**,** 175-185.

FLYNN, J. M., HUBLEY, R., GOUBERT, C., ROSEN, J., CLARK, A. G., FESCHOTTE, C. & SMIT, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families.

Proceedings of the National Academy of Sciences of the United States of America, 117**,** 9451-9457.

FONG, D. & LEE, B. 1988. Beta tubulin gene of the parasitic protozoan Leishmania mexicana. *Molecular and biochemical parasitology,* 31**,** 97-106.

GAVGANI, A. M., HODJATI, M., MOHITE, H. & DAVIES, C. 2002. Effect of insecticide-impregnated dog collars on incidence of zoonotic visceral leishmaniasis in Iranian children: a matchedcluster randomised trial. *The Lancet,* 360**,** 374-379.

GERNHARD, T. 2008. The conditioned reconstructed process. *J Theor Biol,* 253**,** 769-78.

GIJÓN-ROBLES, P., ABATTOUY, N., MERINO-ESPINOSA, G., EL KHALFAOUI, N., MORILLAS-MÁRQUEZ, F., CORPAS-LÓPEZ, V., PORCEL-RODRÍGUEZ, L., JAAOUANI, N., DÍAZ-SÁEZ, V. & RIYAD, M. 2018. Risk factors for the expansion of cutaneous leishmaniasis by Leishmania tropica: Possible implications for control programmes. *Transboundary and emerging diseases,* 65**,** 1615-1626.

GIRIBET, G. & RIBERA, C. 1998. The position of arthropods in the animal kingdom: a search for a reliable outgroup for internal arthropod phylogeny. *Molecular phylogenetics and evolution,* 9**,** 481-488.

GNERRE, S., MACCALLUM, I., PRZYBYLSKI, D., RIBEIRO, F. J., BURTON, J. N., WALKER, B. J., SHARPE, T., HALL, G., SHEA, T. P., SYKES, S., BERLIN, A. M., AIRD, D., COSTELLO, M., DAZA, R., WILLIAMS, L., NICOL, R., GNIRKE, A., NUSBAUM, C., LANDER, E. S. & JAFFE, D. B. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 1513-1518.

GOSSAGE, S. M., ROGERS, M. E. & BATES, P. A. 2003. Two separate growth phases during the development of Leishmania in sand flies: implications for understanding the life cycle. *International journal for parasitology,* 33**,** 1027-1034.

GOTO, H. & LAULETTA LINDOSO, J. A. 2012. Cutaneous and mucocutaneous leishmaniasis. *Infect Dis Clin North Am,* 26**,** 293-307.

GRAHAM, S. W., OLMSTEAD, R. G. & BARRETT, S. C. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Mol Biol Evol,* 19**,** 1769-81.

GREMME, G., STEINBISS, S. & KURTZ, S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform,* 10**,** 645-56.

GRIFFERTY, G., SHIRLEY, H., MCGLOIN, J., KAHN, J., ORRIOLS, A. & WAMAI, R. 2021. Vulnerabilities to and the Socioeconomic and Psychosocial Impacts of the Leishmaniases: A Review. *Research and Reports in Tropical Medicine,* 12**,** 135.

GROVE, S. 1978. The clinical and histological features of South West African cutaneous leishmaniasis. *South African Medical Journal,* 53**,** 712-715.

GROVE, S. S. 1989. Leishmaniasis in South West-Africa Namibia to Date. *South African Medical Journal,* 75**,** 290-292.

GROVE, S. S. & LEDGER, J. A. 1975. Leishmania from a Hyrax in South West-Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 69**,** 523-524.

GRUNING, B., DALE, R., SJODIN, A., CHAPMAN, B. A., ROWE, J., TOMKINS-TINCH, C. H., VALIERIS, R., KOSTER, J. & BIOCONDA, T. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods,* 15**,** 475-476.

GUIGUEMDÉ, R. T., SAWADOGO, O. S., BORIES, C., TRAORE, K. L., NEZIEN, D., NIKIEMA, L., PRATLONG, F., MARTY, P., HOUIN, R. & DENIAU, M. 2003. Leishmania major and HIV co-infection in Burkina Faso. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 97**,** 168-169.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-1075.

HARHAY, M. O., OLLIARO, P. L., VAILLANT, M., CHAPPUIS, F., LIMA, M. A., RITMEIJER, K., COSTA, C. H., COSTA, D. L., RIJAL, S. & SUNDAR, S. 2011. Who is a typical patient with visceral leishmaniasis? Characterizing the demographic and nutritional profile of patients in Brazil, East Africa, and South Asia. *The American journal of tropical medicine and hygiene,* 84**,** 543.

HARKINS, K. M., SCHWARTZ, R. S., CARTWRIGHT, R. A. & STONE, A. C. 2016. Phylogenomic reconstruction supports supercontinent origins for Leishmania. *Infection Genetics and Evolution,* 38**,** 101-109.

HAYANI, K., DANDASHLI, A. & WEISSHAR, E. 2015. Cutaneous leishmaniasis in Syria: clinical features, current status and the effects of war. *Acta dermato-venereologica,* 95.

HERNANDEZ, D., FRANCOIS, P., FARINELLI, L., OSTERAS, M. & SCHRENZEL, J. 2008. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research,* 18**,** 802-809.

HERRER, A. 1971. Leishmania hertigi sp. n., from the Tropical Porcupine, Coendou rothschildi Thomas. *The Journal of Parasitology,* 57**,** 626.

HERWALDT, B. L. 1999. Miltefosine—the long-awaited therapy for visceral leishmaniasis? : Mass Medical Soc.

HIDE, M., BANULS, A. L. & TIBAYRENC, M. 2001. Genetic heterogeneity and phylogenetic status of Leishmania (Leishmania) infantum zymodeme MON-1: epidemiological implications. *Parasitology,* 123**,** 425-32.

HOARE, C. A. & WALLACE, F. G. 1966. Developmental Stages of Trypanosomatid Flagellates - a New Terminology. *Nature,* 212**,** 1385-&.

HODIAMONT, C. J., KAGER, P. A., BART, A., DE VRIES, H. J., VAN THIEL, P. P., LEENSTRA, T., DE VRIES, P. J., VAN VUGT, M., GROBUSCH, M. P. & VAN GOOL, T. 2014. Species-directed therapy for leishmaniasis in returning travellers: a comprehensive guide. *PLoS neglected tropical diseases,* 8**,** e2832.

HOFF, K. J. & STANKE, M. 2019. Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinformatics,* 65**,** e57.

HOLT, C. & YANDELL, M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics,* 12**,** 491.

HOTEZ, P. J. 2018. Modern Sunni-Shia conflicts and their neglected tropical diseases. Public Library of Science San Francisco, CA USA.

HOTEZ, P. J., ALVARADO, M., BASANEZ, M. G., BOLLIGER, I., BOURNE, R., BOUSSINESQ, M., BROOKER, S. J., BROWN, A. S., BUCKLE, G., BUDKE, C. M., CARABIN, H., COFFENG, L. E., FEVRE, E. M., FURST, T., HALASA, Y. A., JASRASARIA, R., JOHNS, N. E., KEISER, J., KING, C. H., LOZANO, R., MURDOCH, M. E., O'HANLON, S., PION, S. D., PULLAN, R. L., RAMAIAH, K. D., ROBERTS, T., SHEPARD, D. S., SMITH, J. L., STOLK, W. A., UNDURRAGA, E. A., UTZINGER, J., WANG, M., MURRAY, C. J. & NAGHAVI, M. 2014. The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. *PLoS Negl Trop Dis,* 8**,** e2865.

HUSON, D. H. & BRYANT, D. 2005. Estimating phylogenetic trees and networks using SplitsTree 4. *Manuscript in preparation, software available from www. splitstree. org*.

IBRAHIM, M. E. 2002. The epidemiology of visceral leishmaniasis in east Africa: hints and molecular revelations. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 96**,** S25-S29.

IOSUP, A., OSTERMANN, S., YIGITBASI, M. N., PRODAN, R., FAHRINGER, T. & EPEMA, D. H. J. 2011. Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing. *IEEE Transactions on Parallel and Distributed Systems,* 22**,** 931-945.

IVENS, A. C., PEACOCK, C. S., WORTHEY, E. A., MURPHY, L., AGGARWAL, G., BERRIMAN, M., SISK, E., RAJANDREAM, M. A., ADLEM, E., AERT, R., ANUPAMA, A., APOSTOLOU, Z., ATTIPOE, P., BASON, N., BAUSER, C., BECK, A., BEVERLEY, S. M., BIANCHETTIN, G., BORZYM, K., BOTHE, G., BRUSCHI, C. V., COLLINS, M., CADAG, E., CIARLONI, L., CLAYTON, C., COULSON, R. M. R., CRONIN, A., CRUZ, A. K., DAVIES, R. M., DE GAUDENZI, J., DOBSON, D. E., DUESTERHOEFT, A., FAZELINA, G., FOSKER, N., FRASCH, A. C., FRASER, A., FUCHS, M., GABEL, C., GOBLE, A., GOFFEAU, A., HARRIS, D., HERTZ-FOWLER, C., HILBERT, H., HORN, D., HUANG, Y. T., KLAGES, S., KNIGHTS, A., KUBE, M., LARKE, N., LITVIN, L., LORD, A., LOUIE, T., MARRA, M., MASUY, D., MATTHEWS, K., MICHAELI, S., MOTTRAM, J. C., MULLER-AUER, S., MUNDEN, H., NELSON, S., NORBERTCZAK, H., OLIVER, K., O'NEIL, S., PENTONY, M., POHL, T. M., PRICE, C., PURNELLE, B., QUAIL, M. A., RABBINOWITSCH, E., REINHARDT, R., RIEGER, M., RINTA, J., ROBBEN, J., ROBERTSON, L., RUIZ, J. C., RUTTER, S., SAUNDERS, D., SCHAFER, M., SCHEIN, J., SCHWARTZ, D. C., 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. G., ZIMMERMANN, W., SMITH, D. F., BLACKWELL, J. M., STUART, K. D., BARRELL, B., et al. 2005. The genome of the kinetoplastid parasite, Leishmania major. *Science,* 309**,** 436-442.

IVES, A., RONET, C., PREVEL, F., RUZZANTE, G., FUERTES-MARRACO, S., SCHUTZ, F., ZANGGER, H., REVAZ-BRETON, M., LYE, L.-F. & HICKERSON, S. M. 2011. Leishmania RNA virus controls the severity of mucocutaneous leishmaniasis. *Science,* 331**,** 775-778.

JACQUES, I., ANDREWS, N. W. & HUYNH, C. 2010. Functional characterization of LIT1, the Leishmania amazonensis ferrous iron transporter. *Molecular and biochemical parasitology,* 170**,** 28-36.

JANG, M. 2006. *Linux Annoyances for Geeks: Getting the Most Flexible System in the World Just the Way You Want It*, " O'Reilly Media, Inc.".

JARIYAPAN, N., DAROONTUM, T., JAIWONG, K., CHANMOL, W., INTAKHAN, N., SOR-SUWAN, S., SIRIYASATIEN, P., SOMBOON, P., BATES, M. D. & BATES, P. A. 2018a. Leishmania (Mundinia) orientalis n. sp (Trypanosomatidae), a parasite from Thailand responsible for localised cutaneous leishmaniasis. *Parasites & Vectors,* 11.

JARIYAPAN, N., DAROONTUM, T., JAIWONG, K., CHANMOL, W., INTAKHAN, N., SOR-SUWAN, S., SIRIYASATIEN, P., SOMBOON, P., BATES, M. D. & BATES, P. A. 2018b. Leishmania (Mundinia) orientalis n. sp. (Trypanosomatidae), a parasite from Thailand responsible for localised cutaneous leishmaniasis. *Parasites & Vectors,* 11.

JIRKU, M., YURCHENKO, V. Y., LUKES, J. & MASLOV, D. A. 2012. New Species of Insect Trypanosomatids from Costa Rica and the Proposal for a New Subfamily within the Trypanosomatidae. *Journal of Eukaryotic Microbiology,* 59**,** 537-547.

JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R. & HUNTER, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-40.

KARSANI, S. A. 2006. Proteomic analysis of Leishmania mexicana differentation.

KASHIF, M., MANNA, P. P., AKHTER, Y., ALAIDAROUS, M. & RUB, A. 2017. Screening of Novel Inhibitors Against Leishmania donovani Calcium ion Channel to Fight Leishmaniasis. *Infect Disord Drug Targets,* 17**,** 120-129.

KAUFER, A., BARRATT, J., STARK, D. & ELLIS, J. 2019. The complete coding region of the maxicircle as a superior phylogenetic marker for exploring evolutionary relationships between members of the Leishmaniinae. *Infection Genetics and Evolution,* 70**,** 90-100.

KAYE, P. & SCOTT, P. 2011. Leishmaniasis: complexity at the host–pathogen interface. *Nature reviews microbiology,* 9**,** 604-615.

KAZEMI, B. 2011. Genomic organization of Leishmania species. *Iranian Journal of Parasitology,* 6**,** 1.

KERBAUGH, M. A. & EVANS, J. B. 1968. Aerococcus viridans in the hospital environment. *Applied Microbiology,* 16**,** 519-523.

KHAMESIPOUR, A. & RATH, B. 2016. Refugee health and the risk of cutaneous leishmaniasis in Europe. *International Journal of Infectious Diseases,* 53**,** 95-96.

KIMURA, M. 1980. A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide-Sequences. *Journal of Molecular Evolution,* 16**,** 111-120.

KIMUTAI, R., MUSA, A. M., NJOROGE, S., OMOLLO, R., ALVES, F., HAILU, A., KHALIL, E. A., DIRO, E., SOIPEI, P. & MUSA, B. 2017. Safety and effectiveness of sodium stibogluconate and paromomycin combination for the treatment of visceral leishmaniasis in eastern Africa: results from a pharmacovigilance programme. *Clinical drug investigation,* 37**,** 259-272.

KITTS, P., MADDEN, T., SICOTTE, H., BLACK, L. & OSTELL, J. 2011. UniVec database. *Available from: ncbi. nlm. nih. gov/VecScreen/UniVec. html*.

KITTS, P. A., CHURCH, D. M., THIBAUD-NISSEN, F., CHOI, J., HEM, V., SAPOJNIKOV, V., SMITH, R. G., TATUSOVA, T., XIANG, C. & ZHERIKOV, A. 2016. Assembly: a resource for assembled genomes at NCBI. *Nucleic acids research,* 44**,** D73-D80.

KOHN, G. C. 2007. *Encyclopedia of plague and pestilence: from ancient times to the present*, Infobase Publishing.

KOLMOGOROV, M., YUAN, J., LIN, Y. & PEVZNER, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology,* 37**,** 540-+.

KOLTAS, I. S., EROGLU, F., ALABAZ, D. & UZUN, S. 2014. The emergence of Leishmania major and Leishmania donovani in southern Turkey. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 108**,** 154-158.

KOSTYGOV, A. Y. & YURCHENKO, V. 2017. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitologica,* 64.

KUMAR, S., STECHER, G., LI, M., KNYAZ, C. & TAMURA, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution,* 35**,** 1547-1549.

KUMAR, S., STECHER, G., SULESKI, M. & HEDGES, S. B. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol,* 34**,** 1812-1819.

KURTZ, S., PHILLIPPY, A., DELCHER, A. L., SMOOT, M., SHUMWAY, M., ANTONESCU, C. & SALZBERG, S. L. 2004. Versatile and open software for comparing large genomes. *Genome Biol,* 5**,** R12.

KWAKYE-NUAKO, G., MOSORE, M. T., DUPLESSIS, C., BATES, M. D., PUPLAMPU, N., MENSAH-ATTIPOE, I., DESEWU, K., AFEGBE, G., ASMAH, R. H., JAMJOOM, M. B., AYEH-KUMI, P. F., BOAKYE, D. A. & BATES, P. A. 2015. First isolation of a new species of Leishmania responsible for human cutaneous leishmaniasis in Ghana and classification in the Leishmania enriettii complex. *Int J Parasitol,* 45**,** 679-84.

LAINSON, R. 1997. On Leishmania enriettii and other enigmatic Leishmania species of the neotropics. *Memorias Do Instituto Oswaldo Cruz,* 92**,** 377-387.

LAINSON, R. & SHAW, J. J. 1987. *Evolution, classification and geographical distribution*, Academic Press.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LAWRENCE, C. J., DAWE, R. K., CHRISTIE, K. R., CLEVELAND, D. W., DAWSON, S. C., ENDOW, S. A., GOLDSTEIN, L. S., GOODSON, H. V., HIROKAWA, N. & HOWARD, J. 2004. A standardized kinesin nomenclature. *The Journal of cell biology,* 167**,** 19-22.

LEELAYOOVA, S., SIRIPATTANAPIPONG, S., HITAKARUN, A., KATO, H., TAN-ARIYA, P., SIRIYASATIEN, P., OSATAKUL, S. & MUNGTHIN, M. 2013. Multilocus characterization and phylogenetic analysis of Leishmania siamensis isolated from autochthonous visceral leishmaniasis cases, southern Thailand. *BMC microbiology,* 13**,** 1-7.

LEGER, A. & LEONARDI, T. 2019. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source Software,* 4**,** 1236.

LEVICK, M. P., BLACKWELL, J. M., CONNOR, V., COULSON, R. M., MILES, A., SMITH, H. E., WAN, K. L. & AJIOKA, J. W. 1996. An expressed sequence tag analysis of a full-length, spliced-leader cDNA library from Leishmania major promastigotes. *Mol Biochem Parasitol,* 76**,** 345-8.

LI, H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics,* 32**,** 2103-2110.

LI, J., TAN, X., CHEN, X., WONG, D. S. & XHAFA, F. 2015. OPoR: Enabling Proof of Retrievability in Cloud Computing with Resource-Constrained Devices. *IEEE Transactions on Cloud Computing,* 3**,** 195-205.

LIGHTHALL, G. & GIANNINI, S. 1992. The chromosomes of Leishmania. *Parasitology Today,* 8**,** 192-199.

LLANES, A., RESTREPO, C. M., DEL VECCHIO, G., ANGUIZOLA, F. J. & LLEONART, R. 2015. The genome of Leishmania panamensis: insights into genomics of the L. (Viannia) subgenus. *Scientific Reports,* 5.

LOBSIGER, L., MULLER, N., SCHWEIZER, T., FREY, C. F., WIEDERKEHR, D., ZUMKEHR, B. & GOTTSTEIN, B. 2010. An autochthonous case of cutaneous bovine leishmaniasis in Switzerland. *Vet Parasitol,* 169**,** 408-14.

LODGE, R. & DESCOTEAUX, A. 2008. Leishmania invasion and phagosome biogenesis. *Molecular Mechanisms of Parasite Invasion***,** 174-181.

LUKES, J., BUTENKO, A., HASHIMI, H., MASLOV, D. A., VOTYPKA, J. & YURCHENKO, V. 2018. Trypanosomatids Are Much More than Just Trypanosomes: Clues from the Expanded Family Tree. *Trends in Parasitology,* 34**,** 466-480.

LUKES, J., MAURICIO, I. L., SCHONIAN, G., DUJARDIN, J. C., SOTERIADOU, K., DEDET, J. P., KUHLS, K., TINTAYA, K. W., JIRKU, M., CHOCHOLOVA, E., HARALAMBOUS, C., PRATLONG, F., OBORNIK, M., HORAK, A., AYALA, F. J. & MILES, M. A. 2007a. Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. *Proc Natl Acad Sci U S A,* 104**,** 9375-80.

LUKES, J., MAURICIO, I. L., SCHONIAN, G., DUJARDIN, J. C., SOTERIADOU, K., DEDET, J. P., KUHLS, K., TINTAYA, K. W. Q., JIRKU, M., CHOCHOLOVA, E., HARALAMBOUS, C., PRATLONG, F., OBORNIK, M., HORAK, A., AYALA, F. J. & MILES, M. A. 2007b. Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 9375-9380.

LUO, R. B., LIU, B. H., XIE, Y. L., LI, Z. Y., HUANG, W. H., YUAN, J. Y., HE, G. Z., CHEN, Y. X., PAN, Q., LIU, Y. J., TANG, J. B., WU, G. X., ZHANG, H., SHI, Y. J., LIU, Y., YU, C., WANG, B., LU, Y., HAN, C. L., CHEUNG, D. W., YIU, S. M., PENG, S. L., ZHU, X. Q., LIU, G. M., LIAO, X. K., LI, Y. R., YANG, H. M., WANG, J., LAM, T. W. & WANG, J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience,* 1.

MANN, M. & PANDEY, A. 2001. Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends in biochemical sciences,* 26**,** 54-61.

MARCILI, A., SPERANCA, M. A., DA COSTA, A. P., MADEIRA, M. D., SOARES, H. S., SANCHES, C. D. C. C., ACOSTA, I. D. L., GIROTTO, A., MINERVINO, A. H. H., HORTA, M. C., SHAW, J. J. & GENNARI, S. M. 2014. Phylogenetic relationships of Leishmania species based on trypanosomatid barcode (SSU rDNA) and gGAPDH genes: Taxonomic revision of Leishmania (L.) infantum chagasi in South America. *Infection Genetics and Evolution,* 25**,** 44-51.

MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics,* 24**,** 133-141.

MARDIS, E. R. 2011. A decade's perspective on DNA sequencing technology. *Nature,* 470**,** 198-203.

MARLOWE, F. W. 2005. Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews,* 14**,** 54-67.

MARTíNEZ-CALVILLO, S., YAN, S., NGUYEN, D., FOX, M., STUART, K. & MYLER, P. J. 2003. Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Molecular cell,* 11**,** 1291-1299.

MASLOV, D. A., OPPERDOES, F. R., KOSTYGOV, A. Y., HASHIMI, H., LUKES, J. & YURCHENKO, V. 2019. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology,* 146**,** 1-27.

MASLOV, D. A., VOTYPKA, J., YURCHENKO, V. & LUKES, J. 2013. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends in Parasitology,* 29**,** 43-52.

MASSINGHAM, T. & GOLDMAN, N. 2005a. Detecting amino acid sites under positive selection and purifying selection. *Genetics,* 169**,** 1753-1762.

MASSINGHAM, T. & GOLDMAN, N. 2005b. Detecting amino acid sites under positive selection and purifying selection. *Genetics,* 169**,** 1753-62.

MCDONAGH, P. D., MYLER, P. J. & STUART, K. 2000. The unusual gene organization of Leishmania major chromosome 1 may reflect novel transcription processes. *Nucleic acids research,* 28**,** 2800-2803.

MCINERNEY, F. A. & WING, S. L. 2011. The Paleocene-Eocene Thermal Maximum: A perturbation of carbon cycle, climate, and biosphere with implications for the future. *Annual Review of Earth and Planetary Sciences,* 39**,** 489-516.

MCKEAN, P. G., TRENHOLME, K. R., RANGARAJAN, D., KEEN, J. K. & SMITH, D. F. 1997. Diversity in repeat-containing surface proteins of Leishmania major. *Molecular and Biochemical Parasitology,* 86**,** 225-235.

MEDLEY, G. F., HOLLINGSWORTH, T. D., OLLIARO, P. L. & ADAMS, E. R. 2015. Health-seeking behaviour, diagnostics and transmission dynamics in the control of visceral leishmaniasis in the Indian subcontinent. *Nature,* 528**,** S102-S108.

MILLER, J. R., KOREN, S. & SUTTON, G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics,* 95**,** 315-327.

MIRÓ, G., MÜLLER, A., MONTOYA, A., CHECA, R., MARINO, V., MARINO, E., FUSTER, F., ESCACENA, C., DESCALZO, M. A. & GÁLVEZ, R. 2017. Epidemiological role of dogs since the human leishmaniosis outbreak in Madrid. *Parasites & vectors,* 10**,** 1-7.

MOAFI, M., REZVAN, H., SHERKAT, R. & TALEBAN, R. 2019. Leishmania vaccines entered in clinical trials: a review of literature. *International journal of preventive medicine,* 10.

MOCK, D. J., HOLLENBAUGH, J. A., DADDACHA, W., OVERSTREET, M. G., LAZARSKI, C. A., FOWELL, D. J. & KIM, B. 2012. Leishmania induces survival, proliferation and elevated cellular dNTP levels in human monocytes promoting acceleration of HIV co-infection. *PLoS pathogens,* 8**,** e1002635.

MOLDER, F., JABLONSKI, K. P., LETCHER, B., HALL, M. B., TOMKINS-TINCH, C. H., SOCHAT, V., FORSTER, J., LEE, S., TWARDZIOK, S. O., KANITZ, A., WILM, A., HOLTGREWE, M., RAHMANN, S., NAHNSEN, S. & KOSTER, J. 2021. Sustainable data analysis with Snakemake. *F1000Res,* 10**,** 33.

MOLYNEUX, D. 1974. Virus-like particles in Leishmania parasites. *Nature,* 249**,** 588-589.

MOLYNEUX, D. H., SAVIOLI, L. & ENGELS, D. 2017. Neglected tropical diseases: progress towards addressing the chronic pandemic. *The Lancet,* 389**,** 312-325.

MONDIALE DE LA SANTÉ, O. & ORGANIZATION, W. H. 2016. Leishmaniasis in high-burden countries: an epidemiological update based on data reported in 2014. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire,* 91**,** 286-296.

MOREIRA, D., LOPEZ-GARCIA, P. & VICKERMAN, K. 2004. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *International Journal of Systematic and Evolutionary Microbiology,* 54**,** 1861-1875.

MULLER, N., WELLE, M., LOBSIGER, L., STOFFEL, M. H., BOGHENBOR, K. K., HILBE, M., GOTTSTEIN, B., FREY, C. F., GEYER, C. & VON BOMHARD, W. 2009. Occurrence of Leishmania sp. in cutaneous lesions of horses in Central Europe. *Vet Parasitol,* 166**,** 346-51.

MUNIARAJ, M. 2014. The lost hope of elimination of Kala-azar (visceral leishmaniasis) by 2010 and cyclic occurrence of its outbreak in India, blame falls on vector control practices or co-infection with human immunodeficiency virus or therapeutic modalities? *Tropical parasitology,* 4**,** 10.

MURRAY, H. W., BERMAN, J. D., DAVIES, C. R. & SARAVIA, N. G. 2005. Advances in leishmaniasis. *Lancet,* 366**,** 1561-1577.

MYLER, P., SISK, E., MCDONAGH, P., MARTINEZ-CALVILLO, S., SCHNAUFER, A., SUNKIN, S., YAN, S., MADHUBALA, R., IVENS, A. & STUART, K. 2000. Genomic organization and gene function in Leishmania. *Biochemical Society Transactions,* 28**,** 527-531.

MYLER, P. J., AUDLEMAN, L., DEVOS, T., HIXSON, G., KISER, P., LEMLEY, C., MAGNESS, C., RICKEL, E., SISK, E. & SUNKIN, S. 1999. Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proceedings of the National Academy of Sciences,* 96**,** 2902-2906.

MYLER, P. J., BEVERLEY, S. M., CRUZ, A. K., DOBSON, D. E., IVENS, A. C., MCDONAGH, P. D., MADHUBALA, R., MARTINEZ-CALVILLO, S., RUIZ, J. C. & SAXENA, A. 2001. The Leishmania genome project: new insights into gene organization and function. *Medical microbiology and immunology,* 190**,** 9-12.

NBIS, S. 2021. *Genome Assembly Annotation Service (GAAS)* [Online]. Available: https://github.com/NBISweden/GAAS [Accessed].

NEGERA, E., GADISA, E., YAMUAH, L., ENGERS, H., HUSSEIN, J., KURU, T., HAILU, A., GEDAMU, L. & ASEFFA, A. 2008. Outbreak of cutaneous leishmaniasis in Silti woreda, Ethiopia: risk factor assessment and causative agent identification. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 102**,** 883-890.

NETO, A. L. C., BRITO, A. N., REZENDE, A. M., MAGALHÃES, F. B. & DE MELO NETO, O. P. 2019. In silico characterization of multiple genes encoding the GP63 virulence protein from Leishmania braziliensis: identification of sources of variation and putative roles in immune evasion. *BMC genomics,* 20**,** 1-17.

NIXON, K. C. & CARPENTER, J. M. 1993. On outgroups. *Cladistics,* 9**,** 413-426.

NODEN, B. H. & VAN DER COLF, B. E. 2013. Neglected tropical diseases of Namibia: Unsolved mysteries. *Acta Tropica,* 125**,** 1-17.

NOYES, H. 1998a. Implications of a Neotropical origin of the genus Leishmania. *Memórias do Instituto Oswaldo Cruz,* 93**,** 657-662.

NOYES, H. 1998b. Implications of a Neotropical origin of the genus Leishmania. *Mem Inst Oswaldo Cruz,* 93**,** 657-61.

NOYES, H., PRATLONG, F., CHANCE, M., ELLIS, J., LANOTTE, G. & DEDET, J. P. 2002. A previously unclassified trypanosomatid responsible for human cutaneous lesions in Martinique (French West Indies) is the most divergent member of the genus Leishmania ss. *Parasitology,* 124**,** 17-24.

NOYES, H. A., ARANA, B. A., CHANCE, M. L. & MAINGON, R. 1997. The Leishmania hertigi (Kinetoplastida; Trypanosomatidae) complex and the lizard Leishmania: their classification and evidence for a neotropical origin of the Leishmania-Endotrypanum clade. *J Eukaryot Microbiol,* 44**,** 511-7.

NOYES, H. A., MORRISON, D. A., CHANCE, M. L. & ELLIS, J. T. 2000. Evidence for a neotropical origin of Leishmania. *Mem Inst Oswaldo Cruz,* 95**,** 575-8.

OBWALLER, A. G., KARAKUS, M., POEPPL, W., TÖZ, S., ÖZBEL, Y., ASPÖCK, H. & WALOCHNIK, J. 2016. Could Phlebotomus mascittii play a role as a natural vector for Leishmania infantum? New data. *Parasites & Vectors,* 9**,** 1-6.

OGG, J. 2020. Geomagnetic polarity time scale. *Geologic Time Scale 2020.* Elsevier.

ORLANDO, T. C., RUBIO, M. A. T., STURM, N. R., CAMPBELL, D. A. & FLOETER-WINTER, L. M. 2002. Intergenic and external transcribed spacers of ribosomal RNA genes in lizard-infecting Leishmania: molecular structure and phylogenetic relationship to mammal-infecting Leishmania in the subgenus Leishmania (Leishmania). *Memórias do Instituto Oswaldo Cruz,* 97**,** 695-701.

ORÓSTICA, K. & VERDUGO, R. 2016. chromPlot: global visualization tool of genomic data. *Bioinformatics,* 32**,** 2366-2368.

PACHECO-FERNANDEZ, T., VOLPEDO, G., GANNAVARAM, S., BHATTACHARYA, P., DEY, R., SATOSKAR, A., MATLASHEWSKI, G. & NAKHASI, H. L. 2021. Revival of Leishmanization and Leishmanin. *Frontiers in Cellular and Infection Microbiology,* 11**,** 127.

PALMER, J. & NEXTGENUSFS, S. J. 2019. Funannotate: Funannotate v1. 6.0. Zenodo.

PANNUNZIO, N. R. & LIEBER, M. R. 2016. Dissecting the Roles of Divergent and Convergent Transcription in Chromosome Instability. *Cell Rep,* 14**,** 1025-1031.

PARANAIBA, L. F., PINHEIRO, L. J., MACEDO, D. H., MENEZES-NETO, A., TORRECILHAS, A. C., TAFURI, W. L. & SOARES, R. P. 2018. An overview on Leishmania (Mundinia) enriettii: biology, immunopathology, LRV and extracellular vesicles during the host-parasite interaction. *Parasitology,* 145**,** 1265-1273.

PAVLI, A. & MALTEZOU, H. C. 2010. Leishmaniasis, an emerging infection in travelers. *International Journal of Infectious Diseases,* 14**,** e1032-e1039.

PAWAR, H., SAHASRABUDDHE, N. A., RENUSE, S., KEERTHIKUMAR, S., SHARMA, J., KUMAR, G. S. S., VENUGOPAL, A., SEKHAR, N. R., KELKAR, D. S. & NEMADE, H. 2012. A proteogenomic approach to map the proteome of an unsequenced pathogen–Leishmania donovani. *Proteomics,* 12**,** 832-844.

PEACOCK, C. S., SEEGER, K., HARRIS, D., MURPHY, L., RUIZ, J. C., QUAIL, M. A., PETERS, N., ADLEM, E., TIVEY, A., ASLETT, M., KERHORNOU, A., IVENS, A., FRASER, A., RAJANDREAM, M. A., 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, S. L., FAULCONBRIDGE, A., JEFFARES, D., DEPLEDGE, D. P., OYOLA, S. O., HILLEY, J. D., BRITO, L. O., TOSI, L. R. O., BARRELL, B., CRUZ, A. K., MOTTRAM, J. C., SMITH, D. F. & BERRIMAN, M. 2007. Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nature Genetics,* 39**,** 839-847.

PEARSON, R. D. & DE QUEIROZ SOUSA, A. 1996. Clinical spectrum of leishmaniasis. *Clinical infectious diseases***,** 1-11.

PENG, Y., LEUNG, H. C. M., YIU, S. M. & CHIN, F. Y. L. 2010. IDBA - A Practical Iterative de Bruijn Graph De Novo Assembler. *Research in Computational Molecular Biology, Proceedings,* 6044**,** 426-440.

PETERS, W. 1977. Chemotherapy of Leishmaniasis. LIVERPOOL SCHOOL OF TROPICAL MEDICINE (ENGLAND) DEPT OF PARASITOLOGY.

PIARROUX, R., FONTES, M., PERASSO, R., GAMBARELLI, F., JOBLET, C., DUMON, H. & QUILICI, M. 1995. Phylogenetic relationships between Old World Leishmania strains revealed by analysis of a repetitive DNA sequence. *Molecular and biochemical parasitology,* 73**,** 249-252.

POINAR, G. 2004. Palaeomyia burmitis (Diptera : Phlebotomidae), a new genus and species of cretaceous sand flies with evidence of blood-sucking habits. *Proceedings of the Entomological Society of Washington,* 106**,** 598-605.

POINAR, G., JR. & POINAR, R. 2004a. Evidence of vector-borne disease of Early Cretaceous reptiles. *Vector Borne Zoonotic Dis,* 4**,** 281-4.

POINAR, G., JR. & POINAR, R. 2004b. Paleoleishmania proterus n. gen., n. sp., (Trypanosomatidae: Kinetoplastida) from Cretaceous Burmese amber. *Protist,* 155**,** 305-10.

POTHIRAT, T., TANTIWORAWIT, A., CHAIWARITH, R., JARIYAPAN, N., WANNASAN, A., SIRIYASATIEN, P., SUPPARATPINYO, K., BATES, M. D., KWAKYE-NUAKO, G. & BATES, P. A. 2014a. First Isolation of Leishmania from Northern Thailand: Case Report, Identification as Leishmania martiniquensis and Phylogenetic Position within the Leishmania enriettii Complex. *Plos Neglected Tropical Diseases,* 8.

POTHIRAT, T., TANTIWORAWIT, A., CHAIWARITH, R., JARIYAPAN, N., WANNASAN, A., SIRIYASATIEN, P., SUPPARATPINYO, K., BATES, M. D., KWAKYE-NUAKO, G. & BATES, P. A. 2014b. First isolation of Leishmania from Northern Thailand: case report, identification as Leishmania martiniquensis and phylogenetic position within the Leishmania enriettii complex. *PLoS neglected tropical diseases,* 8**,** e3339.

POTTER, S., LUCIANI, A., EDDY, S., PARK, Y., LOPEZ, R. & FINN, R. 2018. Nucleic Acids Research Web Server Issue, 46. W200–W204.

PRASAD, A. S. & RAO, S. 2014. A Mechanism Design Approach to Resource Procurement in Cloud Computing. *IEEE Transactions on Computers,* 63**,** 17-30.

PUECHBERTY, J., BLAINEAU, C., MEGHAMLA, S., CROBU, L., PAGÈS, M. & BASTIEN, P. 2007. Compared genomics of the strand switch region of Leishmania chromosome 1 reveal a novel genus-specific gene and conserved structural features and sequence motifs. *BMC genomics,* 8**,** 1-10.

QUARESMA, P. F., RÊGO, F. D., BOTELHO, H. A., DA SILVA, S. R., MOURA, A. J., NETO, R. G. T., MADEIRA, F. M., CARVALHO, M. B., PAGLIA, A. P. & MELO, M. N. 2011. Wild, synanthropic and domestic hosts of Leishmania in an endemic area of cutaneous leishmaniasis in Minas Gerais State, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 105**,** 579-585.

QUINLAN, A. R. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics,* 47**,** 11 12 1-34.

QUINNELL, R. J. & COURTENAY, O. 2009. Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology,* 136**,** 1915-1934.

RAMBAUT, A. 2009. FigTree v1. 3.1: Tree figure drawing tool. Edinburgh, UK.

RAVEL, C., MACARI, F., BASTIEN, P., PAGÈS, M. & BLAINEAU, C. 1995. Convervation among Old World Leishmania species of six physical linkage groups defined in Leishmania infantum small chromosomes. *Molecular and biochemical parasitology,* 69**,** 1-8.

RAYMOND, F., BOISVERT, S., ROY, G., RITT, J. F., LEGARE, D., ISNARD, A., STANKE, M., OLIVIER, M., TREMBLAY, M. J., PAPADOPOULOU, B., OUELLETTE, M. & CORBEIL, J. 2012. Genome sequencing of the lizard parasite Leishmania tarentolae reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Research,* 40**,** 1131-1147.

RAZAVINASAB, S. Z., SHARIFI, I., AFLATOONIAN, M. R., BABAEI, Z., MOHAMMADI, M. A., SALARKIA, E., SHARIFI, F., AGHAEI AFSHAR, A. & BAMOROVAT, M. 2019. Expansion of urban cutaneous leishmaniasis into rural areas of southeastern Iran: Clinical, epidemiological and phylogenetic profiles explored using 7SL high resolution melting-PCR analysis. *Transboundary and emerging diseases,* 66**,** 1602-1610.

REITHINGER, R., DUJARDIN, J.-C., LOUZIR, H., PIRMEZ, C., ALEXANDER, B. & BROOKER, S. 2007. Cutaneous leishmaniasis. *The Lancet Infectious Diseases,* 7**,** 581-596.

REITHINGER, R., ESPINOZA, J. C., LLANOS-CUENTAS, A. & DAVIES, C. R. 2003. Domestic dog ownership: a risk factor for human infection with Leishmania (Viannia) species. *Transactions of the royal society of tropical medicine and hygiene,* 97**,** 141-145.

REUSS, S. M., DUNBAR, M. D., CALDERWOOD MAYS, M. B., OWEN, J. L., MALLICOTE, M. F., ARCHER, L. L. & WELLEHAN, J. F., JR. 2012. Autochthonous Leishmania siamensis in horse, Florida, USA. *Emerg Infect Dis,* 18**,** 1545-7.

RICHARDSON, D. N., SIMMONS, M. P. & REDDY, A. S. 2006. Comprehensive comparative analysis of kinesins in photosynthetic eukaryotes. *BMC genomics,* 7**,** 1-37.

RITMEIJER, K., DAVIES, C., VAN ZORGE, R., WANG, S. J., SCHORSCHER, J., DONGU'DU, S. I. & DAVIDSON, R. N. 2007. Evaluation of a mass distribution programme for fine-mesh impregnated bednets against visceral leishmaniasis in eastern Sudan. *Tropical medicine & international health,* 12**,** 404-414.

RITMEIJER, K., VEEKEN, H., MELAKU, Y., LEAL, G., AMSALU, R., SEAMAN, J. & DAVIDSON, R. 2001. Ethiopian visceral leishmaniasis: generic and proprietary sodium stibogluconate are equivalent; HIV co-infected patients have a poor outcome. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 95**,** 668-672.

ROGERS, M. B., HILLEY, J. D., DICKENS, N. J., WILKES, J., BATES, P. A., DEPLEDGE, D. P., HARRIS, D., HER, Y., HERZYK, P., IMAMURA, H., OTTO, T. D., SANDERS, M., SEEGER, K., DUJARDIN, J. C., BERRIMAN, M., SMITH, D. F., HERTZ-FOWLER, C. & MOTTRAM, J. C. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Research,* 21**,** 2129-2142.

ROOK, D. L. & HUNTER, J. P. 2014. Rooting around the eutherian family tree: the origin and relations of the Taeniodonta. *Journal of Mammalian Evolution,* 21**,** 75-91.

ROSE, K., CURTIS, J., BALDWIN, T., MATHIS, A., KUMAR, B., SAKTHIANANDESWAREN, A., SPURCK, T., LOW CHOY, J. & HANDMAN, E. 2004. Cutaneous leishmaniasis in red kangaroos: isolation and characterisation of the causative organisms. *Int J Parasitol,* 34**,** 655-64.

RUB, A., DEY, R., JADHAV, M., KAMAT, R., CHAKKARAMAKKIL, S., MAJUMDAR, S., MUKHOPADHYAYA, R. & SAHA, B. 2009. Cholesterol depletion associated with Leishmania major infection alters macrophage CD40 signalosome composition and effector function. *Nature immunology,* 10**,** 273-280.

RUOFF, K. 1995. Leuconostoc, Pediococcus, Stomatococcus, and miscellaneous gram-positive cocci that grow aerobically. *Manual of clinical microbiology***,** 315-323.

SALAM, N., AL-SHAQHA, W. M. & AZZI, A. 2014. Leishmaniasis in the Middle East: incidence and epidemiology. *PLoS neglected tropical diseases,* 8**,** e3208.

SAMADY, J. A., JANNIGER, C. K. & SCHWARTZ, R. A. 1996. Cutaneous and mucocutaneous leishmaniasis. *Cutis (New York, NY),* 57**,** 13-20.

SAMARAS, N. & SPITHILL, T. W. 1987. Molecular karyotype of five species of Leishmania and analysis of gene locations and chromosomal rearrangements. *Mol Biochem Parasitol,* 25**,** 279-91.

SANCHIZ, Á., MORATO, E., RASTROJO, A., CAMACHO, E., GONZALEZ-DE LA FUENTE, S., MARINA, A., AGUADO, B. & REQUENA, J. M. 2020. The experimental proteome of Leishmania infantum promastigote and its usefulness for improving gene annotations. *Genes,* 11**,** 1036.

SASIDHARAN, S. & SAUDAGAR, P. 2021. Leishmaniasis: where are we and where are we heading? *Parasitology Research,* 120**,** 1541-1554.

SAXENA, A., LAHAV, T., HOLLAND, N., AGGARWAL, G., ANUPAMA, A., HUANG, Y., VOLPIN, H., MYLER, P. & ZILBERSTEIN, D. 2007. Analysis of the Leishmania donovani transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Molecular and biochemical parasitology,* 152**,** 53-65.

SCOTT, P. & NOVAIS, F. O. 2016. Cutaneous leishmaniasis: immune responses in protection and pathogenesis. *Nature Reviews Immunology,* 16**,** 581-592.

SEBLOVA, V., SADLOVA, J., VOJTKOVA, B., VOTYPKA, J., CARPENTER, S., BATES, P. A. & VOLF, P. 2015. The Biting Midge Culicoides sonorensis (Diptera: Ceratopogonidae) Is Capable of Developing Late Stage Infections of Leishmania enriettii. *Plos Neglected Tropical Diseases,* 9.

SERENO, D. 2019. Leishmania (Mundinia) spp.: from description to emergence as new human and animal Leishmania pathogens. *New Microbes and New Infections,* 30**,** 100540.

SEVA, A. D. P., MAO, L., GALVIS-OVALLOS, F., TUCKER LIMA, J. M. & VALLE, D. 2017. Risk analysis and prediction of visceral leishmaniasis dispersion in São Paulo State, Brazil. *PLoS neglected tropical diseases,* 11**,** e0005353.

SIMPSON, J. T., WONG, K., JACKMAN, S. D., SCHEIN, J. E., JONES, S. J. M. & BIROL, I. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Research,* 19**,** 1117-1123.

SMITH, T., ROSE, K. D. & GINGERICH, P. D. 2006. Rapid Asia–Europe–North America geographic dispersal of earliest Eocene primate Teilhardina during the Paleocene–Eocene thermal maximum. *Proceedings of the National Academy of Sciences,* 103**,** 11223-11227.

SONESON, C., YAO, Y., BRATUS-NEUENSCHWANDER, A., PATRIGNANI, A., ROBINSON, M. D. & HUSSAIN, S. 2019. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature communications,* 10**,** 1-14.

STEIN, L. D. 2010. The case for cloud computing in genome informatics. *Genome Biol,* 11**,** 207.

STRAZZULLA, A., COCUZZA, S., PINZONE, M. R., POSTORINO, M. C., COSENTINO, S., SERRA, A., CACOPARDO, B. & NUNNARI, G. 2013. Mucosal Leishmaniasis: An Underestimated Presentation of a Neglected Disease. *BioMed Research International,* 2013**,** 1-7.

SUKMEE, T., SIRIPATTANAPIPONG, S., MUNGTHIN, M., WORAPONG, J., RANGSIN, R., SAMUNG, Y., KONGKAEW, W., BUMRUNGSANA, K., CHANACHAI, K., APIWATHANASORN, C., RUJIROJINDAKUL, P., WATTANASRI, S., UNGCHUSAK, K. & LEELAYOOVA, S. 2008. A suspected new species of Leishmania, the causative agent of visceral leishmaniasis in a Thai patient. *Int J Parasitol,* 38**,** 617-22.

SUNDAR, S. & CHAKRAVARTY, J. 2010. Antimony toxicity. *International journal of environmental research and public health,* 7**,** 4267-4277.

SUNDAR, S., MORE, D. K., SINGH, M. K., SINGH, V. P., SHARMA, S., MAKHARIA, A., KUMAR, P. C. & MURRAY, H. W. 2000. Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic. *Clinical infectious diseases,* 31**,** 1104-1107.

SUPSRISUNJAI, C., KOOTIRATRAKARN, T., PUANGPET, P., BUNNAG, T., CHAOWALIT, P. & WESSAGOWIT, V. 2017. Case Report: Disseminated Autochthonous Dermal Leishmaniasis Caused by Leishmania siamensis (PCM2 Trang) in a Patient from Central Thailand Infected with Human Immunodeficiency Virus. *American Journal of Tropical Medicine and Hygiene,* 96**,** 1160-1163.

SZPARA, M. L., GATHERER, D., OCHOA, A., GREENBAUM, B., DOLAN, A., BOWDEN, R. J., ENQUIST, L. W., LEGENDRE, M. & DAVISON, A. J. 2014. Evolution and diversity in human herpes simplex virus genomes. *J Virol,* 88**,** 1209-27.

TAMURA, K., NEI, M. & KUMAR, S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences,* 101**,** 11030-11035.

TARR, P. I., ALINE, R. F., SMILEY, B. L., SCHOLLER, J., KEITHLY, J. & STUART, K. 1988. LR1: a candidate RNA virus of Leishmania. *Proceedings of the National Academy of Sciences,* 85**,** 9572-9575.

THOMAIDOU, E., HOREV, L., JOTKOWITZ, D., ZAMIR, M., INGBER, A., ENK, C. D. & MOLHO-PESSACH, V. 2015. Lymphatic dissemination in cutaneous leishmaniasis following local treatment. *The American journal of tropical medicine and hygiene,* 93**,** 770.

THOMAZ-SOCCOL, V., LANOTTE, G., RIOUX, J., PRATLONG, F., MARTINI-DUMAS, A. & SERRES, E. 1993. Phylogenetic taxonomy of new world Leishmania. *Annales de parasitologie humaine et comparee,* 68**,** 104-104.

THOMAZ-SOCCOL, V., PRATLONG, F., LANGUE, R., CASTRO, E., LUZ, E. & DEDET, J. 1996. New isolation of Leishmania enriettii Muniz and Medina, 1948 in Paraná State, Brazil, 50 years after the first description, and isoenzymatic polymorphism of the L. enriettii taxon. *Annals of Tropical Medicine & Parasitology,* 90**,** 491-495.

THOMAZ-SOCCOL, V., VELEZ, I. D., PRATLONG, F., AGUDELOS, S., LANOTTE, G. & RIOUX, J. A. 2000. Enzymatic polymorphism and phylogenetic relationships in Leishmania Ross, 1903 (Sarcomastigophora: Kinetoplastida): a case study in Colombia. *Syst Parasitol,* 46**,** 59-68.

TRANSMISSÍVEIS/AIDS, P. N. D. D. S. 2004. *Manual de recomendações para diagnóstico, tratamento e acompanhamento da co-infecção Leishmania-HIV*, Editora MS.

TSOKANA, C. N., SOKOS, C., GIANNAKOPOULOS, A., MAMURIS, Z., BIRTSAS, P., PAPASPYROPOULOS, K., VALIAKOS, G., SPYROU, V., LEFKADITIS, M., CHATZOPOULOS, D. C., KANTERE, M., MANOLAKOU, K., TOULOUDI, A., BURRIEL, A. R., FERROGLIO, E., HADJICHRISTODOULOU, C. & BILLINIS, C. 2016. First evidence of Leishmania infection in European brown hare (Lepus europaeus) in Greece: GIS analysis and phylogenetic position within the Leishmania spp. *Parasitology Research,* 115**,** 313-321.

UHLEN, M., OKSVOLD, P., FAGERBERG, L., LUNDBERG, E., JONASSON, K., FORSBERG, M., ZWAHLEN, M., KAMPF, C., WESTER, K., HOBER, S., WERNERUS, H., BJORLING, L. & PONTEN, F. 2010. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol,* 28**,** 1248-50.

UNIPROT, C. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res,* 49**,** D480-D489.

UNOARUMHI, Y., BATRA, D., SHETH, M., NARAYANAN, V., LIN, W., ZHENG, Y., ROWE, L. A., POHL, J. & DE ALMEIDA, M. 2021. Chromosome-Level Genome Sequence of Leishmania (Leishmania) tropica Strain CDC216-162, Isolated from an Afghanistan Clinical Case. *Microbiol Resour Announc,* 10.

VALDIVIA, H. O., SCHOLTE, L. L. S., OLIVEIRA, G., GABALDON, T. & BARTHOLOMEU, D. C. 2015. The Leishmania metaphylome: a comprehensive survey of Leishmania protein phylogenetic relationships. *Bmc Genomics,* 16.

VALERO, N. N. H. & URIARTE, M. 2020. Environmental and socioeconomic risk factors associated with visceral and cutaneous leishmaniasis: a systematic review. *Parasitology research,* 119**,** 365-384.

VAN DIJK, E. L., JASZCZYSZYN, Y., NAQUIN, D. & THERMES, C. 2018. The Third Revolution in Sequencing Technology. *Trends in Genetics,* 34**,** 666-681.

VAN GRIENSVEN, J., ZIJLSTRA, E. E. & HAILU, A. 2014. Visceral leishmaniasis and HIV coinfection: time for concerted action. *PLoS neglected tropical diseases,* 8**,** e3023.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A. & HOLT, R. A. 2001. The sequence of the human genome. *science,* 291**,** 1304-1351.

VICKERMAN, K. & PRESTON, T. 1976. Comparative cell biology of the kinetoplastid flagellates. *Biology of Kinetoplastida,* 1**,** 66-67.

WAKI, K., DUTTA, S., RAY, D., KOLLI, B. K., AKMAN, L., KAWAZU, S., LIN, C. P. & CHANG, K. P. 2007. Transmembrane molecules for phylogenetic analyses of pathogenic protists: Leishmania-specific informative sites in hydrophilic loops of trans- endoplasmic reticulum N-acetylglucosamine-1-phosphate transferase. *Eukaryot Cell,* 6**,** 198-210.

WALKER, B. J., ABEEL, T., SHEA, T., PRIEST, M., ABOUELLIEL, A., SAKTHIKUMAR, S., CUOMO, C. A., ZENG, Q. D., WORTMAN, J., YOUNG, S. K. & EARL, A. M. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *Plos One,* 9.

WALL, E. C., WATSON, J., ARMSTRONG, M., CHIODINI, P. L. & LOCKWOOD, D. N. 2012. Epidemiology of imported cutaneous leishmaniasis at the Hospital for Tropical Diseases, London, United Kingdom: use of polymerase chain reaction to identify the species. *The American journal of tropical medicine and hygiene,* 86**,** 115.

WEINA, P. J., NEAFIE, R. C., WORTMANN, G., POLHEMUS, M., ARONSON, N. E. & STRAUSBAUGH, L. J. 2004. Old world leishmaniasis: an emerging infection among deployed US military and civilian workers. *Clinical infectious diseases,* 39**,** 1674-1680.

WELLER, M. G. 2021. The Protocol Gap. *Methods and Protocols,* 4**,** 12.

WHO 2005. Regional Strategic Framework for elimination of kala-azar from the South-East Asia Region.(2011-2015).

WHO 2010. *Control of the leishmaniases: report of a meeting of the WHO Expert Commitee on the Control of Leishmaniases, Geneva, 22-26 March 2010*, World Health Organization.

WHO 2012. London Declaration on neglected tropical diseases. World Health Organization Geneva.

WHO 2018. Recognizing neglected tropical diseases through changes on the skin: a training guide for front-line health workers.

WICK, R. R., JUDD, L. M., GORRIE, C. L. & HOLT, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Plos Computational Biology,* 13.

WILSON, J. T. 1963. Continental drift. *Scientific American,* 208**,** 86-103.

WINCKER, P., RAVEL, C., BLAINEAU, C., PAGES, M., JAUFFRET, Y., DEDET, J. P. & BASTIEN, P. 1996a. The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res,* 24**,** 1688-94.

WINCKER, P., RAVEL, C., BLAINEAU, C., PAGES, M., JAUFFRET, Y., DEDET, J. P. & BASTIEN, P. 1996b. The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Research,* 24**,** 1688-1694.

YADON, Z. E., RODRIGUES, L. C., DAVIES, C. R. & QUIGLEY, M. A. 2003. Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: a retrospective case-control study. *The American journal of tropical medicine and hygiene,* 68**,** 519-526.

YAMADA, K. D., TOMII, K. & KATOH, K. 2016. Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics,* 32**,** 3246-3251.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci,* 13**,** 555-6.

YANG, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol,* 24**,** 1586-91.

YANIK, M., GUREL, M., SIMSEK, Z. & KATI, M. 2004. The psychological impact of cutaneous leishmaniasis. *Clinical and Experimental Dermatology: Clinical Dermatology,* 29**,** 464-467.

YEHIA, H. M., AL-OLAYAN, E. M., EL-KHADRAGY, M. F. & METWALLY, D. M. 2017. In vitro and in vivo control of secondary bacterial infection caused by Leishmania major. *International Journal of Environmental Research and Public Health,* 14**,** 777.

YIMER, M., ABERA, B., MULU, W., ZENEBE, Y. & BEZABIH, B. 2014. Proportion of Visceral leishmaniasis and human immune deficiency virus co-infection among clinically confirmed visceral leishmaniasis patients at the endemic foci of the Amhara National Regional State, north-west Ethiopia. *Am J Biomed Life Sci,* 2**,** 1-7.

ZACARIAS, D. A., ROLÃO, N., DE PINHO, F. A., SENE, I., SILVA, J. C., PEREIRA, T. C., COSTA, D. L. & COSTA, C. H. 2017. Causes and consequences of higher Leishmania infantum burden in patients with kala-azar: a study of 625 patients. *Tropical Medicine & International Health,* 22**,** 679-687.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research,* 18**,** 821-829.

ZHANG, J. R., GUO, X. G., LIU, J. L., ZHOU, T. H., GONG, X., CHEN, D. L. & CHEN, J. P. 2016. Molecular detection, identification and phylogenetic inference of Leishmania spp. in some desert lizards from Northwest China by using internal transcribed spacer 1 (ITS1) sequences. *Acta Tropica,* 162**,** 83-94.

ZHOU, S., KILE, A., KVIKSTAD, E., BECHNER, M., SEVERIN, J., FORREST, D., RUNNHEIM, R., CHURAS, C., ANANTHARAMAN, T. S., MYLER, P., VOGT, C., IVENS, A., STUART, K. & SCHWARTZ, D. C. 2004. Shotgun optical mapping of the entire Leishmania major Friedlin genome. *Mol Biochem Parasitol,* 138**,** 97-106.

ZIAIE, H. & SADEGHIAN, G. 2008. Isolation of bacteria causing secondary bacterial infection in the lesions of cutaneous leishmaniasis. *Indian journal of dermatology,* 53**,** 129.