High-quality genome assemblies and comparative genomics of *Leishmania* (*Viannia*) species

Cián Lynch

MSc by Research

University of York

Biology

August 2025

Abstract

Leishmaniasis affects over 350 million people worldwide, caused by *Leishmania* parasites including the *Viannia* subgenus comprising nine Americas-endemic species causing cutaneous and mucocutaneous disease. High-quality genomic resources remain limited for most *Viannia* species, hindering evolutionary and epidemiological studies.

This study generated chromosome-level genome assemblies for eight *Viannia* strains and five additional *Leishmania* species using Oxford Nanopore Technologies long-read sequencing with Illumina polishing. The assembly pipeline employed NECAT, followed by multi-stage polishing using Medaka, Racon, and Pilon. Comparative genomics across 63 trypanosomatid species used OrthoFinder for ortholog identification and phylogenetic reconstruction. Genomic ancestry was analyzed using ADMIXTURE with 21,530 genome-wide variants, focusing on putative hybrid strain MHOM/VE/1991/PM-H197. Strains were sourced from a biological collection; eight underwent a promastigote–amastigote–promastigote differentiation process with five passages prior to DNA extraction, while the remainder were analysed from pre-extracted DNA of unrecorded passage history.

Generated assemblies substantially improved upon existing references, achieving N50 values of 563-1,144 kb and >90% BUSCO completeness for euglenozoan gene sets. Comparative analysis identified 16,683 orthogroups, revealing minimal core genome conservation (5.06%) and extensive accessory genome diversity (53.76%). Phylogenetic analysis confirmed four major *Leishmania* lineages supporting subgeneric classifications.

Most significantly, strain PM-H197 exhibited 80% *L. mexicana* complex ancestry and 20% *Viannia* ancestry, representing the first genomic characterization of intersubgeneric *Leishmania* hybridization. However, Oxford Nanopore coverage analysis revealed contrasting 2:1 *Viannia* genomic dominance, creating an 8-fold discrepancy with ancestry predictions. Ultra-long read competitive mapping demonstrated 96% *Viannia* preference, supporting unbalanced hybrid or mixed infection scenarios.

This work provides the most comprehensive *Viannia* genomic dataset to date, establishing new parasitic genome assembly benchmarks. The discovery of intersubgeneric genetic exchange challenges current understanding of *Leishmania* reproductive barriers with important implications for population genetics, epidemiology, and diagnostics. These resources provide essential foundations for future leishmaniasis research and therapeutic development.

Author’s Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Acknowledgements

I would like to thank my supervisor, Professor Daniel Jeffares, for his outstanding support and guidance throughout my Master's degree. I also want to thank my secondary and tertiary supervisors, Professor Jeremy Mottram and Dr. Joana Faria, for their unwavering support and advice. My appreciation extends to Professor Anthony Wilkinson for his support and guidance.

Additionally, I wish to acknowledge the constant support and advice of Dr. João Luís Reis-Cunha, Dr. Sally James, and everyone at the Genomics Lab at York, as well as Dr. Cobus Smit and all who worked in the Jeffares group during my time there.

I am grateful for the support of my parents, brother, and partner throughout my degree. Special thanks to Jessica J. Petts for her constant support, wisdom, and kindness.

Table of Contents

1. Introduction

1.1. Leishmaniasis

Leishmaniasis is a complex parasitic disease caused by protozoa of the genus *Leishmania,* part of the Trypanosomatidae family.These parasites are primarily transmitted to humans and other mammals through the bite of infected female sandflies, predominantly species of the Phlebotomus and Lutzomyia genera. The disease has a broad global impact, with an estimated 350 million people at risk in 82 to 98 countries, particularly in tropical and subtropical regions. Over 1 billion people live in areas endemic for leishmaniasis, putting them at risk of infection (World Health Organization, 2025). Each year, there are approximately 700,000 – 1.2 million new CL cases and 200,000 – 400,000 new VL cases globally (Alvar et al., 2012; Alatif, 2021).

1.2. Clinical Manifestations

The two most common clinical manifestations of disease are cutaneous leishmaniasis (CL) and visceral leishmaniasis (VL). CL usually presents as localized skin lesions, which may start as small papules or nodules and can progress to ulcerating sores. These ulcers often have raised edges and a central crater, and may heal spontaneously or become chronic, leading to scarring and social stigma (Reithinger et al., 2007; Saidi et al., 2023). Lesions can also appear in unusual forms, such as eczematous, verrucous, annular, or zosteriform patterns, and may be mistaken for other skin conditions like warts, eczema, or even skin cancers. Lesions can occur on the face, limbs, trunk, and less commonly on the palms, soles, or genital areas (Meireles et al., 2017; Gunasekara et al., 2023). The clinical spectrum ranges from single, self-healing ulcers (localized cutaneous leishmaniasis) to more severe forms, such as diffuse cutaneous leishmaniasis which manifests as multiple non-ulcerating nodules across the body (Yadav and Madke, 2023). There are also Disseminated forms where multiple lesions form across different body areas (Saidi et

al., 2023). CL lesions can progress to mucocutaneous leishmaniasis (ML), leading to degradation of the mucous membranes in the mouth, throat, and nasal cavities (McDowell and Robichaud, 2024). The type and severity of skin lesions depend on the *Leishmania* species and the host's immune response. The balance between inflammatory and anti-inflammatory factors determines whether the disease remains localized or becomes more severe and widespread (Scorza, Carvalho and Wilson, 2017; Volpedo et al., 2021). Early lesions are often mistaken for insect bites, pimples, or other benign skin issues, leading to delays in seeking medical care which can worsen outcomes (Gunasekara et al., 2023).

Visceral leishmaniasis primarily affects internal organs such as the spleen, liver, and bone marrow, and can be fatal if left untreated. The most common signs include prolonged or irregular fever, enlargement of the spleen (splenomegaly) and liver (hepatomegaly), weight loss, pallor, and abdominal swelling. In individuals with weakened immune systems (e.g., HIV infection, organ transplant recipients), VL may present atypically. Classic features like splenomegaly or hyperglobulinemia may be absent, and the disease can involve unusual organs. These patients are at higher risk for severe disease, relapses, and poor response to treatment (Saporito et al., 2013; Diro et al., 2015; Lauletta Lindoso et al., 2016). In individuals that have recovered from VL, they may develop post-kala-azar dermal leishmaniasis (PKDL), which typically presents as a rash with various types of skin lesions in individuals who are otherwise healthy after VL. Lesions often start around the mouth and can spread to the face, upper body, extremities, and sometimes the genitalia and tongue (Zijlstra et al., 2003). Lesions are usually progressive and rarely heal on their own, and PKDL patients can harbor parasites in their skin, serving as a reservoir for ongoing transmission of VL, which is a significant public health concern in endemic areas (Gedda et al., 2020).

1.3. Parasite species

There are currently approximately 53 species described in the *Leishmania* genus, with 20 of these species known to cause disease in humans (Barratt et al., 2017). The species of *Leishmania* is strongly correlated with the type of disease it causes, but the relationship is not absolute. Different *Leishmania* species are typically associated with specific clinical forms (CL, ML, and VL), though host factors and parasite strain differences can also influence disease manifestation. VL is most commonly caused by *Leishmania donovani* complex (including *L. donovani* and *L. infantum*), which are highly associated with visceral disease affecting internal organs (Wilson, Jeronimo and Pearson, 2005; Torres-Guerrero et al., 2017).

CL is typically caused by species such as *L. major*, *L. tropica*, and various *Viannia* subgenus species such as *L. braziliensis, L. guyanensis,* and *L. panamensis* (Torres-Guerrero et al., 2017). There are exceptions to what is classically seen in species-specific interactions. Some species, like *L. tropica* and *L. amazonensis*, can cause both cutaneous and, less commonly, visceral forms, depending on host and parasite factors (Torres-Guerrero et al., 2017; Silva et al., 2021). Similarly, *L. infantum* which is known to cause VL, can cause a non-ulcerated form of cutaneous disease (NUCL) in Honduras (Batista et al., 2025).

1.4. Parasite spatial distribution

Animal reservoirs of *Leishmania* parasites are significant for epidemiology, with dogs being a key reservoir for *L. infantum*, as well as *L. braziliensis*, however for *L. braziliensis*, humans may be more important reservoirs in certain settings (Dantas-Torres, 2007; Brilhante et al., 2019; Ascencio et al., 2020).

Livestock and other wild mammals have been found to harbor a range of species such as *L. donovani* in the Indian subcontinent and rats infected with *L. infantum* and *L. major* in Turkey (Karakuş et al., 2020; Ghouse Peer et al., 2024). The ecoepidemiology of this parasite is complex. Transmission dynamics are influenced by the interactions between different sand fly species, their feeding behaviors, and the availability of multiple vertebrate hosts. Sandfly abundance and activity are significantly correlated with temperature, humidity, and other climatic variables. Some species, like *Lutzomyia longipalpis*, show adaptability and can be found throughout the year, but their numbers fluctuate with temperature changes (Pareyn et al., 2019; de Souza Fernandes et al., 2022).

In Europe, sandfly species are currently concentrated in the Mediterranean, but climate models predict a northward expansion as temperatures rise, increasing the potential range for leishmaniasis transmission (Koch et al., 2017). In China and Spain, sandfly distributions are linked to ecological niches defined by temperature and humidity, with some species favoring cooler or warmer zones, which affects where leishmaniasis cases occur (Aransay et al., 2004; Hong et al., 2024). In South America, *Lutzomyia longipalpis* is the primary vector for visceral leishmaniasis (*L. infantum*) and is widespread in urban and peri-urban areas, with high infection rates detected in several studies (Thomaz-Soccol et al., 2023; Leonel et al., 2024; Michalsky et al., 2024).

Lutzomyia whitmani and *Lutzomyia migonei* are major vectors for CL, especially *L. braziliensis* and *L. guyanensis* and are more commonly found in rural environments as part of a sylvatic cycle (Teles et al., 2016; Barrios et al., 2020). However, this pattern of ecoepidemiology is different in African leishmaniasis cases with the key vector for all forms of disease being *P. papatasi* (Cecílio, Cordeiro-da-Silva and Oliveira, 2022; Labbé et al., 2023).

VL in Africa, particularly in East Africa, is most commonly reported in rural regions, where environmental and socioeconomic conditions favor the sandfly vector and transmission cycles (Ready, 2014; Geto et al., 2024). CL is found throughout all types of environments in Africa with many cases and outbreaks reported across varied anthropogenic areas. Hence an underlying understanding of the species of

sandfly in an area, and the species of *Leishmania* that they harbor, is crucial for understanding pathology and disease dynamics.

1.5. Treatment and prevention

There is no registered or widely approved vaccine for human leishmaniasis. Control relies on drug treatment and vector control measures. Multiple vaccine candidates (e.g., Leish-F1, F2, F3, ChAd63-KH) have reached clinical trials, showing safety and immune response induction, but none are yet approved for general human use (Gillespie et al., 2016; Osman et al., 2017; Iborra et al., 2018). New strategies include peptide vaccines, chimeric proteins, and nanovaccines, aiming for better specificity, stability, and immune activation (De Brito et al., 2018; Iborra et al., 2018).

The only oral drug approved for both VL and CL is miltefosine. It shows high cure rates for VL in South Asia and good efficacy for Old World CL, though resistance and variable results in other regions are concerns. It is also being tested in combination therapies and for maintenance prophylaxis (Dorlo et al., 2012; Monge-Maillo and López-Vélez, 2015). Liposomal amphotericin B is highly effective for VL, especially in the Indian subcontinent, Mediterranean, and South America. Single-dose regimens are preferred where feasible (Sundar and Chakravarty, 2015; Chakravarty and Sundar, 2019). Pentavalent antimonials are historically the mainstay for both VL and CL, but resistance is widespread in some regions (notably India). However, they are still used in Africa and for some CL cases. Paromomycin can be used in combination with antimonials for VL in Africa and as a topical or parenteral option for CL (Croft and Coombs, 2003; Sundar and Chakravarty, 2015).

Emerging therapies include various combinations of treatments already listed, and a combination of miltefosine with nifuratel; a repurposed drug showing strong activity against both VL and CL in animal models, especially when combined with miltefosine (Domínguez-Asenjo et al., 2021; Melcon-Fernandez et al., 2023). Experimental targeted peptides have shown efficacy in mouse models for both CL and VL and the drug fluconazole has demonstrated effectiveness for CL in some studies (Sundar and Rai, 2002; Habib et al., 2024).

1.6. Leishmaniasis in South America

Cutaneous leishmaniasis in South America is predominantly caused by the *Viannia* clade (Table 1), with the exception of a handful of other species (e.g. *L. mexicana, E. colombiensis* etc.). Visceral leishmaniasis is caused by *L. infantum* in many urban areas of Brazil, Bolivia, Venezuela and Honduras (Pasquali et al., 2019; Schwabl et al., 2021; Silveira et al., 2021; Delgado-Noguera et al., 2023). In Honduras it causes an unusual form of CL where there is an absence of obvious ulceration (Batista et al., 2025). *L. infantum* can occasionally cause CL, such as in Argentina and parts of Brazil (Acosta-Soto et al., 2020; Soares et al., 2024), but it is rare. Between 2001 and 2023, 17 countries in the region reported 1,105,545 cases of cutaneous and mucosal leishmaniasis to PAHO, averaging 52,645 cases annually. Since 2005, case numbers have steadily declined, reaching a record low of 34,954 in 2023. Compared to 2022, this represents an 8% overall decrease. However, in 2023 case numbers rose in Argentina (15%), Costa Rica (36%), Ecuador (18%), Guyana (100%), Mexico (11%), and Suriname (40%). Between 2001 and 2023, the Region of the Americas reported 73,092 new cases of visceral leishmaniasis (VL), averaging 3,178 cases annually. Since 2017, cases have declined significantly, with a 164% drop leading to a historic low of 1,604 cases in 2023. Notable decreases were seen in Brazil, Colombia, and Venezuela, where transmission has been ongoing. Conversely, cases rose in 2023 in Argentina and Paraguay by 65% and 68%, respectively (Pan American Health Organization, 2024). The overall decline in cases across the region since 2005 is generally attributed to strengthened vector control programmes, earlier diagnosis and treatment, and improved epidemiological surveillance coordinated through PAHO (Pan American Health Organization, 2024). In contrast, the rises in countries such as Argentina and Paraguay reflect ongoing southward geographic expansion of the primary VL vector *Lutzomyia longipalpis*, driven by a combination of urbanisation, movement of infected dogs across borders, deforestation, and climate change-associated shifts in the vector's climatic range (Cruz et al., 2010; Salomón et al., 2012).

Viannia are widespread across America, and *L. braziliensis* contains many sub-species and/or strains that have been characterised in different environments. The *L. braziliensis* clade is genetically highly heterogeneous, with distinct parasite groups associated with different environments. Most notably, the Amazon and Atlantic Forest biomes, which show different recombination histories and population structures (De Los Santos et al., 2022; Heeren et al., 2024). High intraspecific genetic diversity and evidence of mixed reproductive modes are observed in *Viannia* species, with population structure often correlating with geography (Patino et al., 2020; Llanes et al., 2022). The primary reservoirs for *Leishmania* (*Viannia*) species are sylvatic mammals, including rodents, marsupials, and other wild animals. In some regions, domestic and synanthropic animals, particularly dogs, have been implicated as potential reservoirs, especially for *L. (V.) braziliensis*, but recent evidence suggests that humans may be the most important domestic reservoirs in certain transmission cycles (Dantas-Torres, 2007). The transmission cycle is maintained by phlebotomine sand flies, with specific vector species varying by region and *Leishmania* species (Kato et al., 2016).

In the Amazon Basin, there are high genetic diversity and large effective population sizes seen in Amazonian *L. braziliensis* populations, likely reflecting complex eco-epidemiological dynamics and diverse reservoir and vector communities (Guerra et al., 2011; Heeren et al., 2024). This large population size is likely predicted based on the observed high diversity using the Watterson estimator. Multiple *Viannia* species, including *L. guyanensis, L. lainsoni,* and others, circulate in this region, often with overlapping distributions (Guerra et al., 2011; Kato et al., 2016). In the Atlantic Forest, *L. braziliensis* populations are more clonal, with lower genetic diversity, possibly due to ecological isolation and different transmission dynamics (Heeren et al., 2024). In the Andean and Pacific Regions, *L. guyanensis* is the predominant species in some areas, with evidence of intra- and peri-domestic transmission, especially in children, suggesting adaptation to human-modified environments (Gomez et al., 2018).

Dogs are the principal urban and peri-urban reservoir for *L. infantum*, the main causative agent of VL in the region. Their infection is a key indicator and precursor of human cases, especially as the disease urbanizes (Ferreira et al., 2022a; de Freitas et al., 2024). Seroprevalence in dogs can exceed 50% in endemic areas, with some foci reaching up to 75% (Cruz et al., 2010; Chiyo et al., 2023). Outbreaks in dogs often signal the expansion of VL into new areas, as seen in Uruguay and Argentina, and are associated with the presence of the sandfly vector *Lutzomyia longipalpis* (Cruz et al., 2010; Satragno et al., 2017). The shift from rural to urban transmission is linked to landscape changes, increased dog populations, and the adaptation of vectors to urban settings. Factors such as household income, literacy, vegetation, and open sewage influence the spatial distribution and clustering of canine cases. Infection risk varies by dog sex, breed, coat color, and age, with certain groups more susceptible (Thomaz-Soccol et al., 2023; Chiyo et al., 2023; de Freitas et al., 2024). Differences in immune profiles and parasite loads among dogs in different regions help explain local variations in disease patterns (Guerra et al., 2021).

Table 1. The *Viannia* species, the form of disease they are known to cause, and where they are located.

Species	Disease	Distribution
L. braziliensis	CL & ML	Widely distributed across South & Central America
L. guyanensis	CL & ML	Ecuador, Brazil, Bolivia, Belize, French Guiana, Costa Rica, and Colombia.
L. lainsoni	CL	Ecuador, Brazil, Bolivia, Peru, Suriname, French Guiana, and Colombia.
L. lindenbergi	CL	Brazilian Amazon region
L. naiffi	CL	Brazil, French Guiana, Colombia, Ecuador, and Peru
L. panamensis	CL & ML	Brazil, Colombia, Ecuador, Nicaragua, and Panama
L. peruviana	CL	Peru, Ecuador, and Bolivia
L. shawi	CL	Brazil and Peru
L. utingensis	CL	Brazil, Peru, and Ecuador.

1.7. *Leishmania* genomics

While population genomics falls beyond the scope of this project, an overview of the genomic features of *Leishmania* provides essential context for understanding the parasite's biology and pathogenicity. The *L. major* Friedlin strain genome was sequenced and assembled in 2005, marking a major milestone for *Leishmania* molecular biology and serving as a reference for subsequent studies in the genus. The genome consists of 36 chromosomes and is about 32.8 megabases in size, with predictions of 911 RNA genes, 39 pseudogenes, and 8,272 protein-coding genes. Only about 36% of these protein-coding genes could be assigned a putative function (Ivens et al., 2005). The sequencing of two further species using whole-genome shotgun sequencing in 2007 revealed that while the genomes of *Leishmania major, Leishmania infantum,* and *Leishmania braziliensis* are highly conserved in structure, there are key differences in gene content and mechanisms that may underlie their distinct disease manifestations. The genomes of the three species show marked conservation of synteny (gene order and organization), with only about 200 genes differing in their distribution among the species. Unlike *L. major* and *L. infantum*, *L. braziliensis* uniquely possesses active retrotransposable elements (SLACS and TATE) alongside components of an RNA interference (RNAi) pathway. The transposable elements are themselves a potential source of genomic diversity through insertional mutagenesis and chromosomal rearrangement, and the RNAi machinery is thought to be retained specifically to suppress their activity and protect genome integrity, species that have lost active transposons appear to have subsequently lost the need for RNAi (Peacock et al., 2007; Lye et al., 2010). Thus the co-occurrence of active mobile elements and RNAi in *L. braziliensis* suggests greater genomic plasticity in this species compared to other *Leishmania*, as well as offering the practical advantage of RNAi as an experimental tool for functional genetics (Smith, Peacock and Cruz, 2007).

It was observed that pseudogene formation and gene loss, rather than gene gain, are the main evolutionary forces shaping these genomes (Smith, Peacock and Cruz, 2007; Peacock et al., 2007).

Currently, the genomes of many more species of *Leishmania* have been sequenced and assembled, and many species receive higher quality genomes through advanced techniques such as Illumina short-read sequencing and long read sequencing (Oxford Nanopore Technologies, PacBio HiFi). Most *Leishmania* species have highly similar genome organization and gene content, with only a small number of species-specific genes (often fewer than 200). Major differences between species arise from gene amplification and variation in gene copy number, rather than unique gene content (Rogers et al., 2011; Valdivia et al., 2015a). Variation in chromosome number (aneuploidy) is common, with different species showing different numbers of supernumerary chromosomes. Some species, like *L. braziliensis* and *L. tropica*, show high genomic variability, including structural changes, frequent recombination, and hybridization events (Franssen et al., 2020; Glans et al., 2021; Hadermann et al., 2023c).

Leishmania reproduce sexually, and whole genome sequencing of hybrids generated in sand flies shows inheritance patterns consistent with classical meiosis, including recombination and crossing over, indicating a true sexual cycle in *Leishmania* (Inbar et al., 2019; Shaik et al., 2021). *Leishmania* can undergo self-mating (self-hybridization) as well as interstrain mating, both leading to genetic recombination and karyotype variation (Ferreira et al., 2022b). Field and laboratory studies reveal hybrid genotypes and patterns of genetic diversity that cannot be explained by strict clonality, supporting the occurrence of sexual reproduction in natural populations (Rogers et al., 2014; Rougeron, De Meeûs and Bañuls, 2017). Sexual reproduction in *Leishmania* is facultative and relatively rare, estimated to occur about once every 10,000 generations, with most reproduction being clonal (Rogers et al., 2014; Nicoll et al., 2024). The facultative and relatively rare nature of sexual reproduction in *Leishmania* contrasts with patterns seen in related trypanosomatids. In *Trypanosoma brucei*, meiotic sexual reproduction is well-characterised and occurs obligately in the salivary glands of the tsetse fly vector, with haploid gametes and Mendelian allele segregation clearly demonstrated (Peacock et al., 2011, 2014). In contrast, *T. cruzi* exhibits a predominantly clonal population structure with only rare and facultative genetic exchange, resembling *Leishmania* more closely in this respect; two of its six recognised lineages (TcV and TcVI) are of hybrid origin, demonstrating that even infrequent hybridisation can have lasting evolutionary

consequences (Tibayrenc and Ayala, 2015; Matos et al., 2022). Both inbreeding and outcrossing have been observed, with some populations showing high levels of inbreeding and occasional recombination events (Rougeron et al., 2009; Rogers et al., 2014; Ferreira et al., 2022b).

1.8. *Leishmania* Phylogenetics

Single gene phylogenies have been widely used to study the evolutionary relationships among *Leishmania* and related species. These studies reveal that single gene analyses can clarify species boundaries, evolutionary lineages, and genetic diversity, but also highlight limitations such as non-monophyly in some recognized species and the influence of gene conversion.

Sequencing of the hsp70 gene across 17 *Leishmania* species showed the genus forms a monophyletic group with three main subgenera: *Leishmania, Viannia*, and *Sauroleishmania*. The phylogeny supports the distinctness of several species (e.g., *L. donovani, L. major, L. tropica, L. mexicana, L. lainsoni, L. naiffi, L. guyanensis, L. braziliensis*), but some recognized species (*L. aethiopica, L. garnhami, L. amazonensis*) did not form monophyletic clusters, suggesting taxonomic revisions may be needed (Fraga et al., 2010). Analysis of the gp63 gene family in Old World *Leishmania* revealed that gene conversion is a major evolutionary force, generating both diversity and homogenization. A specific class of gp63 genes (gp63(EXT)) can distinguish genetic groups within the *L. donovani* complex and provides insights into host associations and species definitions, indicating that some species definitions (e.g., *L. infantum, L. archibaldi*) require revision (Mauricio et al., 2007). While single gene phylogenies are informative, combining multiple genes or using genome-wide data provides a more robust framework for resolving complex relationships and addressing cases where single gene trees do not match species boundaries. Furthermore, single gene analyses can be affected by gene conversion, recombination, and incomplete lineage sorting, which may obscure true evolutionary relationships.

Microsatellite phylogenies, which use variation at many highly polymorphic loci, often provide a more comprehensive picture of species relationships than single gene phylogenies, especially for closely related taxa such as *Leishmania*. Multilocus microsatellite data can capture broader genomic patterns and reduce the impact of such anomalies (Harr et al., 1998; Fisher et al., 2000; Alvarez et al., 2001). However, microsatellites have drawbacks, such as high mutation rates and potential for homoplasy (independent evolution of similar allele sizes), which can obscure deeper evolutionary relationships or lead to underestimation of divergence times (Fisher et al., 2000; Ochieng et al., 2007). Multilocus microsatellite typing (MLMT) has been applied to *Leishmania donovani, L. major, L. infantum, L. braziliensis,* and *L. panamensis*, revealing distinct genetic populations, patterns of inbreeding, and geographic structuring. For example, microsatellite phylogenies have clarified the relationships among *L. donovani* strains across continents and identified clonal expansions linked to historical events, while also showing that SNPs provide even finer resolution for population studies (Alam et al., 2009; Downing et al., 2012). In *L. major,* MLMT distinguished three main populations corresponding to Central Asia, the Middle East, and Africa, with further sub-populations identified within these regions (Al-Jawabreh et al., 2008). Studies on *L. infantum* have demonstrated low heterogeneity among New World populations and confirmed their recent Old World origin, as well as differentiated European populations and detected gene flow (Bulle et al., 2002; Kuhls et al., 2008, 2011). In *L. braziliensis,* microsatellite analysis revealed extreme inbreeding and strong population structure at a microgeographic scale (Rougeron et al., 2009).

These microsatellite methods have been used in *Viannia,* where 120 *Leishmania (Viannia*) isolates from Brazil were genotyped using 15 microsatellite markers to assess population structure and reproductive strategy. They found three major genetic clusters: Amazonian *L. guyanensis*, Atlantic Forest *L. braziliensis,* and a diverse northern group. There is evidence for both clonal propagation and genetic recombination, especially in *L. braziliensis* and *L. guyanensis*. The Amazon region showed high genetic diversity likely due to vector and host complexity, while some southeastern *L. braziliensis* populations showed signs of recent clonal expansion (Kuhls et al., 2013). A multilocus sequence analysis was conducted on 96 strains

representing nearly all species within the *Viannia* subgenus, targeting partial sequences from four housekeeping genes (G6PD, 6PGD, MPI, and ICD) to evaluate genetic variation both within and between species. The markers showed different levels of polymorphism, allowing identification of species-specific SNPs, ambiguous heterozygous sites, and a potentially ancestral 6PGD allele for the group. Haplotype networks generated using maximum parsimony generally aligned strains with their traditional species classifications based on multilocus enzyme electrophoresis (MLEE), although some discrepancies were noted. A combined NeighborNet phylogenetic analysis confirmed distinct species groupings but also revealed evidence of recombination in *L. braziliensis* and *L. guyanensis*. Phylogenetic trees placed *L. lainsoni* and *L. naiffi* as the earliest-diverging lineages and grouped *L. shawi* within the *L. guyanensis* cluster, calling into question its classification as a separate species. BURST analysis identified six clonal complexes consistent with recognized species, including two genetically distinct *L. braziliensis* clusters: one broadly distributed and the other limited in range. These findings highlight the value of MLSA in *Viannia* taxonomy, demonstrate the influence of both recombination and clonal propagation, and contribute to a clearer understanding of species boundaries in the subgenus (Boité et al., 2012).

Comprehensive genome-wide analyses have been performed on *Leishmania donovani* and *L. infantum*, revealing global population structure, hybridization events, and extensive genetic diversity that single-gene studies could not capture (Franssen et al., 2020; Lypaczewski and Matlashewski, 2021). Whole-genome sequencing has also been applied directly to clinical samples, allowing for detailed phylogenomic analysis and the identification of population structure and genetic variants associated with disease and treatment outcomes (Domagalska et al., 2019; Pilling et al., 2023). In *L. braziliensis*, whole-genome phylogenomic studies have uncovered significant intraspecific variability and the existence of distinct clades, as well as evidence for hybridization and structural genomic variation (S. L. Figueiredo de Sá et al., 2019; Patino et al., 2020). Comparative genomics across different *Leishmania* subgenera, including *Mundinia*, has further clarified evolutionary relationships and gene family changes using whole-genome data (Butenko et al., 2019; Anuntakarun et al., 2022). These genome-wide approaches provide a much more detailed and accurate picture of *Leishmania* evolution and diversity than traditional gene-based phylogenies,

however there are no studies that attempt to include large numbers of different species using high quality whole-genome data.

1.9. Hybridization in *Leishmania*

Genetic exchange has been documented across trypanosomatids, though mechanisms differ substantially between genera, from the well-characterised obligate meiotic sex of *T. brucei* (Peacock et al., 2014) to the predominantly clonal structure of *T. cruzi* where hybrid lineages are rare but consequential (Tibayrenc and Ayala, 2015). In *Leishmania*, the picture is intermediate and still being resolved. Meiosis-like sexual recombination in *Leishmania* has been robustly demonstrated in laboratory settings, particularly through whole-genome sequencing of experimental hybrids generated in sand flies and in vitro. These studies show that hybrid progeny inherit chromosomes and alleles in patterns consistent with classical meiosis, including Mendelian segregation and evidence of crossing over, although some non-Mendelian inheritance and chromosomal anomalies also occur (Akopyants et al., 2009; Inbar et al., 2019; Ferreira et al., 2022b; Catta-Preta et al., 2023). Key meiosis-related genes, such as HOP1 and HAP2, have been shown to be essential for successful mating and genetic exchange, further supporting the presence of a meiotic mechanism (Monte-Neto, Fernandez-Prada and Moretti, 2022; Catta-Preta et al., 2023). In vitro experiments have also produced full-genome hybrids, especially under stress conditions that induce DNA damage, and these hybrids display genetic signatures of meiotic recombination (Louradour et al., 2020a; Monte-Neto, Fernandez-Prada and Moretti, 2022).

Population genetic studies of natural *Leishmania* isolates reveal high levels of heterozygosity, hybrid genotypes, and patterns of recombination, all consistent with sexual reproduction occurring in the wild (Glans et al., 2021; Hadermann et al., 2023c). However, the frequency and mechanisms of sexual versus parasexual processes may vary among species and populations, with evidence for both meiotic and mitotic recombination contributing to genetic diversity (Sterkers et al., 2014). In *L. donovani* populations, studies showed that linkage disequilibrium (LD) decreases

as the distance between SNPs increases, a classic signature of recombination (Franssen et al., 2020). Significant LD among microsatellite markers in *L. braziliensis* has been detected, but not uniform as specific populations exhibit LD breakdown consistent with recombination (Tibayrenc and Ayala, 2021).

Laboratory crosses of *Leishmania* species can also produce hybrids, both within and between species. Early work showed that hybridization was mainly observed in promastigotes developing in the sand fly vector, but recent advances have enabled the generation of hybrids in vitro under specific conditions. For example, exposing *Leishmania* promastigote cultures to DNA-damaging stress (such as irradiation or chemical agents) dramatically increases the efficiency of hybrid formation, allowing not only intraspecies but also interspecies hybrids, including between strains causing different disease forms, to be generated in the lab. These hybrids are typically full-genome, often polyploid (mostly tetraploid), and inherit chromosomes from both parents, as confirmed by whole-genome sequencing and ploidy analysis. Some studies have also demonstrated that natural IgM antibodies can facilitate hybrid formation by promoting parasite clumping and fusion, both in vitro and in vivo (Serafim et al., 2023). While intraspecies hybrids can sometimes produce further progeny, interspecies hybrids often appear sterile. The ability to generate *Leishmania* hybrids in the lab provides a powerful tool for studying parasite genetics, adaptation, and the mechanisms underlying sexual reproduction in these important pathogens, with implications for understanding disease diversity and the evolution of drug resistance and virulence traits (Inbar et al., 2019; Louradour et al., 2020a, 2022; Gutiérrez-Corbo et al., 2022; Serafim et al., 2023).

Natural hybrids of *Leishmania* have been found in the wild, involving both closely and distantly related species. Documented cases include hybrids between *L. donovani* and *L. aethiopica* in Ethiopia, as well as hybrids between different lineages of *L. donovani* itself, indicating that recombination and gene flow are significant features in natural populations of these parasites (Odiwuor et al., 2011; Cotton et al., 2020b; Hadermann et al., 2023c). In the Neotropics, hybrids between *L. braziliensis* and *L. guyanensis* have been identified in Costa Rica and Panama, and full-genome hybrids between Amazonian and Andean lineages of the *L. braziliensis* species complex have been reported in Peru (Van den Broeck et al., 2020, 2023). There is

also evidence for natural hybrids between *L. major* and *L. arabica* in Saudi Arabia, and between *L. infantum* and *L. donovani* strains in Turkey (Kelly et al., 1991; Rogers et al., 2014). These findings show that hybridization can occur both within species (intraspecific) and between different species (interspecific), sometimes resulting in triploid or diploid hybrids with mixed nuclear and mitochondrial genomes, but the interspecific hybrids are always between two closely related species (Van den Broeck et al., 2020, 2023; Hadermann et al., 2023c). Such genetic exchanges may contribute to the emergence of new traits, such as altered virulence or drug resistance, and play a role in the evolution and epidemiology of leishmaniasis.

1.10. Genome assembly

Genome assembly has advanced significantly in recent years, largely due to improvements in long-read sequencing technologies and the development of sophisticated assembly algorithms, which have enabled the production of more complete and accurate genomes, even for complex organisms. Assessing the quality of these assemblies is crucial, and several metrics have been developed for this purpose. The most commonly used metric, N50, measures assembly contiguity but does not always reflect biological completeness or accuracy, leading to calls for more comprehensive evaluation methods (Thrash, Hoffmann and Perkins, 2020; Jauhal and Newcomb, 2021; Wang and Wang, 2023). A contig (short for contiguous sequence) is a continuous stretch of DNA sequence that has been assembled from overlapping reads, with no gaps. In genome assembly, contigs are the basic building blocks and they represent the raw assembled sequences before any scaffolding is done. The N50 value is the length of the shortest contig or scaffold such that 50% of the total genome assembly is contained in contigs/scaffolds of that length or longer. A higher N50 correlates to fewer, longer contigs, however this is not always the case as it can be inflated by the presence of a few very large contigs, with the remaining genome being heavily fragmented. Tools like BUSCO (Benchmarking Universal Single-Copy Orthologs) assess gene space completeness by checking for the

presence of universal single-copy orthologs, providing a biologically meaningful measure that complements technical metrics like N50 (Manni et al., 2021).

There are many genome assemblers available, each designed to handle the challenges of assembling genomes from long-read sequencing technologies like Oxford Nanopore and PacBio. NECAT is recognized for its reliable assembly and effective contig circularization, and it uses a progressive error correction approach that improves the accuracy of Nanopore read assemblies, making it particularly strong in handling complex errors and producing high-quality assemblies efficiently, even for large genomes (Chen et al., 2020; Wick and Holt, 2021). Combining Oxford Nanopore Technologies (ONT) long-read sequencing with Illumina short-read polishing consistently produces high-quality genome assemblies, often matching or exceeding the accuracy and completeness of reference genomes. ONT reads provide long-range continuity but have higher error rates, especially indels, which can affect gene prediction and downstream analyses. Polishing these assemblies with Illumina data corrects most errors, significantly improving base-level accuracy and reducing mispredicted coding sequences and haplotype switching errors (Zhang et al., 2022; Darian et al., 2024).

Long-read sequencing technologies (such as PacBio and Oxford Nanopore) generate reads tens to thousands of kilobases long, enabling the assembly of highly contiguous and accurate genomes, including complex and repetitive regions that are difficult for short-read methods to resolve. These technologies have enabled the creation of near-complete, gapless, and even telomere-to-telomere assemblies in both microbial and human genomes, as well as large, repeat-rich plant genomes (Jain et al., 2017; Logsdon, Vollger and Eichler, 2020; Mascher et al., 2021; Zhang et al., 2022). Combining long reads with short reads (hybrid assembly) further improves accuracy and completeness, often fully resolving genomes with minimal manual intervention (De Maio et al., 2019). The cost of long-read sequencing has dropped significantly. Automated hybrid assembly using Oxford Nanopore and Illumina reads can now fully resolve most bacterial genomes at a lower consumables cost per isolate compared to previous methods. The cost of finishing a microbial genome with long reads has fallen below $1,000, and similar trends are seen in

larger genomes as technology advances (Koren and Phillippy, 2015; Jain et al., 2017; Mascher et al., 2021).

1.11. Telomeres in *Leishmania*

The canonical telomeric repeat in *Leishmania* is typically 5'-TTAGGG-3', but variations exist. A conserved 100 bp sequence, called LCTAS (*Leishmania* Conserved Telomere-Associated Sequence), is found adjacent to telomeres in all studied *Leishmania* species. However, its organization varies: *L. major* has tandem repeats of LCTAS, while L. braziliensis has a single copy per chromosome end (Fu and Barker, 1998; Assis et al., 2021). *Viannia* parasites typically utilize the canonical TTAGGG telomeric repeat, which can be pattern-matched in de novo genome assemblies to identify telomeric sites and assess chromosome-level completeness. The presence or absence of telomeric repeats at contig termini therefore serves as a practical quality metric for the chromosome-level assemblies generated in this study, with fully capped chromosomes indicating complete end-to-end assembly.

1.12. Aims for this project

The aims of this project are to generate high quality genome data for several understudied *leishmania* species and to produce high quality, whole genome assemblies for these species. We also aim to produce the highest quality phylogeny across the various *leishmania* clades to date using the genomes produced in this work as well as publicly available data.

To date, there have been no attempts at generating a phylogeny using all 9 of the currently described *Viannia* species. Those that exist always lack one or more species in their analysis (Boité et al., 2012; Patino et al., 2020; Fraga et al., 2010). The best attempt at an all-encompassing *Viannia* phylogeny made use of 4 housekeeping gene sequences for 96 strains across 7 of the 9 *Viannia* species, but

did not include *L. panamensis* or *L. peruviana*. This analysis establishes *L. lainsoni* as the most divergent species from those sampled, and suggests that *L. shawi* may exist as a subspecies of *L. guyanensis* (Boité et al., 2012). The authors recognize the critical need for a standardised, globally accepted species definition and typing method for *Viannia*, and that to achieve this a consensus gene set that offers optimal differentiation for *Leishmania* species and/or strains is needed. This set of optimally differentiated genes can be easily identified with whole genome sequences assembled for the type strain of each species.

Different *Leishmania* species exhibit varying levels of infectivity, pathogenicity, and virulence. Accurate phylogenetic trees help identify species-specific adaptations and evolutionary processes that influence these traits, which are critical for understanding disease mechanisms and developing targeted interventions (Valdivia et al., 2015b; Bañuls, Hide and Prugnolle, 2007). Phylogenetic analysis provides insights into the genetic variability and population structure of *Leishmania* species, which is important for epidemiological studies. This information can guide public health strategies and control measures by identifying the geographic distribution and transmission dynamics of different species (Bañuls, Hide and Prugnolle, 2007; Behniafar et al., 2024; Ziaei Hezarjaribi et al., 2021). Accurate phylogenetic trees assist in the taxonomic classification of *Leishmania* species, which is essential for diagnosing and treating different forms of leishmaniasis. This includes distinguishing between species that cause cutaneous, mucocutaneous, and visceral leishmaniasis, each requiring different clinical approaches where possible (Real et al., 2013; Chaouch et al., 2013; Marcili et al., 2014). Phylogenetic studies reveal the evolutionary history and divergence of *Leishmania* species, which can inform hypotheses about their origins and spread. This knowledge is valuable for understanding how these parasites have adapted to different hosts and environments over time (Ibrahim and Barker, 2001; Chaouch et al., 2013).

2. Methods

2.1. Samples

The first round of samples used for genomic assembly were chosen to try to achieve one high-quality assembly for each species of *Viannia*. Moreover, some sequencing had already been done in 2022 from multiple samples, including some species outside of the clade, and so the ONT reads were rebasecalled with the latest basecalling software Dorado version 0.9.1 using the dna_r10.4.1_e8.2_400bps_sup@v4.1.0 model. In total, 13 strains of *Leishmania* were sequenced and assembled (Table 2), 8 *Viannia* species and another 5 species split across multiple clades. Two of the *Viannia* samples, L. braziliensis and *L. lainsoni*, are the type strains of their species.

Strains were obtained from a biological collection where isolates are deposited by different researchers, and passage histories prior to deposition were not recorded. For eight strains (Table 2), DNA was extracted following a promastigote–amastigote–promastigote differentiation cycle using infection of macrophage cell lines, with DNA prepared from cultures after five passages. For the remaining strains, DNA had been previously extracted prior to this project and was provided directly, meaning their passage histories are unknown. Biological cloning was not performed, as strains were obtained from an established collection and cloning was not part of the deposition workflow. Multiclonality cannot be entirely excluded, as Leishmania's genomic plasticity means that even cultures established from a single clone may diverge over time; this is acknowledged as an inherent limitation of collection-based comparative genomics studies.

2.2. DNA extraction and sequencing

2.2.1. Phenol/Chloroform DNA extraction

For long read sequencing, genomic DNA was extracted from cells preserved in DNA/RNA Shield (Zymo Research) using a phenol/chloroform extraction protocol specifically for long read DNA sequencing. Wide-bore pipette tips and gentle mixing by inversion was used to prevent DNA shearing. 100 µl of DNA/RNA shield-stored cells were brought up to 200 µl in Elution Buffer (EB - 10mM Tris HCl, pH 8.5). An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) pH 7.8-8.2 was added and the tube was mixed by inversion and centrifuged at 14,000 rpm for 5 minutes. The top aqueous layer was removed and placed into a new tube with equal volume of chloroform/isoamyl alcohol. The tube was mixed by inversion and centrifuged at 14,000 rpm for 5 minutes, followed by removing the top aqueous layer and placing it into a new tube. Ammonium Acetate (NH40Ac) was added to a final concentration of 0.75 M. DNA was precipitated using 1 µl of glycogen (20 µg/mL) and 2.5 times the total volume of 100% ethanol. This tube was left in ice for 1 hour to allow for DNA precipitation, followed by centrifugation at 14,000 rpm for 20 minutes. The supernatant was discarded carefully without disturbing the DNA pellet, followed by a wash with 300 µl 80% ethanol and centrifugation at 14,000 rpm for 5 minutes. The supernatant was discarded followed by a second wash and centrifugation. The residual ethanol was removed and the DNA was resuspended in 40 µl of elution Buffer at 50°C for 20 minutes.

2.2.2. Qiagen kit DNA extraction for short read sequencing

For short read sequencing the same cell suspension was used to extract DNA using the Qiagen Blood and Tissue DNA extraction kit (Qiagen) according to the manufacturer's instructions. After extraction, DNA was quantified using the Invitrogen Qubit 4 Fluorometer, with samples prepared using the dsDNA High Sensitivity (HS) and Broad Range (BR) Assay Kits.

2.2.3. Sequencing

Long read DNA sequencing libraries were prepared by Sally James at the Technology Facility at University of York using the Oxford Nanopore Technologies (ONT) native DNA barcoding kits (SQK-NBD114-24) at the University of York. Barcoded libraries sequenced on R10.4.1 flowcells ONT MinION and PromethION 24 devices running MinKNOW software version 23.07.12, with superaccuracy basecalling performed using running ONT dorado models for basecalling.

Following Qiagen DNA extraction, short read sequencing libraries were prepared using the NEBNext Ultra II FS DNA library preparation kit for Illumina, with unique dual index primers (New England Biolabs). The resultant library was subject to paired end 150 bp sequencing on an Illumina NovaSeq 6000.

2.3. Genome assembly and annotation pipeline

The ONT and Illumina reads were filtered with fastp version 0.23.4(Chen, 2023) using default filtering options for low quality reads and adapter trimming. This means that any reads with bases of a phred quality score <=Q15 are filtered out. The passing ONT reads were then assembled using the NECAT assembler (version 0.0.1) (Chen et al., 2020) using the parameters specified in the config file.These parameters include a genome size of 32 mb, minimum read length of 3000 bases, and coverage of 40x minimum. The assembly was polished with a single round of polishing with medaka (version 1.9.1), racon (version 1.5.0), and pilon (version 1.23) (Safar, Alatar and Mustafa, 2024). Purge Haplotigs version 1.1.3 was used to remove any redundant haplotigs in the assembly (Roach, Schmidt and Borneman, 2018). The assembly was annotated automatically using Companion developed at University of Glasgow (https://companion.gla.ac.uk/) (Haese-Hill, Crouch and Otto, 2024). The *Leishmania* braziliensis MHOM/BR/75/M2904 genome was used as a reference and the companion run used BRAKER3 (Lomsadze et al., 2005; Stanke et al., 2008; Simão et al., 2015). The assembly was evaluated with QUAST v.5.2.0 (Gurevich et al., 2013) after each assembly & polishing step to assess any effects on

the assembly. The number of contigs and N50 were used to assess any improvements in the assembly at each step.

2.4. Phylogenetic Analysis using Orthofinder

Phylogenetic analysis was performed using translated amino acid sequences rather than nucleotide sequences. This choice was made because the dataset spans 63 trypanosomatid species with divergence times estimated at over 100 million years (Lukeš et al., 2018), a depth at which nucleotide sequences, particularly at synonymous third codon positions, are prone to substitution saturation and homoplasy, which can obscure true evolutionary signal and produce artefactual tree topologies (Nei and Kumar, 2000; Opperdoes and Lemey, 2009). At such evolutionary distances, amino acid sequences are generally considered more reliable for phylogenetic inference due to their slower rate of evolution and larger character-state space of 20 amino acids compared to 4 nucleotides, which reduces the probability of convergent evolution (Opperdoes and Lemey, 2009). However, it should be acknowledged that using amino acid sequences comes at a cost, synonymous nucleotide variation is lost, which may reduce resolution among more closely related taxa within subgenera and could lead to underestimation of divergence at shallower nodes (Brown, 2006). Future analyses focusing specifically on within-subgenus relationships, particularly within *Viannia*, would benefit from nucleotide-level analyses to capture this additional variation.

The CDS .fasta file of every strain was filtered and tagged before being translated to protein .fasta files using seqkit v.2.3.1. (Shen et al., 2016). These protein files were then fed to Orthofinder v.2.5.4 which identifies orthologs (genes that originated from a common ancestor in different species) across multiple genomes. The orthologs that are found only once in every species in one copy (single copy genes, SCGs) were extracted. Orthofinder found 1921 of these SCGs which were used for multiple sequence alignment using MAFFT v.7.505 and a modified version of Paraat (Zhang et al., 2012). These alignments were concatenated to a single .fasta file and fed to

Splitstree v.4.19.2 to generate the splitstree network (Huson, Kloepper and Bryant, 2008).

2.5. Telomere mapping

The telomeric repeats in L. braziliensis are characterized by a novel tandem repeat sequence, specifically CCCTAACCCGTGGA, which is considered unusual compared to other species (Fu and Barker, 1998). The LCTAS (*Leishmania* conserved telomere-associated sequence) is a conserved sequence found in all *Leishmania* species, with variations in its organization between species like *L. braziliensis* and *L. major*, is generally described as TTAGGG (Fu and Barker, 1998). Using the fuzznuc function from the EMBOSS package version 6.6.0, this pattern was searched for in all genome assemblies. This fuzznuc output was converted to BED format using a custom perl script. Using the merge function from BEDtools version 2.31.0, any hits of that searched sequence within 10bp of each other and on the same strand were merged to demonstrate any telomeric repeat clusters, rather than treating each hit as a separate feature. These .bed files were then used in R studio to map their locations onto each assembly.

2.6. Population Structure and Admixture Analysis

To investigate phylogenetic relationships and potential hybridization patterns among *Leishmania* species, we employed a novel application of ADMIXTURE v1.3.0 (Alexander, Novembre and Lange, 2009) analysis using fixed differences derived from orthologous genes rather than traditional single nucleotide polymorphisms (SNPs). Concatenated coding sequences from single-copy orthologous genes identified by OrthoFinder were used as input, resulting in a 115,335 bp alignment containing 63 sequences (44 *Leishmania* species and 19 outgroup taxa). From this alignment, we extracted 21,530 biallelic variant sites representing fixed differences between species at homologous positions using a custom Python script. Sites with >20% missing data were excluded, and nucleotides were coded in a binary format (0/2) appropriate for haploid organisms. These orthologous gene variants were

converted to PLINK format (.ped/.map files) and analyzed using ADMIXTURE with K values ranging from 2 to 63. Cross-validation was performed to determine the optimal number of ancestral populations (K), and ancestry proportions were visualized using custom R scripts. This approach differs from conventional ADMIXTURE applications by utilizing fixed interspecific differences rather than intraspecific polymorphisms, enabling detection of species-level admixture and hybridization events while minimizing issues related to missing data and non-homologous sites that can confound genome-wide SNP analyses across distantly related taxa.

The ADMIXTURE analysis specifically focused on identifying potential admixed individuals showing ancestry from multiple distinct lineages, with particular attention to intersubgeneric hybridization events between *Viannia* and *Leishmania* subgenera. Ancestry proportions were interpreted in the context of known *Leishmania* phylogenetic relationships and geographic distributions.

To determine the optimal number of ancestral populations (K), cross-validation analysis was conducted across K values ranging from 2 to 20. For each K value, ADMIXTURE was run with 10-fold cross-validation using the --cv flag, and the analysis was repeated five times with different random seeds to ensure reproducibility. The optimal K was determined by identifying the value that minimized the cross-validation error, following standard protocols for ADMIXTURE analysis (Alexander and Lange, 2011).

Cross-validation errors were calculated as the average prediction error across all validation sets, providing a robust estimate of model performance. The cross-validation curve was generated to visualize the relationship between K and prediction accuracy, with the minimum cross-validation error indicating the most appropriate number of ancestral populations for the dataset.

Individual ancestry proportions were estimated at the optimal K value, with ancestry coefficients representing the proportion of each individual's genome derived from each of the K ancestral populations. Ancestry assignments were considered significant when the proportion exceeded 0.05 (5%) to account for potential noise in admixture estimates.

All analyses were performed on a high-performance computing cluster (Viking2) with sufficient memory allocation (>32GB RAM) to handle the large dataset. Quality control included verification of convergence across multiple runs and assessment of log-likelihood values to ensure optimal model fitting. Visualization of results was performed using custom R scripts with ggplot2 for publication-quality figure generation.

All statistical analysis and plots were produced with R version 4.4.1.

2.7. Hybrid vs Mixed Infection Discrimination Analysis

To distinguish between true intersubgeneric hybridization and mixed infection in strain MHOM/VE/1991/PM-H197, we employed Oxford Nanopore Technologies (ONT) long-read sequencing analysis with differential reference mapping. ONT reads were mapped against high-quality reference genomes representing both putative parental lineages: *Leishmania* braziliensis M2904 (*Viannia* subgenus) and *Leishmania* mexicana MNYC/BZ/62/M379 (*Leishmania* subgenus) using minimap2 v2.24 with parameters optimized for long-read mapping (-ax map-ont).

Mapping statistics were calculated using samtools flagstat v1.15, and coverage depth analysis was performed using samtools depth with per-chromosome coverage calculations. Coverage consistency was assessed using the coefficient of variation (CV = σ/μ) across chromosomes for each reference mapping. The expected coverage ratio of approximately 4:1 (*Leishmania*:*Viannia*) used as a diagnostic criterion here was derived from preliminary ADMIXTURE ancestry inference (see Results section 3.5), which indicated approximately 80% *L. mexicana* complex ancestry and 20% *Viannia* ancestry in strain PM-H197. Coverage analysis was subsequently performed to test whether this ancestry signal was reflected in actual genomic abundance., while mixed infections were expected to show variable coverage patterns and ratios inconsistent with genome dosage.

Quality control included assessment of mapping rates, read length distributions, and mapping quality scores. Coverage outliers were identified using interquartile range analysis, and chromosomes with extremely low coverage (<5x) were excluded from consistency calculations. All analyses were performed in R using custom scripts for

statistical analysis and visualization.

The diagnostic criteria for hybrid vs mixed infection classification included: (1) coverage ratio consistency with expected genome dosage from ADMIXTURE analysis, (2) chromosome-level coverage uniformity (CV < 0.5), (3) mapping rate thresholds (>70% for dominant genome, >40% for minor genome), and (4) absence of extreme coverage heterogeneity indicative of sample contamination.

To distinguish between hybrid and mixed infection scenarios in strain MHOM/VE/1991/PM-H197, competitive mapping was performed using ultra-long Oxford Nanopore reads (≥100 kb). Ultra-long reads were extracted from the total ONT dataset using seqkit v2.3.1 with length filtering parameters. A composite reference genome was constructed by concatenating high-quality reference sequences from *Leishmania* braziliensis M2904 (*Viannia* subgenus) and *Leishmania* mexicana MNYC/BZ/62/M379 (*Leishmania* subgenus), with species-specific prefixes added to contig identifiers (viannia_ and mexicana_) to enable unambiguous mapping target identification.

Ultra-long reads were mapped to the concatenated reference using minimap2 with Oxford Nanopore-optimized parameters (-ax map-ont). Mapping statistics were calculated using samtools flagstat, and coverage analysis was performed using samtools depth with 10 kb genomic bins to reduce noise while maintaining positional resolution. Species preference was quantified by calculating the proportion of total coverage mapping to viannia-tagged versus mexicana-tagged reference sequences.

Coverage fractions were calculated as the ratio of species-specific coverage to total coverage across genomic positions. Statistical analysis and visualization were performed using R with custom scripts. The competitive approach eliminates reference genome quality bias by providing equivalent access to both species' sequences, ensuring mapping preferences reflect genuine sequence similarity rather than technical artifacts. Ultra-long reads (≥100 kb) minimize mapping ambiguity and can unambiguously span repetitive genomic regions that confound shorter read analyses, providing high-confidence species assignment for complex infections.

3. Results

3.1. Genome Assemblies

Fourteen samples of various Leishmania species were sequenced using Oxford Nanopore Technology (ONT) and Illumina short read sequencing. One sample produced only 1.3x ONT coverage and was excluded from assembly and downstream analysis; the remaining thirteen assembled samples are listed in Table 2. ONT runs produced a mean of 3.5 million reads per sample. There was a mean coverage of 138x for ONT runs, with one sample of *L. panamensis* resulting in lower than average coverage of 56x (Table 2.). This sample was still assembled as studies assembling complete *Leishmania* genomes with long-read technologies (including ONT and PacBio) typically achieve high-quality, chromosome-scale assemblies with coverage in the range of 50x - 100x or higher, especially to resolve repetitive regions and close sequence gaps (Lypaczewski et al., 2018; Camacho et al., 2019; Almutairi et al., 2021). This data, paired with illumina short reads at 123x coverage for polishing, was deemed sufficient for an assembly. One sample produced only 1.3x coverage with ONT and so was not assembled and excluded from the analysis.

Table 2. List of all samples sequenced and assembled. All statistics are for Oxford Nanopore Sequencing outputs only. Million represented by ‘m’. Billion represented by ‘b’

Species	WHO ID	Total reads	Total bases	Bases >=Q20 after filtering	Bases >=Q30 after filtering	Estimated Coverage
L. braziliensis	MHOM/BR/1975/M2903	4.2m	5.4b	84%	74%	155x
L. lainsoni	MHOM/BR/2002/NMT-RBO 027P	3.3m	4.4b	88%	77%	136x
L. lainsoni	MHOM/BR/1981/M6426	3.6m	4.5b	90%	82%	137x
L. naiffi	ISQU/BR/1985/IM2264	4.4m	5.5b	85%	76%	176x
L. panamensis	MHOM/NI/1988/ZE09	4.1m	5.4b	87%	79%	168x
L. panamensis	MHOM/PA/1971/LS94	1m	1.8b	86%	74%	56x
L. shawi	MHOM/BR/1990/IM2842	3.3m	4.4b	83%	79%	136x
Leishmania (Viannia) Hybrid	MHOM/VE/1991/PM-H197	2.8m	3.8b	89%	78%	117x
L. colombiensis	IHAR/CO/1985/CL500	3.5m	4.3b	88%	76%	128x
L. equatorensis	MCHO/EC/1982/LSP1	3.2m	4.4b	87%	79%	132x
L. herreri	MCHO/CR/1974/016;Ch-97;LV344	4.2m	5.4b	88%	76%	156x
L. adleri	RLIZ/KE/1954/LV30	2.9m	4.1b	88%	77%	121x
L. pifanoi	MHOM/VE/1957/LL1	4.4m	5.6b	86%	74%	176x

The sequencing data was passed through a genome assembly pipeline, using NECAT as the assembler, and after two rounds of polishing using the ONT and illumina reads, the final assembly statistics were reported (Table 3.). Generally in genome assembly, a number of contigs closest to the expected number of chromosomes in the genome is desirable, with generally higher N50 values similarly desirable for a contiguous chromosome-level assembly (Wallberg et al., 2019; Human Genome Structural Variation Consortium et al., 2021). In the case of *L. braziliensis,* the current reference genome reports a 35 contig, chromosomal genome assembly with an N50 of 992,961 bases. However, after the initial assembly of the Illumina short reads, there were in fact 1,031 contigs with an N50 of 57,784 bases (Peacock et al., 2007; Rogers et al., 2011). This is because they used scaffolding to join various contigs together to form longer contigs, whereas in the case of the work presented here, contig number and N50 are not a result of scaffolding but rather from assembly and polishing.

Table 3. Comparison between the current highest quality genome published for each species and the genomes generated in this work for each species. The strain of each species sequenced in this work may not correspond with the best available assembly in the literature. Number of contigs and N50 values were used as metrics of quality for each assembly.

Species	Genus	WHO ID	Published: contigs	Published: N50 (kb)	This study: contigs	This study: N50 (kb)
L. braziliensis	*Viannia*	MHOM/BR/1975/M2903	35	1063	54	861
L. lainsoni	*Viannia*	MHOM/BR/2002/NMT-RBO 027P	137	638	53	920
L. lainsoni	*Viannia*	MHOM/BR/1981/M6426	137	638	49	840
L. naiffi	*Viannia*	ISQU/BR/1985/IM2264	1749	42	68	868
L. panamensis	*Viannia*	MHOM/NI/1988/ZE09	35	1043	38	1144
L. panamensis	*Viannia*	MHOM/PA/1971/LS94	35	1043	137	563
L. shawi	*Viannia*	MHOM/BR/1990/IM2842	1543	44	55	913
Leishmania (Viannia) Hybrid	*Viannia*	MHOM/VE/1991/PM-H197	NA	NA	48	1063
L. colombiensis	Leishmania/Endotrypanum	IHAR/CO/1985/CL500	NA	NA	50	913
L. equatorensis	Leishmania/Endotrypanum	MCHO/EC/1982/LSP1	NA	NA	80	758
L. herreri	Leishmania/Endotrypanum	MCHO/CR/1974/016;Ch-97;LV344	NA	NA	44	903
L. adleri	Sauroleishmania	RLIZ/KE/1954/LV30	5785	39	47	845
L. pifanoi	Leishmania	MHOM/VE/1957/LL1	NA	NA	48	1059

Assemblies generated in this study achieved N50 values ranging from 563 kb (L. panamensis LS94) to 1,144 kb (L. panamensis ZE09), with the majority of assemblies exceeding 840 kb. This represents a substantial improvement over previously published references for the most fragmented species: L. naiffi improved from a published N50 of 42 kb to 868 kb, and L. shawi from 44 kb to 913 kb. For species with existing high-quality references, such as L. braziliensis (published N50: 1,063 kb), our assembly achieved a comparable N50 of 861 kb without scaffolding. Contig numbers were similarly reduced, with most assemblies producing 38–80 contigs compared to >1,500 in some published references. A comparison of N50 values across species is shown in Figure 1, and log-transformed contig counts are shown in Figure 2.

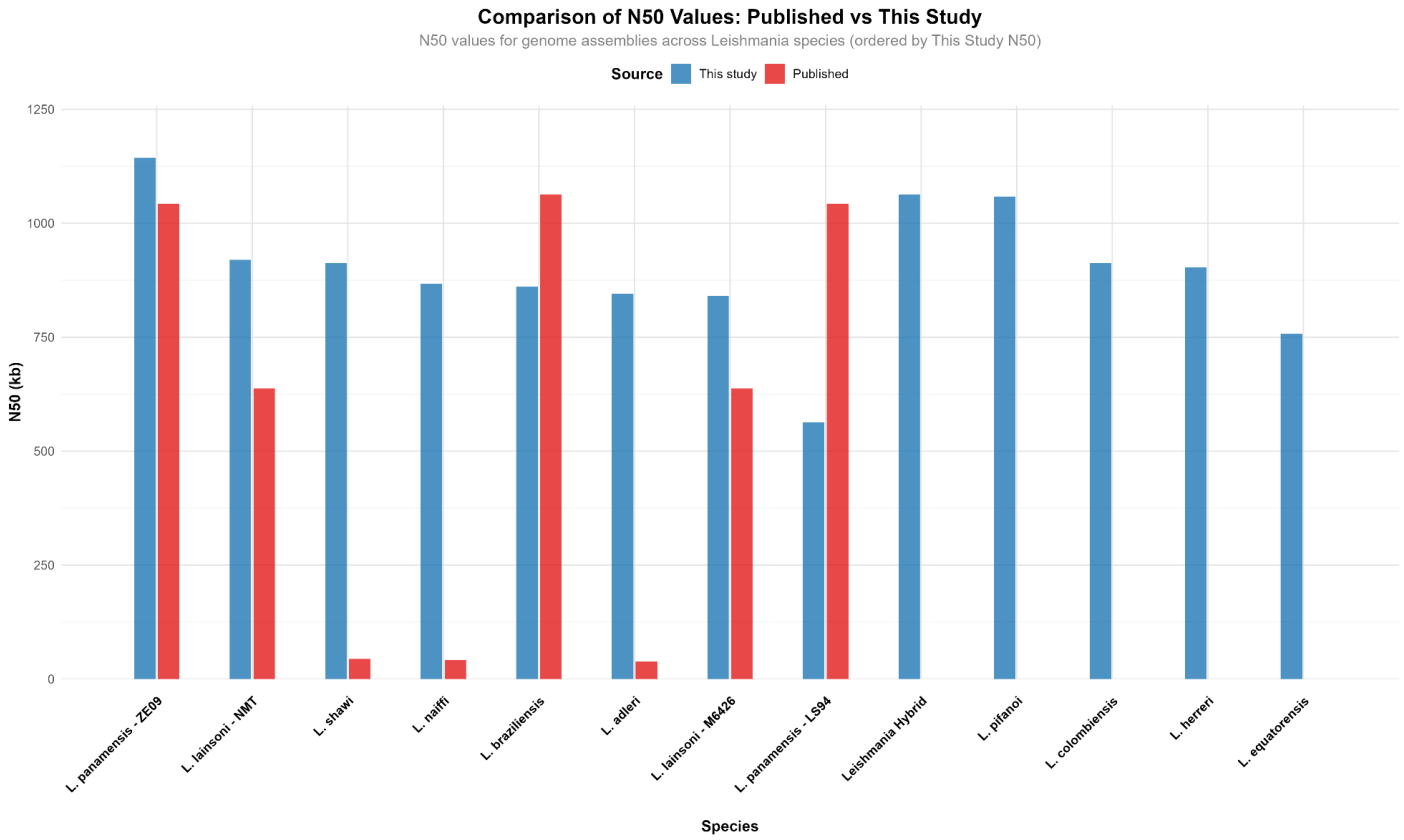


Figure 1. Comparison of N50 values between published genome assemblies and this study across *Leishmania* species. Bar chart showing N50 values (kb) for genome assemblies of *Leishmania* species, comparing published data (red bars) with assemblies generated in this study (blue bars). Species are ordered by descending N50 values from this study, with species having published N50 data displayed first, followed by species with only this study's data. The N50 statistic

represents the contig length at which 50% of the total assembly length is contained in contigs of that size or larger, serving as a measure of assembly contiguity. Higher N50 values indicate more contiguous assemblies with fewer gaps. Data were obtained from NCBI and genome assemblies were generated using ONT read data with polishing using illumina short read data. Error bars are not shown as N50 represents a single calculated value per assembly.

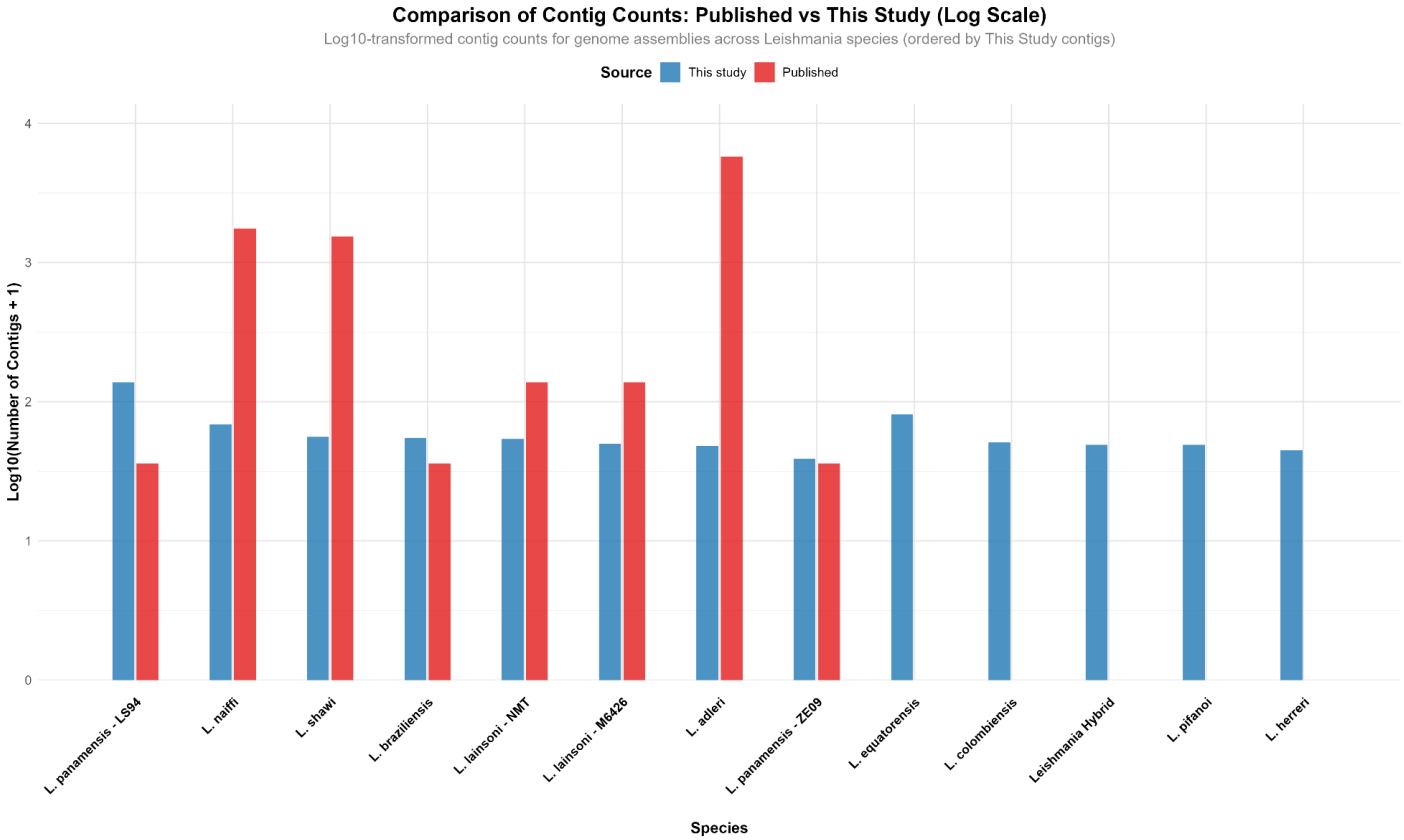


Figure 2. Comparison of contig counts between published genome assemblies and this study across *Leishmania* species (log-transformed scale).
Bar chart showing log10-transformed contig counts for genome assemblies of *Leishmania* species, comparing published data (red bars) with assemblies generated in this study (blue bars). Species are ordered by descending contig counts from this study, with species having published contig data displayed first, followed by species with only this study's data. Contig counts were log10-transformed [log10(count + 1)]

to accommodate the wide range of values and improve visualization of differences across species. Lower contig counts generally indicate more contiguous assemblies with fewer breaks, while higher contig counts suggest more fragmented assemblies. The log transformation allows for better comparison of assemblies with vastly different fragmentation levels. Genome data were obtained from NCBI and genome assemblies were generated using ONT read data with polishing using illumina short read data. Contig count and N50 values assessed using QUAST v.5.2.0.

BUSCO evaluates genome assemblies by searching for sets of Benchmarking Universal Single-Copy Orthologs (Figure 3. & 4.). These are genes expected to be present and single-copy in nearly all members of a lineage. The results include percentages of complete, duplicated, fragmented, and missing genes, providing a biologically meaningful measure of assembly completeness (Manni et al., 2021). While >90% completeness is often cited as a benchmark for high-quality assemblies, the appropriate threshold can vary by organism and available data (Jauhal and Newcomb, 2021; Feron and Waterhouse, 2022). For the eukaryota database, our assemblies achieved approximately 50% BUSCO completeness on average against 255 benchmark genes (Figure 3). This relatively modest score is expected given the broad phylogenetic distance between Leishmania and general eukaryotic gene sets. Against the more relevant euglenozoa database, all assemblies exceeded 90% complete and single-copy BUSCO scores, confirming high assembly completeness for this lineage (Figure 4).

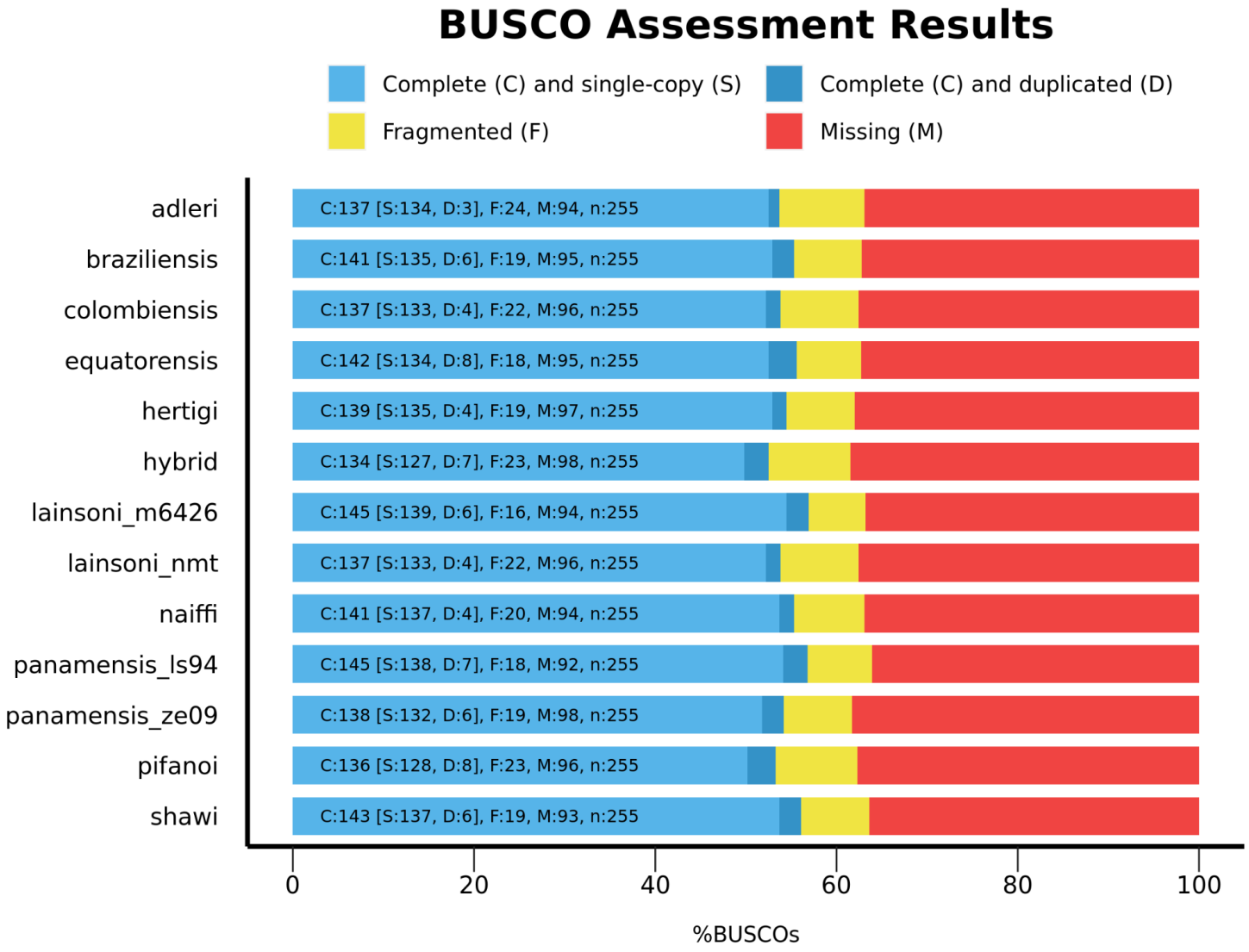


Figure 3. BUSCO assessment of each assembly using the eukaryota database. This compares the gene content of our assemblies to 255 key single-copy orthologs found in all eukaryota as a measure of assembly quality and completeness.

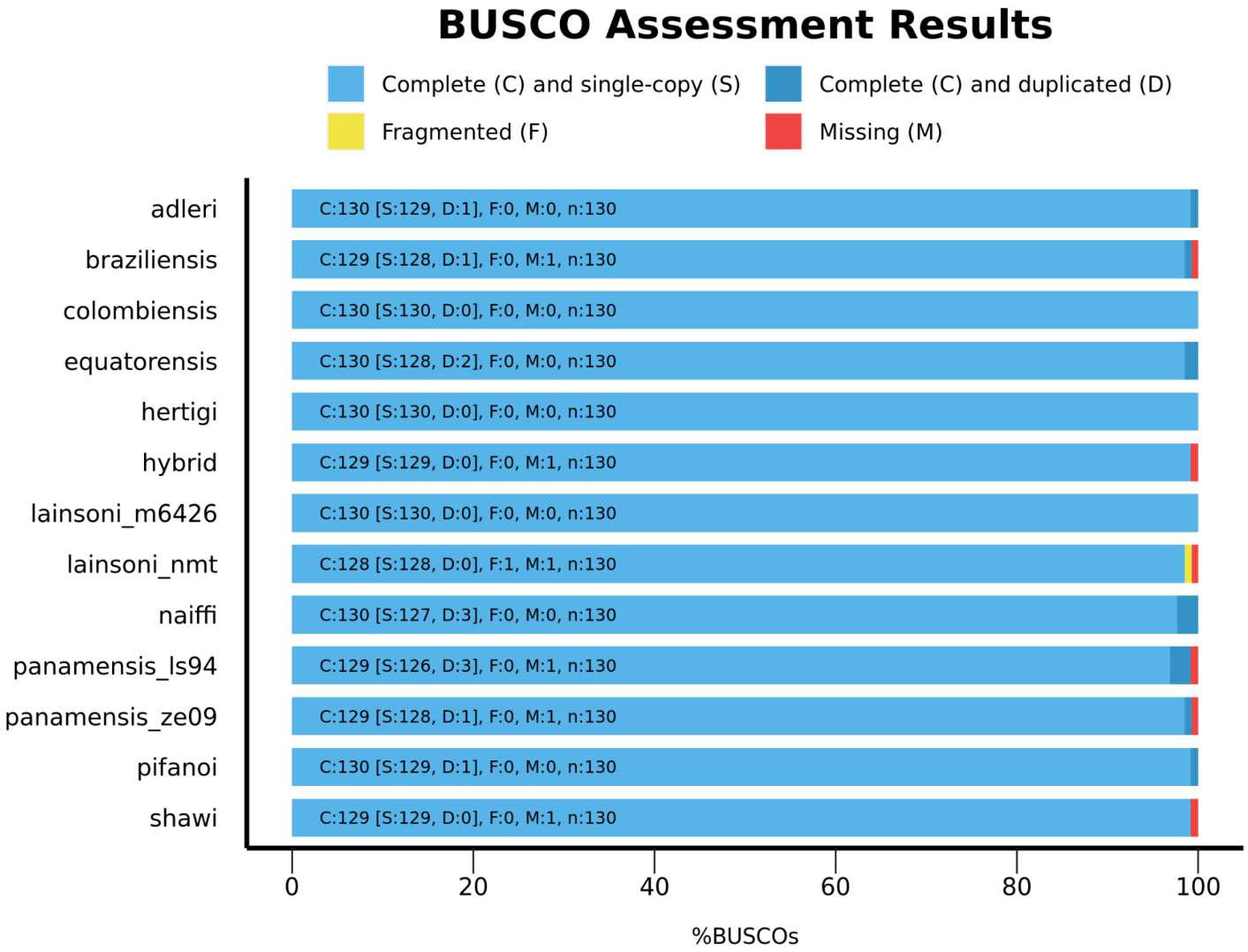
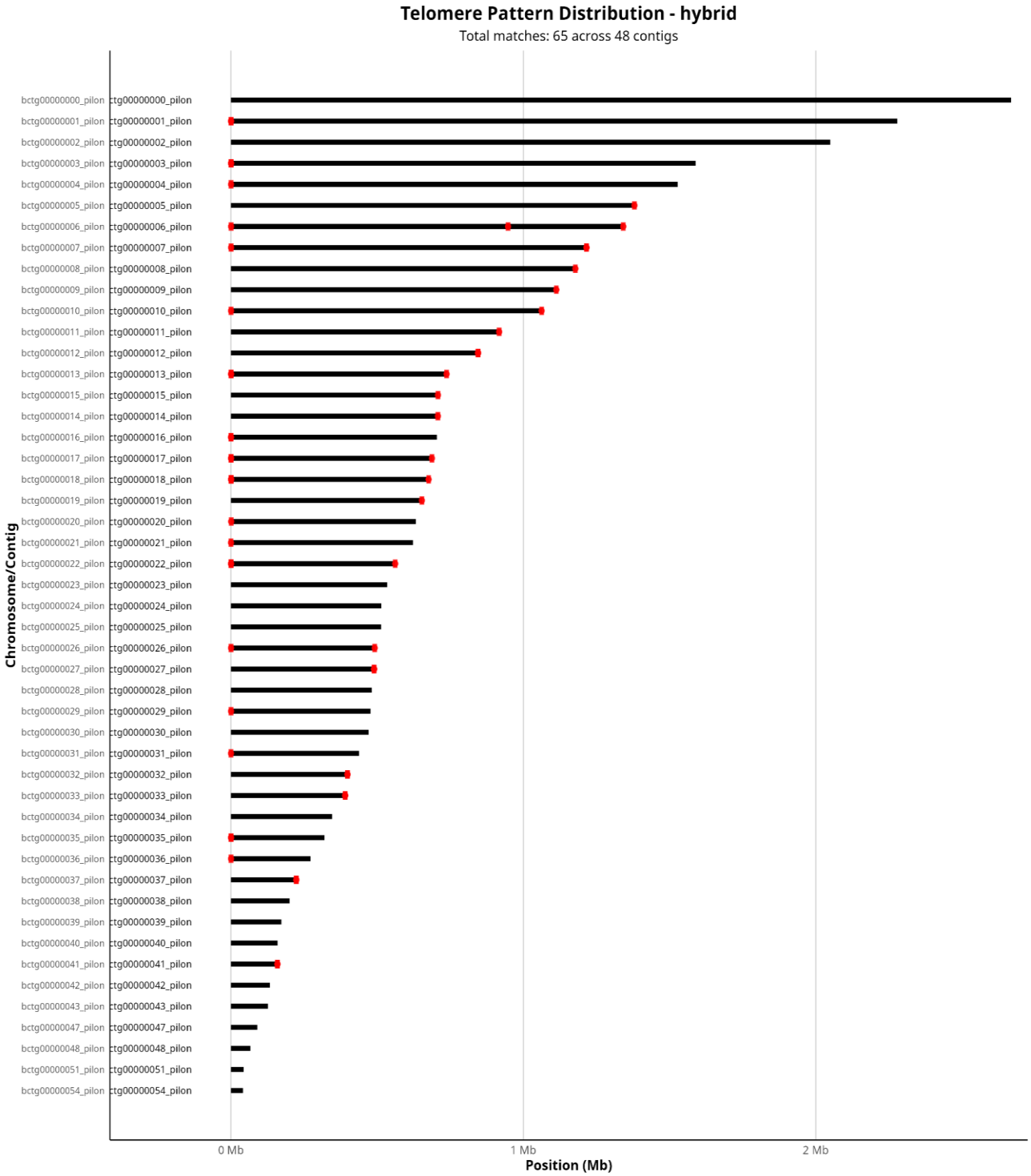


Figure 4. BUSCO assessment of each assembly using the euglenozoa database. This compares the gene content of our assemblies to 255 key single-copy

orthologs found in all euglenozoa as a measure of assembly quality and completeness.

3.2. Telomeric repeats mapped

Telomere patterns are typically enriched at chromosome termini, appearing as clusters of markers at the 5' and 3' ends of chromosomes. The presence of internal telomere-like sequences may indicate chromosomal rearrangements, fusion events, or assembly artifacts. Contig names are displayed on the left, with total contig lengths indicated. Telomere pattern identification was performed using fuzznuc with the pattern TTAGGG (and reverse complement CCCTAA). Across the assemblies, telomeric repeats were predominantly enriched at chromosomal termini, consistent with well-assembled chromosomes. The *L. lainsoni NMT* assembly showed the highest capping percentage at 45.28%, with 24 of 53 contigs fully capped, while *L. naiffi* and *L. panamensis* LS94 showed the lowest at 1.47% and 1.45% respectively, indicating more fragmented terminal regions. Internal telomere-like signals, where present, may indicate chromosomal rearrangements or assembly artifacts. Telomere distribution plots for *L. hybrid* and *L. lainsoni* NMT are shown in Figure 5; plots for all remaining assemblies are provided in supplementary materials.


Telomere Pattern Distribution - hybrid
Total matches: 65 across 48 contigs
Chromosome/Contig
bctg00000000_pilon ctg00000000_pilon
bctg00000001_pilon ctg00000001_pilon
bctg00000002_pilon ctg00000002_pilon
bctg00000003_pilon ctg00000003_pilon
bctg00000004_pilon ctg00000004_pilon
bctg00000005_pilon ctg00000005_pilon
bctg00000006_pilon ctg00000006_pilon
bctg00000007_pilon ctg00000007_pilon
bctg00000008_pilon ctg00000008_pilon
bctg00000009_pilon ctg00000009_pilon
bctg00000010_pilon ctg00000010_pilon
bctg00000011_pilon ctg00000011_pilon
bctg00000012_pilon ctg00000012_pilon
bctg00000013_pilon ctg00000013_pilon
bctg00000015_pilon ctg00000015_pilon
bctg00000014_pilon ctg00000014_pilon
bctg00000016_pilon ctg00000016_pilon
bctg00000017_pilon ctg00000017_pilon
bctg00000018_pilon ctg00000018_pilon
bctg00000019_pilon ctg00000019_pilon
bctg00000020_pilon ctg00000020_pilon
bctg00000021_pilon ctg00000021_pilon
bctg00000022_pilon ctg00000022_pilon
bctg00000023_pilon ctg00000023_pilon
bctg00000024_pilon ctg00000024_pilon
bctg00000025_pilon ctg00000025_pilon
bctg00000026_pilon ctg00000026_pilon
bctg00000027_pilon ctg00000027_pilon
bctg00000028_pilon ctg00000028_pilon
bctg00000029_pilon ctg00000029_pilon
bctg00000030_pilon ctg00000030_pilon
bctg00000031_pilon ctg00000031_pilon
bctg00000032_pilon ctg00000032_pilon
bctg00000033_pilon ctg00000033_pilon
bctg00000034_pilon ctg00000034_pilon
bctg00000035_pilon ctg00000035_pilon
bctg00000036_pilon ctg00000036_pilon
bctg00000037_pilon ctg00000037_pilon
bctg00000038_pilon ctg00000038_pilon
bctg00000039_pilon ctg00000039_pilon
bctg00000040_pilon ctg00000040_pilon
bctg00000041_pilon ctg00000041_pilon
bctg00000042_pilon ctg00000042_pilon
bctg00000043_pilon ctg00000043_pilon
bctg00000047_pilon ctg00000047_pilon
bctg00000048_pilon ctg00000048_pilon
bctg00000051_pilon ctg00000051_pilon
bctg00000054_pilon ctg00000054_pilon
0 Mb
1 Mb
2 Mb
Position (Mb)

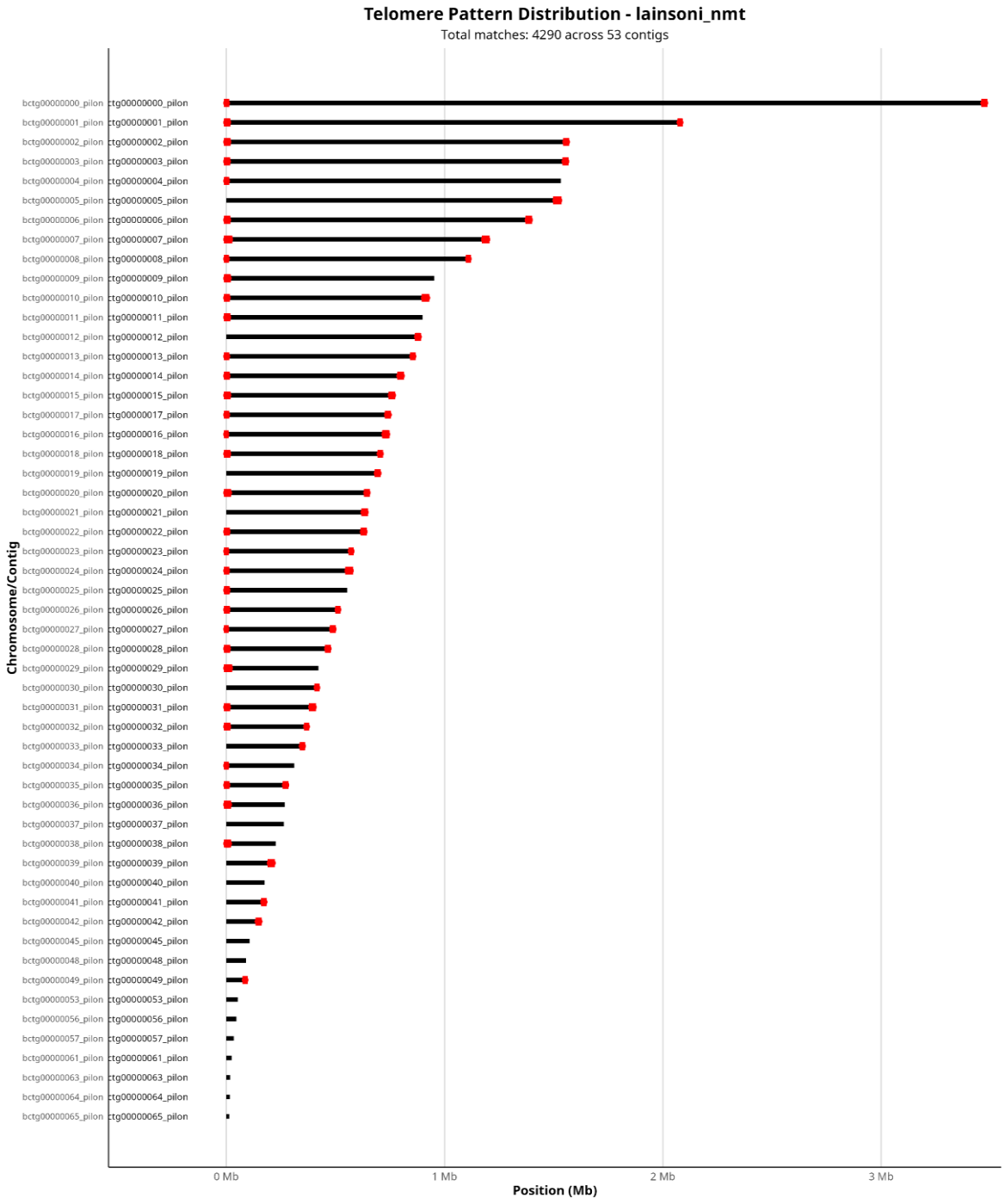

Telomere Pattern Distribution - lainsoni_nmt
Total matches: 4290 across 53 contigs
Chromosome/Contig
bctg00000000_pilon ctg00000000_pilon
bctg00000001_pilon ctg00000001_pilon
bctg00000002_pilon ctg00000002_pilon
bctg00000003_pilon ctg00000003_pilon
bctg00000004_pilon ctg00000004_pilon
bctg00000005_pilon ctg00000005_pilon
bctg00000006_pilon ctg00000006_pilon
bctg00000007_pilon ctg00000007_pilon
bctg00000008_pilon ctg00000008_pilon
bctg00000009_pilon ctg00000009_pilon
bctg00000010_pilon ctg00000010_pilon
bctg00000011_pilon ctg00000011_pilon
bctg00000012_pilon ctg00000012_pilon
bctg00000013_pilon ctg00000013_pilon
bctg00000014_pilon ctg00000014_pilon
bctg00000015_pilon ctg00000015_pilon
bctg00000017_pilon ctg00000017_pilon
bctg00000016_pilon ctg00000016_pilon
bctg00000018_pilon ctg00000018_pilon
bctg00000019_pilon ctg00000019_pilon
bctg00000020_pilon ctg00000020_pilon
bctg00000021_pilon ctg00000021_pilon
bctg00000022_pilon ctg00000022_pilon
bctg00000023_pilon ctg00000023_pilon
bctg00000024_pilon ctg00000024_pilon
bctg00000025_pilon ctg00000025_pilon
bctg00000026_pilon ctg00000026_pilon
bctg00000027_pilon ctg00000027_pilon
bctg00000028_pilon ctg00000028_pilon
bctg00000029_pilon ctg00000029_pilon
bctg00000030_pilon ctg00000030_pilon
bctg00000031_pilon ctg00000031_pilon
bctg00000032_pilon ctg00000032_pilon
bctg00000033_pilon ctg00000033_pilon
bctg00000034_pilon ctg00000034_pilon
bctg00000035_pilon ctg00000035_pilon
bctg00000036_pilon ctg00000036_pilon
bctg00000037_pilon ctg00000037_pilon
bctg00000038_pilon ctg00000038_pilon
bctg00000039_pilon ctg00000039_pilon
bctg00000040_pilon ctg00000040_pilon
bctg00000041_pilon ctg00000041_pilon
bctg00000042_pilon ctg00000042_pilon
bctg00000045_pilon ctg00000045_pilon
bctg00000048_pilon ctg00000048_pilon
bctg00000049_pilon ctg00000049_pilon
bctg00000053_pilon ctg00000053_pilon
bctg00000056_pilon ctg00000056_pilon
bctg00000057_pilon ctg00000057_pilon
bctg00000061_pilon ctg00000061_pilon
bctg00000063_pilon ctg00000063_pilon
bctg00000064_pilon ctg00000064_pilon
bctg00000065_pilon ctg00000065_pilon
0 Mb
1 Mb
2 Mb
3 Mb
Position (Mb)

Figure 5. Telomere pattern distribution across chromosomes in L. hybrid and *L. lainsoni*_NMT genome assemblies. Chromosome plots for the remaining assemblies can be found in supplementary materials. Chromosomal plots showing the distribution of telomere repeat patterns across assembled contigs for each *Leishmania* species. Each horizontal bar represents a contig or chromosome, with telomere pattern matches (TTAGGG/CCCTAA hexamer repeats) indicated by colored markers along the length of each sequence. The x-axis shows genomic coordinates (bp) along each contig, while the y-axis displays individual contigs ordered by size (largest to smallest).

The genomic coordinates of matched telomeric patterns were analyzed to determine chromosome completeness (Table 4.). Contigs were classified as "capped" when telomeric repeats were present within terminal regions (5 kb from both ends), indicating complete chromosome assembly. The analysis categorizes contigs into four groups: (1) capped contigs with telomeres at both terminal regions, (2) contigs with telomeres only at the left terminus, (3) contigs with telomeres only at the right terminus, and (4) contigs lacking telomeres at either terminus. Capping percentage represents the proportion of contigs with telomeres at both ends relative to the total number of contigs per assembly, serving as a metric for chromosome-level assembly completeness. Higher capping percentages indicate more complete genome assemblies with fewer fragmented chromosomes.

Table 4. Chromosome capping analysis of *Leishmania* genome assemblies based on telomeric repeat distribution.

Sample	Total Contigs	Capped Contigs	Uncapped Contigs	Left Only	Right Only	Neither End	Capping Percentage
adleri	47	8	39	8	6	25	17.02
braziliensis	54	3	51	9	8	34	5.55
colombiensis	50	10	40	10	6	24	20
equatorensis	80	18	62	14	18	30	22.5
hertigi	44	12	32	11	12	9	27.27
hybrid	48	8	40	10	13	17	16.66
lainsoni_m6426	49	1	48	3	5	40	2.04
lainsoni_nmt	53	24	29	8	10	11	45.28
naiffi	68	1	67	3	8	56	1.47
panamensis_ls94	137	2	135	21	11	103	1.45
panamensis_ze09	38	3	35	9	5	21	7.89
panamensis_ze09	38	3	35	9	5	21	7.89
pifanoi	48	20	28	7	6	15	41.66
shawi	55	7	48	3	13	32	12.72

3.3. Orthofinder

A comprehensive orthology analysis was performed using OrthoFinder on 63 trypanosomatid species, encompassing 29 (including the 13 assembled in this study) *Leishmania* species, 19 *Trypanosoma* species, and 15 other trypanosomatid taxa. The analysis identified a total of 16,683 orthogroups across all species examined.

The analysis revealed limited core genome conservation across the trypanosomatid phylogeny, with only 844 orthogroups (5.06%) present in all 63 species. The majority of orthogroups were classified as accessory genes (8,969 orthogroups, 53.76%), indicating substantial genomic diversity within the family. Soft-core genes, present in most but not all species, comprised 3,515 orthogroups (21.07%). Notably, 3,355 orthogroups (20.11%) were unique to individual species, suggesting extensive lineage-specific gene evolution.

Considerable variation in gene repertoire size was observed across species, with a 4.41-fold difference between the largest and smallest genomes. *Trypanosoma cruzi CL* exhibited the largest gene complement with 31,539 genes, while *L. guyanensis* possessed the smallest with 7,151 genes. Among the *Viannia* species of primary interest, gene counts ranged from 7,151 (*L. guyanensis*) to 9,282 *(L. panamensis LS94*), with most species containing approximately 8,400-9,000 genes.

The analysis identified only 76 single-copy orthogroups (0.46% of total orthogroups) suitable for phylogenetic reconstruction, representing a significant constraint for molecular phylogenetic studies. The remaining 16,607 orthogroups contained multiple gene copies across species, making them unsuitable for standard phylogenetic inference methods.Comparative analysis revealed that *Trypanosoma* species generally possessed larger gene repertoires than *Leishmania* species. Notably, *T. cruzi* showed a substantially expanded gene complement compared to the average *Leishmania* species, though the magnitude of this difference should be interpreted cautiously as the reference *T. cruzi* CL Brener strain used here is itself a hybrid, and allelic duplications in hybrid genomes can inflate gene count estimates (El-Sayed et al., 2005). More conservative estimates place the *T. cruzi* coding

genome at approximately 12,000 genes compared to around 8,000 in *Leishmania* (Andersson, 2011). This expansion is driven primarily by the massive amplification of multigene surface protein families including trans-sialidases, mucins, and mucin-associated surface proteins, which are thought to facilitate host immune evasion across the diverse tissue environments *T. cruzi* encounters during infection (El-Sayed et al., 2005).

A full list of statistical findings with biological interpretations can be found in supplementary materials (Supplementary Table 1.)

3.4. Phylogenetic tree using Single Copgy Gene (SCG) orthogroups

Our phylogenetic analysis of 63 trypanosomatid species revealed well-supported major clades consistent with established taxonomic classifications (Figure 6). The maximum likelihood tree clearly separated the main genera, with Trypanosoma species forming distinct clades separate from *Leishmania* and related genera. Within the *Leishmania* complex, our analysis recovered four major subgeneric groups that align with current taxonomic understanding.

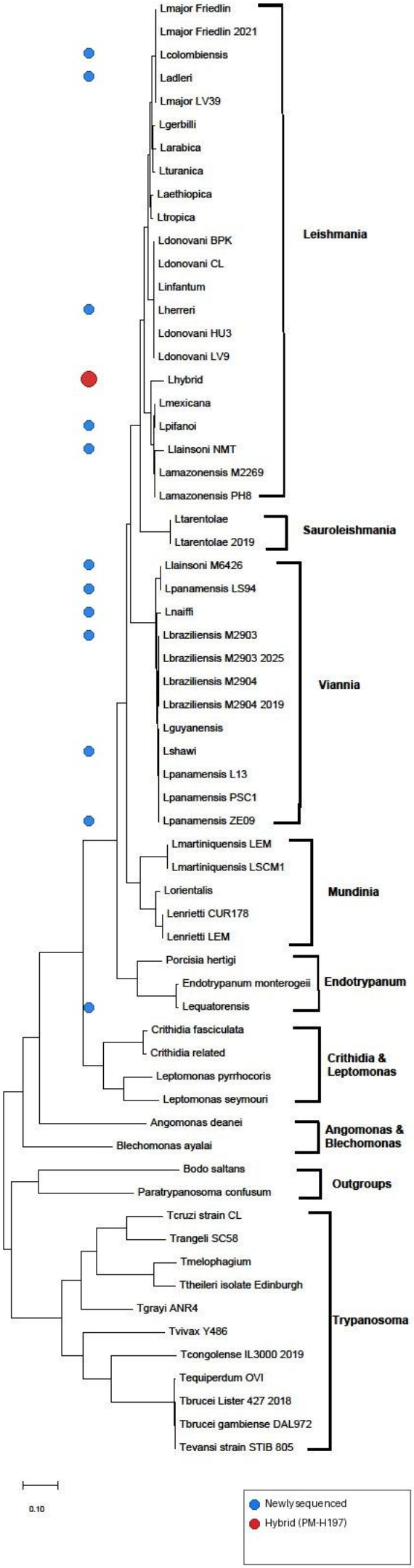

Lmajor Friedlin
Lmajor Friedlin 2021
Lcolombiensis
Ladleri
Lmajor LV39
Lgerbilli
Larabica
Lturanica
Laethiopica
Ltropica
Ldonovani BPK
Ldonovani CL
Linfantum
Lherreri
Ldonovani HU3
Ldonovani LV9
Lhybrid
Lmexicana
Lpifanoi
Llainsoni NMT
Lamazonensis M2269
Lamazonensis PH8
Leishmania
Ltarentolae
Ltarentolae 2019
Sauroleishmania
Llainsoni M6426
Lpanamensis LS94
Lnaiffi
Lbraziliensis M2903
Lbraziliensis M2903 2025
Lbraziliensis M2904
Lbraziliensis M2904 2019
Lguyanensis
Lshawi
Lpanamensis L13
Lpanamensis PSC1
Lpanamensis ZE09
Viannia
Lmartiniquensis LEM
Lmartiniquensis LSCM1
Lorientalis
Lenrietti CUR178
Lenrietti LEM
Mundinia
Porcisia hertigi
Endotrypanum monterogeii
Lequatorensis
Endotrypanum
Crithidia fasciculata
Crithidia related
Leptomonas pyrrhocoris
Leptomonas seymouri
Crithidia & Leptomonas
Angomonas deanei
Blechomonas ayalai
Angomonas & Blechomonas
Bodo saltans
Paratrypanosoma confusum
Outgroups
Tcruzi strain CL
Trangeli SC58
Tmelophagium
Ttheileri isolate Edinburgh
Tgrayi ANR4
Tvivax Y486
Tcongolense IL3000 2019
Tequiperdum OVI
Tbrucei Lister 427 2018
Tbrucei gambiense DAL972
Tevansi strain STIB 805
Trypanosoma
0.10
Newly sequenced
Hybrid (PM-H197)

Figure 6. Phylogenetic tree of 63 trypanosomatid species. Maximum likelihood phylogenetic tree constructed from molecular sequence data showing evolutionary relationships among trypanosomatid parasites. The tree includes representatives from major genera including *Leishmania, Trypanosoma*, *Crithidia*, *Leptomonas*, *Endotrypanum*, and other related kinetoplastid species. Branch lengths are proportional to evolutionary distance (scale bar = 0.10 substitutions per site). Species names follow standard nomenclature with strain identifiers where applicable.

Species within the *Trypanosoma* genus formed well-separated clades distinct from *Leishmania,* with clear separation between African and American trypanosome species. Other kinetoplastid genera, including *Crithidia*, *Leptomonas*, *Endotrypanum*, and various monoxenous species, occupied basal positions, consistent with their role as early-diverging lineages within the trypanosomatid phylogeny.

The genus *Leishmania* formed a well-supported monophyletic group that included the traditional *Leishmania* species as well as the more recently recognized subgenera. Within this broader *Leishmania* clade, we identified four distinct evolutionary lineages which are well defined in the Splitstree network (Figure 7.): the *Leishmania leishmania* group (highlighted in green), containing species such *as L. major, L. tropica, L. donovani* complex species, and *L. mexicana*; the *Sauroleishmania* subgenus (blue circle), represented by *L. tarentolae* strains; the *Mundinia* subgenus (purple circle), including *L. enrietti* and *L. martiniquensis*; and the *Viannia* subgenus (orange circle), comprising New World species such as *L. braziliensis, L. panamensis, L. guyanensis*, and *L. shawi*. It is worth noting that *L. adleri* exhibits an unexpected placement in our phylogenetic analysis, possibly due to poor sample quality or contamination. More striking is the discordant placement of the two *L. lainsoni* strains: while *L. lainsoni* M6426 placed within the *Viannia* subgenus as expected, *L. lainsoni* NMT was positioned outside *Viannia* entirely, clustering within the Leishmania leishmania group alongside *L. amazonensis* and *L. mexicana.* This intersubgeneric discordance between two strains of the same species suggests the NMT strain may harbour a composite genotype, potentially arising from contamination, mixed infection, or genuine interspecific genetic exchange. This pattern is reminiscent of the ancestry signals observed in the PM-

H197 hybrid strain investigated in section 3.5 and 3.6, though a full investigation was beyond the scope of this study. Future ADMIXTURE analysis or competitive read mapping of the NMT strain would help distinguish between these possibilities.

The SplitsTree network analysis (Figure 7.) provided additional insights into the evolutionary complexity of trypanosomatid relationships. While the network largely confirmed the major groupings observed in the phylogenetic tree, it also revealed evidence of conflicting phylogenetic signals, particularly at nodes connecting the major *Leishmania* subgenera. The reticulate patterns observed in the network suggest either ancient hybridization events, horizontal gene transfer, or incomplete lineage sorting during the rapid diversification of these parasite lineages.

Notably, the network analysis highlighted the distinctiveness of each major *Leishmania* subgenus, with *Sauroleishmania, Mundinia*, and *Viannia* each forming tight clusters with minimal internal reticulation. In contrast, the connections between these major groups showed more complex patterns, indicating uncertainty in the precise branching order of these ancient divergence events. The positioning of *Viannia* as a sister group to other *Leishmania* species was consistent across both analytical approaches, supporting its recognition as a distinct evolutionary lineage.

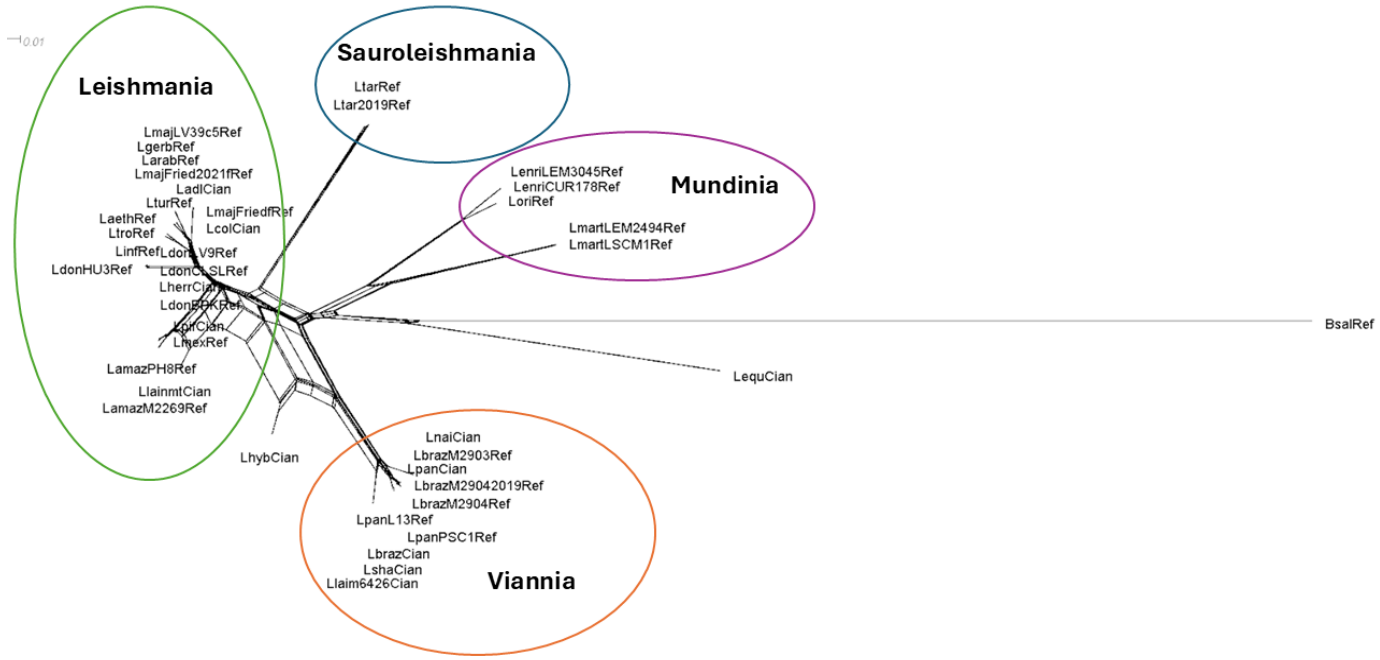


Figure 7. SplitsTree network analysis of *Leishmania* relationships. Neighbor-net network showing evolutionary relationships and potential conflicting phylogenetic signals among 29 *Leishmania* species and an outgroup (Bodo saltans). Major taxonomic groups are highlighted with colored circles: *Leishmania* (green), *Sauroleishmania* (blue), *Mundinia* (purple), and *Viannia* (orange). The network visualization reveals reticulate patterns indicative of complex evolutionary histories, including possible horizontal gene transfer events or incomplete lineage sorting. Branch lengths reflect evolutionary distances between taxa. Key for species names can be found in supplementary materials.

The concordance between the maximum likelihood tree and the network analysis for major taxonomic groupings, combined with the revelation of complex evolutionary patterns at deeper nodes, provides robust support for current taxonomic classifications while highlighting areas where additional molecular markers may be needed to fully resolve ancient evolutionary relationships within this medically and veterinary important group of parasites.

3.5. ADMIXTURE Analysis

To unfurl the ancestry of the hybrid, ADMIXTURE v1.3.0 was used to investigate genetic ancestry patterns across 63 trypanosomatid species, including 29 *Leishmania* species, 19 *Trypanosoma* species, and 15 other trypanosomatid taxa. The analysis utilized 21,530 fixed differences derived from the Orthofinder single copy gene outputs, through a VCF-to-PLINK conversion pipeline.

Cross-validation analysis was performed across K values ranging from K=2 to K=63 to determine the optimal number of ancestral populations. The cross-validation error curve revealed a clear minimum at K=6 (CV error = 0.44950), indicating this as the most statistically supported population structure model (Figure 8). Cross-validation errors for the initial range showed: K=2 (0.53378), K=3 (0.58314), K=4 (0.46266), K=5 (0.46935), K=6 (0.44950), K=7 (0.47005), K=8 (0.46222), K=9 (0.69341), and K=10 (0.60125).

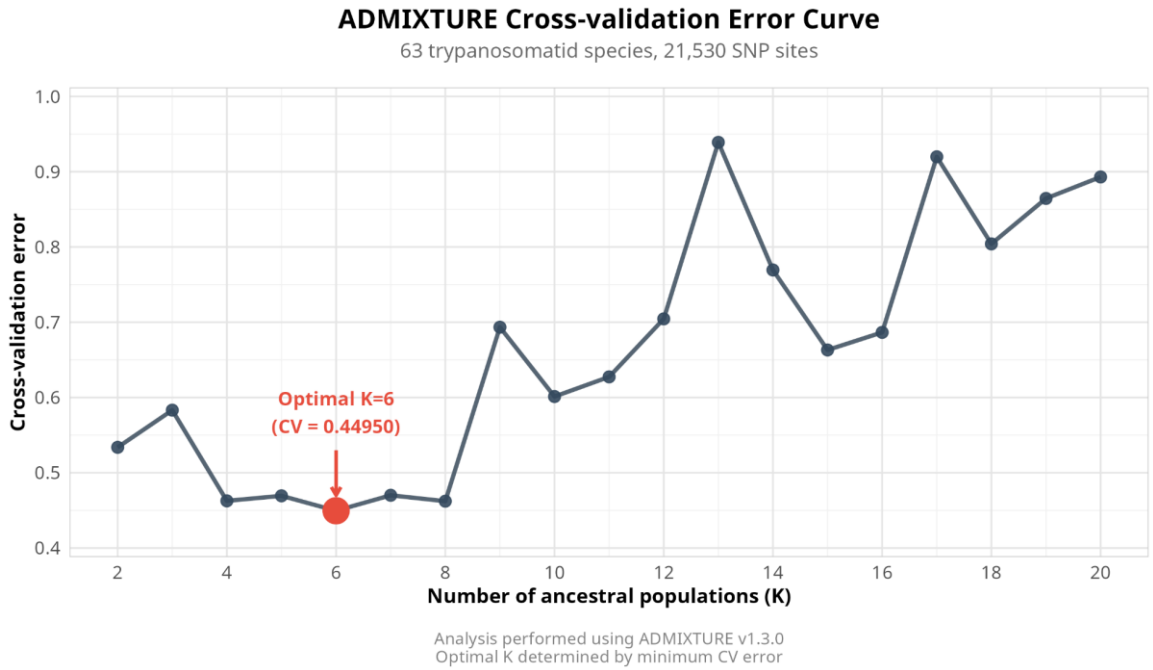


Figure 8.Figure 8. Cross-validation error analysis for ADMIXTURE population structure inference across 63 trypanosomatid species. The curve shows cross-validation (CV) error values for models tested across K=2 to K=20 ancestral populations, derived from 21,530 genome-wide SNP sites. Lower CV error indicates a better-supported model. A clear global minimum is observed

at K=6 (CV error = 0.44950), highlighted in red, indicating this as the optimal number of ancestral populations for this dataset. CV error rises sharply at K=7 and beyond up to K=10, before increasing further at higher K values, confirming that K=6 provides the best balance between model fit and complexity..

Extended analysis through K=20 confirmed that higher K values consistently produced elevated cross-validation errors, with dramatic increases observed at K=9 and beyond. Notable CV errors included K=13 (0.93903), K=17 (0.91981), and K=20 (0.89301), demonstrating substantial model overfitting at these higher values. Analysis was extended to K=63 to explore species-level resolution, revealing ultra-low CV errors at very high K values (K=47: 0.01694, K=52: 0.01532), suggesting that while individual species can be distinguished, these fine-scale subdivisions do not represent biologically meaningful population structure for comparative analyses.

The K=6 model represents the optimal balance between biological interpretability and statistical support, avoiding both underfitting at lower K values and overfitting at higher K values. This six-population structure likely corresponds to major evolutionary lineages within the trypanosomatid family, capturing fundamental taxonomic divisions while maintaining sufficient resolution for epidemiological and evolutionary analyses.

The selection of K=6 as optimal is supported by multiple criteria: the global minimum in cross-validation error, the sharp increase in CV error at K=7 and higher values, and biological plausibility given the known taxonomic structure of trypanosomatids. The six ancestral populations identified at K=6 provide a framework for understanding major evolutionary relationships while remaining practically interpretable for downstream analyses.

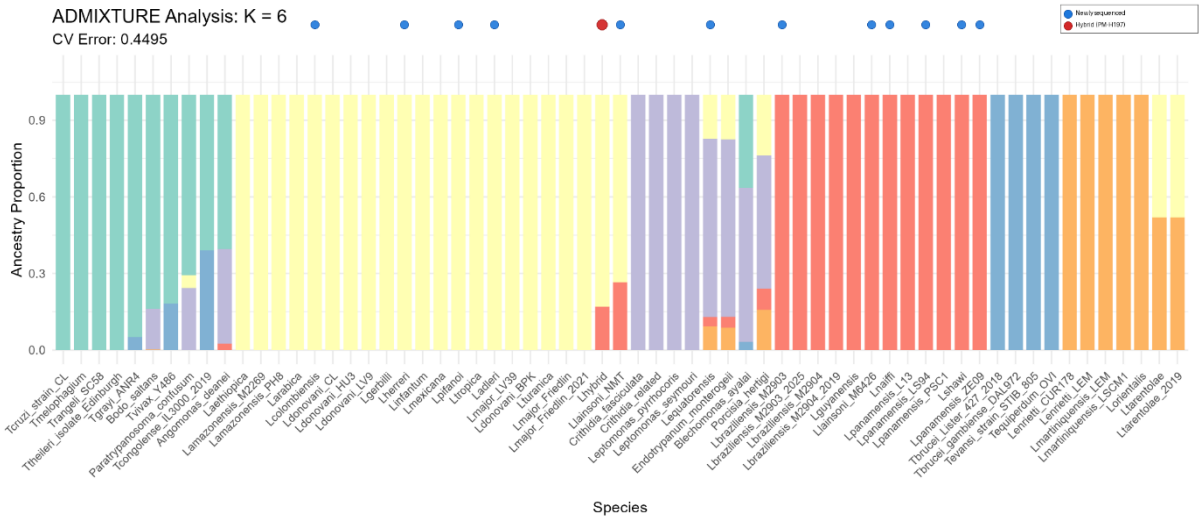


Figure 9. Admixture analysis of 63 trypanosomatid species based on variant call format (VCF) data with *K* = 6 species groups. Each vertical bar represents an individual strain or species, and colors indicate the proportion of inferred ancestry from each genetic cluster. The suspected hybrid isolate “Lhybrid” shows a mixed ancestry profile, consistent with contributions from multiple clusters. Species are ordered along the x-axis by taxonomic identity, with ancestry proportions plotted on the y-axis. Cross-validation (CV) error for *K* = 6 was 0.4495.

The ancestry of Lhybrid at k = 6 surmounts to approximately 0.25 *Viannia* and 0.75 *Leishmania leishmania* (Figure. 9).. As the K values increase, the ancestry clustering narrows to include approximately 80% *Leishmania* mexicana complex (consisting of *L. mexicana* and *L. amazonensis*) and 20% *Viannia*.

While K=6 provides optimal population structure for comparative analysis, exploration of higher K values revealed important insights into species-level genetic differentiation. The dramatic reduction in cross-validation errors at very high K values (K≥40) indicates that ADMIXTURE can achieve near-perfect species-level classification when sufficient ancestral populations are specified. However, these high-K models, while statistically powerful for species identification, do not provide meaningful insights into population structure patterns relevant for understanding evolutionary relationships or epidemiological dynamics.

The contrast between moderate K values (K=6-20) showing high CV errors and extreme K values (K≥40) showing ultra-low errors suggests two distinct analytical regimes: population structure analysis (optimal at K=6) and species classification (optimal at K≥40). This finding demonstrates the importance of selecting appropriate K values based on analytical objectives rather than solely minimizing cross-validation error.

The analysis included 12,694 missing data points out of 1,356,390 total genotype calls (0.94% missing data rate), indicating high-quality fixed differences across the diverse trypanosomatid dataset. This low missing data rate supports the reliability of ancestry proportion estimates and reduces potential bias in population structure inference.

3.6. ONT-based Hybrid vs Mixed Infection Analysis

3.6.1. Oxford Nanopore Technology Coverage Analysis Reveals Genomic Abundance Patterns Contradicting ADMIXTURE Ancestry Inference

Long-read sequencing analysis of strain MHOM/VE/1991/PM-H197 using Oxford Nanopore Technology (ONT) generated 7.5 GB of sequence data, revealing significant discrepancies with population genetic ancestry inference. While ADMIXTURE analysis using 21,530 orthologous gene variants indicated 80% *L. mexicana* complex ancestry and 20% *Viannia* ancestry, separate mapping of ONT reads to individual reference genomes demonstrated markedly different abundance patterns.

Coverage analysis revealed a consistent 2:1 abundance ratio favoring *Viannia* genomic content (Figure 10.). Mapping to the *Viannia* reference genome yielded 97.1x mean coverage across 54 chromosomes with remarkably uniform distribution (coefficient of variation [CV] = 0.474, mapping rate = 93.5%). In contrast, mapping to the *L. mexicana* reference produced 48.5x mean coverage across 300 chromosomes with highly variable distribution (CV = 3.292, mapping rate = 58.0%).

The observed coverage ratio of 0.5 (*L. mexicana*:*Viannia* = 1:2) directly contradicted ADMIXTURE predictions, which would predict approximately 4:1 *L. mexicana* dominance based on ancestry proportions.

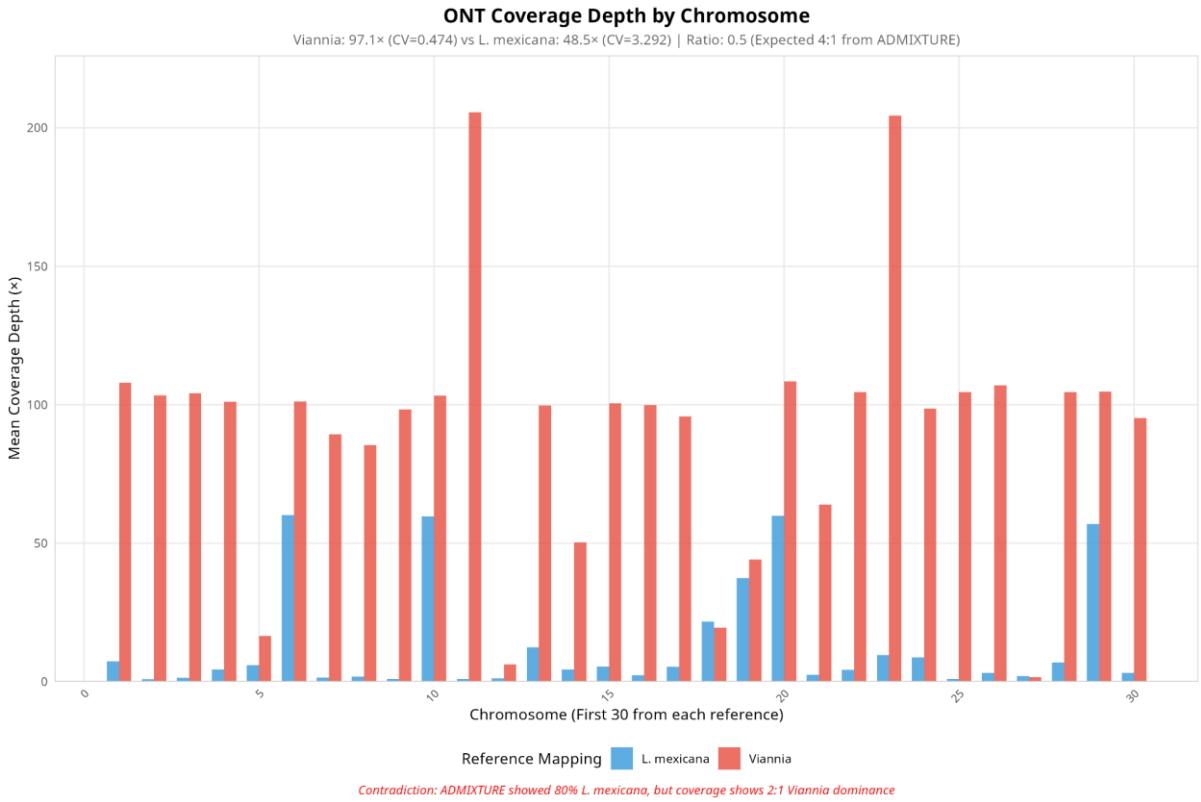


Figure 10. coverage depth comparison by chromosome. Bar chart showing mean coverage depth across the first 30 chromosomes from separate mapping of ONT reads to *L. mexicana* (blue) and *Viannia* (red) reference genomes. *Viannia* mapping demonstrates consistent high coverage (97.1× mean, CV = 0.474) while *L. mexicana* mapping shows variable coverage (48.5× mean, CV = 3.292). The observed 1:2 coverage ratio contradicts ADMIXTURE ancestry predictions of 4:1 *L. mexicana* dominance. Error bars represent standard deviation across chromosomal positions.

Coverage consistency analysis demonstrated fundamentally different mapping patterns between references (Figure 11.). *Viannia* mapping exhibited low inter-chromosomal variability (CV = 0.474), consistent with stable genomic representation. Conversely, *L. mexicana* mapping showed extreme variability (CV = 3.292), with coverage ranging from near-zero to >100x across different chromosomes. This 7-fold

difference in coefficient of variation suggested either highly unstable genomic architecture in hybrid regions or the presence of independent parasite populations.

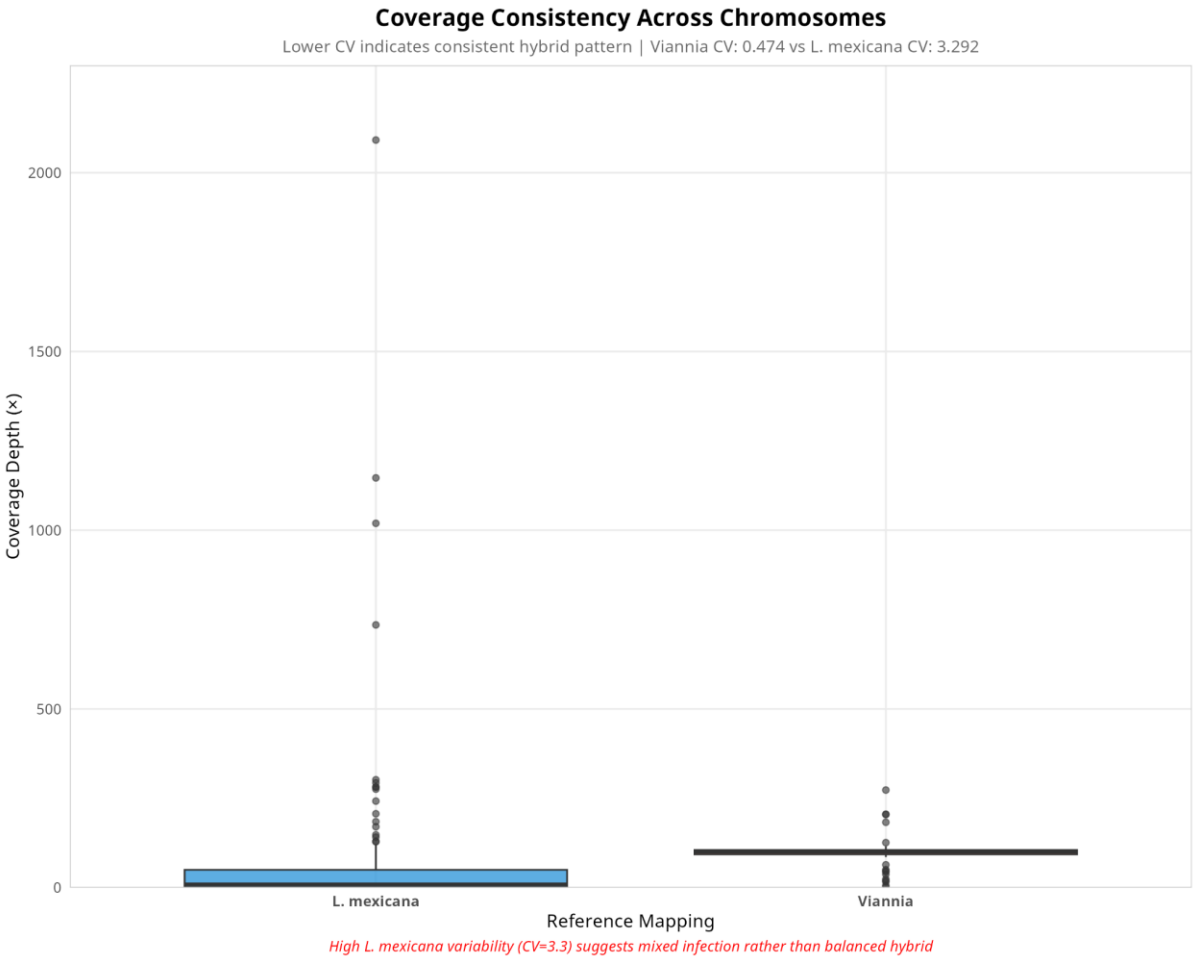


Figure 11. Coverage consistency analysis across chromosomes. Boxplot comparing coverage depth distributions between *Viannia* and *L. mexicana* reference mappings. *Viannia* mapping exhibits low inter-chromosomal variability (CV = 0.474), indicating stable genomic representation, while *L. mexicana* mapping shows extreme variability (CV = 3.292) with coverage ranging from near-zero to >100x across different chromosomes. The 7-fold difference in coefficient of variation suggests either genomic instability in hybrid regions or independent parasite populations. Box boundaries represent 25th and 75th percentiles, center lines indicate median values, and whiskers extend to 1.5x interquartile range.

The 8-fold discrepancy between ADMIXTURE ancestry predictions and observed genomic abundance prompted investigation using competitive mapping approaches

to distinguish between true hybridization with unequal genomic contributions versus mixed infection scenarios.

3.6.2. Ultra-Long Read Competitive Mapping Provides High-Confidence Species Assignment

To resolve the ADMIXTURE-coverage contradiction, competitive mapping was implemented using ultra-long reads (≥100 kb) for maximum mapping specificity and unbiased species discrimination. Ultra-long reads were extracted from the ONT dataset and competitively mapped to a concatenated reference genome containing tagged *L. mexicana* (mexicana_) and *Viannia* (*viannia*_) sequences, allowing reads to select optimal mapping targets without reference bias.

Competitive mapping analysis demonstrated extreme species preference patterns (Figure 12.). Ultra-long reads showed overwhelming preference for *Viannia* reference sequences, with approximately 96% of reads ≥100 kb mapping preferentially to *viannia*-tagged contigs and only 4% selecting mexicana-tagged sequences. This 24:1 preference ratio provided the highest confidence evidence for genomic composition, as ultra-long reads minimize mapping ambiguity and can unambiguously span repetitive genomic regions that confound shorter read analyses.

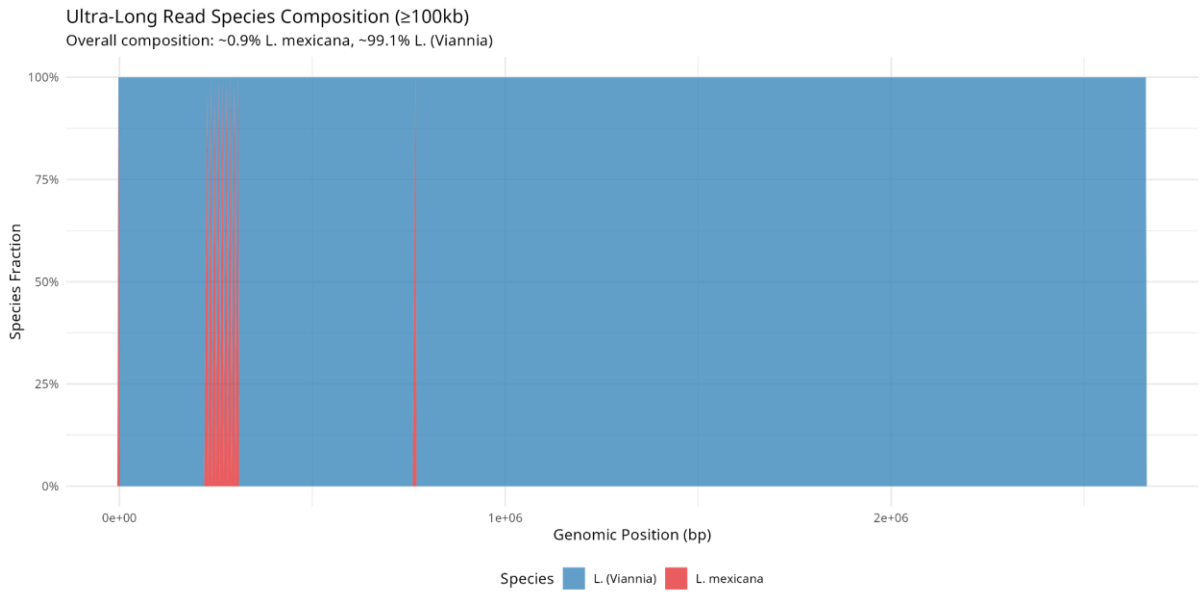


Figure 12. Ultra-long read competitive mapping demonstrates extreme *Viannia* preference. Coverage fraction analysis of ultra-long reads (≥100 kb) competitively mapped to concatenated reference genome containing tagged *L. mexicana* (mexicana_) and *Viannia* (*viannia*_) sequences. Line plots show coverage fractions across genomic positions for the top three contigs by total coverage. *Viannia* sequences (red line) consistently dominate coverage with ~96% of ultra-long reads showing preferential mapping, while *L. mexicana* sequences (blue line) attract only ~4% of reads. The 24:1 preference ratio provides high-confidence evidence for genomic composition, as ultra-long reads minimize mapping ambiguity and can unambiguously span repetitive regions. Analysis performed using 10 kb genomic bins to reduce noise while maintaining positional resolution.

The competitive approach eliminated potential reference genome quality bias by providing equivalent access to both species' sequences, ensuring that mapping preferences reflected genuine sequence similarity rather than technical artifacts. The extreme bias toward *Viannia* sequences persisted across multiple read length thresholds, indicating that the observed pattern was not dependent on specific read length cutoffs.

3.6.3. Reconciling ADMIXTURE Ancestry with Genomic Abundance Patterns

The integration of ADMIXTURE ancestry inference, ONT coverage analysis, and ultra-long read competitive mapping revealed a complex genomic architecture that resists straightforward interpretation. Three distinct analytical approaches yielded conflicting results: ADMIXTURE indicated 80% L. mexicana ancestry, ONT coverage demonstrated a consistent 2:1 Viannia dominance (~67% Viannia), and competitive mapping showed 96% Viannia preference. These conflicting signals necessitate careful evaluation of two alternative hypotheses.

In a mixed infection scenario, the sample contains two genetically distinct and co-existing parasite populations, a dominant Viannia population and a minor L. mexicana population. The competitive mapping result (96% Viannia, 4% L. mexicana) could superficially support this, as it suggests the vast majority of sequencing reads derive from Viannia parasites. However, this hypothesis faces two serious contradictions. First, if genuine L. mexicana parasites comprised only 4% of the sample, they would be present in very low abundance and would not be expected to generate the strong and consistent ADMIXTURE ancestry signal of 80% L. mexicana observed across 21,530 orthologous loci, a signal of that magnitude requires a substantial and uniform genomic contribution, not a minor contaminant population. Second, a mixed infection of two independent parasite populations would produce highly variable and uneven coverage patterns as reads from each population map inconsistently across chromosomes. Instead, ONT coverage analysis revealed a remarkably uniform 2:1 Viannia:L. mexicana ratio across the entire genome (coefficient of variation = 0.474), a level of consistency incompatible with stochastic co-infection dynamics.

In an unbalanced hybrid hypothesis, PM-H197 is a single organism carrying a stably integrated genome derived asymmetrically from both parental lineages, approximately 80% Viannia and 20% L. mexicana. This interpretation accounts for all three analytical results more coherently. The strong genome-wide ADMIXTURE signal of 80% L. mexicana ancestry is consistent with fixed genetic differences stably integrated across thousands of loci, as expected in a true hybrid rather than a contaminating population. The uniform 2:1 coverage ratio across all chromosomes is

likewise consistent with a stably integrated genome present in every cell, rather than a mixed population producing variable read depths. The apparent discrepancy with competitive mapping (which shows 96% Viannia rather than the ~ 67% predicted by coverage) may reflect a methodological sensitivity difference: ultra-long reads are assigned to one species based on overall sequence similarity across their entire length, and may therefore under-represent the L. mexicana component if that ancestry exists as smaller introgressed segments distributed across an otherwise Viannia genomic background.

Taken together, the weight of evidence favours the unbalanced hybrid interpretation. The genome-wide uniformity of both the coverage ratio and the ADMIXTURE ancestry signal is most parsimoniously explained by stable genomic integration rather than co-infection, and the mixed infection hypothesis cannot adequately account for the strength of the L. mexicana ADMIXTURE signal given its apparent rarity in competitive mapping. Nevertheless, definitive resolution of these hypotheses would require single-cell sequencing to determine whether L. mexicana genetic material is present within individual cells or confined to a minority parasite sub-population, and kDNA haplotyping to establish the maternal parent of the putative cross.

Notably, the geographic origin of strain PM-H197 is consistent with the inferred hybrid ancestry. The strain was isolated in Venezuela in 1991 (WHO ID: MHOM/VE/1991/PM-H197), a region where both Viannia species and L. mexicana complex species are known to co-circulate. The geographic co-occurrence of the putative parental lineages in Venezuela supports the biological plausibility of a natural hybridisation event, as physical sympatry is a prerequisite for interspecific genetic exchange. This study further demonstrates that integrating multiple analytical approaches is essential for accurate characterisation of complex Leishmania infections, as different methodologies may yield conflicting results that require careful interpretation to distinguish between hybrid and mixed infection scenarios.

4. Discussion

4.1. Genome assembly

The comprehensive genome assembly project presented here represents a significant advancement in trypanosomatid genomics, delivering chromosome-level assemblies for multiple medically important *Leishmania* (*Viannia*) species that substantially exceed the quality of existing reference genomes. Through the integration of Oxford Nanopore long-read sequencing with advanced basecalling algorithms and multi-stage polishing protocols, we have established new benchmarks for parasitic genome assembly quality while providing unprecedented genomic resources for understanding the evolution, pathogenesis, and epidemiology of leishmaniasis causative agents.

Our assembly pipeline incorporated several critical technological advances that collectively enabled the production of superior genomic resources. The implementation of Dorado duplex basecalling using the dna_r10.4.1_e8.2_400bps_sup@v4.1.0 model represented a substantial improvement over conventional basecalling approaches, providing enhanced accuracy particularly in the repetitive and GC-rich regions characteristic of trypanosomatid genomes.

The multi-stage polishing strategy, incorporating NECAT for initial assembly correction, Medaka for long-read consensus polishing, Racon for iterative improvement, and Pilon for short-read validation, addressed the historical challenges associated with long-read assembly accuracy. This approach proved particularly effective for resolving the complex repetitive elements and telomeric regions that have traditionally been assembly gaps in *Leishmania* genomes (González-de la Fuente et al., 2017; Pita et al., 2019; Lypaczewski et al., 2018). The integration of Purge Haplotigs for redundancy removal further refined assembly quality by eliminating spurious duplications that commonly arise in long-read assemblies of heterozygous regions (Roach, Schmidt and Borneman, 2018). The systematic

evaluation of assembly quality at each pipeline stage through QUAST metrics (Gurevich et al., 2013) enabled optimization of processing parameters and identification of assembly steps providing maximal improvement. This iterative refinement approach represents a reproducible framework applicable to other trypanosomatid genome projects.

The comprehensive quality assessment conducted across multiple metrics consistently demonstrated the superiority of our assemblies compared to existing *Leishmania* reference genomes. In the majority of cases, N50 values for our *Viannia* assemblies substantially exceeded those of published genomes, indicating improved contiguity and chromosomal integrity. BUSCO completeness scores approached theoretical maxima, suggesting near-complete gene space recovery across all assembled species. In certain assemblies, the successful recovery of telomeric sequences and chromosome-level organization represented a marked improvement over fragmented assemblies typical of short-read approaches.

The systematic comparison of assembly metrics demonstrates the substantial improvements achieved through modern long-read sequencing approaches. Our assemblies consistently achieved N50 values comparable to or exceeding published reference genomes, with several species showing marked improvements in chromosomal contiguity. Notably, *L. naiffi* showed dramatic improvement from a highly fragmented reference assembly (1,749 contigs, N50 = 42 kb) to a chromosome-level assembly (68 contigs, N50 = 868 kb), representing a 20-fold improvement in contiguity.

Similarly, *L. shawi* assemblies improved from extremely fragmented references (1,543 contigs, N50 = 44 kb) to substantially more contiguous assemblies (55 contigs, N50 = 913 kb). These improvements reflect the power of long-read sequencing to span repetitive regions and resolve complex genomic architectures that have historically challenged short-read assembly approaches (De Maio et al., 2019; Logsdon, Vollger and Eichler, 2020).

The variation in assembly quality across species, particularly evident in the *L. panamensis* assemblies, highlights the importance of sequencing depth and read

quality in achieving optimal results. The LS94 strain, with lower ONT coverage (56x), produced a more fragmented assembly (137 contigs, N50 = 563 kb) compared to the ZE09 strain with higher coverage (168x), which achieved superior contiguity (38 contigs, N50 = 1,144 kb). This coverage-dependent assembly quality reinforces established principles in long-read genome assembly while providing practical guidance for future *Leishmania* genome projects (Venter, Smith and Adams, 2015; Sharma et al., 2021).

The BUSCO assessments provide complementary evidence for assembly completeness beyond simple contiguity metrics. While the eukaryotic BUSCO scores (~50% completeness) might initially appear concerning, this reflects the taxonomic distance between *Leishmania* and the broader eukaryotic gene set, which includes highly divergent lineages with different gene repertoires (Marques et al., 2015; Black et al., 2023). The substantially higher euglenozoan BUSCO scores (>90% completeness) provide a more relevant assessment of gene space recovery, confirming that our assemblies successfully capture the core genetic machinery essential for trypanosomatid biology.

The consistently high euglenozoan BUSCO completeness across all assemblies indicates that the genome assembly pipeline successfully recovered conserved metabolic pathways, cellular machinery, and regulatory elements critical for parasite survival and pathogenesis. This level of gene space completeness enables confident downstream analyses including comparative genomics, phylogenetic reconstruction, and functional annotation.

When contextualized against existing *Leishmania* reference genomes, our assemblies demonstrate that modern long-read approaches can substantially improve upon previous assembly efforts, particularly for understudied species lacking high-quality references. The chromosome-level assemblies generated for species such as *L. colombiensis, L. equatorensis*, and *L. herreri* represent the first high-quality genomic resources for these medically important taxa.

The improved genomic resources provide essential foundations for understanding the *Viannia* subgenus diversity and evolution. Complete chromosomal assemblies

enable accurate assessment of gene synteny, identification of species-specific genomic regions, and characterization of chromosomal rearrangements that may contribute to virulence differences among *Viannia* species (Liu, Hunt and Tsai, 2018; O'Connor et al., 2018; Rhie et al., 2021).

The establishment of these genomic resources for the complete *Viannia* subgenus enables systematic comparative analyses that were previously impossible due to incomplete or fragmented reference sequences. This capability is particularly important for understanding the evolutionary dynamics of medically important *Viannia* species and their adaptation to different transmission cycles and geographic environments.

While our assemblies represent substantial improvements over existing resources, several limitations should be acknowledged. Assembly gaps may persist in highly repetitive regions or regions with complex secondary structures. Additionally, the haploid assembly approach does not capture heterozygous variation that may be epidemiologically or functionally significant. Future assembly efforts would benefit from integration with chromosome conformation capture (Hi-C) technologies to validate chromosomal organization and resolve any remaining structural uncertainties. The development of diploid-aware assembly approaches could provide insights into heterozygous variation patterns within *Leishmania* populations.

4.2. Telomere analysis

The comprehensive analysis of telomere-associated sequences across our *Leishmania* (*Viannia*) genome assemblies provides critical insights into assembly completeness and chromosomal integrity. By systematically screening for canonical telomeric repeats and Leishmania-specific telomere-associated sequences, we have established a robust framework for evaluating the quality of long-read genome assemblies in trypanosomatid parasites.

Our analysis targeted the telomere-associated CCCTAACCCTAACCCTAA repeat. The sequence 5′-CCCTAA-3′ is the main telomeric repeat in *Leishmania,* identical to the telomeric repeats found in related protozoa like Trypanosoma brucei. It is organized in tandem arrays at chromosome ends and hybridizes to all *Leishmania* chromosomes, indicating its universal presence across the genus (Ellis and Crampton, 1988).While CCCTAA is the standard repeat, some *Leishmania* species, such as L. braziliensis, can have additional or slightly modified telomeric repeats (e.g., CCCTAACCCGTGGA), but CCCTAA remains a core component (Fu and Barker, 1998).The telomeric repeat CCCTAA is highly conserved and is associated with telomere-binding proteins and telomerase, which are essential for chromosome end protection and genome stability (Paiva et al., 2025). The presence of this sequence at contig termini serves as a powerful indicator of chromosome-level assembly completeness, as genuine telomere-to-telomere assemblies should retain these critical structural elements.

The distribution of telomere-capped versus uncapped contigs across our *Viannia* assemblies reveals significant variation in assembly completeness (Table 4.). Species showing high proportions of telomere-capped chromosomes indicate successful resolution of chromosomal termini, a historically challenging aspect of genome assembly that has been substantially improved by long-read sequencing technologies (Logsdon, Vollger and Eichler, 2020; Lang et al., 2020; van Dijk et al., 2023). Conversely, assemblies with predominantly uncapped contigs may reflect either incomplete assembly processes or genuine biological variation in telomere structure maintenance.

The differential distribution of telomere-associated sequences across *Viannia* species provides novel insights into chromosomal evolution within this medically important subgenus. The presence of the *L. braziliensis*-specific repeat CCCTAACCCGTGGA (Fu and Barker, 1998) in our analysis extends previous observations of species-specific telomere sequence divergence, suggesting that telomere architecture may represent a rapidly evolving genomic feature with potential implications for chromosome stability and species adaptation.

Our chromosome capping analysis demonstrates the substantial improvements achieved through modern long-read assembly approaches compared to previous short-read methodologies. The recovery of telomere-associated sequences in multiple *Viannia* assemblies represents a significant technical achievement, as these repetitive, GC-rich regions have historically been assembly gaps in trypanosomatid genomes. Many older assemblies use short read sequencing, and short reads often cannot span these repeats, leading to fragmented assemblies and missing telomeric regions (Pita et al., 2019; Díaz-Viraqué et al., 2021; Albanaz et al., 2023).

The variation in telomere recovery across our assemblies likely reflects differences in sequencing depth, read length distribution, and assembly algorithm performance rather than genuine biological absence. This technical heterogeneity provides valuable benchmarks for optimizing future assembly protocols and establishes minimum quality thresholds for chromosome-level genome completion.

When contextualized against existing *Leishmania* reference genomes, our telomere analysis reveals that several *Viannia* assemblies achieve or exceed the chromosomal completeness of established reference sequences. This finding has significant implications for comparative genomics studies, as it suggests that species-specific reference genomes may provide more accurate foundations for evolutionary and functional analyses than cross-species mapping approaches.

The establishment of telomere-based assembly quality metrics fills a critical gap in current genome assessment protocols. While traditional metrics such as N50, BUSCO completeness, and contig number provide valuable assembly statistics, they do not directly assess chromosomal integrity. Our approach complements these measures by providing direct evidence of chromosome-level completion.

Furthermore, the chromosome-level assemblies validated through telomere analysis provide essential foundations for population genomic studies, enabling accurate variant calling and structural variation detection that require properly assembled chromosomal contexts (Huddleston et al., 2017; Human Genome Structural Variation Consortium et al., 2021). This capability is particularly important for understanding the evolutionary dynamics of medically important *Viannia* species and their

adaptation to different transmission cycles and host environments (Coughlan et al., 2018; Patino et al., 2020; De Los Santos et al., 2022).

Our integrated approach combining fuzznuc pattern recognition, BED file processing, and statistical visualization represents a reproducible framework applicable to other trypanosomatid genome projects. The automation of telomere detection and assembly quality assessment addresses a significant bottleneck in genome annotation pipelines, enabling systematic evaluation of large-scale genome sequencing projects.

While our telomere analysis provides robust evidence for assembly completeness, several limitations should be acknowledged. The presence of telomere-associated sequences at contig ends does not guarantee complete chromosomal assembly, as internal gaps or rearrangements may still exist. Additionally, the biological significance of telomere sequence variation remains to be fully characterized through functional studies.

4.3. Orthofinder

4.3.1. Pan-Genome Architecture and Evolutionary Dynamics

The identification of 16,683 orthogroups across the 63 analyzed species reveals a remarkably dynamic and plastic genome architecture characteristic of rapidly evolving parasitic organisms. The distribution of these orthogroups across functional categories provides critical insights into the evolutionary forces shaping trypanosomatid genomes and their adaptation to diverse ecological niches and host environments.

The severely limited core genome, comprising only 844 orthogroups (5.06% of total orthogroups) present in all species, represents one of the most striking findings of

this analysis. This extraordinarily small core genome proportion is substantially lower than observed in most bacterial taxa, where core genomes typically represent 20-40% of total gene content (Vernikos et al., 2015; McInerney, McNally and O'Connell, 2017), and even lower than many eukaryotic parasite groups (Croll and McDonald, 2017; Silva et al., 2018). The minimal core genome in trypanosomatids likely reflects the ancient diversification of this family, estimated to have occurred over 100 million years ago (Lukeš et al., 2018), combined with the extreme selective pressures imposed by diverse host environments, transmission strategies, and ecological adaptations.

The 844 core genes represent the essential genetic machinery required for fundamental cellular processes shared across all trypanosomatid lineages. These likely include core metabolic pathways, basic cellular machinery for DNA replication, transcription, translation, and essential structural components. The conservation of these genes across such divergent taxa underscores their critical importance for parasite survival and suggests that these represent non-dispensable genetic elements that cannot be lost without compromising cellular viability (Morris et al., 2011; Wolf and Koonin, 2013).

In stark contrast to the minimal core genome, the accessory gene pool comprises 8,969 orthogroups (53.76% of total), representing the largest functional category identified in this analysis. This massive accessory genome reflects the extraordinary genomic plasticity that has enabled trypanosomatids to colonize diverse ecological niches, adapt to different host species, and evolve distinct pathogenic strategies. The accessory genome likely contains genes involved in host-specific adaptations, virulence factors, immune evasion mechanisms, and metabolic adaptations required for survival in different tissue environments (Shapiro and Polz, 2014; Brockhurst et al., 2019).

The substantial representation of unique orthogroups (3,355 orthogroups, 20.11% of total) indicates extensive lineage-specific gene evolution within the Trypanosomatidae. This pattern suggests that species-specific adaptations have played a major role in trypanosomatid diversification, with individual lineages acquiring novel genetic capabilities through gene duplication, horizontal gene

transfer, or de novo gene evolution (Andersson, 2005; Morris, Lenski and Zinser, 2012; Van Oss and Carvunis, 2019). The prevalence of species-specific genes may partially explain the diverse clinical manifestations, host preferences, and transmission strategies observed across different trypanosomatid species.

The soft-core orthogroups, present in most but not all species (3,515 orthogroups, 21.07%), represent an intermediate category that may include genes that are functionally important but not absolutely essential, or genes that have been lost in specific lineages under particular selective pressures. This category likely contains genes involved in pathogenesis, host interaction, and environmental adaptation that provide selective advantages in specific contexts but are not universally required (Wolf and Koonin, 2013; Cosentino Lagomarsino et al., 2015).

4.3.2. Genome Size Variation and Evolutionary Implications

The dramatic variation in genome size across trypanosomatid species, spanning a 4.41-fold range from 7,151 genes in *L. guyanensis* to 31,539 genes in *T. cruzi*, represents one of the most extreme examples of genome size plasticity documented in any parasitic organism family. This variation far exceeds that typically observed within closely related bacterial genera (1.5-3 fold variation) and rivals the genome size variation seen across entire bacterial phyla (Mira, Ochman and Moran, 2001; Lynch, 2006; Kuo and Ochman, 2009).

The extraordinary genome expansion observed in *T. cruzi*, containing 3.57 times more genes than the average *Leishmania* species, likely reflects the complex life cycle and diverse host interactions characteristic of this species. *T. cruzi* alternates between invertebrate vectors (triatomine bugs) and mammalian hosts, encounters diverse tissue environments during infection, and must adapt to highly variable nutritional and physiological conditions (Tyler and Engman, 2001; Docampo and Moreno, 2003). This ecological complexity may have driven the retention and expansion of gene families involved in metabolic flexibility, stress response, immune evasion, and tissue-specific adaptations.

The relatively compact genomes characteristic of most *Leishmania* species (average ~8,400 genes) may reflect their more specialized ecological niches and streamlined life cycles compared to *T. cruzi*. *Leishmania* parasites typically alternate between sand fly vectors and mammalian hosts, with more restricted tissue tropisms and less complex developmental programs (Kamhawi, 2006; Bates, 2007). This ecological specialization may have permitted genome reduction through the loss of genes required for survival in diverse environments, following patterns observed in other obligate parasites (Mira, Ochman and Moran, 2001; McCutcheon and Moran, 2011).

Within the *Leishmania* genus, the 30% variation in genome size between the largest (*L. panamensis* LS94: 9,282 genes) and smallest (*L. guyanensis*: 7,151 genes) genomes suggests significant evolutionary pressures affecting gene content even within closely related species. *L. guyanensis*, possessing the smallest genome in the *Viannia* subgenus, may represent a lineage that has undergone particular genome streamlining, possibly reflecting adaptation to a more restricted ecological niche or simplified life cycle requirements (Wolf and Koonin, 2013; Morris, Lenski and Zinser, 2012).

The intermediate genome sizes observed in most *Viannia* species (8,400-9,000 genes) suggest a relatively stable genome architecture within this subgenus, despite the species' diverse geographic distributions across the Americas and clinical manifestations. This genomic stability may reflect shared ecological constraints within the Neotropical transmission cycles characteristic of *Viannia* parasites, or may indicate that the major diversification within this subgenus has occurred through gene regulation, copy number variation, and chromosomal rearrangements rather than wholesale gene gain or loss (Rogers et al., 2011; Franssen et al., 2020).

4.3.3. Phylogenomic Constraints and Methodological Implications

The identification of only 76 single-copy orthogroups (0.46% of total orthogroups) suitable for phylogenetic reconstruction represents a severe constraint for molecular

phylogenetic studies of trypanosomatid relationships. This extremely low proportion of single-copy genes reflects the extensive gene duplication, copy number variation, and genomic rearrangements that characterize trypanosomatid evolution (El-Sayed et al., 2005; Peacock et al., 2007; Rogers et al., 2011).

The prevalence of multi-copy gene families (16,607 orthogroups, 99.54% of total) indicates that gene duplication has been a major evolutionary force shaping trypanosomatid genomes. Extensive gene duplication can provide raw material for evolutionary innovation, enable rapid adaptation to new environments, and facilitate the evolution of novel functions through subfunctionalization and neofunctionalization (Ohno, 2013; Lynch and Conery, 2000; Kondrashov et al., 2002). However, the prevalence of multi-copy genes also complicates phylogenetic inference, as gene trees may not reflect species trees due to differential gene loss, convergent evolution, and horizontal gene transfer (Maddison, 1997; Degnan and Rosenberg, 2009).

The limited number of single-copy orthogroups available for phylogenetic analysis necessitates careful evaluation of phylogenetic signal and may require novel analytical approaches to extract maximum information from available data. Concatenated analyses using the 76 single-copy genes may provide sufficient phylogenetic signal for resolving major taxonomic relationships, but may lack the resolution needed for addressing fine-scale evolutionary questions or recent diversification events (Philippe et al., 2011; Kocot et al., 2013).

The extensive gene duplication observed in trypanosomatids may reflect adaptive responses to the complex selective pressures imposed by alternating between different host environments. Gene families involved in surface antigens, proteases, transporters, and metabolic enzymes are known to be extensively duplicated in trypanosomatids, potentially facilitating rapid adaptation to changing host environments and immune pressures(Barry et al., 2003; Jackson, 2010; Clayton, 2016).

4.3.4. *Viannia* Subgenus Genomic Characteristics

The detailed analysis of *Viannia* species reveals important patterns of genomic diversity within this medically important subgenus. The 17 *Viannia* species analyzed represent the most comprehensive genomic sampling of this group to date, enabling the first systematic comparative analysis of genome evolution within the subgenus responsible for mucocutaneous and cutaneous leishmaniasis in the Americas.

The gene content variation within *Viannia* species (7,151-9,282 genes, 30% range) is substantial but lower than the variation observed across the entire *Leishmania* genus, suggesting that *Viannia* represents a relatively cohesive evolutionary unit with shared genomic constraints. The intermediate position of most *Viannia* species in terms of genome size (8,400-9,000 genes) may reflect adaptation to the particular ecological conditions and transmission cycles characteristic of Neotropical environments. This cohesion is particularly apparent when *Viannia* is considered in the broader trypanosomatid context: while *Leishmania* species as a whole show remarkable genome size conservation (coefficient of variation ~15%), *Trypanosoma* species exhibit far greater variation, especially due to the massive genome expansion in *T. cruzi*. The strong selective constraints maintaining genomic architecture within *Leishmania* likely reflect shared ecological pressures or fundamental biological constraints absent in the more ecologically diverse *Trypanosoma*.

L. guyanensis, with the smallest genome in the *Viannia* subgenus (7,151 genes), represents an interesting case of potential genome reduction that warrants further investigation. This species, which causes cutaneous leishmaniasis in the Amazon basin, may have undergone genome streamlining in response to adaptation to particular host species, vector requirements, or ecological constraints (Silveira et al., 2010; Shaw, 2007). The reduced genome size could reflect loss of genes required for survival in diverse environments, consistent with patterns observed in other specialized parasites. This is consistent with broader patterns of parasitic genome evolution: the free-living trypanosomatid *Bodo saltans* possesses 13,254 genes, retaining metabolic pathways and cellular machinery lost in obligate parasites,

illustrating the scale of genomic streamlining that accompanies the transition to parasitism.

The largest *Viannia* genomes (*L. panamensis* LS94: 9,282 genes; *L. braziliensis*: 9,136 genes) may retain expanded gene repertoires that facilitate survival in diverse ecological contexts or enable the complex pathogenic processes characteristic of these species. *L. braziliensis*, in particular, is notable for causing severe mucocutaneous disease and possessing unique genomic features including RNA interference machinery and retrotransposable elements not found in other *Leishmania* species (Peacock et al., 2007; Jensen, Englund and Parsons, 2009).

The presence of hybrid genomes within the *Viannia* dataset (8,538 genes) provides evidence for ongoing genetic exchange within the subgenus and highlights the importance of recombination in shaping genomic diversity. Natural hybridization events between *Viannia* species have been documented in field populations and may contribute to the emergence of novel pathogenic traits and adaptation to new environments (Van den Broeck et al., 2020, 2023).

4.3.5. Implications for Understanding Leishmaniasis Pathogenesis and Evolution

The genomic diversity revealed by this comparative analysis has important implications for understanding the evolution of pathogenicity and clinical diversity within the *Leishmania* genus. The substantial accessory genome (53.76% of orthogroups) likely contains many genes involved in host-pathogen interactions, virulence, and immune evasion that contribute to the diverse clinical manifestations of leishmaniasis.

Species-specific genes (20.11% of orthogroups) may include virulence factors that determine tissue tropism, disease severity, and clinical outcomes. For example, the unique genomic features of *L. braziliensis,* including expanded gene families and

retrotransposable elements, may contribute to its ability to cause severe mucocutaneous disease (Peacock et al., 2007). Similarly, species-specific adaptations in other *Viannia* species may explain their distinct geographic distributions, vector associations, and clinical presentations.

The limited core genome suggests that fundamental aspects of *Leishmania* biology are controlled by a relatively small set of essential genes, while the diverse clinical and epidemiological characteristics of different species result from the variable accessory genome. This genomic architecture may facilitate rapid adaptation to new environments and hosts while maintaining the core cellular machinery required for parasitic survival.

4.3.6. Technical Considerations and Future Directions

The successful analysis of 63 trypanosomatid genomes using OrthoFinder demonstrates the power of modern comparative genomic approaches for understanding evolutionary relationships in complex parasitic organisms. However, several technical considerations should be acknowledged that may influence the interpretation of results.

Assembly quality variation across species may affect gene content estimates and orthology assignments. While the assemblies generated in this study represent substantial improvements over existing references for many *Viannia* species, remaining assembly gaps or fragmentation could lead to underestimation of gene content in some species. The use of chromosome-level assemblies generated through long-read sequencing provides increased confidence in gene content estimates compared to fragmented short-read assemblies.

Gene annotation quality represents another potential source of variation in comparative analyses. Differences in annotation pipelines, reference databases, and prediction algorithms can lead to variation in predicted gene content that may not

reflect genuine biological differences. The use of consistent annotation approaches across all species in this analysis helps minimize such technical artifacts.

The identification of orthogroups relies on sequence similarity and may be influenced by rapid evolution of particular gene families or horizontal gene transfer events. Genes that have undergone rapid evolutionary change may be incorrectly classified as species-specific rather than homologous, potentially inflating estimates of genomic diversity. Conversely, genes acquired through horizontal transfer may be incorrectly grouped with non-homologous genes, affecting orthology assignments.

Future comparative genomic studies would benefit from several methodological advances. The integration of synteny information could improve orthology inference and provide insights into chromosomal rearrangements and genome evolution. Functional enrichment analysis of OrthoFinder orthogroups, including GO term annotation of expanded or contracted gene families across Viannia and broader Leishmania taxa, was not performed in the present study. Such analyses were considered beyond the scope of this Masters thesis, which prioritised the generation and validation of high-quality genome assemblies and the characterisation of phylogenetic relationships and population structure. Nevertheless, functional enrichment approaches represent a valuable avenue for future work. By integrating GO term annotations, available for Leishmania through resources such as TriTrypDB, with the orthogroup membership data produced by OrthoFinder, it would be possible to identify biological processes and molecular functions disproportionately represented in lineage-specific or expanded gene families, providing insight into adaptive evolution across the Viannia subgenus. This type of analysis has recently been applied in a comparable study of Leishmania (Viannia) comparative genomics by (Gonzalez-Garcia et al., 2025), who identified consistent copy number variation between species for different gene families, including notably larger and more diverse amastin gene families than previously reported, demonstrating the potential of such approaches to uncover biologically meaningful patterns of gene family evolution in this group.Population genomic approaches incorporating multiple strains per species could distinguish between species-specific adaptations and intraspecific variation.

4.3.7. Evolutionary and Medical Significance

The patterns of genomic diversity revealed by this analysis have profound implications for understanding both the evolutionary history of trypanosomatids and the medical significance of leishmaniasis. The extensive accessory genome and high proportion of species-specific genes suggest that much of the functional diversity among *Leishmania* species results from recent evolutionary innovations rather than ancient divergence in core cellular machinery.

This genomic architecture may facilitate rapid adaptation to new hosts, vectors, and environments, potentially enabling the emergence of new disease patterns or geographic expansion of existing species. Understanding the genetic basis of such adaptations could inform predictions about disease emergence and guide surveillance efforts in regions experiencing environmental change or vector range expansion.

The limited phylogenetic signal available from single-copy genes underscores the importance of developing alternative approaches for resolving evolutionary relationships among closely related *Leishmania* species. Whole-genome approaches incorporating information from synteny, gene order, and copy number variation may provide improved resolution for phylogenetic questions that cannot be addressed using traditional sequence-based methods.

From a medical perspective, the substantial genomic diversity identified within and among *Leishmania* species highlights the importance of species-specific diagnostic and therapeutic approaches. The presence of species-specific gene content may influence drug susceptibility, immune responses, and clinical outcomes, necessitating continued research into the functional significance of genomic diversity for treatment and prevention strategies.

The identification of core genes conserved across all trypanosomatid species provides potential targets for broad-spectrum therapeutic development, while species-specific genes may represent targets for species-specific interventions. The

extensive gene duplication observed in most species may contribute to drug resistance through gene dosage effects and should be considered in drug development strategies.

This comprehensive comparative genomic analysis establishes a foundation for future research into trypanosomatid evolution, pathogenesis, and therapeutic development. The genomic resources and evolutionary insights generated through this work provide essential context for understanding the complex relationships between genomic diversity and the medical and evolutionary significance of these important parasitic organisms.

4.4. SCG Phylogenetic analysis

4.4.1 Phylogenetic Relationships and Taxonomic Implications

Our comprehensive phylogenetic analysis of 63 trypanosomatid species provides robust support for the current taxonomic framework while revealing important insights into the evolutionary complexity of this medically significant group of parasites. The clear monophyly of the major *Leishmania* subgenera*: Leishmania leishmania, Sauroleishmania, Mundinia,* and *Viannia*, confirms the validity of recent taxonomic revisions that have elevated these groups to subgeneric status (Akhoundi et al., 2016; Espinosa et al., 2018a).

The positioning of *Viannia* as a distinct sister lineage to other *Leishmania* species is consistent with previous molecular phylogenetic studies and supports its recognition as representing an ancient divergence within the *Leishmania* complex (Fraga et al., 2010; Kuhls et al., 2011). This finding has important implications for understanding the biogeographical history of leishmaniasis, as *Viannia* species are exclusively found in the Neotropical region and represent a monophyletic radiation of New World cutaneous and mucocutaneous leishmaniasis agents.

The distinct clustering of *Mundinia* species, including *L. enrietti* and *L. martiniquensis*, reinforces the unique evolutionary trajectory of this recently recognized subgenus (Espinosa et al., 2018b). These species, which infect diverse mammalian hosts including guinea pigs and humans in the Caribbean, represent an independent colonization of vertebrate hosts that differs significantly from the transmission patterns observed in other *Leishmania* subgenera.

4.4.2. Evolutionary Complexity and Conflicting Phylogenetic Signals

The reticulate patterns observed in our SplitsTree network analysis highlight the complex evolutionary history of trypanosomatids, particularly at nodes representing ancient divergences. Such conflicting phylogenetic signals are not uncommon in rapidly diversifying parasite lineages and may result from several evolutionary processes (Bapteste et al., 2013; Debevec and Whitfield, 2013).

Incomplete lineage sorting during rapid speciation events is likely a major contributor to the observed phylogenetic incongruence. The relatively short internal branches connecting major *Leishmania* subgenera suggest a period of rapid evolutionary radiation, which is consistent with paleontological evidence suggesting that the major diversification of *Leishmania* occurred during the Cenozoic era coincident with the radiation of mammalian hosts (Valdivia et al., 2015a; Harkins et al., 2016).

Horizontal gene transfer, while less commonly documented in trypanosomatids compared to other microbial groups, cannot be entirely excluded as a contributing factor to phylogenetic complexity. Recent studies have identified potential instances of genetic exchange between trypanosomatid species, particularly involving repetitive DNA elements and some metabolic genes (Lachaud et al., 2014; Jackson, 2010).

4.4.3. Methodological Considerations and Future Directions

The concordance between our maximum likelihood phylogenetic analysis and the network-based approach provides confidence in the major taxonomic groupings while acknowledging areas of uncertainty. The use of SplitsTree analysis has proven particularly valuable in identifying regions of the tree where additional molecular data may be needed to resolve ambiguous relationships (Huson and Bryant, 2006).

Future phylogenetic studies of trypanosomatids would benefit from the inclusion of genome-scale datasets, which can provide thousands of orthologous genes for analysis and help resolve challenging nodes through increased statistical power (Ryan et al., 2021). Phylogenomic approaches have already begun to shed light on deep evolutionary relationships within kinetoplastid parasites and will likely be essential for resolving remaining uncertainties in trypanosomatid systematics (Kaufer et al., 2017).

The application of coalescent-based species tree methods, which can account for incomplete lineage sorting and conflicting gene tree topologies, represents another promising avenue for future research (Liu, Wu and Yu, 2015). Such approaches may be particularly valuable for understanding the rapid radiation events that appear to characterize trypanosomatid evolutionary history.

4.4.4. Medical and Epidemiological Implications

Our phylogenetic findings have important implications for understanding disease epidemiology and drug development strategies. The clear evolutionary distinctiveness of the major *Leishmania* subgenera correlates with differences in clinical manifestations, vector associations, and drug sensitivities (Torres-Guerrero et al., 2017; Burza, Croft and Boelaert, 2018). For instance, the unique phylogenetic position of *Viannia* species aligns with their distinct clinical presentations and their association with specific sandfly vectors in the Neotropics.

The evolutionary relationships revealed in our analysis also provide a framework for predicting drug cross-resistance patterns and identifying targets for broad-spectrum therapeutic development. Species within well-supported monophyletic groups are more likely to share similar drug susceptibility profiles, while phylogenetically distant species may require different therapeutic approaches (Croft, Sundar and Fairlamb, 2006; Sundar and Chakravarty, 2015).

4.5. ADMIXTURE

4.5.1. Population Structure and Taxonomic Relationships in Trypanosomatids

The ADMIXTURE analysis of 63 trypanosomatid species using 21,530 genome-wide fixed differences provides the most comprehensive population structure assessment of this medically important parasite family to date. The cross-validation analysis strongly supports K=6 as the optimal number of ancestral populations (CV error = 0.44950), representing a statistically robust framework for understanding evolutionary relationships within the Trypanosomatidae. This six-population structure likely corresponds to major taxonomic divisions within the family, including the three recognized *Leishmania* subgenera *(Leishmania, Viannia*, and *Sauroleishmania*), distinct *Trypanosoma* lineages, and other trypanosomatid groups such as *Crithidia* and related genera.

The sharp increase in cross-validation errors beyond K=6, particularly at K=9 (0.69341) and K=13 (0.93903), demonstrates clear evidence of model overfitting, indicating that while higher K values may provide finer taxonomic resolution, they do not represent biologically meaningful population structure for comparative evolutionary analyses. This finding has important implications for future population genetic studies of trypanosomatids, as it establishes an empirical basis for selecting appropriate K values that balance statistical rigor with biological interpretability.

4.5.2. Novel Insights into Natural Hybridization

One of the most significant findings of this study is the genetic composition of the *Leishmania (Viannia)* hybrid strain MHOM/VE/1991/PM-H197, which exhibits approximately 20% *Viannia* ancestry and 80% *Leishmania (Leishmania)* ancestry, specifically from the *L. mexicana* complex. This ancestry pattern represents the first genomic characterization of a natural intersubgeneric *Leishmania* hybrid and provides unprecedented insights into the evolutionary dynamics of these parasites.

The detection of such extensive admixture between distantly related *Leishmania* subgenera challenges traditional assumptions about reproductive barriers within the genus. Previous studies of natural *Leishmania* hybrids have documented genetic exchange exclusively between closely related species within the same subgenus, such as *L. braziliensis* × *L. guyanensis* hybrids in Central America or *L. donovani* complex hybrids in the Old World (Van den Broeck et al., 2020; Cotton et al., 2020b; Hadermann et al., 2023a). The identification of a *Viannia* × *Leishmania* (*Leishmania*) hybrid demonstrates that reproductive compatibility can extend across much greater evolutionary distances than previously recognized.

The asymmetric ancestry pattern observed in this hybrid (80% *L. mexicana complex*, 20% *Viannia*) may reflect differential fitness effects of genetic material from the two parental lineages or could indicate multiple generations of backcrossing following an initial hybridization event. The predominance of *L. mexicana* complex ancestry suggests either stronger selective pressure favoring these genetic backgrounds or more frequent backcrossing with *L. mexicana* complex populations in the geographic region where this strain was isolated (Venezuela).

One important limitation of the current genomic characterisation is the absence of kinetoplast DNA (kDNA) analysis. In Leishmania, kDNA, the mitochondrial genome, is inherited uniparentally, meaning hybrids carry the kDNA of only one parental lineage rather than a mixture of both. This makes kDNA haplotype a powerful independent marker for confirming hybridisation and identifying the maternal parent of a cross. In previously characterised Leishmania hybrids, kDNA typing has been

used alongside nuclear markers to resolve the directionality of hybridisation events (Van den Broeck et al., 2020; Hadermann et al., 2023b). For PM-H197, determining whether the kDNA is of Viannia or L. mexicana complex origin would help establish which lineage contributed the maternal genome, providing independent corroboration of the intersubgeneric hybridisation hypothesis. Future work should therefore include kDNA haplotyping of this strain, alongside targeted sequencing of mitochondrial markers such as cytochrome b or the kDNA minicircle sequences, to fully characterise the parentage of this putative hybrid.

4.5.3. Evolutionary and Epidemiological Implications

The demonstration of intersubgeneric hybridization capability has profound implications for understanding *Leishmania* evolution and the potential for emergence of novel pathogenic phenotypes. Genetic exchange between divergent lineages can facilitate rapid adaptation through the introduction of beneficial alleles from alternative genetic backgrounds, potentially leading to enhanced virulence, altered host specificity, or modified drug resistance profiles (Inbar et al., 2019; Louradour et al., 2020b).

From an epidemiological perspective, the existence of natural intersubgeneric hybrids raises important questions about disease surveillance and diagnostic accuracy. Traditional molecular diagnostic approaches that rely on subgenus-specific markers may misclassify hybrid strains, potentially leading to inappropriate treatment protocols given the distinct clinical manifestations typically associated with different *Leishmania* subgenera. The *L. mexicana* complex is primarily associated with cutaneous leishmaniasis, while *Viannia* species can cause both cutaneous and mucocutaneous disease forms. Hybrid strains may exhibit intermediate or novel pathogenic characteristics that do not conform to established clinical paradigms.

4.5.4. Methodological Advances in Trypanosomatid Population Genetics

This study demonstrates the power of genome-wide data for resolving complex population structure patterns that would be undetectable using traditional molecular markers. The ability to identify and quantify admixture proportions in hybrid individuals represents a significant advance over previous approaches that relied on limited gene sets or microsatellite markers. The comprehensive sampling across the trypanosomatid phylogeny provides a robust comparative framework for future studies of population structure and hybridization in these parasites.

The establishment of K=6 as the optimal population structure model for trypanosomatids also provides a standardized framework for comparative studies across research groups. This consistency will facilitate meta-analyses and enable more robust comparisons of population genetic patterns across different geographic regions and host systems.

4.6. Hybrid vs. Mixed infection

4.6.1. Discordant Analytical Approaches Reveal Complex Genomic Architecture in *Leishmania* Clinical Isolates

The striking discordance between ADMIXTURE ancestry inference (80% *L. mexicana*, 20% *Viannia*) and genomic abundance measurements (67% *Viannia* dominance) in strain PM-H197 highlights fundamental limitations in applying single analytical approaches to complex *Leishmania* infections. This 8-fold contradiction between methods reflects the distinct biological processes each approach measures: ADMIXTURE detects historical recombination signatures preserved in contemporary genomes, while coverage analysis quantifies current DNA abundance in sequenced samples (Alexander, Novembre and Lange, 2009; Lawson et al., 2012).

4.6.2. Mechanistic Explanations for ADMIXTURE-Coverage Discordance

Several biological mechanisms could account for the observed discrepancy between ancestry and abundance measurements. ADMIXTURE analysis may disproportionately detect minority genomic components if they contributed high-diversity allelic variants during historical hybridization events, as the algorithm weights polymorphic sites more heavily in ancestry inference (Porras-Hurtado et al., 2013). In unbalanced hybrid scenarios, even small genomic contributions from divergent lineages can generate strong ancestry signals if those regions harbor species-diagnostic variants.

Furthermore, post-hybridization genomic instability may have altered the contemporary genomic composition relative to ancestral recombination signatures. *Leishmania* parasites exhibit remarkable chromosomal plasticity, including aneuploidy, chromosomal loss, and copy number variation (Sterkers et al., 2011; Dumetz et al., 2017). If PM-H197 originated as a more balanced hybrid, subsequent preferential loss of *L. mexicana* chromosomes could explain the current *Viannia* dominance while preserving historical recombination signatures in surviving genomic regions.

Finally, the extreme coefficient of variation observed in *L. mexicana* coverage (CV = 3.292) compared to *Viannia* coverage (CV = 0.474) suggests fundamental differences in genomic stability or representation. This pattern is inconsistent with balanced tetraploid hybrids, which would be expected to show relatively uniform coverage ratios across chromosomes (Romano et al., 2014; Inbar et al., 2019).

The integration of multiple analytical approaches supports two primary interpretations of PM-H197's genomic architecture, each with distinct biological implications and supporting evidence.

4.6.3. Evidence Supporting Mixed Infection

The mixed infection hypothesis is supported by several key observations. The extreme ultra-long read preference for *Viannia* sequences (96:4 ratio) resembles

patterns expected from independent parasite populations with vastly different abundances rather than integrated hybrid genomes. Additionally, the high variability in *L. mexicana* coverage across chromosomes could reflect stochastic sampling from a minor population component rather than stable genomic integration (Rougeron et al., 2009).

Mixed infections are epidemiologically plausible in regions where multiple *Leishmania* species co-circulate, and have been documented using microsatellite and multilocus enzyme electrophoresis approaches (Bañuls, Hide and Prugnolle, 2007). The ADMIXTURE signal in mixed infections could arise from shared ancestry between co-infecting lineages that diverged after ancient hybridization events, creating detectable recombination signatures without requiring contemporary genomic integration.

4.6.4. Evidence Supporting Unbalanced Hybrid

Conversely, the unbalanced hybrid hypothesis explains several observations that are difficult to reconcile with random co-infection. The remarkably consistent 2:1 coverage ratio observed across the entire genome suggests coordinated inheritance rather than stochastic population dynamics. If PM-H197 represented independent parasite populations, greater variation in abundance ratios would be expected due to differential growth rates, immune responses, or sampling effects.

Furthermore, the stable ADMIXTURE ancestry signatures (80:20) combined with consistent genomic abundance patterns (67:33) suggest integrated genomic architecture rather than independent populations. True hybrids with unequal parental contributions have been documented in other eukaryotic systems, including fungi and plants, where asymmetric genome loss or preferential chromosome inheritance creates viable organisms with skewed genomic compositions (Schardl and Craven, 2003; Soltis and Soltis, 2009).

The biological feasibility of unbalanced *Leishmania* hybrids is supported by experimental evidence demonstrating successful hybridization between divergent

species under laboratory conditions (Akopyants et al., 2009; Romano et al., 2014). While most documented hybrids exhibit relatively balanced genomic contributions, the selective pressures operating in natural infections may favor organisms with optimal genomic compositions for specific host environments.

4.6.5. Implications for *Leishmania* Population Genetics and Clinical Management

These findings have important implications for both fundamental understanding of *Leishmania* population structure and clinical parasite characterization. The demonstration that ADMIXTURE analysis can yield misleading results in complex infections emphasizes the importance of multi-approach validation when characterizing field isolates (Rogers et al., 2014; Cotton et al., 2020a).

From a clinical perspective, distinguishing between mixed infections and unbalanced hybrids has therapeutic relevance, as these scenarios may respond differently to treatment regimens. Mixed infections could potentially be cleared by targeting the dominant species, while stable hybrids might require approaches effective against both parental lineages (Dujardin et al., 1995; Antinori, Schifanella and Corbellino, 2012).

The methodological framework developed here, combining ADMIXTURE analysis with coverage-based abundance measurements and competitive ultra-long read mapping, provides a robust approach for characterizing complex *Leishmania* infections that should be applicable to other clinical isolates and epidemiological studies.

4.6.6. Future Directions

Definitive resolution of the mixed infection versus unbalanced hybrid question requires additional analytical approaches. Single-cell genomic analysis could determine whether *Viannia* and *L. mexicana* sequences co-occur within individual parasites or represent separate populations. Chromosomal-level assembly and

karyotyping could identify recombinant chromosomes diagnostic of true hybridization versus intact parental chromosomes indicative of mixed infection.

Long-term culture studies examining genomic stability over multiple passages could distinguish between dynamic mixed populations and stable hybrid genomes. Additionally, experimental crosses between characterized *L. mexicana* and *Viannia* strains could provide controlled comparisons for interpreting field isolate patterns.

5. Conclusion

This comprehensive genomic study of *Leishmania* (*Viannia*) species represents a significant advancement in our understanding of one of the world's most important neglected tropical disease pathogens. Through the integration of cutting-edge long-read sequencing technologies, sophisticated assembly algorithms, and comprehensive comparative genomic analyses, this research has generated unprecedented insights into the evolutionary biology, population structure, and genetic diversity of the *Viannia* subgenus.

The production of high-quality, chromosome-level genome assemblies for eight *Viannia* species and five additional *Leishmania* taxa addresses a critical gap in genomic resources for these medically important parasites. The superior assembly quality achieved through Oxford Nanopore Technologies sequencing, evidenced by substantially improved N50 values, BUSCO completeness scores, and successful telomere recovery, establishes new standards for parasitic genome assembly and provides robust foundations for downstream genomic analyses. These resources enable precise characterization of species-specific genomic features, accurate phylogenetic inference, and detailed investigations of evolutionary processes that shape pathogen diversity.

The comprehensive comparative genomic analysis across 63 trypanosomatid species reveals the remarkable evolutionary plasticity that characterises this ancient parasite family. The identification of minimal core genome conservation (5.06% of orthogroups) combined with extensive accessory genome diversity (53.76% of

orthogroups) demonstrates that the clinical and epidemiological diversity of leishmaniasis is underpinned by a highly flexible genomic architecture, where a small conserved core maintains fundamental cellular machinery while an expansive accessory genome drives species-specific adaptation. The 20.11% of orthogroups unique to individual species likely encode the virulence factors, host interaction molecules, and environmental adaptations that determine tissue tropism, disease severity, and geographic range. These are findings with direct implications for identifying novel drug targets and understanding why different *Leishmania* species cause such distinct clinical outcomes. These patterns of genomic adaptation represent a contribution of equal significance to the hybrid finding, establishing a comprehensive framework for understanding how this parasite family has diversified across the Americas.

The methodological innovations developed in this research, particularly the integration of ADMIXTURE ancestry inference with Oxford Nanopore coverage analysis and competitive ultra-long read mapping, provide powerful new approaches for characterizing complex parasite infections. The identification of an 8-fold discrepancy between ancestry predictions and genomic abundance measurements highlights the importance of employing multiple analytical strategies when investigating field isolates and demonstrates that single-method approaches may yield misleading conclusions about infection composition.

From a clinical perspective, the genomic resources and analytical frameworks established here provide essential tools for improving leishmaniasis diagnosis, treatment, and epidemiological surveillance. The demonstration that hybrid strains may exhibit genomic compositions that differ substantially from their ancestry signatures emphasizes the need for comprehensive characterization approaches that can accurately distinguish between true hybridization and mixed infection scenarios, as these may require different therapeutic strategies.

The phylogenomic analyses presented in this study provide the most robust evolutionary framework for *Leishmania* systematics to date, confirming the validity of current subgeneric classifications while revealing complex evolutionary patterns that reflect the ancient diversification and rapid adaptive radiation characteristic of trypanosomatid parasites. The clear monophyly of major *Leishmania* subgenera,

combined with evidence for complex evolutionary processes including incomplete lineage sorting and potential horizontal gene transfer, provides crucial context for understanding the biogeographical history and evolutionary dynamics of leishmaniasis agents.

This research establishes important foundations for future investigations into *Leishmania* biology and leishmaniasis control. The high-quality genomic assemblies enable detailed functional genomics studies to identify virulence factors, drug resistance mechanisms, and vaccine targets. The comprehensive comparative genomic framework facilitates investigations of host-pathogen interactions, vector specificity, and environmental adaptation. The population genetic insights provide crucial context for epidemiological studies and inform surveillance strategies in regions experiencing environmental change or vector range expansion.

Several important research directions emerge from this work. Single-cell genomic approaches could definitively resolve whether complex strains like PM-H197 represent stable hybrid organisms or dynamic mixed populations. Experimental crosses between characterized strains could elucidate the mechanisms and frequency of intersubgeneric hybridization under controlled conditions. Population genomic studies incorporating multiple strains per species could distinguish between species-specific adaptations and intraspecific variation patterns. Functional characterization of species-specific and accessory genes could identify the biological processes driving genomic diversity and their relevance for pathogenesis and drug development.

In conclusion, this thesis advances our fundamental understanding of *Leishmania* evolution through three interconnected contributions: the generation of the most comprehensive high-quality genomic resource for the *Viannia* subgenus to date; the characterisation of genomic adaptation patterns across 63 trypanosomatid species that explain the remarkable clinical and ecological diversity of these parasites; and the first genomic characterisation of intersubgeneric hybridisation in *Leishmania*, which challenges existing assumptions about reproductive barriers within the genus. Together these findings provide essential genomic resources and analytical frameworks for addressing the ongoing challenges posed by leishmaniasis as a

neglected tropical disease affecting hundreds of millions of people worldwide, contributing to the ultimate goal of developing more effective strategies for prevention, diagnosis, and treatment.

6. References

Acosta-Soto, L. et al. (2020). Autochthonous Outbreak of Cutaneous Leishmaniasis due to Leishmania infantum in Corrientes Province, Argentina. *The American journal of tropical medicine and hygiene*, 102 (3), pp.593–597.

Akhoundi, M. et al. (2016). A historical overview of the classification, evolution, and dispersion of Leishmania parasites and sandflies. *PLoS neglected tropical diseases*, 10 (3), p.e0004349.

Akopyants, N. S. et al. (2009). Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. *Science*, 324 (5924), pp.265–268.

Alam, M. Z. et al. (2009). Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of Leishmania donovani strains in the Indian subcontinent. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 9 (1), pp.24–31.

Alatif, H. (2021). Burden and Trends of Leishmaniasis over the last one Decade Across the Globe: Trend Analysis of WHO Regions. *Integrative Journal of Medical Sciences*. [Online]. Available at: doi:10.15342/ijms.2021.295.

Albanaz, A. T. S. et al. (2023). Shining the spotlight on the neglected: new high-quality genome assemblies as a gateway to understanding the evolution of Trypanosomatidae. *BMC genomics*, 24 (1). [Online]. Available at: doi:10.1186/s12864-023-09591-z.

Alexander, D. H. and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*, 12 (1), p.246.

Alexander, D. H., Novembre, J. and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19 (9), pp.1655–1664.

Al-Jawabreh, A. et al. (2008). Identification of geographically distributed sub-populations of Leishmania (Leishmania) majorby microsatellite analysis. *BMC evolutionary biology*, 8 (1). [Online]. Available at: doi:10.1186/1471-2148-8-183.

Almutairi, H. et al. (2021). Chromosome-scale genome sequencing, assembly and annotation of six genomes from subfamily Leishmaniinae. *Scientific data*, 8 (1). [Online]. Available at: doi:10.1038/s41597-021-01017-3.

Alvar, J. et al. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS one*, 7 (5), p.e35671.

Alvarez, A. E. et al. (2001). Use of microsatellites to evaluate genetic diversity and species relationships in the genus Lycopersicon. *Theoretical and applied genetics*, 103 (8), pp.1283–1292.

Andersson, B. (2011). The Trypanosoma cruzi genome; conserved core genes and extremely variable surface molecule families. *Research in Microbiology*, 162 (6), pp.619–625.

Andersson, J. O. (2005). Lateral gene transfer in eukaryotes. *Cellular and molecular life sciences: CMLS*, 62 (11), pp.1182–1197.

Antinori, S., Schifanella, L. and Corbellino, M. (2012). Leishmaniasis: new insights from an old and neglected disease. *European journal of clinical microbiology & infectious diseases: official publication of the European Society of Clinical Microbiology*, 31 (2), pp.109–118.

Anuntakarun, S. et al. (2022). Genome assembly and genome annotation of Leishmania martiniquensis isolated from a leishmaniasis patient in Thailand. *Journal of parasitology research*, 2022, pp.1–7.

Aransay, A. M. et al. (2004). Distribution of sandfly species in relation to canine leishmaniasis from the Ebro Valley to Valencia, northeastern Spain. *Parasitology research*, 94 (6), pp.416–420.

Ascencio, M. E. et al. (2020). Molecular diagnosis of *Leishmania* spp. in dogs of a subtropical locality of Argentina. *Transboundary and emerging diseases*, 67 (S2), pp.106–110.

Assis, L. H. C. et al. (2021). Cell cycle, telomeres, and telomerase in Leishmania spp.: What do we know so far? *Cells (Basel, Switzerland)*, 10 (11), p.3195.

Bañuls, A.-L., Hide, M. and Prugnolle, F. (2007). Leishmania and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in parasitology*, 64, pp.1–109.

Bapteste, E. et al. (2013). Networks: expanding evolutionary thinking. *Trends in genetics: TIG*, 29 (8), pp.439–441.

Barratt, J. et al. (2017). 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 (1), p.e0005215.

Barrios, S. P. G. et al. (2020). Phlebotominae (Diptera: Psychodidae) and Biomes in the State of Mato Grosso do Sul, Brazil. *Journal of medical entomology*, 57 (6), pp.1882–1904.

Barry, J. D. et al. (2003). Why are parasite contingency genes often associated with telomeres? *International journal for parasitology*, 33 (1), pp.29–45.

Bates, P. A. (2007). Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies. *International journal for parasitology*, 37 (10), pp.1097–1106.

Batista, L. F. S. et al. (2025). Molecular insights into cell-mediated immunity in atypical non-ulcerated cutaneous leishmaniasis. *Microorganisms*, 13 (2), p.413.

Behniafar, H. et al. (2024). Detection and phylogenetic analysis of kinetoplast DNA of Leishmania infantum infected humans, domestic dogs and sandflies in Northwest Iran. *PloS one*, 19 (3), p.e0296777.

Black, J. A. et al. (2023). Life in plastic, it's fantastic! How Leishmania exploit genome instability to shape gene expression. *Frontiers in cellular and infection microbiology*, 13. [Online]. Available at: doi:10.3389/fcimb.2023.1102462.

Boité, M. C. et al. (2012). New insights on taxonomy, phylogeny and population genetics of Leishmania (Viannia) parasites based on multilocus sequence analysis. *PLoS neglected tropical diseases*, 6 (11), p.e1888.

Brilhante, A. F. et al. (2019). Leishmania (Viannia) braziliensis type 2 as probable etiological agent of canine cutaneous leishmaniasis in Brazilian Amazon. *PloS one*, 14 (4), p.e0216291.

Brockhurst, M. A. et al. (2019). The ecology and evolution of pangenomes. *Current biology: CB*, 29 (20), pp.R1094–R1103.

Brown, T. A. (2006). *Genomes*. 3rd ed. New York, NY: Garland Publishing.

Bulle, B. et al. (2002). Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis. *Journal of clinical microbiology*, 40 (9), pp.3391–3397.

Burza, S., Croft, S. L. and Boelaert, M. (2018). Leishmaniasis. *The Lancet*, 392 (10151), pp.951–970.

Butenko, A. et al. (2019). Comparative genomics of Leishmania (Mundinia). *BMC genomics*, 20 (1), p.726.

Camacho, E. et al. (2019). Complete assembly of the Leishmania donovani (HU3 strain) genome and transcriptome annotation. *Scientific reports*, 9 (1). [Online]. Available at: doi:10.1038/s41598-019-42511-4.

Cantanhêde, L. M. et al. (2018). New insights into the genetic diversity of Leishmania RNA Virus 1 and its species-specific relationship with Leishmania parasites. *PloS one*, 13 (6), p.e0198727.

Catta-Preta, C. M. C. et al. (2023). HOP1 and HAP2 are conserved components of the meiosis-related machinery required for successful mating in Leishmania. *Nature communications*, 14 (1). [Online]. Available at: doi:10.1038/s41467-023-42789-z.

Cecílio, P., Cordeiro-da-Silva, A. and Oliveira, F. (2022). Sand flies: Basic information on the vectors of leishmaniasis and their interactions with Leishmania parasites. *Communications biology*, 5 (1). [Online]. Available at: doi:10.1038/s42003-022-03240-z.

Chakravarty, J. and Sundar, S. (2019). Current and emerging medications for the treatment of leishmaniasis. *Expert opinion on pharmacotherapy*, 20 (10), pp.1251–1265.

Chaouch, M. et al. (2013). Identification of Tunisian Leishmania spp. by PCR amplification of cysteine proteinase B (cpb) genes and phylogenetic analysis. *Acta tropica*, 125 (3), pp.357–365.

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, 2 (2), p.e107.

Chen, Y. et al. (2020). Fast and accurate assembly of Nanopore reads via progressive error correction and adaptive read selection. *bioRxiv*. [Online]. Available at: doi:10.1101/2020.02.01.930107.

Chiyo, L. et al. (2023). Cross-sectional spatial and epidemiological analysis of canine visceral leishmaniasis cases in the triple border region, Brazil, Argentina and Paraguay, between 2015 and 2020. *Acta tropica*, 239 (106811), p.106811.

Clayton, C. E. (2016). Gene expression in kinetoplastids. *Current opinion in microbiology*, 32, pp.46–51.

Cosentino Lagomarsino, M. et al. (2015). Hierarchy feedback evolution E. coli transcription network. *Proceedings National Academy Sciences*, 104 (13), pp.5516–5520.

Cotton, J. et al. (2020a). Multi-genome alignment annotation Leishmania spp. comparative genomics. *Wellcome Open Research*, 5.

Cotton, J. A. et al. (2020b). Genomic analysis of natural intra-specific hybrids among Ethiopian isolates of Leishmania donovani. *PLoS neglected tropical diseases*, 14 (4), p.e0007143.

Coughlan, S. et al. (2018). Leishmania naiffi and Leishmania guyanensis reference genomes highlight genome structure and gene evolution in the Viannia subgenus. *Royal Society open science*, 5 (4), p.172212.

Croft, S. L. and Coombs, G. H. (2003). Leishmaniasis– current chemotherapy and recent advances in the search for novel drugs. *Trends in parasitology*, 19 (11), pp.502–508.

Croft, S. L., Sundar, S. and Fairlamb, A. H. (2006). Drug resistance in leishmaniasis. *Clinical microbiology reviews*, 19 (1), pp.111–126.

Croll, D. and McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular ecology*, 26 (7), pp.2027–2040.

Cruz, I. et al. (2010). A canine leishmaniasis pilot survey in an emerging focus of visceral leishmaniasis: Posadas (Misiones, Argentina). *BMC infectious diseases*, 10 (1). [Online]. Available at: doi:10.1186/1471-2334-10-342.

Dantas-Torres, F. (2007). The role of dogs as reservoirs of Leishmania parasites, with emphasis on Leishmania (Leishmania) infantum and Leishmania (Viannia) braziliensis. *Veterinary parasitology*, 149 (3–4), pp.139–146.

Darian, J. C. et al. (2024). Constructing telomere-to-telomere diploid genome by polishing haploid nanopore-based assembly. *Nature methods*, 21 (4), pp.574–583.

De Brito, R. C. F. et al. (2018). Peptide Vaccines for Leishmaniasis. *Frontiers in immunology*, 9. [Online]. Available at: doi:10.3389/fimmu.2018.01043.

De Los Santos, M. B. et al. (2022). Genetic diversity and population structure of Leishmania (Viannia) braziliensis in the Peruvian jungle. *PLoS neglected tropical diseases*, 16 (5), p.e0010374.

De Maio, N. et al. (2019). Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial genomics*, 5 (9). [Online]. Available at: doi:10.1099/mgen.0.000294.

Debevec, A. H. and Whitfield, J. B. (2013). Introduction to phylogenetic networks.—David A. morrison. *Systematic biology*, 62 (1), pp.177–178.

Degnan, J. H. and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in ecology & evolution*, 24 (6), pp.332–340.

Delgado-Noguera, L. A. et al. (2023). Diversity and geographical distribution of Leishmania species and the emergence of Leishmania (Leishmania) infantum and L. (Viannia) panamensis in Central-Western Venezuela. *Acta tropica*, 242 (106901), p.106901.

Díaz-Viraqué, F. et al. (2021). Nanopore long read DNA sequencing of protozoan parasites: Hybrid genome assembly of Trypanosoma cruzi. In: *Methods in Molecular Biology*. Methods in molecular biology (Clifton, N.J.). New York, NY: Springer US. pp.3–13.

van Dijk, E. L. et al. (2023). Genomics in the long-read sequencing era. *Trends in genetics: TIG*, 39 (9), pp.649–671.

Diro, E. et al. (2015). Atypical manifestations of visceral leishmaniasis in patients with HIV in north Ethiopia: a gap in guidelines for the management of opportunistic infections in resource poor settings. *The Lancet infectious diseases*, 15 (1), pp.122–129.

Docampo, R. and Moreno, S. N. (2003). Current chemotherapy Trypanosoma cruzi infections. *Parasitology Today*, 19 (12), pp.527–532.

Domagalska, M. A. et al. (2019). Genomes of Leishmania parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. *PLoS neglected tropical diseases*, 13 (12), p.e0007900.

Domínguez-Asenjo, B. et al. (2021). Ex vivo phenotypic screening of two small repurposing drug collections identifies nifuratel as a potential new treatment against visceral and cutaneous leishmaniasis. *ACS infectious diseases*, 7 (8), pp.2390–2401.

Dorlo, T. P. C. et al. (2012). Miltefosine: a review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *The journal of antimicrobial chemotherapy*, 67 (11), pp.2576–2597.

Downing, T. et al. (2012). Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the Leishmania donovani species complex. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 12 (1), pp.149–159.

Dujardin, J. C. et al. (1995). Putative Leishmania hybrids in the Eastern Andean valley of Huanuco, Peru. *Acta tropica*, 59 (4), pp.293–307.

Dumetz, F. et al. (2017). Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio*, 8 (3). [Online]. Available at: doi:10.1128/mBio.00599-17.

Ellis, J. and Crampton, J. (1988). Characterisation of a simple, highly repetitive DNA sequence from the parasite Leishmania donovani. *Molecular and biochemical parasitology*, 29 (1), pp.9–17.

El-Sayed, N. M. et al. (2005). The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science (New York, N.Y.)*, 309 (5733), pp.409–415.

Espinosa, O. et al. (2018a). evolution Leishmania Endotrypanum parasites: phylogenetic analysis based orthologous groups. *Genes*, 9 (4).

Espinosa, O. A. et al. (2018b). An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as Leishmania and Endotrypanum. *Parasitology*, 145 (4), pp.430–442. [Accessed 7 February 2025].

Feron, R. and Waterhouse, R. M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *GigaScience*, 11. [Online]. Available at: doi:10.1093/gigascience/giac006.

Ferreira, B. A. et al. (2022a). Isolation, typing, and drug susceptibility of Leishmania (Leishmania) infantum isolates from dogs of the municipality of Embu das Artes, an endemic region for canine leishmaniasis in Brazil. *Parasitology research*, 121 (9), pp.2683–2695.

Ferreira, T. R. et al. (2022b). Self-Hybridization in Leishmania major. *mBio*, 13 (6). [Online]. Available at: doi:10.1128/mbio.02858-22.

Fisher, M. C. et al. (2000). A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus Coccidioides immitis. *Molecular biology and evolution*, 17 (8), pp.1164–1174.

Fraga, J. et al. (2010). Phylogeny of Leishmania species based on the heat-shock protein 70 gene. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 10 (2), pp.238–245.

Franssen, S. U. et al. (2020). Global genome diversity of the Leishmania donovani complex. *eLife*, 9. [Online]. Available at: doi:10.7554/elife.51243.

de Freitas, N. D. A. et al. (2024). Spatial analysis and Socio-environmental determinants of canine visceral leishmaniasis in an urban area in northeastern Brazil. *Tropical medicine and infectious disease*, 10 (1), p.6.

Fu, G. and Barker, D. C. (1998). Characterisation of Leishmania telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. *Nucleic acids research*, 26 (9), pp.2161–2167.

Gedda, M. R. et al. (2020). Post kala-azar dermal leishmaniasis: A threat to elimination program. *PLoS neglected tropical diseases*, 14 (7), p.e0008221.

Geto, A. K. et al. (2024). Prevalence of human visceral leishmaniasis and its risk factors in Eastern Africa: a systematic review and meta-analysis. *Frontiers in public health*, 12. [Online]. Available at: doi:10.3389/fpubh.2024.1488741.

Ghouse Peer, G. D. et al. (2024). A systematic assessment of Leishmania donovani infection in domestic and wild animal reservoir hosts of zoonotic visceral leishmaniasis in India. *Microbiology research*, 15 (3), pp.1645–1654.

Gillespie, P. M. et al. (2016). Status of vaccine research and development of vaccines for leishmaniasis. *Vaccine*, 34 (26), pp.2992–2995.

Glans, H. et al. (2021). High genome plasticity and frequent genetic exchange in Leishmania tropica isolates from Afghanistan, Iran and Syria. *PLoS neglected tropical diseases*, 15 (12), p.e0010110.

Gomez, E. A. et al. (2018). Leishmaniasis caused by Leishmania (Viannia) guyanensis in north-central Pacific region of Ecuador: A clinico-epidemiological feature. *Acta tropica*, 185, pp.204–211.

González-de la Fuente, S. et al. (2017). Resequencing of the Leishmania infantum (strain JPCM5) genome and de novo assembly into 36 contigs. *Scientific reports*, 7 (1). [Online]. Available at: doi:10.1038/s41598-017-18374-y.

Gonzalez-Garcia, L. N. et al. (2025). Genetic diversity and comparative genomics across Leishmania (Viannia) species. *Communications biology*, 8 (1), p.925. [Accessed 10 July 2025].

Guerra, J. A. de O. et al. (2011). Mucosal Leishmaniasis Caused by Leishmania (Viannia) braziliensis and Leishmania (Viannia) guyanensis in the Brazilian Amazon. *PLoS neglected tropical diseases*, 5 (3), p.e980.

Guerra, J. M. et al. (2021). Cytokine profile and parasite load in lymph nodes of dogs naturally infected with Leishmania infantum from distinct epidemiological scenarios in São Paulo State, Brazil. *Veterinary immunology and immunopathology*, 233 (110198), p.110198.

Gunasekara, S. D. et al. (2023). ‘We do not rush to the hospital for ordinary wounds (suḷu tuvāla)’: A qualitative study on the early clinical manifestations of cutaneous leishmaniasis and associated health behaviours in rural Sri Lanka. *PLoS neglected tropical diseases*, 17 (5), p.e0010939.

Gurevich, A. et al. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* , 29 (8), pp.1072–1075.

Gutiérrez-Corbo, C. et al. (2022). Axenic interspecies and intraclonal hybrid formation in Leishmania: Successful crossings between visceral and cutaneous strains. *PLoS neglected tropical diseases*, 16 (2), p.e0010170.

Habib, M. et al. (2024). A targeted and protease-activated genetically encoded melittin-containing particle for the treatment of cutaneous and visceral leishmaniasis. *ACS applied materials & interfaces*, 16 (37), pp.49148–49163.

Hadermann et al. (2023a). Genome-wide diversity, population structure signatures inbreeding African trypanosomiasis vector Glossina fuscipes. *Molecular Ecology*, 32 (12), pp.3205–3221.

Hadermann, A. et al. (2023b). Exploring the Leishmania donovani complex evolution and population structure towards an integrated taxonomy. Infection. *Genetics and Evolution*, 113.

Hadermann, A. et al. (2023c). Genome diversity of Leishmania aethiopica. *Frontiers in cellular and infection microbiology*, 13. [Online]. Available at: doi:10.3389/fcimb.2023.1147998.

Haese-Hill, W., Crouch, K. and Otto, T. D. (2024). Annotation and visualization of parasite, fungi and arthropod genomes with Companion. *Nucleic acids research*, 52 (W1), pp.W39–W44. [Accessed 23 July 2024].

Hamilton, P. B., Gibson, W. C. and Stevens, J. R. (2007). Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Molecular phylogenetics and evolution*, 44 (1), pp.15–25.

Harkins, K. M. et al. (2016). Phylogenomic reconstruction supports supercontinent origins for Leishmania. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 38, pp.101–109.

Harr, B. et al. (1998). A microsatellite-based multilocus phylogeny of the Drosophila melanogaster species complex. *Current biology: CB*, 8 (21), pp.1183–1187.

Heeren, S. et al. (2024). Evolutionary genomics of Leishmania braziliensis across the neotropical realm. *Communications biology*, 7 (1). [Online]. Available at: doi:10.1038/s42003-024-07278-z.

Hong, X.-G. et al. (2024). Mapping the distribution of sandflies and sandfly-associated pathogens in China. *PLoS neglected tropical diseases*, 18 (7), p.e0012291.

Huddleston, J. et al. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome research*, 27 (5), pp.677–685.

Human Genome Structural Variation Consortium et al. (2021). Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. *Nature biotechnology*, 39 (3), pp.302–308.

Huson, D. H. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23 (2), pp.254–267.

Huson, D. H., Kloepper, T. and Bryant, D. (2008). SplitsTree 4.0-Computation of phylogenetic trees and networks. *Bioinformatics* , 14, pp.68–73.

Iborra, S. et al. (2018). Vaccine candidates against leishmania under current research. *Expert review of vaccines*, 17 (4), pp.323–334.

Ibrahim, M. E. and Barker, D. C. (2001). The origin and evolution of the Leishmania donovani complex as inferred from a mitochondrial cytochrome oxidase II gene sequence. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 1 (1), pp.61–68.

Inbar, E. et al. (2019). Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in Leishmania. *PLoS genetics*, 15 (5), p.e1008042.

Ivens, A. C. et al. (2005). The Genome of the Kinetoplastid Parasite, *Leishmania major*. *Science (New York, N.Y.)*, 309 (5733), pp.436–442.

Jackson, A. P. (2010). The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Molecular biology and evolution*, 27 (1), pp.33–45.

Jain, M. et al. (2017). Nanopore sequencing and assembly of a human genome with ultra-long reads. *bioRxiv*. [Online]. Available at: doi:10.1101/128835.

Jauhal, A. A. and Newcomb, R. D. (2021). Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Molecular ecology resources*, 21 (5), pp.1416–1421.

Jensen, R. E., Englund, P. T. and Parsons, M. (2009). RNA interference in Trypanosoma brucei: a review. *International Journal for Parasitology*, 39 (1), pp.1–9.

Kamhawi, S. (2006). Phlebotomine sand flies and Leishmania parasites: friends or foes? *Trends in parasitology*, 22 (9), pp.439–445.

Karakuş, M. et al. (2020). First molecular detection and identification of *Leishmania* species in small wild rodents from Turkey. *Parasitology*, 147 (10), pp.1088–1093.

Kato, H. et al. (2016). First Human Cases of Leishmania (Viannia) lainsoni Infection and a Search for the Vector Sand Flies in Ecuador. *PLoS neglected tropical diseases*, 10 (5), p.e0004728.

Kaufer, A. et al. (2017). The evolution of trypanosomatid taxonomy. *Parasites & vectors*, 10 (1), p.287.

Kelly, J. M. et al. (1991). Evidence of genetic recombination in Leishmania. *Molecular and biochemical parasitology*, 46 (2), pp.253–263.

Koch, L. K. et al. (2017). Modeling the climatic suitability of leishmaniasis vector species in Europe. *Scientific reports*, 7 (1). [Online]. Available at: doi:10.1038/s41598-017-13822-1.

Kocot, K. M. et al. (2013). PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary bioinformatics online*, 9, pp.429–435.

Kondrashov, F. A. et al. (2002). Selection in the evolution of gene duplications. *Genome biology*, 3 (2), p.RESEARCH0008.

Koren, S. and Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current opinion in microbiology*, 23, pp.110–120.

Kostygov, A. Y. et al. (2021). Analyses of Leishmania-LRV co-phylogenetic patterns and evolutionary variability of viral proteins. *Viruses*, 13 (11), p.2305.

Kuhls, K. et al. (2008). Differentiation and Gene Flow among European Populations of Leishmania infantum MON-1. *PLoS neglected tropical diseases*, 2 (7), p.e261.

Kuhls, K. et al. (2011). Comparative microsatellite typing of new world Leishmania infantum reveals low heterogeneity among populations and its recent old world origin. *PLoS neglected tropical diseases*, 5 (6), p.e1155.

Kuhls, K. et al. (2013). Population Structure and Evidence for Both Clonality and Recombination among Brazilian Strains of the Subgenus Leishmania (Viannia). *PLoS neglected tropical diseases*, 7 (10), p.e2490.

Kuo, C.-H. and Ochman, H. (2009). Deletional bias across the three domains of life. *Genome biology and evolution*, 1, pp.145–152.

Labbé, F. et al. (2023). Genomic analysis of two phlebotomine sand fly vectors of Leishmania from the New and Old World. *PLoS neglected tropical diseases*, 17 (4), p.e0010862.

Lachaud, L. et al. (2014). Constitutive mosaic aneuploidy is a unique genetic feature widespread in the Leishmania genus. *Microbes and infection*, 16 (1), pp.61–66.

Lang, D. et al. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *GigaScience*, 9 (12). [Online]. Available at: doi:10.1093/gigascience/giaa123.

Lauletta Lindoso, J. A. et al. (2016). Leishmaniasis–HIV coinfection: current challenges. *HIV/AIDS*, 8, pp.147–156.

Lawson, D. J. et al. (2012). Inference of population structure using dense haplotype data. *PLoS genetics*, 8 (1), p.e1002453.

Leonel, J. A. F. et al. (2024). Species, natural *Leishmania* spp. Detection and blood meal sources of Phlebotomine sandflies (Diptera: Psychodidae: Phlebotominae) in

peridomiciles from a leishmaniases endemic area of Brazil. *Transboundary and emerging diseases*, 2024 (1). [Online]. Available at: doi:10.1155/2024/9932530.

Liu, D., Hunt, M. and Tsai, I. J. (2018). Inferring synteny between genome assemblies: a systematic evaluation. *BMC bioinformatics*, 19 (1). [Online]. Available at: doi:10.1186/s12859-018-2026-4.

Liu, L., Wu, S. and Yu, L. (2015). Coalescent methods for estimating species trees from phylogenomic data. *Journal of systematics and evolution*, 53 (5), pp.380–390.

Llanes, A. et al. (2022). Genomic diversity and genetic variation of Leishmania panamensis within its endemic range. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 103 (105342), p.105342.

Logsdon, G. A., Vollger, M. R. and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature reviews. Genetics*, 21 (10), pp.597–614.

Lomsadze, A. et al. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic acids research*, 33 (20), pp.6494–6506.

Louradour, I. et al. (2020a). In vitro generation of Leishmania hybrids. *Cell reports*, 31 (2), p.107507.

Louradour, I. et al. (2020b). Leishmania hybridization mammalian hosts. *Cell Host & Microbe*, 28 (5), pp.788–800.

Louradour, I. et al. (2022). Stress conditions promote Leishmania hybridization in vitro marked by expression of the ancestral gamete fusogen HAP2 as revealed by single-cell RNA-seq. *eLife*, 11. [Online]. Available at: doi:10.7554/elife.73488.

Lukeš, J. et al. (2018). Trypanosomatids are much more than just trypanosomes: Clues from the expanded family tree. *Trends in parasitology*, 34 (6), pp.466–480.

Lye, L.-F. et al. (2010). Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathogens*, 6 (10), p.e1001161.

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. *Annual review of microbiology*, 60 (1), pp.327–349.

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)*, 290 (5494), American Association for the Advancement of Science (AAAS)., pp.1151–1155.

Lypaczewski, P. et al. (2018). A complete Leishmania donovani reference genome identifies novel genetic variations associated with virulence. *Scientific reports*, 8 (1). [Online]. Available at: doi:10.1038/s41598-018-34812-x.

Lypaczewski, P. and Matlashewski, G. (2021). Leishmania donovani hybridisation and introgression in nature: a comparative genomic investigation. *The Lancet. Microbe*, 2 (6), pp.e250–e258.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic biology*, 46 (3), pp.523–536.

Manni, M. et al. (2021). BUSCO: Assessing genomic data quality and beyond. *Current protocols*, 1 (12). [Online]. Available at: doi:10.1002/cpz1.323.

Marcili, A. et al. (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: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 25, pp.44–51.

Marques, C. A. et al. (2015). Genome-wide mapping reveals single-origin chromosome replication in Leishmania, a eukaryotic microbe. *Genome biology*, 16 (1). [Online]. Available at: doi:10.1186/s13059-015-0788-9.

Mascher, M. et al. (2021). Long-read sequence assembly: a technical evaluation in barley. *The plant cell*, 33 (6), pp.1888–1906.

Matos, G. M. et al. (2022). Microevolution of Trypanosoma cruzi reveals hybridization and clonal mechanisms driving rapid genome diversification. *eLife*, 11 (e75237). [Online]. Available at: doi:10.7554/eLife.75237.

Mauricio, I. L. et al. (2007). Glycoprotein 63 (gp63) genes show gene conversion and reveal the evolution of Old World Leishmania. *International journal for parasitology*, 37 (5), pp.565–576.

McCutcheon, J. P. and Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nature reviews. Microbiology*, 10 (1), pp.13–26.

McDowell, M. A. and Robichaud, J. (2024). *Concepts in Animal Parasitology, Chapter 12: Leishmania (Genus) and Leishmaniasis*. [Online]. Available at: https://digitalcommons.unl.edu/parasittext/27/ [Accessed 22 July 2024].

McInerney, J. O., McNally, A. and O'Connell, M. J. (2017). Why prokaryotes have pangenomes. *Nature microbiology*, 2 (4), p.17040.

Meireles, C. B. et al. (2017). Atypical presentations of cutaneous leishmaniasis: A systematic review. *Acta tropica*, 172, pp.240–254.

Melcon-Fernandez, E. et al. (2023). Miltefosine and nifuratel combination: A promising therapy for the treatment of Leishmania donovani visceral leishmaniasis. *International journal of molecular sciences*, 24 (2), p.1635.

Michalsky, É. M. et al. (2024). Abundance of *Leishmania* (Kinetoplastida: Trypanosomatidae) vectors in Baldim, a municipality in the Espinhaço Range Biosphere Reserve in Brazil. *Journal of medical entomology*, 61 (5), pp.1126–1139.

Mira, A., Ochman, H. and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends in genetics: TIG*, 17 (10), pp.589–596.

Monge-Maillo, B. and López-Vélez, R. (2015). Miltefosine for visceral and cutaneous leishmaniasis: Drug characteristics and evidence-based treatment

recommendations. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 60 (9), pp.1398–1404.

Monte-Neto, R. L., Fernandez-Prada, C. and Moretti, N. S. (2022). Sex under pressure: stress facilitates Leishmania in vitro hybridization. *Trends in parasitology*, 38 (4), pp.274–276.

Morris, J. J. et al. (2011). Dependence of the Cyanobacterium Prochlorococcus on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PloS one*, 6 (2), p.e16805.

Morris, J. J., Lenski, R. E. and Zinser, E. R. (2012). The Black Queen Hypothesis: Evolution of dependencies through adaptive gene loss. *mBio*, 3 (2). [Online]. Available at: doi:10.1128/mbio.00036-12.

Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York, NY: Oxford University Press.

Nicoll, A. G. et al. (2024). Estimating the rate of sexual reproduction and the inbreeding rate in*Leishmania*. *bioRxiv*. [Online]. Available at: doi:10.1101/2024.10.15.618469.

Ochieng, J. W. et al. (2007). Microsatellites retain phylogenetic signals across genera in eucalypts (Myrtaceae). *Genetics and molecular biology*, 30 (4), pp.1125–1134.

O'Connor, R. E. et al. (2018). Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome biology*, 19 (1). [Online]. Available at: doi:10.1186/s13059-018-1550-x.

Odiwuor, S. et al. (2011). Natural Leishmania donovani/Leishmania aethiopica hybrids identified from Ethiopia. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 11 (8), pp.2113–2118.

Ohno, S. (2013). *Evolution by gene duplication*. 1970th ed. Berlin, Germany: Springer. [Online]. Available at: doi:10.1007/978-3-642-86659-3.

Opperdoes, F. R. and Lemey, P. (2009). Phylogenetic analysis using protein sequences. In: Lemey, P., Salemi, M. and Vandamme, A.-M. (Eds). *The Phylogenetic Handbook*. Cambridge: Cambridge University Press. pp.313–342. [Accessed 1 March 2026].

Osman, M. et al. (2017). A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63-KH. *PLoS neglected tropical diseases*, 11 (5), p.e0005527.

Paiva, S. C. et al. (2025). Exploring the structure and nucleic acid interactions of the Leishmania sp. telomerase reverse transcriptase N-terminal region. *Archives of biochemistry and biophysics*, 764 (110289), p.110289.

Pan American Health Organization. (2024). Leishmaniasis: Epidemiological report for the Americas. No.13, December 2024. *Leishmaniases: Epidemiological Report of*

the Americas; [Online]. Available at: https://iris.paho.org/handle/10665.2/63165 [Accessed 6 June 2025].

Pareyn, M. et al. (2019). Ecology and seasonality of sandflies and potential reservoirs of cutaneous leishmaniasis in Ochollo, a hotspot in southern Ethiopia. *PLoS neglected tropical diseases*, 13 (8), p.e0007667.

Pasquali, A. K. S. et al. (2019). Dispersion of Leishmania (Leishmania) infantum in central-southern Brazil: Evidence from an integrative approach. *PLoS neglected tropical diseases*, 13 (8), p.e0007639.

Patino, L. H. et al. (2020). Genomic Diversification, Structural Plasticity, and Hybridization in Leishmania (Viannia) braziliensis. *Frontiers in cellular and infection microbiology*, 10. [Online]. Available at: doi:10.3389/fcimb.2020.582192.

Peacock, C. S. et al. (2007). Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nature genetics*, 39 (7), pp.839–847.

Peacock, L. et al. (2011). Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (9), pp.3671–3676.

Peacock, L. et al. (2014). Meiosis and haploid gametes in the pathogen Trypanosoma brucei. *Current Biology*, 24 (2), pp.181–186.

Philippe, H. et al. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, 9 (3), p.e1000602.

Pilling, O. A. et al. (2023). Selective whole-genome amplification reveals population genetics of Leishmania braziliensis directly from patient skin biopsies. *PLoS pathogens*, 19 (3), p.e1011230.

Pita, S. et al. (2019). The tritryps comparative repeatome: Insights on repetitive element evolution in trypanosomatid pathogens. *Genome biology and evolution*, 11 (2), pp.546–551.

Porras-Hurtado, L. et al. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in genetics*, 4, p.98.

Ready, P. (2014). Epidemiology of visceral leishmaniasis. *Clinical epidemiology*, p.147.

Real, F. et al. (2013). The genome sequence of Leishmania (Leishmania) amazonensis: functional annotation and extended analysis of gene models. *DNA research: an international journal for rapid publication of reports on genes and genomes*, 20 (6), pp.567–581.

Reithinger, R. et al. (2007). Cutaneous leishmaniasis. *The Lancet infectious diseases*, 7 (9), pp.581–596.

Rhie, A. et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592 (7856), pp.737–746.

Roach, M. J., Schmidt, S. A. and Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics*, 19 (1), p.460.

Rogers, M. B. et al. (2011). Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome research*, 21 (12), pp.2129–2142.

Rogers, M. B. et al. (2014). Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated Leishmania population. *PLoS genetics*, 10 (1), p.e1004092.

Romano, A. et al. (2014). Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (47), pp.16808–16813.

Rougeron, V. et al. (2009). Extreme inbreeding in *Leishmania braziliensis*. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (25), pp.10224–10229.

Rougeron, V., De Meeûs, T. and Bañuls, A.-L. (2017). Reproduction in Leishmania : A focus on genetic exchange. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 50, pp.128–132.

Ryan, S. et al. (2021). Advances parasite genomics: from sequences genotype-phenotype associations. *Current Opinion Microbiology*, 63, pp.106–114.

S. L. Figueiredo de Sá, B. et al. (2019). Identification of divergent Leishmania (Viannia) braziliensis ecotypes derived from a geographically restricted area through whole genome analysis. *PLoS neglected tropical diseases*, 13 (6), p.e0007382.

Safar, H. A., Alatar, F. and Mustafa, A. S. (2024). Three rounds of read correction significantly improve eukaryotic protein detection in ONT reads. *Microorganisms*, 12 (2). [Online]. Available at: doi:10.3390/microorganisms12020247.

Saidi, N. et al. (2023). Clinical and immunological spectra of human cutaneous leishmaniasis in North Africa and French Guiana. *Frontiers in immunology*, 14, p.1134020.

Salomón, O. D. et al. (2012). Leishmaniasis and climate change-case study: Argentina. *Journal of Tropical Medicine*, 2012, p.601242.

Saporito, L. et al. (2013). Visceral leishmaniasis: host–parasite interactions and clinical presentation in the immunocompetent and in the immunocompromised host. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, 17 (8), pp.e572–e576.

Satragno, D. et al. (2017). Autochthonous outbreak and expansion of canine visceral leishmaniasis, Uruguay. *Emerging infectious diseases*, 23 (3), pp.536–538.

Schardl, C. L. and Craven, K. D. (2003). Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Molecular ecology*, 12 (11), pp.2861–2873.

Schwabl, P. et al. (2021). Colonization and genetic diversification processes of Leishmania infantum in the Americas. *Communications biology*, 4 (1), pp.1–13. [Accessed 17 July 2024].

Scorza, B. M., Carvalho, E. M. and Wilson, M. E. (2017). Cutaneous manifestations of human and Murine leishmaniasis. *International journal of molecular sciences*, 18 (6). [Online]. Available at: doi:10.3390/ijms18061296.

Serafim, T. D. et al. (2023). Leishmania genetic exchange is mediated by IgM natural antibodies. *Nature*, 623 (7985), pp.149–156.

Shaik, J. S. et al. (2021). Leishmania sexual reproductive strategies as resolved through computational methods designed for aneuploid genomes. *Genes*, 12 (2), p.167.

Shapiro, B. J. and Polz, M. F. (2014). Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in microbiology*, 22 (5), pp.235–247.

Sharma, P. et al. (2021). Improvements in the sequencing and assembly of plant genomes. *bioRxiv*. [Online]. Available at: doi:10.1101/2021.01.22.427724.

Shaw, J. J. (2007). The leishmaniases-survival and expansion in a changing world. A mini-review. *Memórias do Instituto Oswaldo Cruz*, 102, pp.541–547.

Shen, W. et al. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS one*, 11 (10), p.e0163962.

Silva, M. S. et al. (2018). Differential regulation Toll-like receptor 2 Trypanosoma cruzi-associated myocarditis. *Frontiers Immunology*, 9.

Silva, T. F. et al. (2021). Phenotypical and genotypical differences among Leishmania (Leishmania) amazonensis isolates that caused different clinical frames in humans and dogs: A systematic review. *Acta tropica*, 221 (106018), p.106018.

Silveira, F. T. et al. (2010). A prospective study on the dynamics of the clinical and immunological evolution of human Leishmania (L.) infantum chagasi infection in the Brazilian Amazon region. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104 (8), pp.529–535.

Silveira, F. T. et al. (2021). Whole-Genome Sequencing of Leishmania infantum chagasi Isolates from Honduras and Brazil. *Microbiology resource announcements*, 10 (48). [Online]. Available at: doi:10.1128/mra.00471-21.

Simão, F. A. et al. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, 31 (19), pp.3210–3212.

Smith, D. F., Peacock, C. S. and Cruz, A. K. (2007). Comparative genomics: from genotype to disease phenotype in the leishmaniases. *International journal for parasitology*, 37 (11), pp.1173–1186.

Soares, G. H. C. et al. (2024). Isolation and characterisation of Leishmania (Leishmania) infantum from cutaneous leishmaniasis patients in northeast Brazil. *Memorias do Instituto Oswaldo Cruz*, 119, p.e240026.

Soltis, P. S. and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review of plant biology*, 60 (1), pp.561–588.

de Souza Fernandes, W. et al. (2022). Phlebotomine Sandfly (Diptera: Psychodidae) fauna and the association between climatic variables and the abundance of *Lutzomyia longipalpis* sensu lato in an intense transmission area for visceral leishmaniasis in central western Brazil. *Journal of medical entomology*, 59 (3), pp.997–1007.

Stanke, M. et al. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics (Oxford, England)*, 24 (5), pp.637–644.

Sterkers, Y. et al. (2011). FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in Leishmania major. *Cellular microbiology*, 13 (2), pp.274–283.

Sterkers, Y. et al. (2014). Parasexuality and mosaic aneuploidy in Leishmania: alternative genetics. *Trends in parasitology*, 30 (9), pp.429–435.

Stevens, J. et al. (2001). The molecular evolution of trypanosomatidae. In: *Advances in Parasitology*. Advances in parasitology. Elsevier. pp.1–53.

Sundar, S. and Chakravarty, J. (2015). An update on pharmacotherapy for leishmaniasis. *Expert opinion on pharmacotherapy*, 16 (2), pp.237–252.

Sundar, S. and Rai, M. (2002). Advances in the treatment of leishmaniasis. *Current opinion in infectious diseases*, 15 (6), pp.593–598.

Teles, C. B. G. et al. (2016). Phlebotomine sandfly (Diptera: Psychodidae) diversity and their Leishmania DNA in a hot spot of American Cutaneous Leishmaniasis human cases along the Brazilian border with Peru and Bolivia. *Memorias do Instituto Oswaldo Cruz*, 111 (7), pp.423–432.

Thomaz-Soccol, V. et al. (2023). One piece of the puzzle: Modeling vector presence and environment reveals seasonality, distribution, and prevalence of sandflies and Leishmania in an expansion area. *One health (Amsterdam, Netherlands)*, 17 (100581), p.100581.

Thrash, A., Hoffmann, F. and Perkins, A. (2020). Toward a more holistic method of genome assembly assessment. *BMC bioinformatics*, 21 (S4). [Online]. Available at: doi:10.1186/s12859-020-3382-4.

Tibayrenc, M. and Ayala, F. J. (2015). The population genetics of Trypanosoma cruzi revisited in the light of the predominant clonal evolution model. *Acta tropica*, 151, pp.156–165.

Tibayrenc, M. and Ayala, F. J. (2021). Leishmania and the model of predominant clonal evolution. *Microorganisms*, 9 (11), p.2409.

Tirera, S. et al. (2017). Unraveling the genetic diversity and phylogeny of Leishmania RNA virus 1 strains of infected Leishmania isolates circulating in French Guiana. *PLoS neglected tropical diseases*, 11 (7), p.e0005764.

Torres-Guerrero, E. et al. (2017). Leishmaniasis: a review. *F1000Research*, 6, p.750.

Tyler, K. M. and Engman, D. M. (2001). The life cycle of Trypanosoma cruzi revisited. *International journal for parasitology*, 31 (5–6), pp.472–481.

Valdivia, H. O. et al. (2015a). Comparative genomic analysis of Leishmania (Viannia) peruviana and Leishmania (Viannia) braziliensis. *BMC genomics*, 16 (1), p.715.

Valdivia, H. O. et al. (2015b). The Leishmania metaphylome: a comprehensive survey of Leishmania protein phylogenetic relationships. *BMC genomics*, 16 (1), p.887.

Van den Broeck, F. et al. (2020). Ecological divergence and hybridization of Neotropical Leishmania parasites. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (40), pp.25159–25168.

Van den Broeck, F. et al. (2023). Genome analysis of triploid hybrid *Leishmania* parasite from the neotropics. *Emerging infectious diseases*, 29 (5). [Online]. Available at: doi:10.3201/eid2905.221456.

Van Oss, S. B. and Carvunis, A.-R. (2019). De novo gene birth. *PLoS genetics*, 15 (5), p.e1008160.

Venter, J. C., Smith, H. O. and Adams, M. D. (2015). The sequence of the human genome. *Clinical chemistry*, 61 (9), pp.1207–1208.

Vernikos, G. et al. (2015). Ten years of pan-genome analyses. *Current opinion in microbiology*, 23, pp.148–154.

Volpedo, G. et al. (2021). Mechanisms of immunopathogenesis in cutaneous leishmaniasis and Post Kala-azar Dermal Leishmaniasis (PKDL). *Frontiers in cellular and infection microbiology*, 11, p.685296.

Wallberg, A. et al. (2019). A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. *BMC genomics*, 20 (1). [Online]. Available at: doi:10.1186/s12864-019-5642-0.

Wang, P. and Wang, F. (2023). A proposed metric set for evaluation of genome assembly quality. *Trends in genetics: TIG*, 39 (3), pp.175–186.

Wick, R. R. and Holt, K. E. (2021). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research*, 8, p.2138.

Wilson, M. E., Jeronimo, S. M. B. and Pearson, R. D. (2005). Immunopathogenesis of infection with the visceralizing Leishmania species. *Microbial pathogenesis*, 38 (4), pp.147–160.

Wolf, Y. I. and Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 35 (9), pp.829–837.

World Health Organization. (2025). *Leishmaniasis*. [Online]. Available at: https://www.who.int/health-topics/leishmaniasis?utm_source=chatgpt.com [Accessed 23 May 2025].

Yadav, N. and Madke, B. (2023). Case report: Diffuse cutaneous leishmaniasis successfully treated with a combination of oral rifampicin and fluconazole. *The American journal of tropical medicine and hygiene*, 109 (2), pp.315–318.

Zakharova, A. et al. (2022). Leishmania guyanensis M4147 as a new LRV1-bearing model parasite: Phosphatidate phosphatase 2-like protein controls cell cycle progression and intracellular lipid content. *PLoS neglected tropical diseases*, 16 (6), p.e0010510.

Zhang, T. et al. (2022). Complex genome assembly based on long-read sequencing. *Briefings in bioinformatics*, 23 (5). [Online]. Available at: doi:10.1093/bib/bbac305.

Zhang, Z. et al. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochemical and biophysical research communications*, 419 (4), pp.779–781.

Ziaei Hezarjaribi, H. et al. (2021). Phylogenetic position of Leishmania tropica isolates from an old endemic focus in south-eastern Iran; relying on atypical cutaneous leishmaniasis. *Transboundary and emerging diseases*, 68 (3), pp.1493–1503.

Zijlstra, E. E. et al. (2003). Post-kala-azar dermal leishmaniasis. *The Lancet infectious diseases*, 3 (2), pp.87–98.

7. Supplementary materials

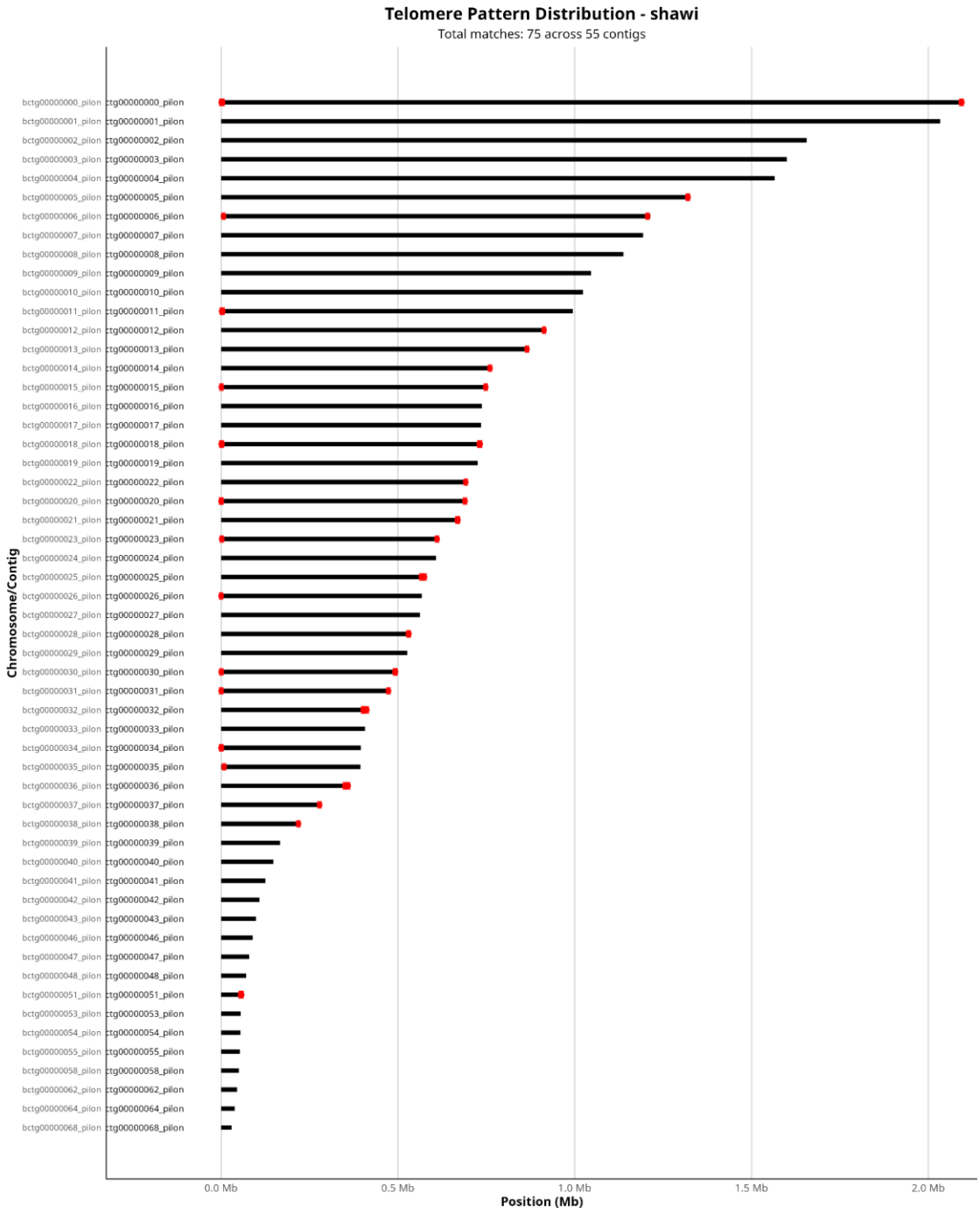


Figure S1. Telomere pattern distribution across chromosomes in *L. shawi*

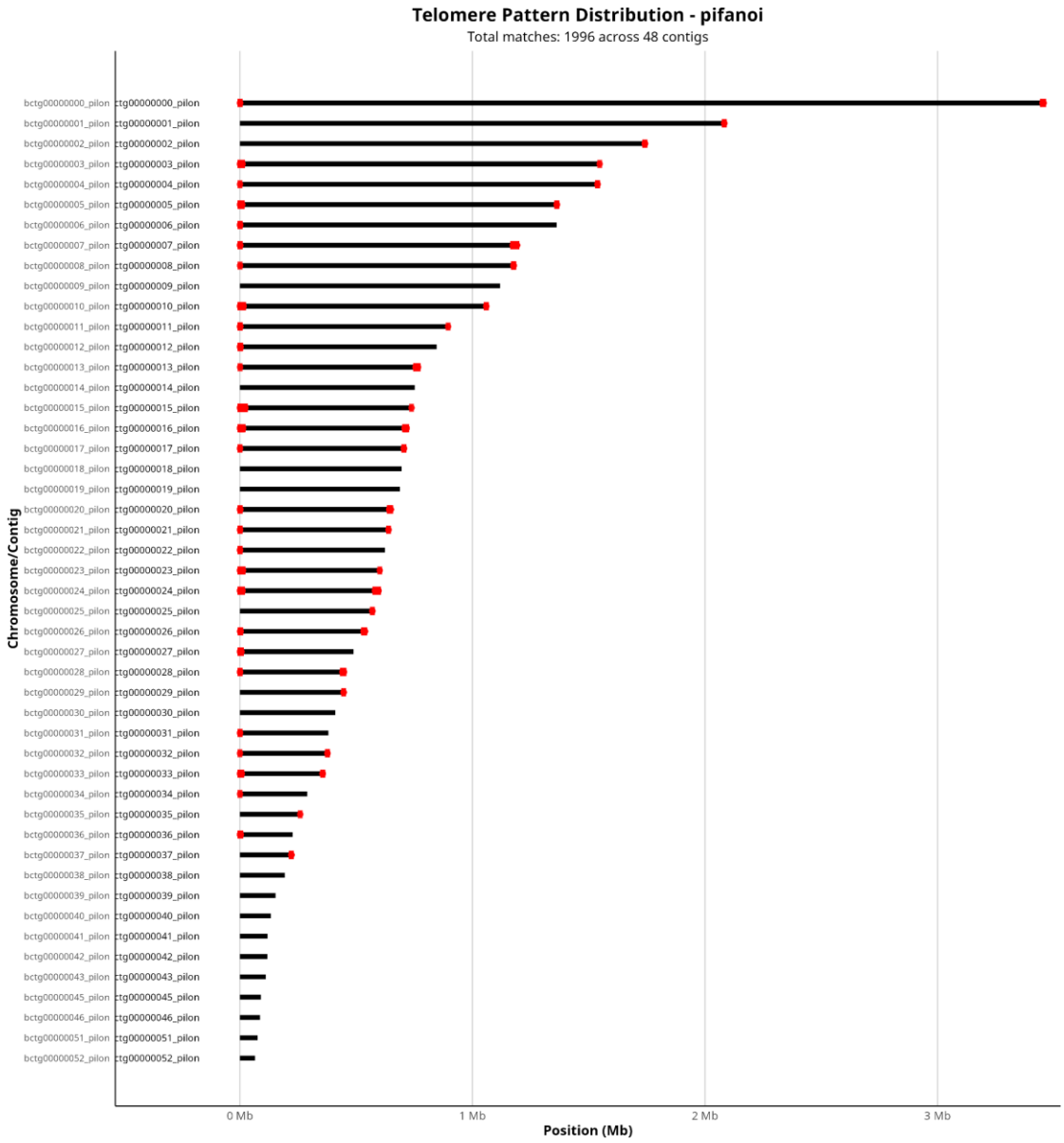


Figure S2. Telomere pattern distribution across chromosomes in *L. pifanoi*

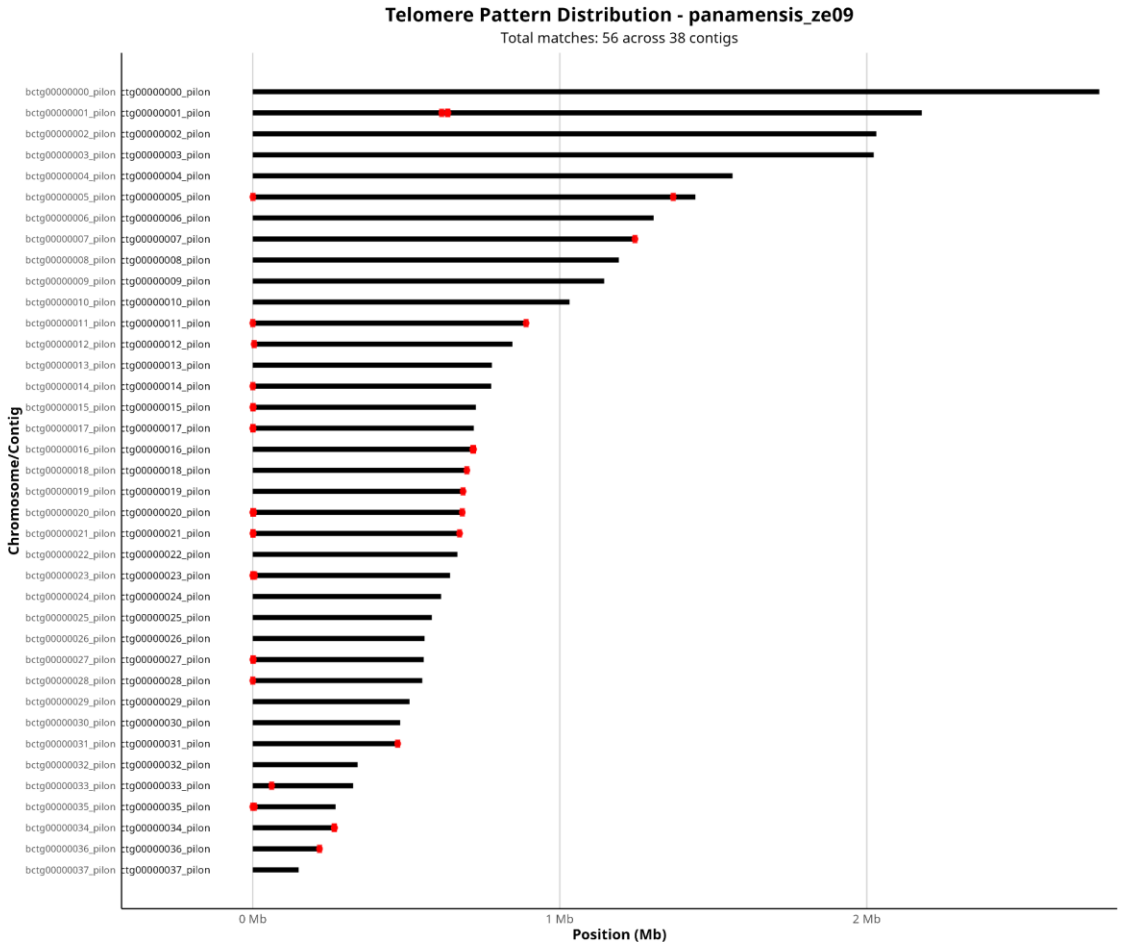


Figure S3. Telomere pattern distribution across chromosomes in *L. panamensis z309.*

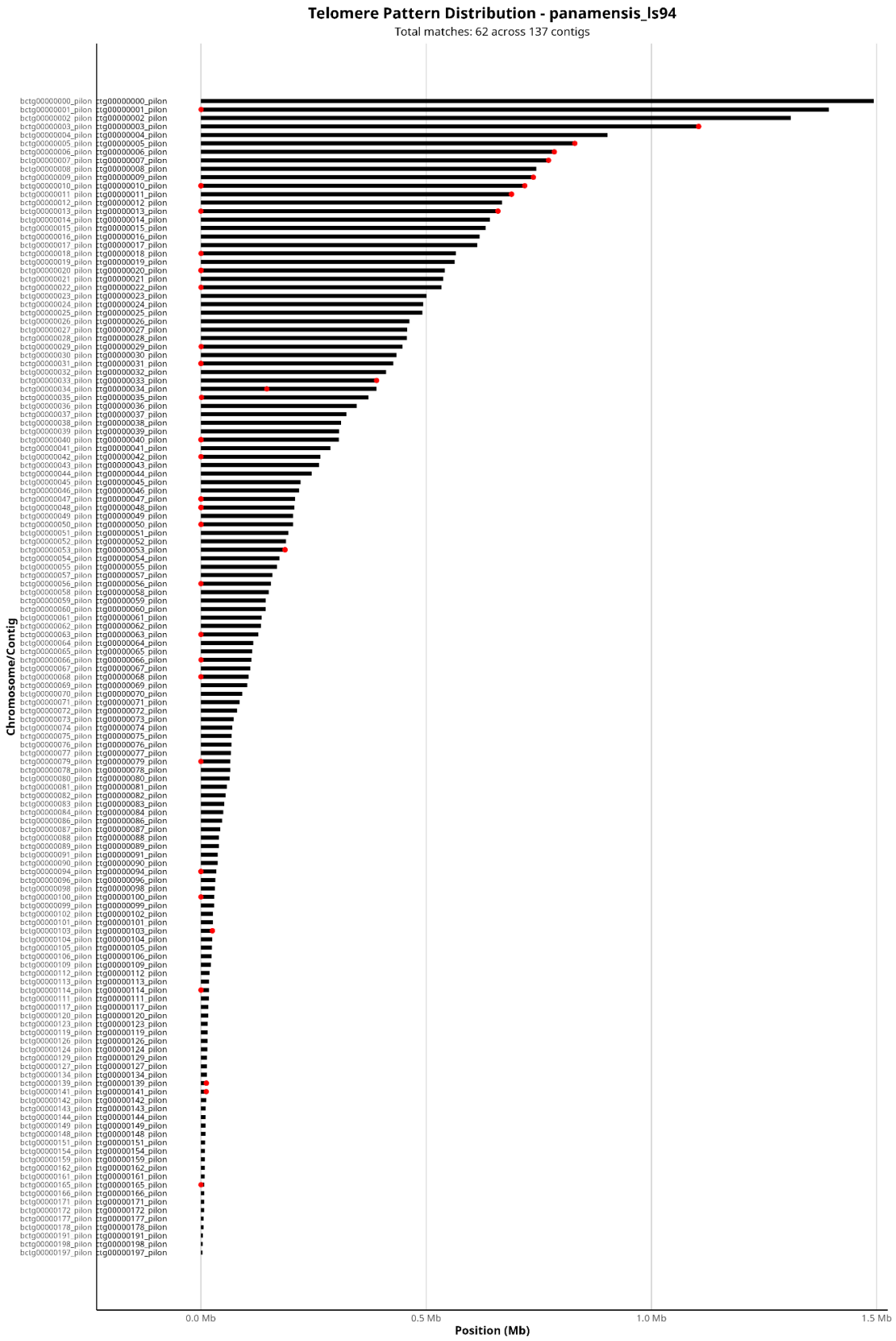


Figure S4. Telomere pattern distribution across chromosomes in *L. panamensis ls94*

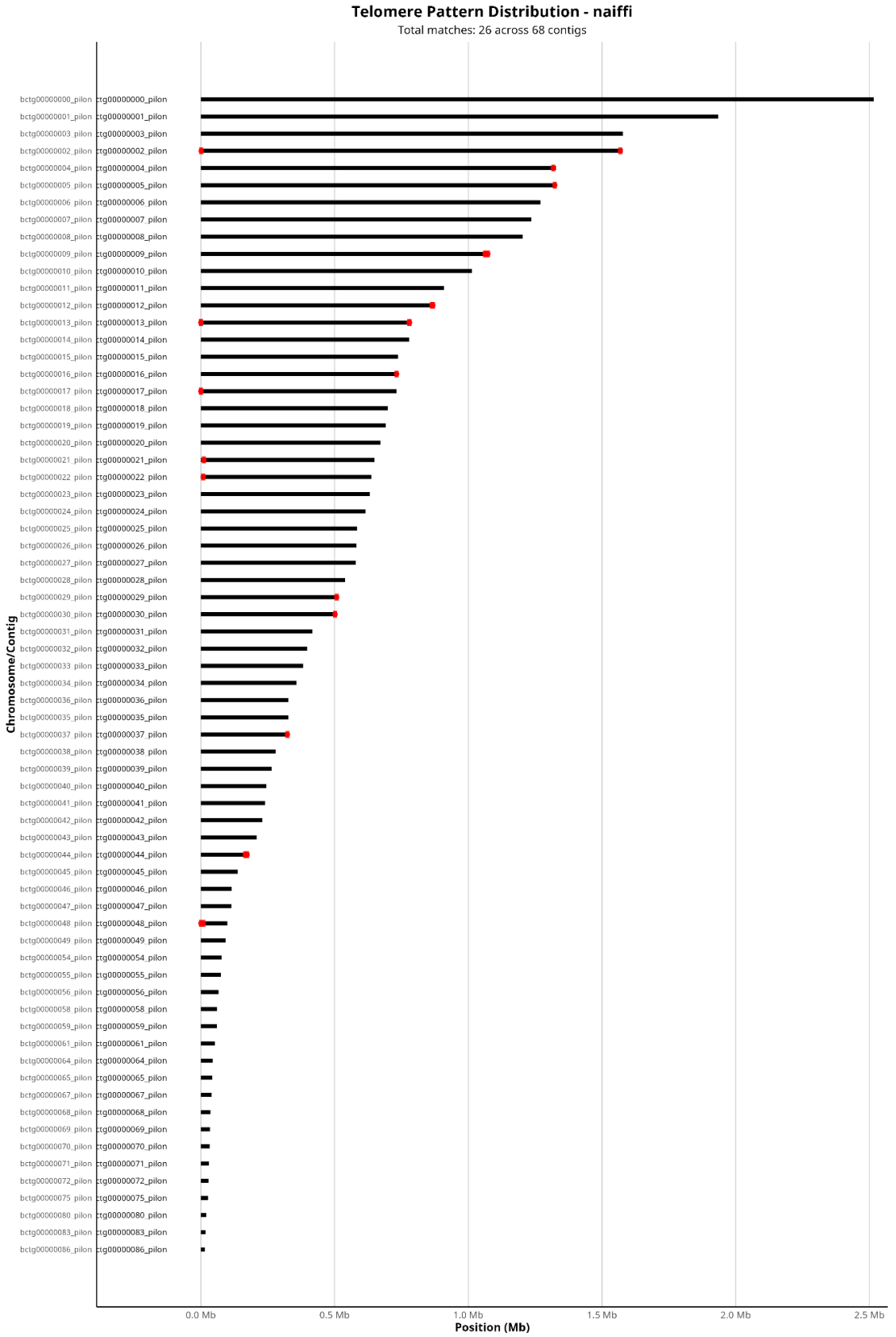


Figure S5. Telomere pattern distribution across chromosomes in *L. naiffi.*

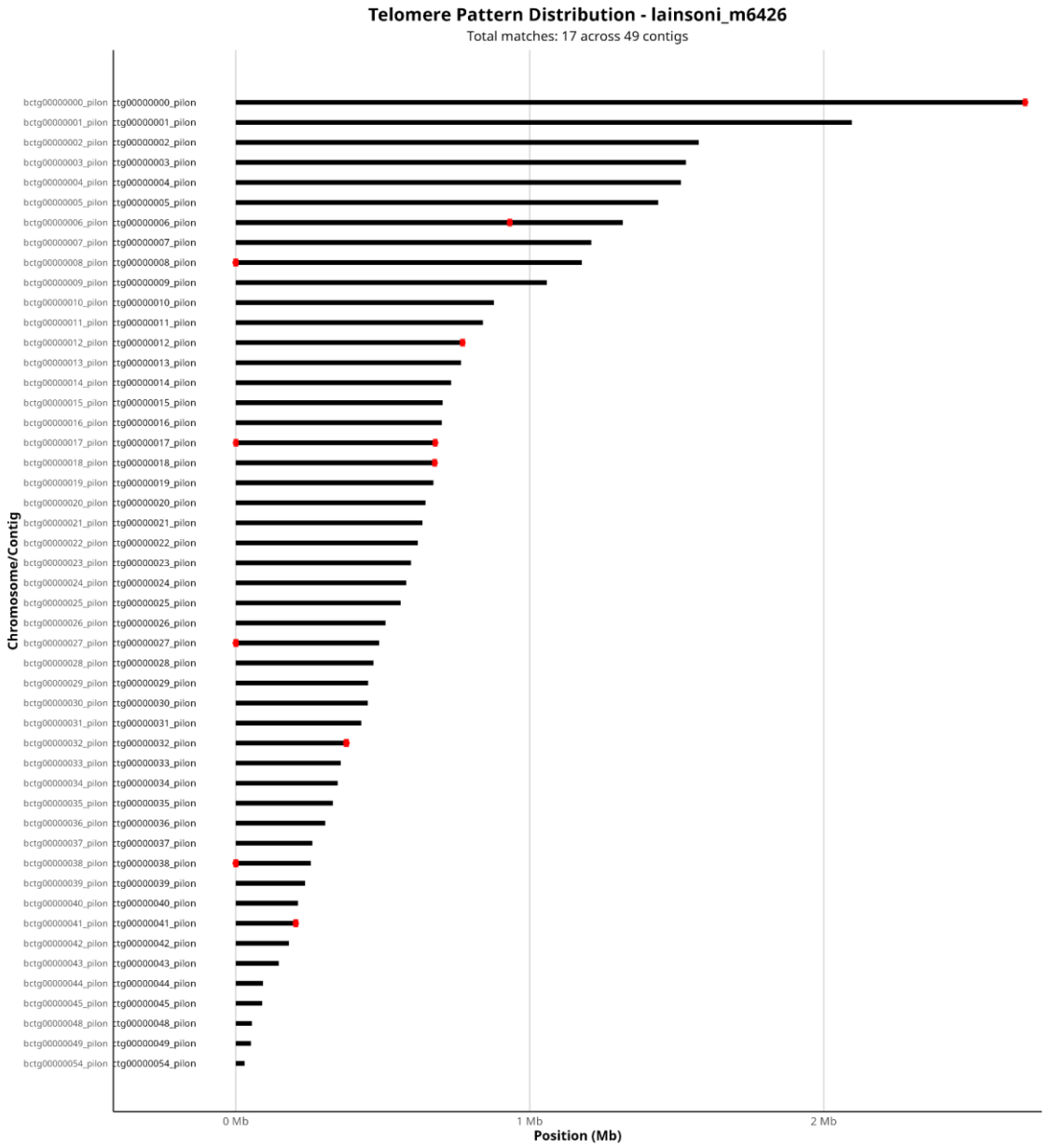


Figure S6. Telomere pattern distribution across chromosomes in *L. lainsoni m6426.*

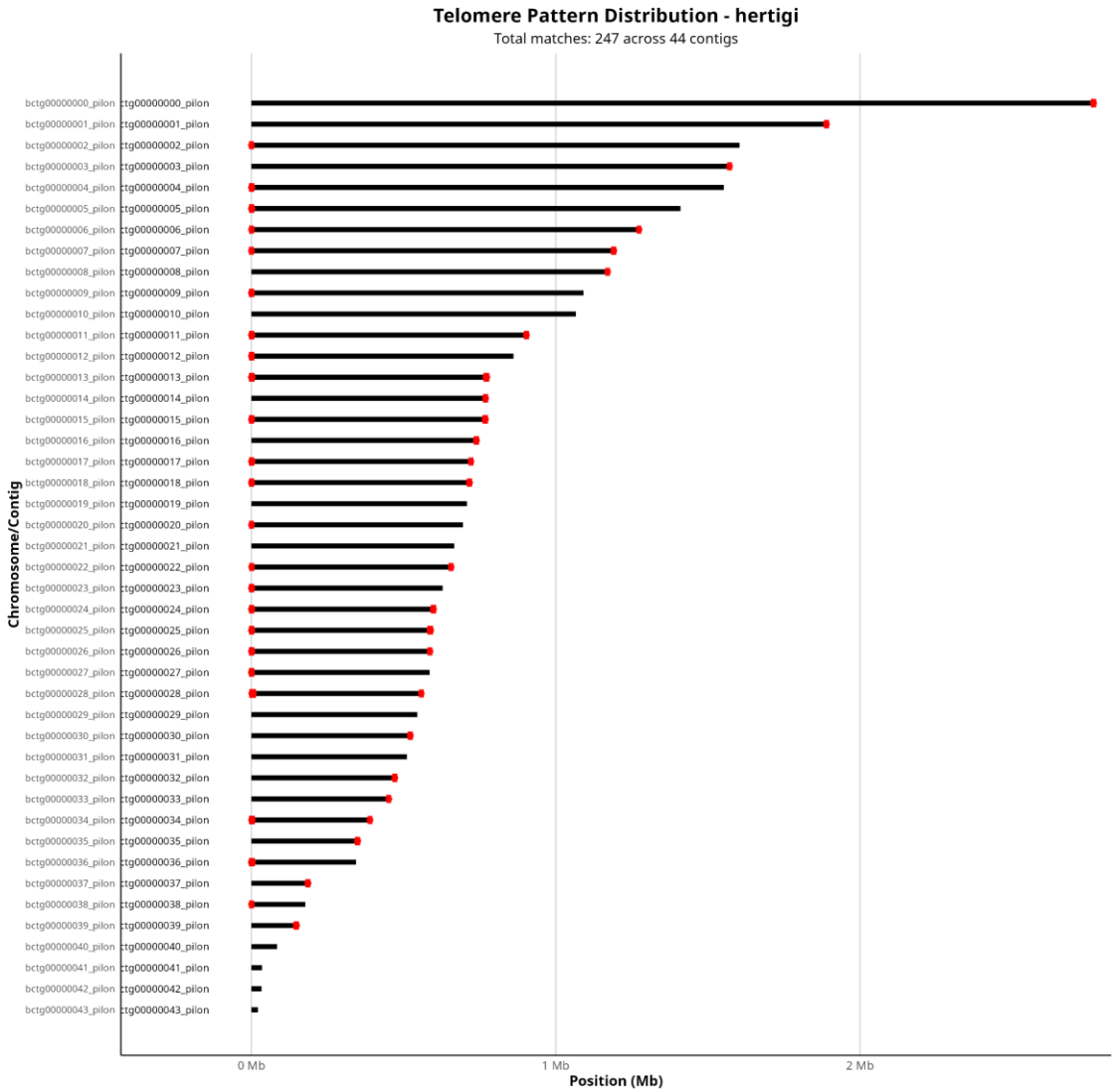


Figure S7. Telomere pattern distribution across chromosomes in *L. hertigi.*

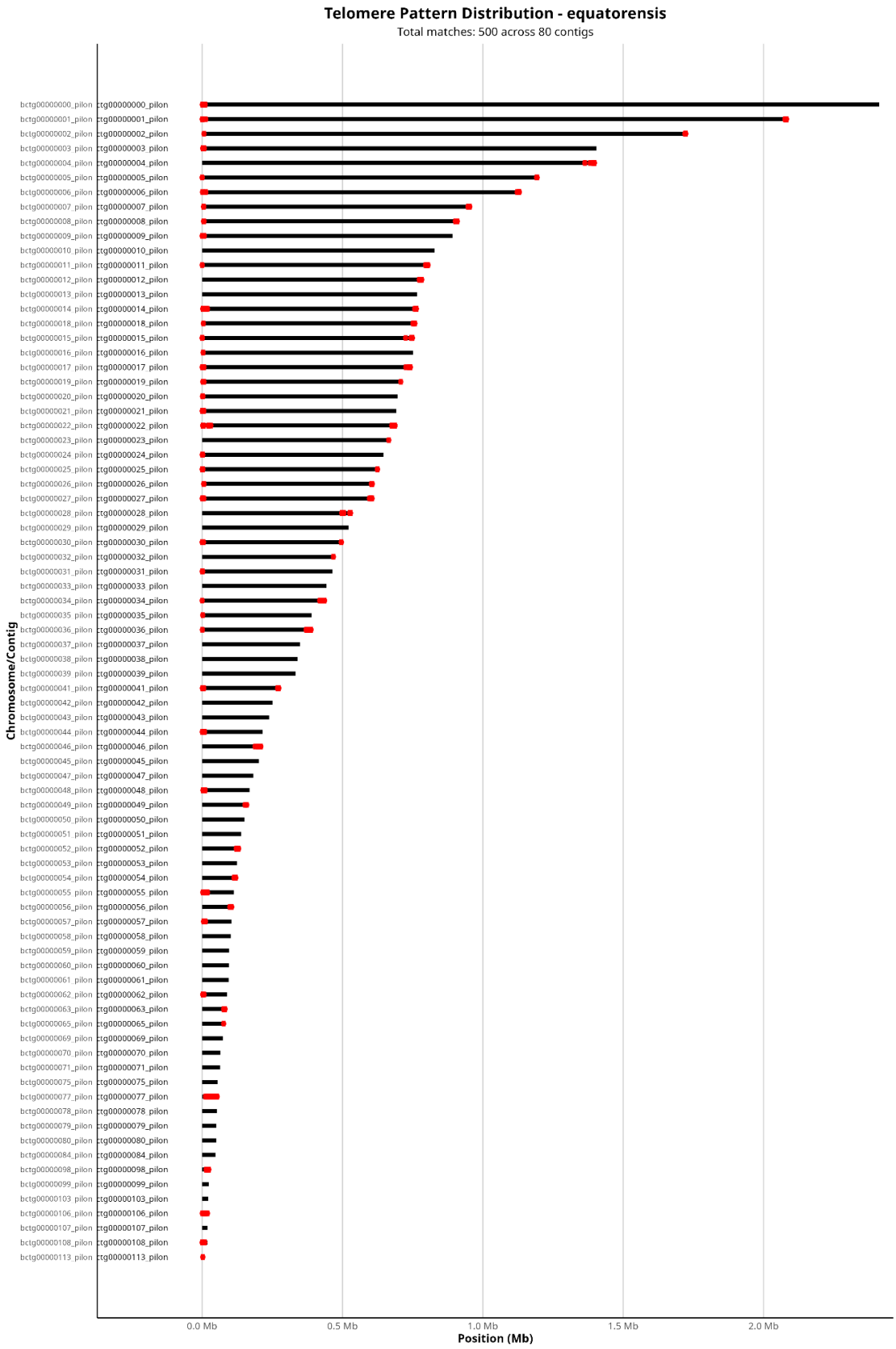


Figure S8. Telomere pattern distribution across chromosomes in *L. equatorensis*.

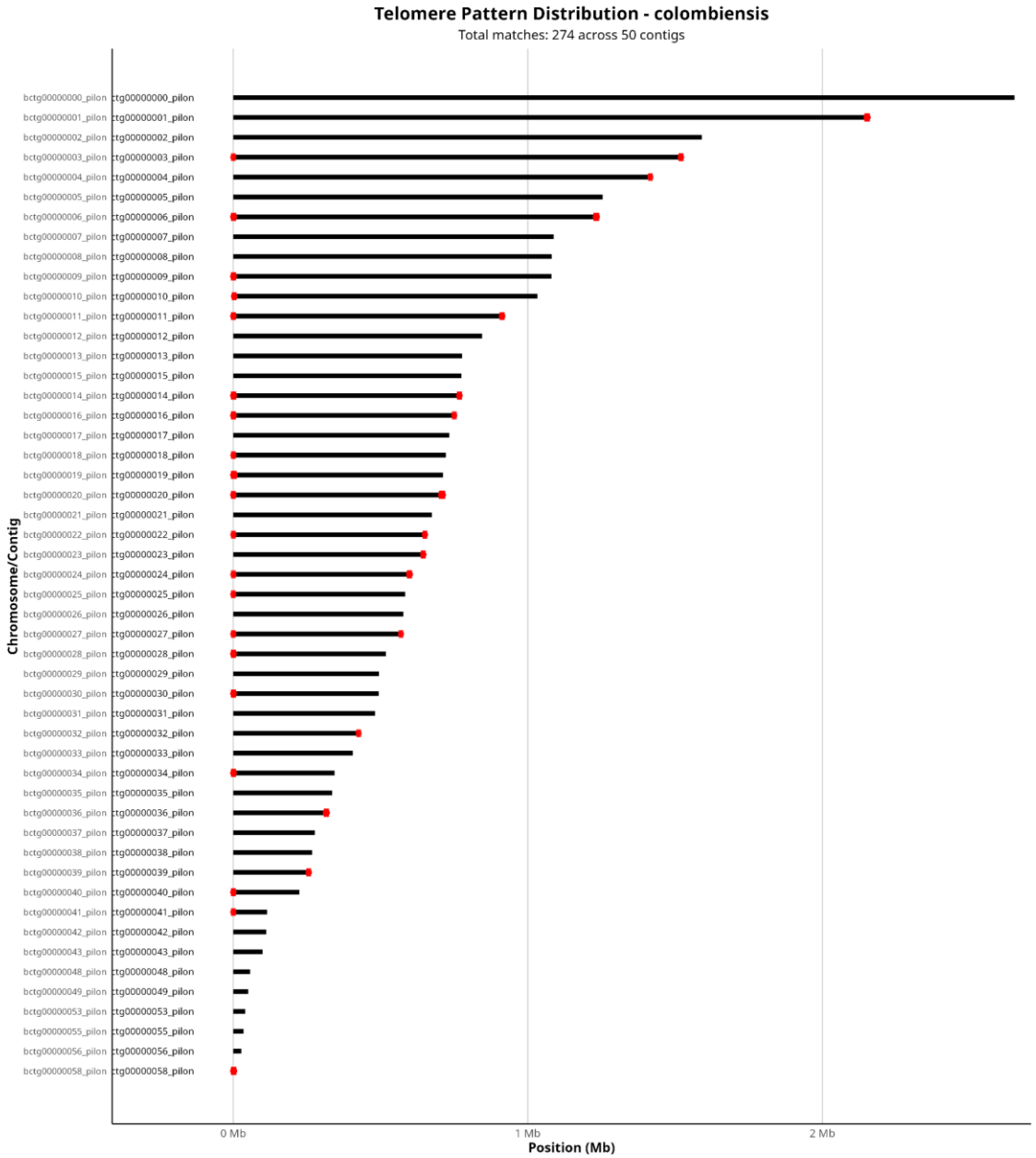


Figure S9. Telomere pattern distribution across chromosomes in *L. colombiensis*.

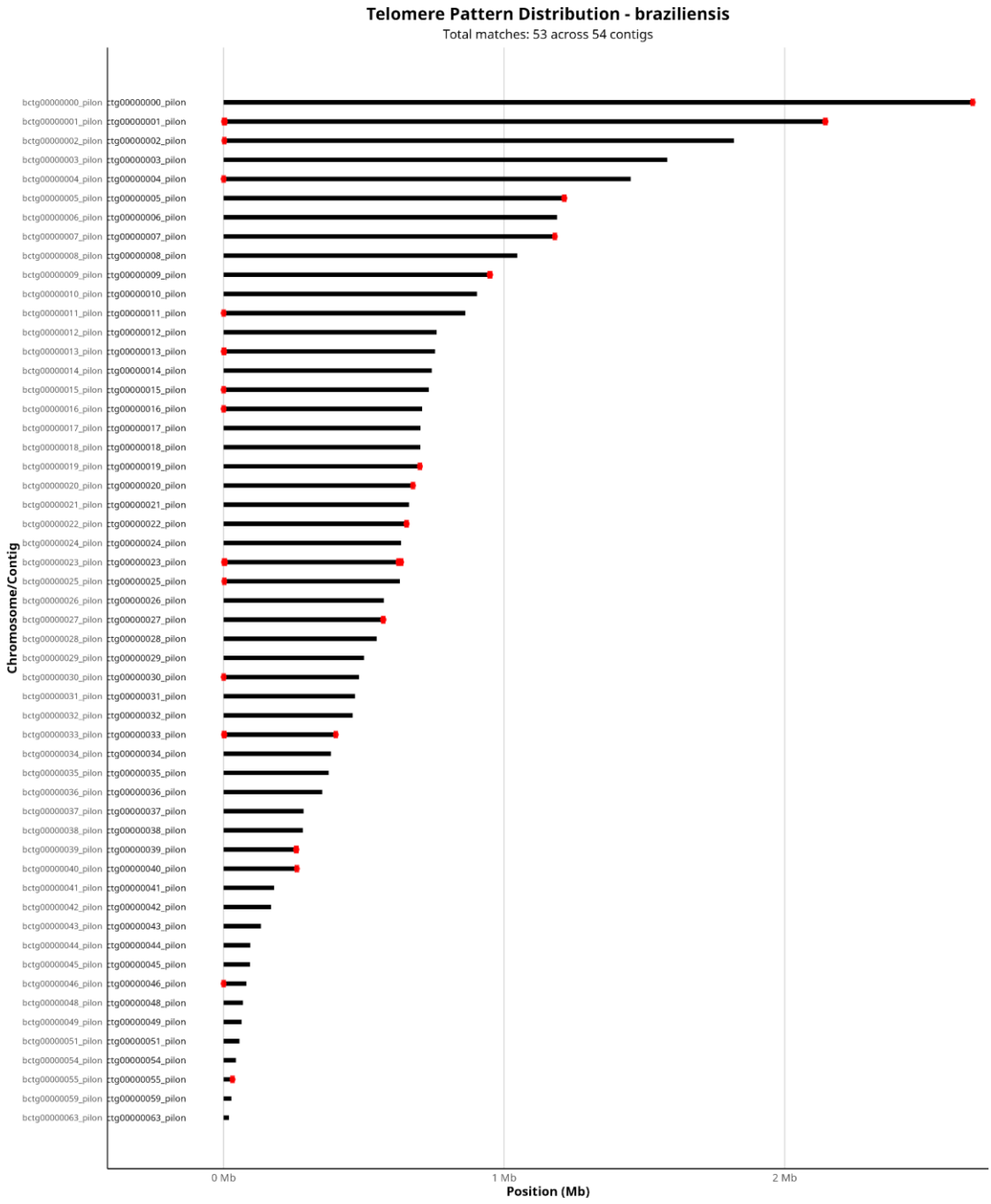


Figure S10. Telomere pattern distribution across chromosomes in *L. braziliensis*.

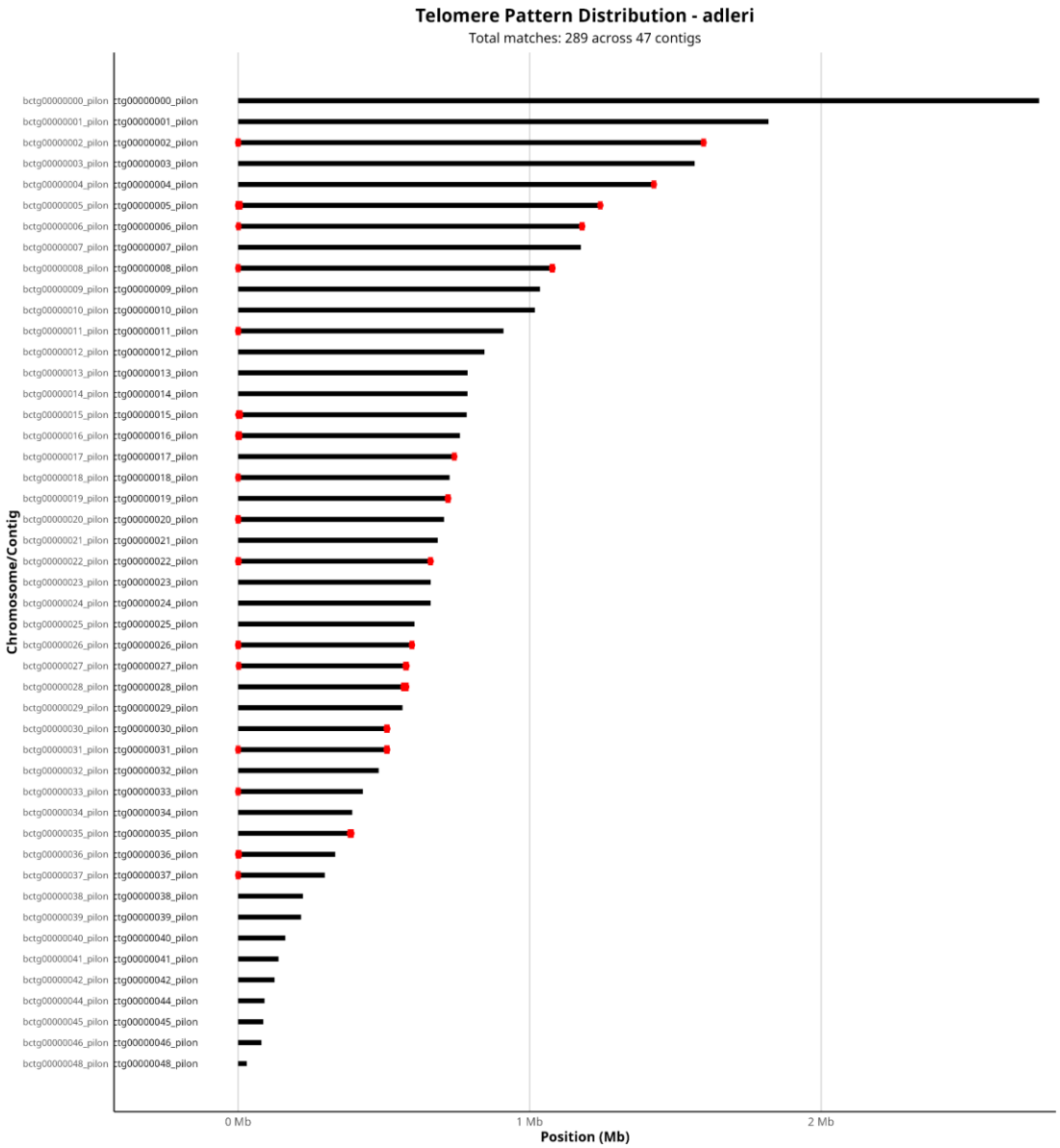


Figure S11. Telomere pattern distribution across chromosomes in *L. adleri.*

Table S1. Comparative analysis of orthologous gene families across trypanosomatid genomes.

This table presents a comprehensive comparative analysis of orthologous gene families identified through OrthoFinder analysis of 63 trypanosomatid genomes, including 29 *Leishmania* species (17 from the *Viannia* subgenus), 19 Trypanosoma species, and 15 other trypanosomatids. The analysis encompasses 16,683 orthogroups and examines multiple dimensions of genomic organization and evolutionary patterns.

Comparison Type	Group A	Group B	Metric	Value A	Value B	Ratio A to B	Significance	Biological Interpretation
Genome size	*Viannia* Average	Major Average	Gene Count	8400	8610	0.98	Minimal	*Viannia* and Major subgenera have similar genome sizes
Genome size	*Viannia* Average	Donovani Average	Gene Count	8400	8072	1.04	Minimal	*Viannia* slightly larger than Donovani complex
Genome size	*Leishmania* Average	Trypanosoma Average	Gene Count	8400	9200	0.91	Moderate	Trypanosoma genomes are ~9% larger than Leishmania
Genome size	All Leishmania	T cruzi	Gene Count	8400	31539	0.27	Extreme	T. cruzi genome is 3.75x larger than average Leishmania
Genome size	Largest *Viannia*	Smallest *Viannia*	Gene Count	9282	7151	1.3	Significant	30% variation within *Viannia* subgenus
Evolutionary conservation	Core Genes	Total Orthogroups	Percentage	5.06	100	0.05	Critical	Only 5% of genes are conserved across all species
Evolutionary conservation	Conserved Genes	Variable Genes	Percentage	26.13	73.87	0.35	Critical	Most genes (74%) are not universally conserved
Evolutionary conservation	Species Specific	Shared Genes	Percentage	20.11	79.89	0.25	Significant	1 in 5 orthogroups are unique to single species
Phylogenetic signal	Single Copy	Multi Copy	Count	76	16607	0.005	Limiting	Very few genes suitable for phylogenetic analysis
Phylogenetic signal	Single Copy	Total Orthogroups	Percentage	0.46	100	0.005	Limiting	Less than 0.5% of orthogroups are single-copy
Taxonomic diversity	*Viannia* Species	Other Leishmania	Count	17	12	1.42	High	*Viannia* subgenus well-represented in this study
Taxonomic diversity	New World Leishmania	Old World Leishmania	Count	20	9	2.22	High	New World bias in species representation

Taxonomic diversity	*Leishmania* Total	Trypanosoma Total	Count	29	19	1.53	Moderate	Good representation of both major genera
Assembly quality	This Study *Viannia*	Reference Leishmania	Average Genes	8400	8300	1.01	Excellent	This study assemblies comparable to references
Assembly quality	*Viannia* Range	Reference Range	Gene Count Variation	2131	1486	1.43	Moderate	Slightly more variation in this study
Biogeographic pattern	Amazon Basin	Other Regions	Species Count	8	9	0.89	Balanced	Good representation of Amazon diversity
Clinical relevance	Pathogenic Leishmania	Non Pathogenic	Count	25	4	6.25	High	Strong focus on medically important species
Clinical relevance	Cutaneous Causative	Visceral Causative	Count	17	8	2.13	Moderate	More cutaneous than visceral species
Genomic plasticity	Accessory Genes	Core Genes	Percentage	53.76	5.06	10.62	Extreme	Accessory genome 10x larger than core genome
Genomic plasticity	Variable Genes	Conserved Genes	Count	12324	4359	2.83	High	2.8x more variable than conserved genes
Species divergence	Unique Orthogroups	Shared Orthogroups	Percentage	20.11	79.89	0.25	Significant	Substantial species-specific gene content
Species divergence	*Viannia* Unique	*Viannia* Shared	Estimated Percentage	15	85	0.18	Moderate	Estimated 15% *Viannia*-specific genes

Table S2. Key for species names on Splitstree in figure 7. Including species, strain and source of CDS files.

Splitstree Name	Species	Strain	Source
BsalRef	Bodo saltans	Bodo saltans strain Lake Konstanz	TriTrypdb
LmartLSCM1Ref	*Leishmania* martiniquensis	MHOM/TH/2012/LSCM1	TriTrypdb
LmartLEM2494Ref	*Leishmania* martiniquensis	MHOM/MQ/92/MAR1; LEM2494	TriTrypdb
LoriRef	*Leishmania* orientalis	MHOM/TH/2014/LSCM4	TriTrypdb
LenriCUR178Ref	*Leishmania* enriettii	MCAV/BR/2001/CUR178	TriTrypdb
LenriLEM3045Ref	*Leishmania* enriettii	LEM3045	TriTrypdb
LtarRef	*Leishmania* tarentolae	Parrot-TarII	TriTrypdb
Ltar2019Ref	*Leishmania* tarentolae	Parrot Tar II 2019	TriTrypdb
LmajFriedfRef	*Leishmania* major	Friedlin	TriTrypdb
LmajFried2021fRef	*Leishmania* major	Friedlin 2021	TriTrypdb
LmajLV39c5Ref	*Leishmania* major	LV39c5	TriTrypdb
LarabRef	*Leishmania* arabica	LEM1108	TriTrypdb
LturRef	*Leishmania* turanica	LEM423	TriTrypdb
LgerbRef	*Leishmania* gerbilli	LEM452	TriTrypdb
LaethRef	*Leishmania* aethiopica	L147	TriTrypdb
LtroRef	*Leishmania* tropica	L590	TriTrypdb
LinfRef	*Leishmania* infantum	JPCM5	TriTrypdb
LdonBPKRef	*Leishmania* donovani	BPK282A1	TriTrypdb
LdonCLSLRef	*Leishmania* donovani	CL-SL	TriTrypdb
LdonHU3Ref	*Leishmania* donovani	HU3	TriTrypdb
LdonLV9Ref	*Leishmania* donovani	LV9	TriTrypdb
LmexRef	*Leishmania* mexicana	MHOM/GT/2001/U1103	TriTrypdb

LamazM2269Ref	*Leishmania* amazonensis	MHOM/BR/71973/M2269	TriTrypdb
LamazPH8Ref	*Leishmania* amazonensis	PH8	TriTrypdb
LpanL13Ref	*Leishmania* panamensis	MHOM/COL/81/L13	TriTrypdb
LpanPSC1Ref	*Leishmania* panamensis	MHOM/PA/94/PSC-1	TriTrypdb
LbrazM2903Ref	*Leishmania* braziliensis	MHOM/BR/75/M2903	TriTrypdb
LbrazM2904Ref	*Leishmania* braziliensis	MHOM/BR/75/M2904	TriTrypdb
LbrazM29042019 Ref	*Leishmania* braziliensis	MHOM/BR/75/M2904 2019	TriTrypdb
LbrazCian	*Leishmania* braziliensis	MHOM/BR/1975/M2903	This Study
LshaCian	*Leishmania* shawi	MHOM/BR/1990/IM2842	This Study
LequCian	*Leishmania* equatorensis	MCHO/EC/1982/LSP1	This Study
LcolCian	*Leishmania* colombiensis	IHAR/CO/1985/CL500	This Study
LadlCian	*Leishmania* adleri	RLIZ/KE/1954/LV30	This Study
LherrCian	*Leishmania* herreri	MCHO/CR/1974/016;Ch-97;LV344	This Study
LpifCian	*Leishmania* pifanoi	MHOM/VE/1957/LL1	This Study
LlainmtCian	Leishmani lainsoni	MHOM/BR/2002/NMT-RBO 027P	This Study
Llaim6426	Leishmani lainsoni	MHOM/BR/1981/M6426	This Study
LhybCian	*Viannia* hybrid	MHOM/VE/1991/PM-H197	This Study
LpanCian	*Leishmania* panamensis	MHOM/NI/1988/ZE09	This Study
LnaiCian	*Leishmania* naiffi	ISQU/BR/1985/IM2264	This Study