The complete coding region of the maxicircle as a superior phylogenetic marker for exploring evolutionary relationships between members of the Leishmaniinae

Alexa Kaufer[a*], Joel Barratt[a], Damien Stark[b], John Ellis[a]

[a] School of Life Sciences, University of Technology Sydney, Ultimo, NSW 2007, Australia

[b] Department of Microbiology, St Vincent's Hospital Sydney, Darlinghurst, NSW 2010, Australia

[*] *Corresponding author*

E-mail addresses:

A. Kaufer: Alexa.Kaufer@student.uts.edu.au

J. Barratt: joelbarratt43@gmail.com

D. Stark: Damien.Stark@svha.org.au

J. Ellis: John.Ellis@uts.edu.au

Abstract

The mitochondrial DNA (mtDNA) is a potentially valuable phylogenetic marker given its presence across all eukaryotic taxa and its relative conservation in structure and sequence. In trypanosomatids, a homologue of the mtDNA referred to as the maxicircle DNA, is located within a specialised structure in the single mitochondrion of the trypanosomatids called the kinetoplast; a high molecular weight network of DNA composed of thousands of catenated minicircles and a smaller number of larger maxicircles. Unique to the kinetoplastid protists, the maxicircle component of this complex network could represent a desirable target for taxonomic inquiry that may also facilitate exploration of the evolutionary history of this important group of parasites. The aim of this study was to investigate the phylogenetic value of the trypanosomatid maxicircle for these applications. Maxicircle sequences were obtained either by assembling raw sequence data publicly accessible in online databases (i.e., NCBI), or by amplification of novel maxicircle sequences from trypanosomatid DNA using long-range (LR) PCR with subsequent Illumina sequencing. This procedure facilitated the generation of nearly complete maxicircle sequences (i.e., excluding the divergent region) for numerous dixenous and monoxenous trypanosomatid species. Annotation of each maxicircle sequence confirmed that their structure was conserved across all taxa examined. Phylogenetic analyses confirmed that *Z. australiensis* showed a greater genetic relatedness with the dixenous trypanosomatids of the genera *Leishmania* and *Endotrypanum*, as opposed to members of the monoxenous genera *Crithidia* and *Leptomonas*. Additionally, molecular clock analysis supported that the dixenous Leishmaniinae appeared approximately 75 million years ago during the breakup of Gondwana. In line with previous studies, our results support the Supercontinents hypothesis regarding the origin of dixenous Leishmaniinae. Ultimately, we demonstrate that the maxicircle represents an excellent phylogenetic marker for studying the evolutionary history of trypanosomatids, resulting in trees with very high bootstrap support values.

Keywords: *Leishmania*; kinetoplast; maxicircle; Long-range PCR; Next-generation sequencing; phylogenetics

1. Introduction

Leishmaniasis remains one of the most important neglected tropical diseases, affecting some of the poorest populations worldwide (Torres-Guerrero et al., 2017). Endemic in 97 countries, 700 000 – 1 million new cases are documented per annum, with a further 350 million people at risk of acquiring the disease (WHO, 2018). Of the approximate 53 species within the *Leishmania* genus, 20 have been identified as the aetiological agents of human leishmaniasis. Depending on the species in question, *Leishmania* infections manifest as three distinct clinical forms; cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL) and visceral leishmaniasis (VL or Kala Azar) (Galluzzi et al., 2018) . In recent years, the taxonomy and evolutionary history of the trypanosomatid parasites has been discussed at length, particularly with respect to establishing a consensus on the procedures for classifying novel species which is partially dependent on the application of robust phylogenetic approaches (Espinosa et al., 2016; Kaufer et al., 2017; Maslov et al., 2018; Votýpka et al., 2015). Additionally, the origin and evolutionary history of the dixenous Leishmaniinae has been rigorously debated as a matter of intrigue and philosophical interest. In any case, both of these pursuits rely on rigorous phylogenetic analysis.

The mitochondrial DNA of the Trypanosomatidae exists as a large, complex network of catenated DNA circles organised into a disk-shaped structure known as the kinetoplast (see Fig. 1) (Lin et al., 2015). The interlocking network referred to as the kinetoplast DNA (kDNA) is comprised of approximately 10, 000 minicircles and 20-50 maxicircles, representing 20-25% of the trypanosomatid's total DNA (Gerasimov et al., 2017; Telleria et al., 2006). Minicircles are circular DNA molecules with species-specific sizes ranging from 0.5 to 10 kb and account for 95-99% of the total kDNA mass (Flegontov et al., 2009). Maxicircles are considerably larger circular DNA molecules ranging from 20 – 40kb in size, depending on the species.

In most trypanosomatid species, the kDNA contains multiple minicircle classes of varied abundance in a single network (Flegontov et al., 2009). Minicircles encode one or more small non-

coding RNAs that act as guide RNAs (gRNA), that are involved in the RNA editing of maxicircle transcripts (Lin et al., 2015). The gRNAs encoded by the minicircles contain information for the number of uridine-insertion/deletions required for the correction of DNA-encoded mRNA frameshifts at the RNA level (Simpson et al., 2015). This unique genetic function of post-transcriptional modification is the most distinguishing characteristic of the kinetoplast (Lin et al., 2015) and is thus crucial for trypanosomatid viability. Minicircle kDNA has been successfully used for the molecular detection of *Leishmania* parasites (Ceccarelli et al., 2014); their high copy number (approx. 10, 000 per cell) makes it ideal as a highly sensitive diagnostic marker. Despite this, minicircles have a high level of nucleotide polymorphisms amongst their several thousand copies, making them unsuitable for resolving phylogenetic relationships between closely related trypanosomatid taxa (de Oliveira et al., 2013). Due to their abundance and low level of sequence conservation, the drawbacks of kDNA minicircles seemingly outweigh their benefits as a target for phylogenetic analyses.

The maxicircle kDNA is comprised of two regions; a coding region with short intergenic spacers and a variable non-coding region termed the divergent region (DR) (Flegontov et al., 2009). The coding region contains mitochondrial gene homologues typical of other eukaryotes, encoding mitochondrial proteins involved in energy production and ribosomal RNAs (see Table 1) (Yatawara et al., 2008). This coding segment accounts for 50-75% of the maxicircle kDNA length, containing a region of 15-17kb that is actively transcribed and conserved between species (Lee et al., 1992). The non-coding divergent region is a non-transcribed segment of various repeats that has been poorly studied (Flegontov et al., 2006a). The divergent region consists almost entirely of repeats and is highly variable at the species level (Flegontov et al., 2006b). Due to its variability, the divergent region represents the main source of size and sequence varatiation in maxicircles of different species (Lee et al., 1992).

Despite the significant developments made in trypanosomatid taxonomy due to advances in molecular biology, issues surrounding the robustness of phylogenetic trees and the choice of an

appropriate taxonomic marker remain (Kaufer et al., 2017; Maslov et al., 2018; Yurchenko et al., 2014). The kinetoplast is an organelle exclusive to the Kinetoplastida, unique in its structure, function and mode of replication and thus its maxicircle genome may represent a valuable taxonomic marker (Shapiro and Englund, 1995). Mitochondrial genomes of other eukaryotic cells have been vital to the analysis of evolutionary relationships between related organisms. The ubiquitous use of mtDNA can be traced to their desirable properties, particularly their relatively fast rate of evolution, high copy number and small size (approx. 15-20kb) (Messenger et al., 2012). Additionally, variation in the kDNA has been shown to significantly impact parasite development and the course of infection (Lin et al., 2015), making it a desirable target for trypanosomatid taxonomy and phylogenetics.

Due to these desirable attributes which are shared with mtDNA, phylogenies based on the complete maxicircle genome should inexplicably improve the robustness of the phylogenetic trees generated, making them ideal markers for phylogenetic inference. The superiority of maxicircle-based phylogenies was demonstrated in the analyses of trypanosomes, providing novel insights into the biological features of this genus (Botero et al., 2018; Hong et al., 2017; Lin et al., 2015; Messenger et al., 2012; Simpson and Simpson, 1980). It is proposed that the use of the entire coding region of the maxicircle genome will provide a much-needed framework for the taxonomic classification of the Trypanosomatidae, specifically the *Leishmania* genus which is studied here.

The use of concatenated sequences from multiple phylogenetically informative loci has been described as standard practice to ensure robust phylogenetic investigations involving *Leishmania* spp. and related trypanosomatids (Kaufer et al., 2017; Maslov et al., 2018). However, given the universally conserved structure of the mtDNA, as well as its possession of several genes varying in function and rates of evolution, we sought to examine the value of the kDNA maxicircle coding region as a target for investigating the evolutionary history of trypanosomatid parasites. Using long-range (LR) PCR and Illumina sequencing technology, we amplified and sequenced the maxicircle (excluding the divergent region) of six trypanosomatid species including the dixenous, *Leishmania*

braziliensis, *Leishmania herreri* (hereafter called *Endotrypanum herreri*), *Leishmania major*, *Leishmania tropica*, *Leishmania tarentolae* and the putatively monoxenous *Zelonia australiensis*. Additionally, the maxicircle was extracted and subsequently assembled from whole genome sequence data of 22 additional trypanosomatid species. Subsequent phylogenetic analyses confirmed that *Z. australiensis* has a greater affinity to the dixenous members of the Leishmaniinae than to the monoxenous trypanosomatids. Additionally, we found that the maxicircle DNA is an excellent target for the phylogenetic analyses of the *Leishmania* genus given the high bootstrap values obtained. Furthermore, our analyses confirm the taxonomic validity of *Leishmania shawi* and support the reclassification of *Endotrypanum herreri* (previously *L. herreri*). Finally, as part of these analyses, we consider the timeframe of evolution for the *Leishmania* genus using the divergence of the common ancestor of *Trypanosoma cruzi* and *Trypanosoma brucei* as a calibration date to further explore the origin of the dixenous parasitism within the Leishmaniinae. These analyses suggest that a common ancestor of the dixenous genera *Leishmania*, *Endotrypanum* and *Porcisia* diverged from a common monoxenous ancestor approximately 75 MYA, providing support for the emergence of dixenous parasitism in the Leishmaniinae during the late cretaceous, coinciding with the breakup of Gondwana.

2. Materials and Methods

2.1. Samples

The various *Leishmania* and trypanosomatid species used in this study are listed in Supplementary Table 1 (S1 file).

2.2. DNA Extraction

DNA extraction was performed on *Z. australiensis* and *L. tropica*. Cultures of *Z. australiensis* were grown over three days in a modified liquid haemoglobin (M3) medium (M199, 10 % inactivated horse serum, 1X penicillin-streptomycin, IsoVitaleX and 0.99 g/L haemoglobin) (Barratt et al., 2017). Cultures of *L. tropica* were first cultured on NNN slopes and subsequently transferred to Minimum Essential Medium (MEM) with 20% foetal calf serum (FCS) (Chouihi et al., 2009). Parasite cultures were centrifuged at 4000 g for 15 mins to pellet the cells. The supernatant was removed and cell pellets were resuspended in 1 ml of DNA Extraction Buffer (0.2 M Tris-HCL, 0.025M EDTA, 0.5% EDTA, 0.25M NaCl, 0.3 mg/ml proteinase K), followed by incubation overnight at 55°C. Samples were centrifuged at 4000 g for 3 minutes and the supernatant was transferred to a new tube. The DNA was then extracted from the resulting lysate using the phenol-chloroform method. Briefly, 500 µl of TE-Saturated phenol was added to the lysate and vortexed for approximately 1 minute. Next, 500 µl of chloroform was added and vortexed for an additional minute. The mixture was then centrifuged at 13 000 g for 1 minute. The aqueous layer was carefully removed and extracted twice more as previously described. This was followed by a final extraction of the aqueous phase once more with 500 µl of chloroform. DNA was then precipitated overnight at -20°C with the addition of 8µl 5 M NaCl and 1 X volume of isopropanol. The tubes were centrifuged at 13 000 g for 15 minutes to pellet the precipitated DNA. Once the supernatant was removed, the pellet was rinsed 3 times with 1 mL of 70% ethanol. Next, the ethanol was decanted, and the DNA pellet was air dried for 10 minutes at room temperature followed by addition of 50 µl ddH_2O. All DNA extracts were stored at -20°C until assayed.

2.3. Preliminary Illumina sequencing of total cell DNA and subsequent maxicircle extraction

To obtain the entire maxicircle sequence, including the divergent region, Illumina MiSeq whole genome sequencing (WGS) of *Z. australiensis* was performed twice on duplicate samples. Illumina shotgun libraries were prepared for *Z. australiensis* at the Australian Genome Research Facility

which were sequenced using MiSeq, yielding 300 bp paired-end reads. Since the GC content was high, the data was hard trimmed and quality control was performed using the software Trim Galore! version 0.5.0 (Krueger, 2018). Additionally, the complete maxicircle sequence of seventeen trypanosomatid species were obtained from the WGS data freely available through the Sequence Read Archive (SRA) on NCBI (S3 file).

Processed reads were assembled into contigs using SPAdes version 3.12.0 (Bankevich et al., 2012). For the purposes of this study, we were only interested in the maxicircle kDNA and additional analyses of the assemblies were outside the scope of this study. The maxicircle contigs were identified through BLAST analysis using NCBI BLAST software (NCBI, 2008). To remove redundancy and close the maxicircle sequences, these larger contigs, sometimes representing fragmented maxicircle sequences, were subsequently assembled with CAP3 (Huang and Madan, 1999) to generate a complete maxicircle genome sequence.

2.4. Long-range Polymerase Chain Reaction (LR-PCR) and cell preparation

Long-range PCR primers were designed to amplify two large regions of the kDNA maxicircle (subsequently referred to here as PCR product A and B), that were approximately 10 kb or greater (species dependant). LR-PCR assays were performed on a PTC-200 Peltier Thermal Cycler with each PCR prepared using the RANGER DNA polymerase kit (Bioline) in a total volume of 50 µl, according to manufacturer's instructions. For a detailed description of the assay conditions for each primer pair, see Supplementary file S2. The PCR products were visualised under UV light following electrophoresis on 1% agarose gel stained with GelRed. After visualisation, amplified LR-PCR fragments were purified with ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific).

2.5. Illumina sequencing of LR-PCR fragments, assembly and genome annotation

Library preparation using a Nextera library prep system and sequencing (Illumina MiSeq) of each separate sample (PCR product A and B) for all six species was performed at the Australian Genome Research Facility. The 250 bp-long paired end reads obtained were trimmed using Trim Galore! version 0.5.0 to remove low-quality reads and adapter content from the reads (Krueger, 2018). The trimmed reads were then analysed with FASTQC version 0.11.7 software, for quality control (QC) (Andrews, 2018). The processed reads were assembled into scaffolds using SPAdes version 3.12.0 (Bankevich et al., 2012). Maxicircle contigs were identified using NCBI BLAST software with the published maxicircle of *L. tarentolae* (GenBank: M10126.1) used as a query sequence. Final partial maxicircle genomes (excluding the divergent region) were assembled from the contiguous sequences of PCR products A and B using CAP3 (Huang and Madan, 1999).

2.6. Conventional PCR to fill the gaps in read assembly

The Illumina MiSeq reads of *L. major* resulted in assemblies possessing two gaps that could not be closed. Primers were designed to amplify these regions of the *L. major* maxicircle to close these gaps (S2 file). Conventional PCR assays were performed on PTC-200 Peltier Thermal Cycler. Each PCR was prepared using the BIOTaq PCR Kit (Bioline) with a total reaction volume of 50 µl, according to manufacturer's instructions. The PCR products were visualised under UV light following electrophoresis on a 2% agarose gel stained with GelRed. The PCR products of the correct size were excised from the agarose gel using a sterile scalpel blade. The amplicons were extracted from gel slices using a QIAquick® Gel Extraction Kit (QIAGEN) following the manufacturer's protocol. Standard Sanger sequencing was performed by the service provider Macrogen Inc. (South Korea) on an ABI 3730XL capillary sequencer. Low-quality bases were trimmed from the ends of the sequence chromatograms with the application SeqTrace (Stucky, 2012), and these sequences were then assembled using CAP3. The Sanger contigs were then

assembled with the *L. major* MiSeq contigs using CAP3 to close the gaps and construct a single contiguous maxicircle sequence for *L. major.*

2.8. Gene identification, annotation and data analysis

Annotation of the *Z. australiensis* maxicircle from whole-genome sequencing and the LR-PCR trypanosomatid maxicircle sequences generated was completed using Geneious version 11.0.2 (Kearse et al., 2012) with the annotated *L. tarentolae* maxicircle sequence (Genbank: M10126.1) as the reference.

2.9. Phylogenetic analysis

Phylogenetic trees were constructed to infer the evolutionary relationships between mono- and dixenous trypanosomatids. Multiple alignments were performed using the MUSCLE algorithmic approach implemented in the Seaview software package (Gouy et al., 2010) and then manually curated to improve accuracy. Phylogenetic trees were constructed using PAUP* version 4.0 and PhyML (Guindon et al., 2005; Swofford, 1993). Phylogenies were inferred with heuristic searches using three methods: parsimony, distance and maximum likelihood (ML) (Ajawatanawong, 2017). Each search involved random stepwise addition with TBR branch swapping and 1000 random replicates (Swofford and Charles, 2017). Bootstrap support for clade topologies was estimated following the analysis of 1000 pseudo-replicate datasets using a heuristic tree search. For ML trees, the best-fit model of evolution, GTR+I+G was selected using jModelTest 2.1 under the Bayesian information criterion (Posada, 2008). Distances from the nucleotide sequences were determined with the General Time Reversible (GTR) method which were computed by PAUP* 4.0.

2.10 Estimating divergence time

To estimate divergence times of various trypanosomatid taxa based on their maxicircle sequences, the NJ method was applied to pairwise hamming distances calculated using the phangorn package in R. The maximum likelihood of this tree was calculated using the Jukes-Cantor model with 1000 bootstrap replicates. The timetree was computed using 1 calibration constraint; the divergence of *T. cruzi* and *T. brucei* approximately 100 million years ago (Harkins et al., 2016; Lukeš et al., 2007). The MEGA7 package (Kumar et al., 2016) was used to infer a timetree using the Reltime method (Tamura et al., 2012) and the Tamura-Nei model (Tamura and Nei, 1993) from the original NJ tree. The maxicircle sequence of the monoxenous trypanosomatid *Paratrypanosoma confusum* served as an outgroup.

3. Results

3.1. Assembly of complete maxicircle genome from Z. australiensis and various trypanosomatid species from whole genome sequencing

From the whole-genome sequencing of *Z. australiensis*, the resulting libraries contained 19 692 760 and 16 875 783 reads respectively. The data used for the additional trypanosomatid sequences are summarised in S1 file. Following the assembly, a BLASTN search was performed using the published *L. tarentolae* (GenBank: M10126.1) sequence as a query, to identify the contigs that corresponded with the maxicircle kDNA. The majority of the remaining contigs derived from the nuclear DNA and kDNA minicircles are not considered further here. The final CAP3 assembly of *Z. australiensis* resulted in a consensus sequence 19, 973 bp in size (Fig. 2). Synonymous with *L. tarentolae*, the maxicircle coding region of *Z. australiensis* encodes 20 genes, accounting for 85% of the total maxicircle. The non-coding divergent region was approximately 3087 bp in size and contained a highly repetitive sequence. The maxicircle sequences of the remaining trypanosomatid species ranged from approximately 16 kb to 30 kb. Given its repetitive nature, it is difficult to

estimate the true length of the divergent region, thus Fig. 2 and S1 file provide an estimate based on the assemblies. The structure of the maxicircles generated from LR-PCR analyses are described in section 3.4 (below).

3.2. LR-PCR amplification and extraction

To obtain the target region spanning the 12S rRNA to COII (PCR product A), one primer pair was successful in producing an amplicon ~10 kb in size for all six species. For the region ranging from the CYb to ND5 (PCR product B), 5 additional species-specific primer pairs were designed to produce a 10 kb amplicon (Fig. 2). For a detailed description of the species-specific primers and PCR assay conditions used, see S2 file. In total, twelve long-range PCRs (for product A and B of each species) were optimised and the products subjected to Illumina sequencing (Fig. 3). The amplicons in the lanes labelled '1' and '4' were ExoSAP-IT extracted according to the manufacturer's instructions for next-generation sequencing due to their band intensity and resolution compared to that of their duplicates (Fig.3).

3.3. Illumina sequencing and assembly

Twelve Nextera libraries were generated with an average fragment length of 250 bp. All twelve were successfully sequenced and the maxicircle genome (excluding the divergent region) was subsequently assembled.

A *de novo* assembly of this data was performed using SPAdes (Bankevich et al., 2012). The maxicircle genomes (excluding the divergent region) of *L. braziliensis*, *E. herreri*, *L. major*, *L. tarentolae*, *L. tropica* and *Z. australiensis* were assembled from these shorts reads to generate the initial contigs. The initial assembly of *L. major* product A resulted in three contigs that could not be assembled to create a final concatenated sequence. To bridge these areas of zero read coverage, two

conventional PCR assays were optimised to successfully generate amplicons of 560 bp (Gap 1) and 650 bp (Gap 2) in size. The final assembly for all species was performed using CAP3, which combined the contigs from product A & B to generate a concatenated sequence approximately 15 kb in size for each species. The quality of the genomes generated were well supported based on the high coverage and percentage of reads used to build the assembly. The depth of coverage exceeded 700 X for all samples. When the reads were mapped to its own assembled genome using BWA, the median percentage of reads mapping back to the assemblies was 95% for *L. braziliensis*, 99% for *E. herreri*, 94% for *L. major*, 96% for *L. tarentolae*, 99% for *L. tropica* and 96% for *Z. australiensis*.

3.4. General features of the maxicircle coding region

All 28 trypanosomatid species had the same overarching maxicircle structure and a schematic diagram of the coding region for the six trypanosomatid species maxicircles generated by long-range PCR is shown in Fig. 4. The sequenced maxicircle genomes consist of an approximately 15 kb region spanning from the 12S rRNA through to the ND5 gene. The total sizes of the sequences obtained for the kDNA maxicircle of *L. braziliensis*, *E. herreri*, *L. major*, *L. tarentolae*, *L. tropica* and *Z. australiensis* are provided in Table 2. The sequences were deposited in GenBank under the accession numbers XXX.

According to previous studies, the gene order and nucleotide sequences of the maxicircle coding region are highly conserved amongst the trypanosomatid family (Lin et al., 2015). This was confirmed in this study, allowing a straightforward annotation, by comparison to the previously published *L. tarentolae* maxicircle kDNA (M10126.1) as a reference (de la Cruz et al., 1984). Both schematic diagrams and annotations were built using the software Geneious version 11.0.2. Our data indicated that the maxicircle sequences of trypanosomatid species belonging to the *Leishmania* and *Zelonia* genera encode 20 genes, with very similar gene structure between all novel maxicircle sequences generated in this study (see Fig. 4). Based on these data, the 12S rRNA, 9S rRNA, ND8,

ND7, COIII, CYb, MURF4 (ATPase 6), G3, COII, MURF2, ND4, RPS12 and ND5 genes of the studied trypanosomatids are transcribed from the forward strand while ND9, MURF5, MURF1, ND1, COI, G4 and ND3 are transcribed from the reverse strand.

3.5. Phylogenetic analysis

Phylogenetic trees were constructed from the coding region of the maxicircle genome to infer the genetic relationships between the trypanosomatid species under investigation. For each alignment, phylogenies inferred using the parsimony, distance and likelihood methods showed the same overall topology with robust structures. Furthermore, topologies were synonymous across the different platforms (PAUP and PhyML).

The parsimony principle states that the simplest explanation i.e. the one that requires the fewest evolutionary changes, is preferred (Kannan and Wheeler, 2012). From the heuristic search of the maxicircle, the most parsimonious tree was found to be 30 093. Of the 18 247 characters analysed under the parsimony optimality criterion, 8759 characters were constant, 2639 variable characters were parsimony-uninformative and a final 6849 characters were considered parsimony-informative.

The evolutionary relationships showing the genetic distance between members of the Leishmaniinae is shown in Fig. 5. In the inferred consensus phylogenies (parsimony, distance and likelihood), all *Leishmania* spp. formed a strongly supported monophyletic group (Fig. 6.). In agreement with a recent study (Barratt et al., 2017), *Z. australiensis* was more closely related to the dixenous trypanosomatids, clustering with species of the *Leishmania* and *Endotrypanum* genera with 100% confidence (Fig. 6.). All phylogenies positioned *Z. australiensis* as a possible intermediate between the dixenous members of the Leishmaninnae subfamily and related

monoxenous trypanosomatids. Another trend observed amongst the trees, was that the species previously known as *L. herreri* is more closely related to *Endotrypanum*, clustering with *E. monterogeii* with 100% bootstrap confidence. Hence, we adopted the use of the name *E. herreri* for this species.

GTR distances among the *Leishmania* and *Endotrypanum* species (except *E. herreri*) ranged from 0.005 (between *L. braziliensis* and *L. peruviana*) to 0.196 (between *L. aethiopica* and *L. enriettii*) (S3 file). The genetic distance between *E. herreri* and other *Leishmania*/*Endotrypanum* species ranged from 0.005 (between *E. herreri* and *E. monterogeii*) to 0.192 (between *E. herreri* and *L. arabica*). In addition, the genetic distance between the monoxenous *Z. australiensis* and other trypanosomatid species ranged from 0.247 (between *Z. australiensis* and *E. montergeii*) and 0.358 (between *Z. australiensis* and *Blechomonas ayalai*).

3.6 Divergence time estimates

The node representing the divergence of the common ancestor of *T. cruzi* and *T. brucei* was selected as a calibration point. This node was set at an average of 100 MYA, which is the estimated time period that Africa and South American became separated, representing a minimum time of separation. Using this as the calibration marker, a common ancestor to the Leishmaniinae subfamily was predicted to have appeared approximately 75 MYA, corresponding to the Late Cretaceous period of earth's geological history.

4. Discussion

The kinetoplast is a diagnostic feature of the Kinetoplastids, a group of organisms characterised by the presence of a single unique network of catenated DNA circles (kDNA) (Cavalcanti and de Souza, 2018). The largest molecule in this network, the maxicircle DNA, is homologous to

mammalian mitochondrial DNA (Simpson et al., 1985). Protozoan parasites of the trypanosomatid family (order Kintetoplastida), predominately infect only insects (i.e. have a monoxenous lifecycle) (Maslov et al., 2013). However, some genera including *Leishmania* are transmitted by insects and are pathogenic to humans (i.e. possess a dixenous lifecycle), being the aetiological agents of the clinically important disease leishmaniasis, which is a severely debilitating and often-fatal diseases (WHO, 2018). While different *Leishmania* species are morphologically very similar and not readily distinguished by morphology (Lee et al., 2000), leishmaniasis includes a broad-spectrum of diseases that can present with a multitude of clinical manifestations. The course of a human *Leishmania* infection is largely determined by the causative species (Rodgers et al., 1990), which despite being morphologically similar, are divided into several phylogenetically supported subgenera.

Elucidating the complex biology, phylogenetics and taxonomy of *Leishmania* spp. requires a clear understanding of the parasite's genetic diversity. Here we undertook an in-depth analysis of maxicircle kDNA from various trypanosomatid species. Sequencing the coding region of the maxicircle allowed us to explore the phylogenetic relationships between members of the *Leishmania* genus, with previous studies traditionally relying on single-gene phylogenies and more recently concatenated sequences of a few phylogenetically informative loci (Asato et al., 2009; Croan et al., 1997; Yang et al., 2013). In this study we present a comprehensive analysis of the maxicircle from several trypanosomatids and further investigate the phylogenetic relationships of the mono- and dixenous species using these maxicircle sequences. This improved the resolution of the trees generated compared to previous studies using single gene and small concatenated gene phylogenies. The work flow for evolutionary analysis in our investigation combined LR-PCR and Illumina MiSeq to assemble a 15 kb-long region of the maxicircle (excluding the DR) of six trypanosomatid species. Additionally, we present the complete maxicircle genome of *Z. australiensis* and 22 trypanosomatid species assembled from previously sequenced whole genome sequencing libraries.

In a previous review, we highlighted three systematic issues that represent the main source of discrepancies in trypanosomatid taxonomy, particularly in the *Leishmania* genus (Kaufer et al., 2017). In summary the main issues were the result of: -

1. The use of slow-evolving genes (such as the 18S rRNA gene) to construct phylogenies
2. The dependence of tree structure on the choice of locus
3. The number of species/isolates used for analysis

A great advantage of using the entire maxicircle coding region is that it not only addresses these biases, but subsequently resolves many of the issues present in trypanosomatid phylogenetics. Phylogenetic trees constructed from the maxicircle kDNA represent an alternative to previous approaches that provides a superior model (based on the strong bootstrap support values) to investigate genetic relationships and also avoids the biases that come with phylogenies based on single-gene and concatenated gene analyses (Som, 2015).

Mitochondrial DNA has a relatively fast rate of mutation compared to nuclear DNA (Messenger et al., 2012). The use of slow evolving genes in the analysis of closely related species is often the downfall of traditional phylogenetic reconstructions. Trees based on slow-evolving genes such as the 18S rDNA are unable to delineate relationships (Deschamps et al., 2011). Consequently, maxicircle kDNA is particularly useful in phylogenetic analyses for species within the same family i.e. the trypanosomatids. The higher-rate of mutations result in a greater number of sites with phylogenetically-informative characters from which the trees are built, ultimately providing a superior molecular target than those presently documented in the literature.

It is widely accepted that different loci possess different genetic histories, resulting in phylogenetic trees that are prone to sampling bias (Yang et al., 2013). In this study, for each method of inference and software package used, the likelihood, parsimony and distance methods all showed the same structure and overall topology in the trees generated (Fig. 6.). We propose this is the direct result of using a larger number of phylogenetically informative characters that fall within the 15 kb

region of the maxicircle sequenced. By published standards in the test of robustness (i.e. bootstrapping), the percentile method justifies the accuracy of a clade, with a confidence interval of >60% in support of the observed clade (Felsenstein, 1985). It is clear that the use of large datasets (approximately 15 kb) such as the maxicircle kDNA is an effective method to alleviate sampling bias, resulting in extremely robust trees, thereby eliminating the interchangeable structure of trees due to the loci chosen for analysis. The maxicircle sequences of *Z. australiensis* generated from both LR-PCR and whole-genome sequencing were 100% identical. Although the LR-PCR assays described here do not amplify the divergent region, the highly variable nature of this repetitive feature of the maxicircle is not conducive to phylogenetic inference. Thus, as an alternative to the often-time-consuming assembly and extraction of maxicircle sequences from whole genome sequence data, LR-PCR amplification offers a simpler and cost-effective method to obtain the maxicircle sequence.

The Leishmaniinae subfamily was originally established for a group of monoxenous (*Leptomonas* and *Crithidia*) and dixenous (*Leishmania*) trypanosomatid parasites (Jirků et al., 2012) and was recently revised to include the newly established monoxenous species *Zelonia* and *Novymonas* (Espinosa et al., 2016). The analysis presented here included sequences from species of the subgenera *Leishmania* (*Leishmania*), *Leishmania* (*Viannia*), *Leishmania* (*Mundinia*) and *Leishmania* (*Sauroleishmania*). Additionally, the genera *Endotrypanum*, *Porcisia*, *Zelonia*, *Leptomonas*, *Crithidia, Herpetomonas* and *Blechomonas* were all represented. In the phylogenies, all *Leishmania*, *Endotrypanum* and *Porcisia* spp. formed a strongly supported monophyletic group (98% bootstrap confidence). The genetic distance analysis (S3 file) and phylogenetic trees (Fig. 6.) all suggest that the monoxenous *Z. australiensis* is genetically closer to the dixenous species of the Leishmaniinae subfamily than to the monoxenous trypanosomatids.

From our previous analyses, we suggested the common ancestor of the dixenous Euleishmania (*L.* (*Leishmania*) and *L.* (*Viannia*)) and Paraleishmania (*Endotrypanum* and *Porcisia*) appeared during the breakup of Gondwana in the Mesozoic era approximately 91 million years ago

(Barratt et al., 2017) as proposed by the Supercontinents hypothesis (Harkins et al., 2016). Based on our molecular data these genera have emerged as distinct monophyletic lineages, strongly supported by phylogenetic analyses. However, based on the maxicircle phylogenies presented here, the ancestor of the dixenous *Leishmania*, *Endotrypanum* and *Porcisia* emerged from monoxenous parasites approximately 75 MYA (Fig. 7.). Despite the report of a more recent emergence of *Leishmania*, our results still place the appearance of dixenous parasitism within the Leishmaniinae in the Late Cretaceous period, which aligns to the Supercontinents theory of *Leishmania* evolution. According to this theory, the divergence of the dixenous genera of the Leishmaniinae coincided with the adaptive radiation of mammals during this period (90-65 MYA) (Cox, 2000). The appearance of this common ancestor to the Euleishmania and Paraleishmania at approximately 75 MYA remains within this timeframe, which still supports a Gondwanan origin. Based on the present study and the previous work of Barratt et al. (2017) and Harkins et al. (2016), we propose by consensus that the earliest dixenous Leishmaniinae parasites arose in the Cretaceous period between 77 – 140 MYA, during the protracted breakup of Gondwana.

Two alternative scenarios have been proposed for the divergence of Old and New World species within the Euleishmania; the first scenario being the presence of Old World and New World species in the *L*. (*Leishmania*) subgenus suggests migration of the Old World to the New. The second scenario is that land bridges existed in the northern hemisphere 50 MYA connected Europe, North America and Asia allowing the movement of host and vector species between the Old and New World until their disappearance during the Eocene-Oligocene boundary approximately 30 MYA (Barratt et al., 2017; Harkins et al., 2016; Momen and Cupolillo, 2000; Ren et al., 2013). The inferred emergence of the New World *L*. (*Leishmania*) spp. coincides with the events of the latter, supporting the disappearance of these land bridges ultimately driving the species in Northern Europe towards Africa and South East Asia in the Old World and forcing the tropical *Leishmania* species towards the Neotropics in the New World (Barratt et al., 2017).

The southern-supercontinent hypothesis which suggests that *T. cruzi* evolved in the New World and *T. brucei* in the Old World following the split of South America and Africa 100 MYA has been widely accepted by these interested the evolution of trypanosomes for the last 30 years (Harkins et al., 2016; Lukeš et al., 2007; Stevens et al., 1999; Stevens et al., 2001). The *T. cruzi* clade is composed of two main sister lineages; (i) the *Schizotrypanum* lineage, formed by *T. cruzi* and bat-restricted trypanosomes and (ii) Tra (Tve – Tco) formed by *Trypanosoma rangeli*, *Trypanosoma vespertilionis* and *Trypanosomas conorhini*. Species of both lineages are associated with *Cimicidae* and *Triatominae* of the order Hemiptera. These vectors are believed to have played a crucial role in the evolution of these trypanosomes. Fossil evidence shows the presence of ancient cimicids and the relatively younger triatomes dating back approximately 100 MYA and 32 MYA respectively, inferring that Old World cimicids were the vector of *T. cruzi* ancestors.

Although the trypanosome southern-supercontinent hypothesis is widely accepted, recent evidence supports an alternate 'bat-seeding' origin where the common ancestor of the *T. cruzi* clade (*T. cruzi* and *T. rangeli*) was a bat trypanosome that made the transition into mammals. This Old-World bat trypanosome is likely to have evolved sometime after bats underwent major diversification approximately 70 – 58 MYA and through successive host switching into terrestrial mammals, gave origin to *T. rangeli* and *T. cruzi* lineages of the *T. cruzi* clade (Espinosa-Alvarez et al., 2018). The key implication of the 'bat-seeding' origin is that *T. cruzi* may have evolved more recently than previously thought (Hamilton et al., 2012). Using the period coinciding with the diversification of the *T. cruzi* clade (i.e., 70 to 58 MYA) rather than the split of the common ancestor of *T. cruzi* from *T. brucei* would likely result in an earlier prediction for the appearance of dixenous Leishmaniinae parasites, although this would still coincide with the adaptive radiation of mammals. However, using the estimated divergence times of a host species (i.e., bats – 70 to 58 MYA) rather than a geological time point is problematic as calibrations based on molecular estimates (i.e. secondary calibrations) may skew the analyses (Sauquet et al., 2012). Based on the scenario proposed by the 'bat-seeding' hypothesis, the last common ancestor of *T. cruzi* was

493 transmitted by ancient cimicids. Fossil evidence shows that the *Cimicidae* were present in the Old
494 World 100 MYA, predating the well-documented vicariance biogeography of South America and
495 Africa. Thus, using the geological isolation and fossil evidence to analyse the separation of *T.*
496 *brucei* from the last common ancestor of *T. cruzi*, the calibration of 100 MYA used in this study
497 remains a suitable based on current understanding.

498 All clades observed support a recent appraisal of the classification of *Leishmania*,
499 *Endotrypanum* and *Porcisia* (Espinosa et al., 2016). *Leishmania* spp. of the *Viannia* subgenus are
500 restricted to the Neotropics (New World), whereas the subgenus *Leishmania* occurs in both the New
501 and Old World (Fig. 6.). The species at the crown of our phylogenetic trees (*Leishmania aethiopica,*
502 *Leishmania tropica, Leishmania arabica, Leishmania turanica, Leishmania major, Leishmania*
503 *donovani, Leishmania infantum, Leishmania mexicana* and *Leishmania pifanoi*) cluster with 100%
504 confidence to form the *L.* (*Leishmania*) subgenus. Immediately below this (*L. tarentolae*) sits the *L.*
505 (*Sauroleishmania*), followed by species restricted to the New World (*L. braziliensis, L. peruviana,*
506 *L. guyanensis, L. panamensis and L. shawi*) that correspond to the *L.* (*Viannia*) subgenus. The most
507 basal *Leishmania* sp. included in our analysis (*L. enriettii*) represents *L.* (*Mundinia)*.

508 The taxonomic validity of *L. shawi* has come under scrutiny, with reports stating it is not a
509 distinct species from *L. guyanensis* (Boité et al., 2012). Phylogenetic analyses indicate that the
510 designation of *L. shawi* as a distinct species is warranted, having emerged from a common ancestor
511 shared with *L. guyanensis* and *L. panamensis* approximately 2.7 MYA (Fig. 7.). However, our
512 analyses challenge the status of additional species of the *Leishmania* (*Viannia*) subgenus (Fig. 5.).
513 Traditionally separated by geographic distribution, the genetic basis for the separation of *L. (V.)*
514 *braziliensis* and *L. (V.) peruviana* has been hotly debated over the years (Fraga et al., 2013;
515 Valdivia et al., 2015). Three arguments have been pursued in the literature with regards to the
516 controversy surrounding *L. braziliensis* and *L. peruviana*: whether or not *L. braziliensis* and *L.*
517 *peruviana* can be considered sole species*;* they are in fact heterogenous species, with *L. peruviana*
518 being a subspecies of *L. braziliensis*; and thirdly that they are two distinct species (Banuls et al.,

519 2000; Fraga et al., 2013; Garcia et al., 2005). Separated by a genetic distance of 0.005 (S3 file), our
520 analyses show these *Viannia* species are very closely related. Following this rationale, the same
521 argument can in theory be used when discussing *L. (V.) guyanensis* and *L. (V.) panamensis,*
522 separated by a genetic distance of only 0.001 (Fig. 5. and S3 File). This result calls in to question
523 whether these species of the *Viannia* subgenus warrant speciation as distinct organisms.

524 Recent revisions of the current taxonomy have established that *Leishmania donovani* in the
525 Old World and *Leishmania infantum* in both the Old World and New World are the only recognised
526 species of the *L. donovani* complex (Jamjoom et al., 2004; Lukeš et al., 2007). Ambiguities
527 concerning this complex have often arisen from phylogenies based on insufficient markers that are
528 unable to detect the DNA polymorphisms (if any) capable of discriminating between these
529 extremely similar species. However, despite using a large dataset (approximately 18 000 characters
530 used in the final analysis), the maxicircle coding region detected very few polymorphisms between
531 the two species, separated by a genetic distance of only 0.007 (S3 file and Fig.5.). This data from
532 also calls into question whether these parasites truly represent distinct species.

533 Basal to the major clades of the *Leishmania* subgenus, our phylogenetic analyses confirm
534 the recent proposal to elevate the previous *L. hertigi*/*L. deanei* complex to generic status (Espinosa
535 et al., 2016). The status of this complex has often been debated and labelled unstable due to the
536 lack of an in-depth genetic analysis involving this group of organisms (Akhoundi et al., 2016;
537 Marcili et al., 2014). These *Leishmania*-like parasites of porcupines' cluster to form a sister clade,
538 long separated (approximately 59 MYA) from *Leishmania* species, with 100% bootstrap confidence
539 (Fig. 6. and Fig. 7.). Thus, the analyses of the maxicircle coding region support the establishment of
540 *Porcisia* as the new genus to accommodate these species. Particularly important is the strong
541 clustering of *E. herreri* (previously *L. herreri*) with *E. monterogeii* and *E. schaudinni*, forming a
542 monophyletic clade basal to all *Leishmania* spp. (100% confidence). It cannot be ignored that based
543 on genetic distance and phylogenetic analysis *E. herreri* is more closely related to *Endotrypanum*
544 than to *Leishmania*. Our results are congruent with the recent suggestion (Espinosa et al., 2016)

545 that the Neotropical trypanosomatid known as *L. herreri* should be placed in the *Endotrypanum*
546 genus (Franco and Grimaldi, 1999; Noyes et al., 1996).

547 In conclusion, given the inconsistencies that exist in trypanosomatid systematics discussed
548 previously (Kaufer et al., 2017; Som, 2015), we propose the use of maxicircle DNA sequences as
549 the taxonomic marker of choice for phylogenetic analyses involving this group of parasites.
550 Specifically, the use of the entire coding region of the maxicircle genome provides more robust
551 evolutionary insight than the single gene-based phylogenies or phylogenies generated by
552 concatenating a small number of gene sequences, such as those commonly reported in the literature
553 (Barratt et al., 2017; Grybchuk et al., 2018; Yazaki et al., 2017). Further research resulting in the
554 generation of additional maxicircle sequences from trypanosomatids, particularly those from the
555 monoxenous-dixenous boundary (e.g. *Zelonia costaricensis*) will provide greater insights into the
556 evolutionary relationships between trypanosomatid taxa including the relationship between
557 pathogenic and non-pathogenic trypanosomatid species. We propose that future investigators
558 aiming to understand the evolutionary relationship between closely related trypanosomatids should
559 consider using the approach described herein as opposed to single-gene based phylogenies.
560 Ultimately, this work highlights the importance of the maxicircle as a valuable tool for the
561 taxonomic and phylogenetic analyses of *Leishmania* spp. and other related trypanosomatids.

562

563 **Acknowledgements**

564 We would like to kindly thank Dr. Harry Noyes and Dr. Rogan Lee for providing the *E. herreri* and
565 *L. tropica* samples respectively. This study was completed by AK in partial fulfilment of the PhD
566 degree at UTS. This research did not receive any specific grant from funding agencies in the public,
567 commercial, or not-for-profit sectors.

568

Declaration of interest

Declarations of interest: none

573 **References**

574 Ajawatanawong, P., 2017. Molecular Phylogenetics: Concepts for a Newcomer. Advances in
575 biochemical engineering/biotechnology 160, 185-196.
576 Akhoundi, M., Kuhls, K., Cannet, A., Votypka, J., Marty, P., Delaunay, P., Sereno, D., 2016. A
577 Historical Overview of the Classification, Evolution, and Dispersion of *Leishmania* Parasites and
578 Sandflies. PLoS Negl Trop Dis 10, e0004349.
579 Andrews, S., 2018. FastQC.
580 Aphasizheva, I., Maslov, D.A., Aphasizhev, R., 2013. Kinetoplast DNA-encoded ribosomal protein
581 S12 A possible functional link between mitochondrial RNA editing and translation in *Trypanosoma*
582 *brucei*. RNA Biol. 10, 1679-1688.
583 Asato, Y., Oshiro, M., Myint, C.K., Yamamoto, Y., Kato, H., Marco, J.D., Mimori, T., Gomez,
584 E.A.L., Hashiguchi, Y., Uezato, H., 2009. Phylogenic analysis of the genus *Leishmania* by
585 cytochrome b gene sequencing. Exp. Parasitol. 121, 352-361.
586 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
587 Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G.,
588 Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its
589 Applications to Single-Cell Sequencing. Journal of Computational Biology 19, 455-477.
590 Banuls, A.L., Dujardin, J.C., Guerrini, F., De Doncker, S., Jacquet, D., Arevalo, J., Noel, S., Le
591 Ray, D., Tibayrenc, M., 2000. Is *Leishmania* (*Viannia*) *peruviana* a distinct species? A
592 MLEE/RAPD evolutionary genetics answer. The Journal of eukaryotic microbiology 47, 197-207.
593 Barratt, J., Kaufer, A., Peters, B., Craig, D., Lawrence, A., Roberts, T., Lee, R., McAuliffe, G.,
594 Stark, D., Ellis, J., 2017. Isolation of Novel Trypanosomatid, *Zelonia australiensis* sp. nov.
595 (Kinetoplastida: Trypanosomatidae) Provides Support for a Gondwanan Origin of Dixenous
596 Parasitism in the Leishmaniinae. Plos Neglect. Trop. Dis. 11, e0005215.

597 Boité, M.C., Mauricio, I.L., Miles, M.A., Cupolillo, E., 2012. New Insights on Taxonomy,
598 Phylogeny and Population Genetics of *Leishmania* (Viannia) Parasites Based on Multilocus
599 Sequence Analysis. Plos Neglect. Trop. Dis. 6, 14.
600 Botero, A., Kapeller, I., Cooper, C., Clode, P.L., Shlomai, J., Thompson, R.C.A., 2018. The
601 kinetoplast DNA of the Australian trypanosome, *Trypanosoma copemani*, shares features with
602 *Trypanosoma cruzi* and *Trypanosoma lewisi*. Int J Parasitol 48, 691-700.
603 Cavalcanti, D.P., de Souza, W., 2018. The Kinetoplast of Trypanosomatids: From Early Studies of
604 Electron Microscopy to Recent Advances in Atomic Force Microscopy. Scanning, 10.
605 Ceccarelli, M., Galluzzi, L., Migliazzo, A., Magnani, M., 2014. Detection and Characterization of
606 *Leishmania* (*Leishmania*) and *Leishmania* (*Viannia*) by SYBR Green-Based Real-Time PCR and
607 High Resolution Melt Analysis Targeting Kinetoplast Minicircle DNA. PLoS One 9, e88845.
608 Chouihi, E., Amri, F., Bouslimi, N., Siala, E., Selmi, K., Zallagua, N., Ben Abdallah, R.,
609 Bouratbine, A., Aoun, K., 2009. Cultures on NNN medium for the diagnosis of leishmaniasis.
610 Pathologie Biologie 57, 219-224.
611 Cox, C.B., 2000. Plate tectonics, seaways and climate in the historical biogeography of mammals.
612 Mem Inst Oswaldo Cruz 95, 509-516.
613 Croan, D.G., Morrison, D.A., Ellis, J.T., 1997. Evolution of the genus *Leishmania* revealed by
614 comparison of DNA and RNA polymerase gene sequences. Mol. Biochem. Parasitol. 89, 149-159.
615 de la Cruz, V.F., Neckelmann, N., Simpson, L., 1984. Sequences of six genes and several open
616 reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. The Journal of
617 biological chemistry 259, 15136-15147.
618 de Oliveira, L., Pereira, R., Brandão, A., 2013. An analysis of trypanosomatids kDNA minicircle by
619 absolute dinucleotide frequency. Parasitol. Int. 62, 397-403.
620 Deschamps, P., Lara, E., Marande, W., López-García, P., Ekelund, F., Moreira, D., 2011.
621 Phylogenomic Analysis of Kinetoplastids Supports That Trypanosomatids Arose from within
622 Bodonids. Mol. Biol. Evol. 28, 53-58.

623 Espinosa-Alvarez, O., Ortiz, P.A., Lima, L., Costa-Martins, A.G., Serrano, M.G., Herder, S., Buck,
624 G.A., Camargo, E.P., Hamilton, P.B., Stevens, J.R., Teixeira, M.M.G., 2018. *Trypanosoma rangeli*
625 is phylogenetically closer to Old World trypanosomes than to *Trypanosoma cruzi*. Int J Parasitol 48,
626 569-584.
627 Espinosa, O., Serrano, M., Camargo, E., Teixeira, M., Shaw, J., 2016. An appraisal of the taxonomy
628 and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*
629 Parasitology Submitted.
630 Felsenstein, J., 1985. Confidence-limits on phylogenies - an approach using the bootstrap.
631 Evolution 39, 783-791.
632 Flegontov, P.N., Guo, Q., Ren, L., Strelkova, M.V., Kolesnikov, A.A., 2006a. Conserved repeats in
633 the kinetoplast maxicircle divergent region of *Leishmania* sp. and *Leptomonas seymouri*. Molecular
634 genetics and genomics : MGG 276, 322-333.
635 Flegontov, P.N., Strelkova, M.V., Kolesnikov, A.A., 2006b. The *Leishmania major* maxicircle
636 divergent region is variable in different isolates and cell types. Mol. Biochem. Parasitol. 146, 173-
637 179.
638 Flegontov, P.N., Zhirenkina, E.N., Gerasimov, E.S., Ponirovsky, E.N., Strelkova, M.V.,
639 Kolesnikov, A.A., 2009. Selective amplification of maxicircle classes during the life cycle of
640 *Leishmania major*. Mol. Biochem. Parasitol. 165, 142-152.
641 Fraga, J., Montalvo, A.M., Van der Auwera, G., Maes, I., Dujardin, J.C., Requena, J.M., 2013.
642 Evolution and species discrimination according to the *Leishmania* heat-shock protein 20 gene.
643 Infect. Genet. Evol. 18, 229-237.
644 Franco, A.M.R., Grimaldi, G., 1999. Characterization of *Endotrypanum* (Kinetoplastida :
645 Trypanosomatidae), a unique parasite infecting the neotropical tree sloths (Edentata). Mem. Inst.
646 Oswaldo Cruz 94, 261-268.
647 Galluzzi, L., Ceccarelli, M., Diotallevi, A., Menotta, M., Magnani, M., 2018. Real-time PCR
648 applications for diagnosis of leishmaniasis. Parasites Vectors 11, 13.

649 Garcia, A.L., Kindt, A., Quispe-Tintaya, K.W., Bermudez, H., Llanos, A., Arevalo, J., Banuls, A.L.,
650 De Doncker, S., Le Ray, D., Dujardin, J.C., 2005. American tegumentary leishmaniasis: antigen-
651 gene polymorphism, taxonomy and clinical pleomorphism. Infection, genetics and evolution :
652 journal of molecular epidemiology and evolutionary genetics in infectious diseases 5, 109-116.
653 Gerasimov, E.S., Gasparyan, A.A., Litus, I.A., Logacheva, M.D., Kolesnikov, A.A., 2017.
654 Minicircle Kinetoplast Genome of Insect Trypanosomatid *Leptomonas pyrrhocoris*. Biochemistry.
655 Biokhimiia 82, 572-578.
656 Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user
657 interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27, 221-224.
658 Grybchuk, D., Kostygov, A.Y., Macedo, D.H., Votypka, J., Lukes, J., Yurchenko, V., 2018. RNA
659 Viruses in *Blechomonas* (Trypanosomatidae) and Evolution of *Leishmaniavirus*. mBio 9.
660 Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML Online - a web server for fast
661 maximum likelihood-based phylogenetic inference. Nucleic Acids Research 33, W557-W559.
662 Hamilton, P.B., Teixeira, M.M., Stevens, J.R., 2012. The evolution of *Trypanosoma cruzi*: the 'bat
663 seeding' hypothesis. Trends Parasitol 28, 136-141.
664 Harkins, K.M., Schwartz, R.S., Cartwright, R.A., Stone, A.C., 2016. Phylogenomic reconstruction
665 supports supercontinent origins for *Leishmania*. Infect. Genet. Evol. 38, 101-109.
666 Hong, X.K., Zhang, X., Fusco, O.A., Lan, Y.G., Lun, Z.R., Lai, D.H., 2017. PCR-based
667 identification of *Trypanosoma lewisi* and *Trypanosoma musculi* using maxicircle kinetoplast DNA.
668 Acta Trop. 171, 207-212.
669 Horvath, A., Kingan, T.G., Maslov, D.A., 2000. Detection of the mitochondrially encoded
670 cytochrome c oxidase subunit I in the trypanosomatid protozoan *Leishmania tarentolae*. Evidence
671 for translation of unedited mRNA in the kinetoplast. The Journal of biological chemistry 275,
672 17160-17165.
673 Huang, X., Madan, A., 1999. CAP3: A DNA Sequence Assembly Program. Genome Research 9,
674 868-877.

675 Jamjoom, M.B., Ashford, R.W., Bates, P.A., Chance, M.L., Kemp, S.J., Watts, P.C., Noyes, H.A.,
676 2004. *Leishmania donovani* is the only cause of visceral leishmaniasis in East Africa; previous
677 descriptions of *L. infantum* and "*L. archibaldi*" from this region are a consequence of convergent
678 evolution in the isoenzyme data. Parasitology 129, 399-409.
679 Jirků, M., Yurchenko, V.Y., Lukeš, J., Maslov, D.A., 2012. New species of insect trypanosomatids
680 from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. The Journal of
681 eukaryotic microbiology 59, 537-547.
682 Kannan, L., Wheeler, W.C., 2012. Maximum Parsimony on Phylogenetic networks. Algorithms.
683 Mol. Biol. 7, 10.
684 Kaufer, A., Ellis, J., Stark, D., Barratt, J., 2017. The evolution of trypanosomatid taxonomy.
685 Parasites Vectors 10, 287.
686 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper,
687 A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious
688 Basic: An integrated and extendable desktop software platform for the organization and analysis of
689 sequence data. Bioinformatics 28, 1647-1649.
690 Krueger, F., 2018. Trim Galore!
691 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis
692 Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.
693 Lee, J., Leedale, G., Bradbury, P., 2000. Illustrated Guide to the Protozoa. Wiley-Blackwell.
694 Lee, S.T., Tarn, C., Wang, C.Y., 1992. Characterization of sequence changes in kinetoplast DNA
695 maxicircles of drug-resistant *Leishmania*. Mol Biochem Parasitol 56, 197-207.
696 Lin, R.H., Lai, D.H., Zheng, L.L., Wu, J., Lukes, J., Hide, G., Lun, Z.R., 2015. Analysis of the
697 mitochondrial maxicircle of *Trypanosoma lewisi*, a neglected human pathogen. Parasites & vectors
698 8, 665.
699 Lukeš, J., Mauricio, I.L., Schönian, G., Dujardin, J.C., Soteriadou, K., Dedet, J.P., Kuhls, K.,
700 Tintaya, K.W.Q., Jirků, M., Chocholová, E., Haralambous, C., Pratlong, F., Oborník, M., Horák,

701 A., Ayala, F.J., Miles, M.A., 2007. Evolutionary and geographical history of the *Leishmania*
702 *donovani* complex with a revision of current taxonomy. Proc. Natl. Acad. Sci. U. S. A. 104, 9375-
703 9380.
704 Marcili, A., Sperança, M.A., da Costa, A.P., Madeira, M.D., Soares, H.S., Sanches, C., Acosta,
705 I.D.L., Girotto, A., Minervino, A.H.H., Horta, M.C., Shaw, J.J., Gennari, S.M., 2014. Phylogenetic
706 relationships of *Leishmania* species based on trypanosomatid barcode (SSU rDNA) and gGAPDH
707 genes: Taxonomic revision of *Leishmania* (*L.*) *infantum chagasi* in South America. Infect. Genet.
708 Evol. 25, 44-51.
709 Maslov, D.A., Opperdoes, F.R., Kostygov, A.Y., Hashimi, H., Lukes, J., Yurchenko, V., 2018.
710 Recent advances in trypanosomatid research: genome organization, expression, metabolism,
711 taxonomy and evolution. Parasitology, 1-27.
712 Maslov, D.A., Sharma, M.R., Butler, E., Falick, A.M., Gingery, M., Agrawal, R.K., Spremulli,
713 L.L., Simpson, L., 2006. Isolation and characterization of mitochondrial ribosomes and ribosomal
714 subunits from *Leishmania tarentolae*. Mol. Biochem. Parasitol. 148, 69-78.
715 Maslov, D.A., Votýpka, J., Yurchenko, V., Lukeš , J., 2013. Diversity and phylogeny of insect
716 trypanosomatids: all that is hidden shall be revealed. Trends Parasitol. 29, 43-52.
717 Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzén, O., Lewis, M.D., Ramírez, J.D.,
718 Carrasco, H.J., Andersson, B., Miles, M.A., 2012. Multiple Mitochondrial Introgression Events and
719 Heteroplasmy in *Trypanosoma cruzi* Revealed by Maxicircle MLST and Next Generation
720 Sequencing. Plos Neglect. Trop. Dis. 6, e1584.
721 Momen, H., Cupolillo, E., 2000. Speculations on the origin and evolution of the genus *Leishmania*.
722 Mem Inst Oswaldo Cruz 95, 583-588.
723 NCBI, 2008. BLAST® Command Line Applications User Manual Bethesda (MD): National Center
724 for Biotechnology Information US.
725 Neboháčová, M., Kim, C.E., Simpson, L., Maslov, D.A., 2009. RNA editing and mitochondrial
726 activity in promastigotes and amastigotes of *Leishmania donovani*. Int. J. Parasit. 39, 635-644.

727 Noyes, H.A., Camps, A.P., Chance, M.L., 1996. *Leishmania herreri* (Kinetoplastida;
728 trypanosomatidae) is more closely related to *Endotrypanum* (Kinetoplastida; trypanosomatidae)
729 than to *Leishmania*. Mol. Biochem. Parasitol. 80, 119-123.
730 Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol 25, 1253-1256.
731 Ren, Z., Zhong, Y., Kurosu, U., Aoki, S., Ma, E., von Dohlen, C.D., Wen, J., 2013. Historical
732 biogeography of Eastern Asian-Eastern North American disjunct Melaphidina aphids (Hemiptera:
733 Aphididae: Eriosomatinae) on Rhus hosts (Anacardiaceae). Mol Phylogenet Evol 69, 1146-1158.
734 Rodgers, M.R., Popper, S.J., Wirth, D.F., 1990. Amplification of kinetoplast DNA as a tool in the
735 detection and diagnosis of *Leishmania*. Exp Parasitol 71, 267-275.
736 Sauquet, H., Ho, S.Y.W., Gandolfo, M.A., Jordan, G.J., Wilf, P., Cantrill, D.J., Bayly, M.J.,
737 Bromham, L., Brown, G.K., Carpenter, R.J., Lee, D.M., Murphy, D.J., Sniderman, J.M.K.,
738 Udovicic, F., 2012. Testing the Impact of Calibration on Molecular Divergence Times Using a
739 Fossil-Rich Group: The Case of Nothofagus (Fagales). Syst. Biol. 61, 289-313.
740 Shapiro, T.A., Englund, P.T., 1995. The structure and replication of kinetoplast DNA. Annual
741 review of microbiology
742 49, 117-143.
743 Simpson, A.M., Neckelmann, N., Cruz, V.F., Muhich, M.L., Simpson, L., 1985. Mapping and 5'
744 End Determination of Kinetoplast NMaxicircle Gene Transcripts from *Leishmania tarentolae*
745 Nucleic Acids Research 13, 5977-5993.
746 Simpson, A.M., Simpson, L., 1980. Kinetoplast DNA and RNA of *Trypanosoma brucei* Mol.
747 Biochem. Parasitol. 2, 93-108.
748 Simpson, L., Douglass, S.M., Lake, J.A., Pellegrini, M., Li, F., 2015. Comparison of the
749 Mitochondrial Genomes and Steady State Transcriptomes of Two Strains of the Trypanosomatid
750 Parasite, *Leishmania tarentolae*. Plos Neglect. Trop. Dis. 9, e0003841.
751 Som, A., 2015. Causes, consequences and solutions of phylogenetic incongruence. Briefings in
752 bioinformatics 16, 536-548.

753 Stevens, J.R., Noyes, H., Dover, G.A., Gibson, W.C., 1999. The ancient and divergent origins of the
754 human pathogenic trypanosomes, *Trypanosoma brucei* and *T-cruzi*. Parasitology 118, 107-116.
755 Stevens, J.R., Noyes, H.A., Schofield, C.J., Gibson, W., 2001. The molecular evolution of
756 Trypanosomatidae. Adv.Parasitol. 48, 1-56.
757 Stucky, B.J., 2012. SeqTrace: A Graphical Tool for Rapidly Processing DNA Sequencing
758 Chromatograms. Journal of Biomolecular Techniques : JBT 23, 90-93.
759 Swofford, D.L., 1993. PAUP - A Computer-Program for Phylogenetic Inference Using Maximum
760 Parsiomy J. Gen. Physiol. 102, A9-A9.
761 Swofford, D.L., Charles, D., B., 2017. PAUP Manual
762 Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S., 2012.
763 Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A 109, 19333-
764 19338.
765 Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control
766 region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10, 512-526.
767 Telleria, J., Lafay, B., Virreira, M., Barnabe, C., Tibayrenc, M., Svoboda, M., 2006. Trypanosoma
768 cruzi: sequence analysis of the variable region of kinetoplast minicircles. Exp Parasitol 114, 279-
769 288.
770 Torres-Guerrero, E., Quintanilla-Cedillo, M.R., Ruiz-Esmenjaud, J., Arenas, R., 2017.
771 Leishmaniasis: a review. F1000Research 6, 750.
772 Valdivia, H.O., Reis-Cunha, J.L., Rodrigues-Luiz, G.F., Baptista, R.P., Baldeviano, G.C., Gerbasi,
773 R.V., Dobson, D.E., Pratlong, F., Bastien, P., Lescano, A.G., Beverley, S.M., Bartholomeu, D.C.,
774 2015. Comparative genomic analysis of *Leishmania* (*Viannia*) *peruviana* and *Leishmania* (*Viannia*)
775 *braziliensis*. BMC Genomics 16, 10.
776 Votýpka, J., d'Avila-Levy, C.M., Grellier, P., Maslov, D.A., Lukeš , J., Yurchenko, V., 2015. New
777 Approaches to Systematics of Trypanosomatidae: Criteria for Taxonomic (Re)description. Trends
778 Parasitol. 31, 460-469.

WHO, 2018. Leishmaniasis. WHO.

Yang, B.B., Chen, D.L., Chen, J.P., Liao, L., Hu, X.S., Xu, J.N., 2013. Analysis of kinetoplast cytochrome b gene of 16 *Leishmania* isolates from different foci of China: different species of *Leishmania* in China and their phylogenetic inference. Parasites Vectors 6, 12.

Yatawara, L., Le, T.H., Wickramasinghe, S., Agatsuma, T., 2008. Maxicircle (mitochondrial) genome sequence (partial) of *Leishmania major:* gene content, arrangement and composition compared with *Leishmania tarentolae*. Gene 424, 80-86.

Yazaki, E., Ishikawa, S.A., Kume, K., Kumagai, A., Kamaishi, T., Tanifuji, G., Hashimoto, T., Inagaki, Y., 2017. Global Kinetoplastea phylogeny inferred from a large-scale multigene alignment including parasitic species for better understanding transitions from a free-living to a parasitic lifestyle. Genes & genetic systems 92, 35-42.

Yurchenko, V., Votýpka, J., Tesarová, M., Klepetková, H., Kraeva, N., Jirků, M., Lukeš, J., 2014. Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. Folia Parasitologica 61, 97-112.

Legends to Figures

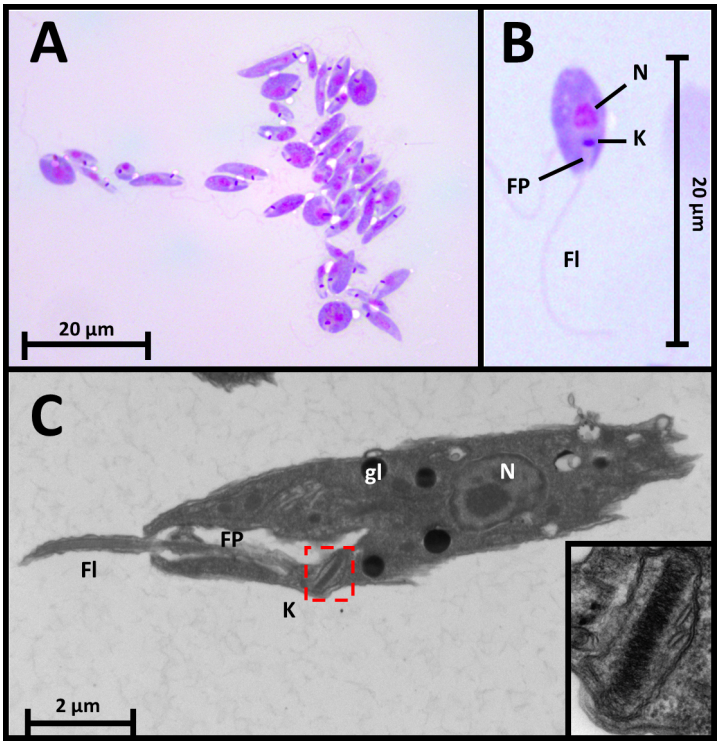
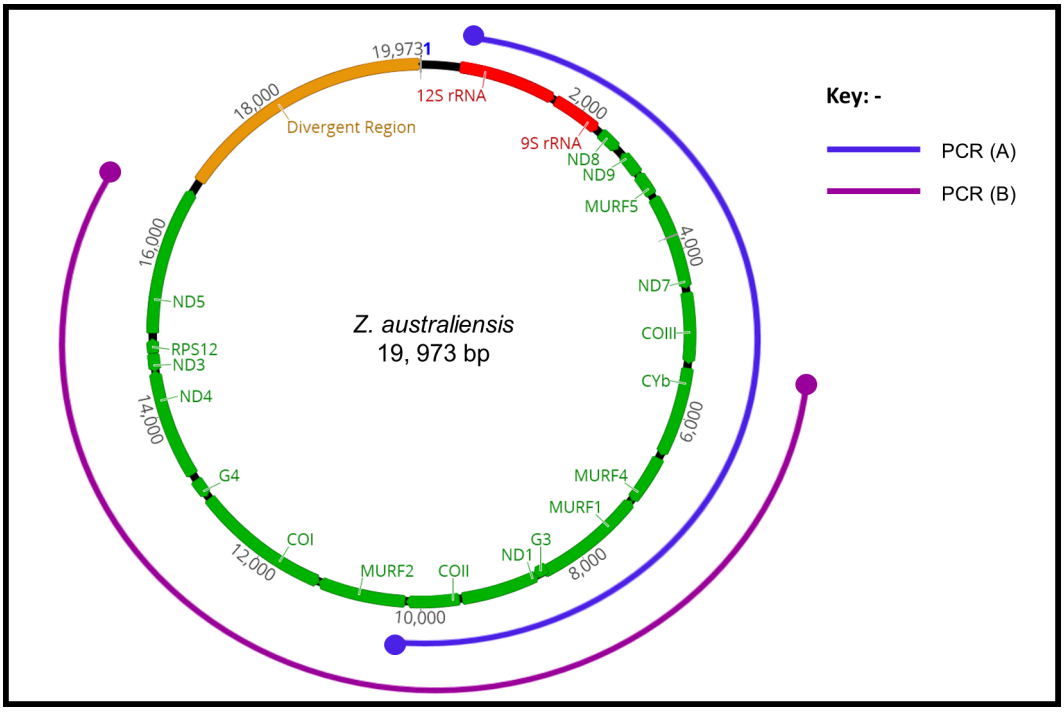


Fig. 1. Light and electron micrographs of *Zelonia australiensis* showing the kinetoplast.

(A) and (B) Light micrographs in a Leishman stained smear showing the morphological features of *Zelonia australiensis* promastigotes including the nucleus (N), kinetoplast (K), flagellar pocket (FP) and flagella (Fl). (C) Transmission electron micrograph showing additional gross morphological features of *Z. australiensis* including the glycosomes (gl) and a zoomed in micrograph of the kinetoplast in the bottom right corner of the lower panel.



804

805 **Fig. 2. Graphical map of *Z. australiensis* maxicircle genome assembled from Illumina**

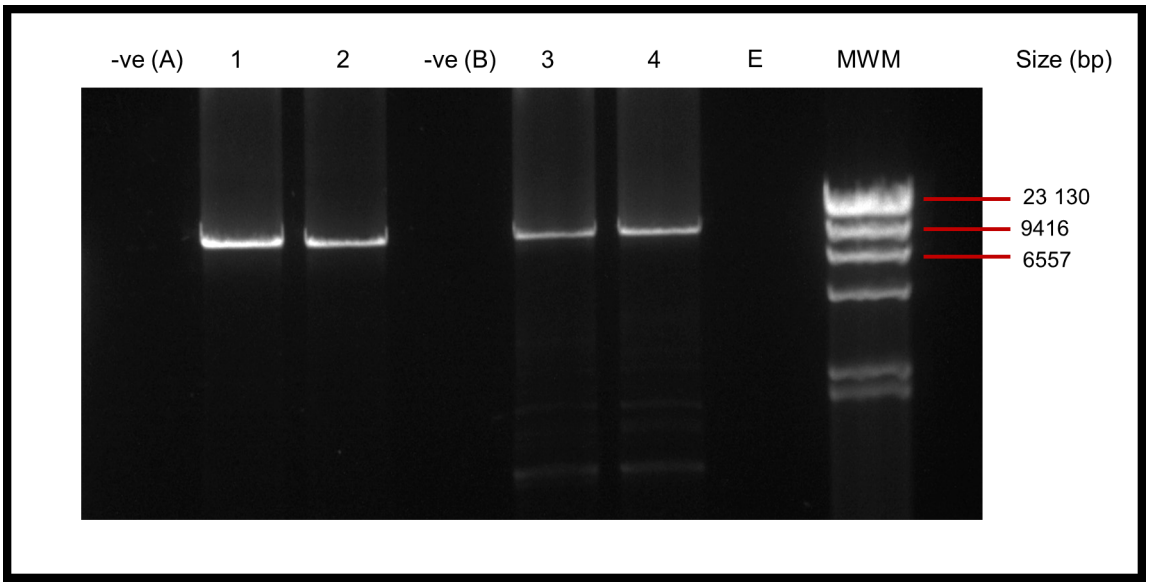
806 **sequencing of total DNA with the regions targeted by long-range PCR highlighted.**

807 The blue line (PCR product A), targets the genes from 12S rRNA to the end of COII and the purple

808 line (PCR product B), targets the genes from CYb to NADH5. A description of the function of the

809 genes shown in this figure is provided in Table 1.

810



811

Fig. 3. DNA electrophoresis of PCR products generated through optimised LR-PCR assays. Samples were run alongside a Lambda DNA *Hind* III Digest molecular weight marker (MWM) (Sigma Aldrich). PCR product A (lanes 1 & 2) and PCR product B (lanes 3 & 4) were run in duplicates, each against a negative control void of DNA, -ve (A) & -ve (b) respectively.

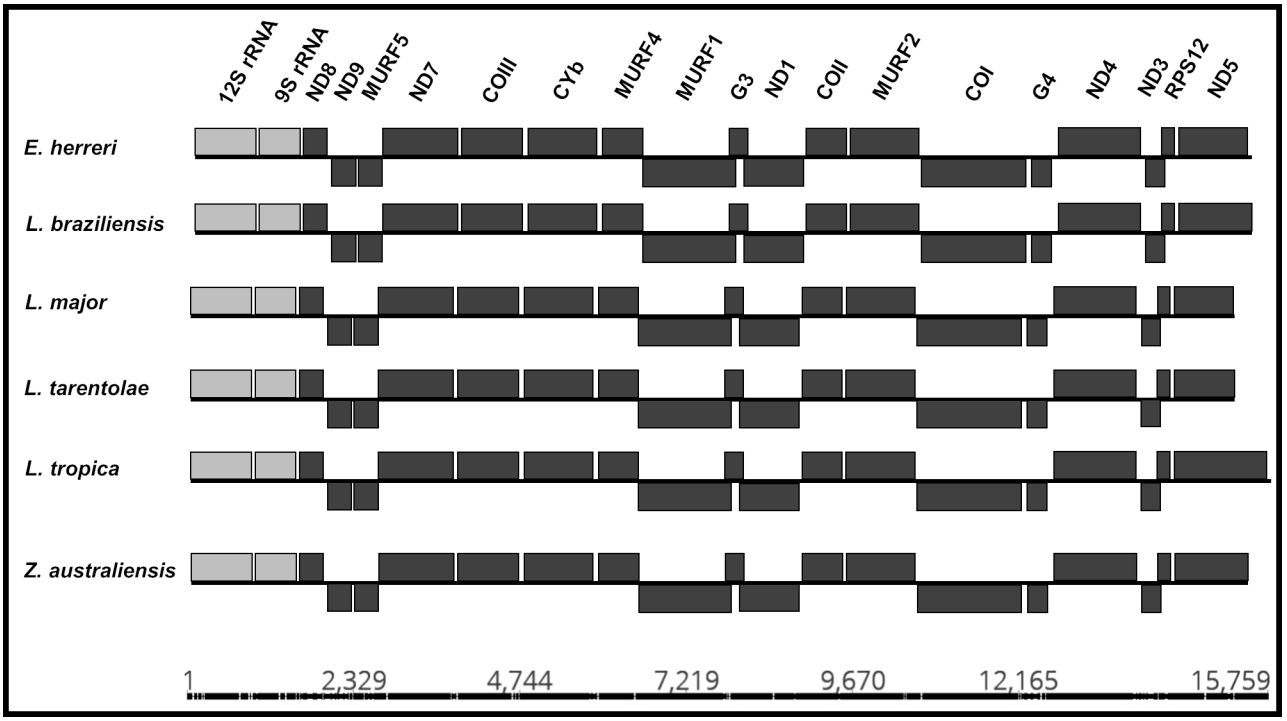
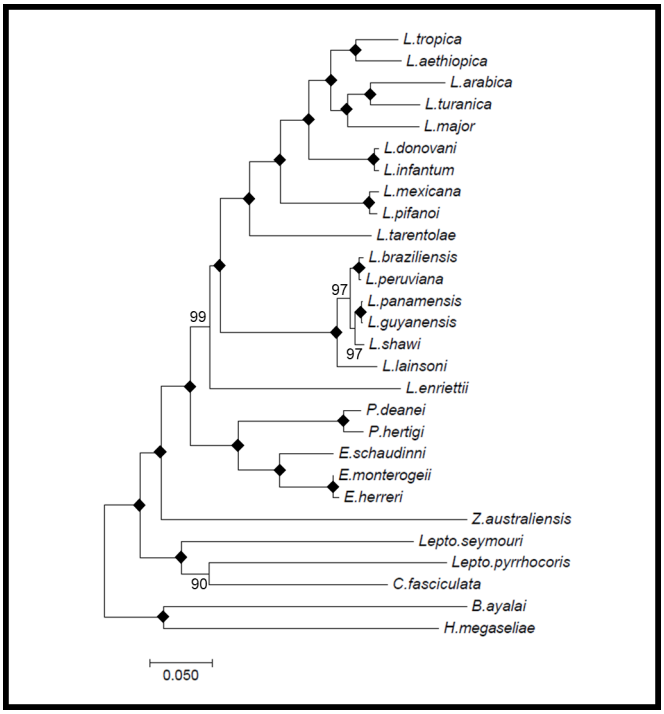


Fig. 4. Schematic diagram of the maxicircle genome sequence of various trypanosomatid spp. generated in this study.

The diagram is composed of the sequences assembled from the Illumina Miseq reads. Gene order and structure is shown of *L. braziliensis, E. herreri*, *L. major*, *L. tarentolae*, *L. tropica* and *Z. australiensis*. Light grey blocks represent rRNA genes and dark grey blocks represent protein-coding genes. Blocks above the line represent genes transcribed on the forward strand and blocks below represent genes transcribed on the reverse strand.



825

826 **Fig. 5. Inferred evolutionary relationship showing genetic distance between *Z. australiensis***

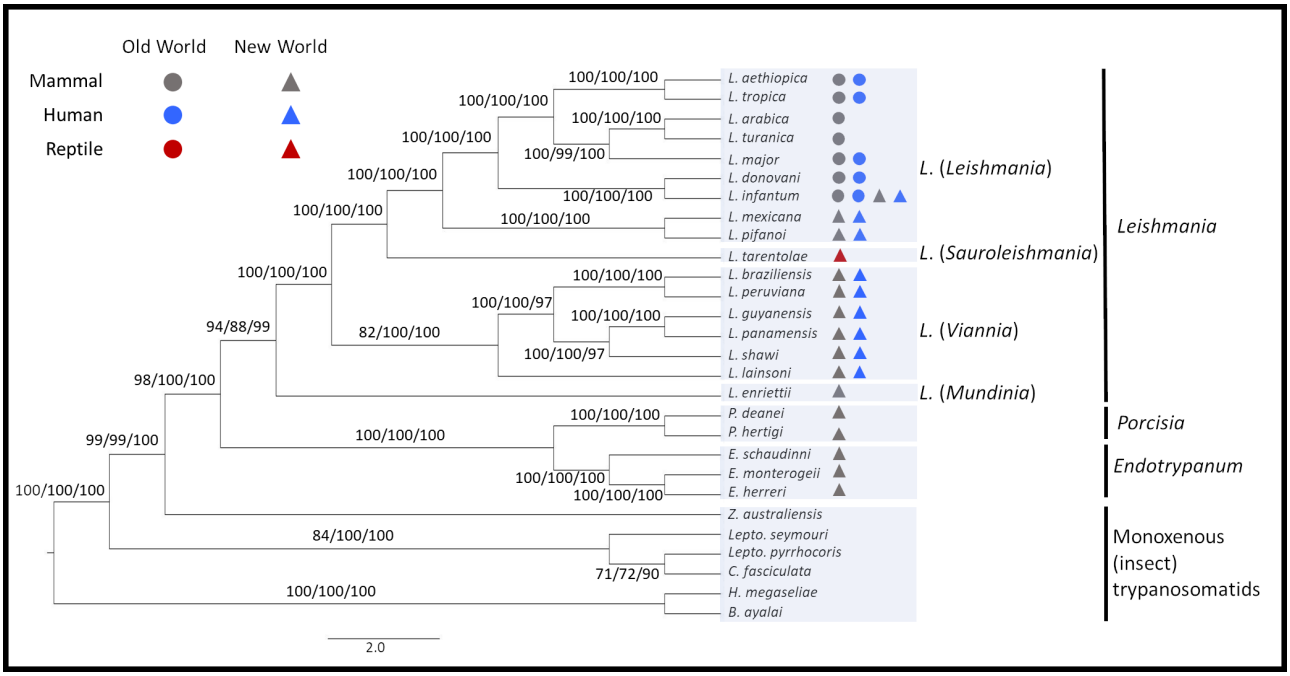
827 **and other trypanosomatids using the maxicircle coding region.**

828 The structure of this tree was inferred using the maximum likelihood based on the GTR+I+G model

829 with 1000 bootstrap replicates. A solid diamond indicates a node that obtained a bootstrap value of

830 100%. The scale bar represents the number of nucleotide substitutions per site.

831



832

833 **Fig. 6. Inferred evolutionary relationship between *Z. australiensis* and other trypanosomatids**

834 **using the maxicircle coding region.**

835 The structure of this tree was inferred using three statistical methods; the parsimony, distance and

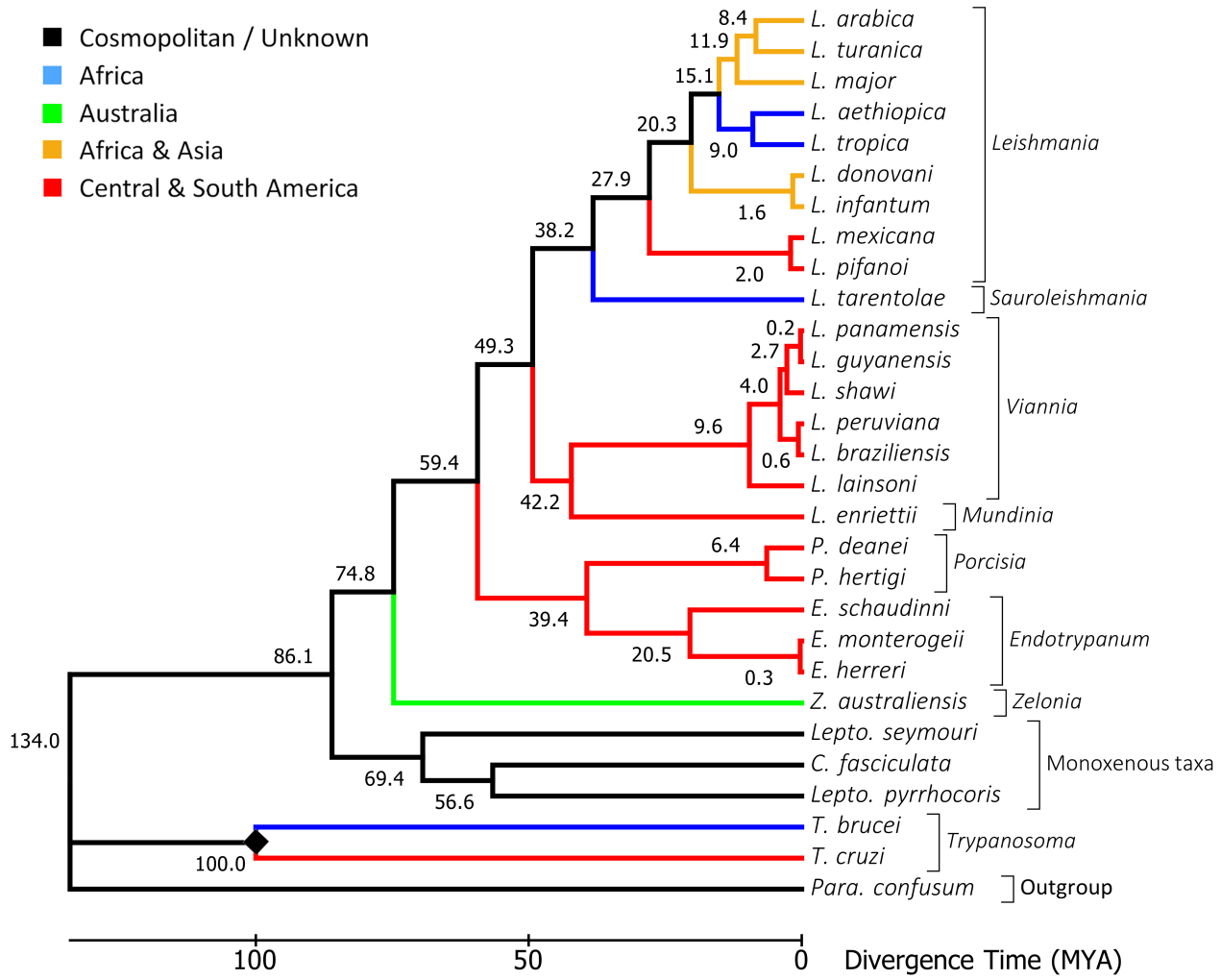
836 maximum likelihood based on the GTR+I+G model. The same tree structure was predicted using

837 each method. The first value at each node is the confidence interval using the parsimony method

838 based on 1,000 bootstrap replicates. The second and third number are the bootstrap support (1,000

839 replicates) values for the distance and ML methods respectively. The scale bar represents the

840 number of nucleotide substitutions per site.



841

842 **Fig. 7. Phylogenetic time tree inferring the evolutionary relationships between the**

843 **Leishmaniinae and other trypanosomatids using the maxicircle coding region.**

844 The structure of this tree was inferred using the NJ method from pairwise hamming distances

845 calculated using the phangorn package in R. The maximum likelihood of this tree was calculated

846 using the Jukes-Cantor model with 1000 bootstrap replicates (log likelihood: -188084.7), achieving

847 a bootstrap support value of 100% for each node. This tree includes several important dixenous

848 (*Leishmania*, *Endotrypanum*, *Porcisia* and *Trypanosoma* sp.) and monoxenous taxa (*Leptomonas*

849 and *Crithidia* sp.), as well as one representative of the genus *Zelonia* which sits on the

850 monoxenous/dixenous boundary. This timetree was computed using 1 calibration constraint,

851 indicated by a diamond (the divergence of *T. cruzi* and *T. brucei* approximately 100 million years

852 ago). Predicted divergence times are displayed on nodes.

853 **Tables with their legends**

854 **Table 1**

855 Main genetic information contained in the maxicircle.

Gene/Region	Description
9S rRNA and 12S RNA	The unusually small ribosomal RNAs of trypanosomatids are significantly smaller than both mammalian mitochondrial and eubacterial rRNAs (Maslov et al., 2006).
MURFs (MURF1, MURF2, MURF4 and MURF5)	MURFs are unidentified open-reading frames whose function is unknown (Yatawara et al., 2008).
ND (ND1, ND3, ND4, ND5, ND7, ND8, ND9)	The NADH dehydrogenase complex is comprised of the subunits involved in the mitochondrial membrane respiratory chain.
Cytochrome Oxidase I, II and III (COI-COIII)	Cytochrome Oxidase subunits I-III constitute the functional core of the enzyme complex. COI is the catalytic component of the respiratory chain responsible for the reduction of oxygen to water. COII transfers the electrons from cytochrome oxidase to the centre of the catalytic COI (Horvath et al., 2000).
G3 – G4	Pan edited cryptogenes, distinguished by intergenic G-rich regions (Neboháčová et al., 2009).
Cytochrome b (CYb)	Cyt *b* is the main redox catalytic subunit of the ubiquinol-cytochrome c reductase complex, which is a component of

	the mitochondrial respiratory chain (Asato et al., 2009).
RPS12	The single ribosomal protein encoded by the kDNA. The RPS12 gene function is ambiguous but is involved in the translation initiation step and its transcript undergoes extensive U-insertion/deletion editing (Aphasizheva et al., 2013).
Divergent region (DR)	The most variable region of the kinetoplast maxicircle. The non-coding segment consists almost entirely of repeats and is highly variable at the species-specific sequence level (Flegontov et al., 2006b).

856

857

858 **Table 2**

859 Data generated from LR-PCR following QC grooming in this study.

Species	PCR Target	Data type	Number and length of QC-trimmed paired-reads	Total combined size (bp)
Endotrypanum herreri	12S rRNA → COII	Reads	240 967 (250 paired-end)	15 306
	Cyt b → ND5		164 501 (250 paired-end)	
Leishmania braziliensis	12S rRNA → COII	Reads	28 543 (250 paired-end)	15 180
	Cyt b → ND5		261 561 (250 paired-end)	
Leishmania major	12S rRNA → COII	Reads	303 013 (250 paired-end)	14 821
	Cyt b → ND5		95 051 (250 paired-end)	
Leishmania tarentolae	12S rRNA → COII	Reads	56 313 (250 paired-end)	15 193
	Cyt b → ND5		55 236 (250 paired-end)	
Leishmania tropica	12S rRNA → COII	Reads	324 870 (250 paired-end)	15 559
	Cyt b → ND5		224 739 (250 paired-end)	
Zelonia australiensis	12S rRNA →	Reads	139 528 (250 paired-end)	15 104

COII	
Cyt b → ND5	92 588 (250 paired-end)

860 **Supplementary Data**

861 **S1 - Table**

862 List of trypanosomatid species used in this study.

863

864 **S2 – Table**

865 Long-range and conventional PCR primers used in this study.

866

867 **S3 - Table**

868 Number of nucleotide differences and genetic differences between maxicircle kDNA of various
869 trypanosomatid species. Below diagonal: genetic distance, above diagonal: number of nucleotide
870 bases which are not identical.

871

872