

Article

Evolutionary Insight into the Trypanosomatidae Using Alignment-Free Phylogenomics of the Kinetoplast

Alexa Kaufer ^{1,*}, Damien Stark ² and John Ellis ¹

¹ School of Life Sciences, University of Technology Sydney, Ultimo, NSW 2007, Australia; john.ellis@uts.edu.au

² Department of Microbiology, St Vincent's Hospital Sydney, Darlinghurst, NSW 2010, Australia; damien.stark@svha.org.au

* Correspondence: Alexa.Kaufer@student.uts.edu.au

Received: 23 August 2019; Accepted: 13 September 2019; Published: 18 September 2019

Abstract: Advancements in next-generation sequencing techniques have led to a substantial increase in the genomic information available for analyses in evolutionary biology. As such, this data requires the exponential growth in bioinformatic methods and expertise required to understand such vast quantities of genomic data. Alignment-free phylogenomics offer an alternative approach for large-scale analyses that may have the potential to address these challenges. The evolutionary relationships between various species within the trypanosomatid family, specifically members belonging to the genera *Leishmania* and *Trypanosoma* have been extensively studied over the last 30 years. However, there is a need for a more exhaustive analysis of the Trypanosomatidae, summarising the evolutionary patterns amongst the entire family of these important protists. The mitochondrial DNA of the trypanosomatids, better known as the kinetoplast, represents a valuable taxonomic marker given its unique presence across all kinetoplastid protozoans. The aim of this study was to validate the reliability and robustness of alignment-free approaches for phylogenomic analyses and its applicability to reconstruct the evolutionary relationships between the trypanosomatid family. In the present study, alignment-free analyses demonstrated the strength of these methods, particularly when dealing with large datasets compared to the traditional phylogenetic approaches. We present a maxicircle genome phylogeny of 46 species spanning the trypanosomatid family, demonstrating the superiority of the maxicircle for the analysis and taxonomic resolution of the Trypanosomatidae.

Keywords: Trypanosomatidae; kinetoplast; second-generation sequencing; third-generation sequencing; alignment-free phylogenetics

1. Introduction

Protozoan flagellates of the trypanosomatid family (syn. Trypanosomatidae) are obligate, unicellular parasites that infect a wide array of vertebrates, invertebrates and plants [1,2]. Protozoan parasites of this diverse family are predominately monoxenous (i.e., those restricted to a single, mainly invertebrate lifecycle), however the better known dixenous members (i.e., those with an invertebrate and vertebrate host) such as *Leishmania* and *Trypanosoma* are the causative agents of some of the most important neglected tropical diseases (NTD) including leishmaniasis, human African sleeping sickness (HAT) and Chagas disease [3–5]. The trypanosomatid family belongs to a distinct evolutionary lineage of eukaryotes within the class Kinetoplastida [6]. The current stance on kinetoplastid phylogeny is that the dixenous organisms evolved from the monoxenous members of

the Trypanosomatidae several times throughout history, leading to the independent emergence of the genera *Trypanosoma*, *Phytomonas* and a group that unites *Leishmania*, *Porcisia* and *Endotrypanum* [5,7,8].

Traditionally, phylogenetic relationships within the kinetoplastids, including the trypanosomatid family have been predominately based on the analysis of the SSU rRNA genes [7,9]. The evolutionary rate of substitutions in genes is considered one of the most important factors that influence the informativeness and robustness of phylogenetic analyses [10]. Advancements in molecular biology have demonstrated that slow-evolving genes (like the SSU rRNA) are not reliable markers for the deep level resolution of species in order to determine the exact branching within the trypanosomatid family [7]. Nowadays, a multi-marker approach using concatenated sequences of multiple genes are becoming the preferred choice and are routinely employed for the phylogenetic inference of related organisms [5].

Exclusive to kinetoplastid protozoans, the mitochondrial DNA of trypanosomatids is an extensive network of DNA circles which are condensed into a periflagellar structure known as the kinetoplast (kDNA) [11,12]. The kDNA consists of approximately 10,000 minicircles ranging from 0.5 kb to 10 kb and 20–50 larger maxicircles ranging from 20 to 40 kb [13,14]. Recent analyses of *Trypanosoma brucei* demonstrated that the kDNA constitutes only 4.18 megabase pairs (Mbp) of the trypanosomatids' total genome size of 77.7 Mbp [12]. Containing the mitochondrial homologues common to other eukaryotes, the maxicircle kDNA consists of two regions; a coding region containing the protein-coding genes and a highly repetitive non-coding region termed the divergent region (DR) [15–17]. In recent years, the maxicircle genome has become well-established as a superior taxonomic marker for the evolutionary analyses between related organisms of the Leishmaniinae and *Trypanosoma* [11,18,19]. Following suit with this rationale, the maxicircle kDNA should provide a more resolute model to investigate the genetic relationships between the entire trypanosomatid family, providing an all-encompassing analysis on the origins and biology of the Trypanosomatidae.

Over the past decade, second-generation sequencing (SGS) and third-generation sequencing (TGS) have become the leading technologies in the typing and evolutionary analyses of related organisms [20]. While short-read SGS techniques such as Illumina platforms have revolutionised biomedical research, their limitations, specifically their short-read (SR) lengths, make them inadequately suited for the assembly of complex and highly repetitive genomic regions [21]. Long-read TGS techniques such as PacBio offer longer read lengths (average >10 kb) than those of SGS, making it an ideal candidate for complex genomes [22]. However, a long-read (LR) length is hindered by a higher error rate of approximately 11%–15% compared to that of SGS [22,23]. Despite this, the advantages of SGS and TGS are complementary, offering an alternative 'hybrid' strategy that makes use of both technologies to overcome the drawbacks of each method alone. These hybrid methodologies using long (and inaccurate) and short (and accurate) have proven to be extremely useful in producing high-quality, accurate assemblies [23,24].

Next-generation sequence data has paved the way for the use of phylogenomic approaches for the analysis of evolutionary relationships. Phylogenomics is the junction between evolution and genomics, using the comparative analysis of genome scale data for the reconstruction of evolutionary histories between organisms [25]. However, analyses with larger datasets across a wider breadth of taxa are becoming increasingly computationally infeasible [26]. The vast increase in genetic information necessitates the exponential growth in bioinformatic methods and expertise required to understand such immense quantities of genome-scale data. Thus, for large-scale analyses of genomes (i.e., phylogenomics), alignment-free (AF) methods of phylogenetic inference have been increasingly employed over the last few years [27]. Alignment-free software was first introduced nearly a decade ago but received little attention due to the traditional belief of its inferior resolution to multiple sequence alignment (MSA) based methods. However, due to the immense quantities of genome-scale data being produced, recent years have seen a surge in publications using AF applications for the phylogenetic analysis of organisms [20,26–32].

Traditionally, MSA-phylogenetics is based on the correspondence of individual nucleotides or amino acids that are in the same order between the species analysed [33]. Alignment-based methods

for the most part yield excellent results when the dataset of sequences can be reliably aligned, however in certain instances, alignment-based sequence analyses can become problematic [34]. When sequences are divergent (i.e., analysis of the entire trypanosomatid family), the accuracy of sequence alignments decreases rapidly when the sequence identity falls below a certain critical point [33]. Second, alignment-based protocols are computationally intensive on a genome-scale [33], which is problematic with the increasing trend of whole-genome sequencing (WGS) replacing the use of single or few genes for the phylogenetic inference of related organisms.

The increasing trend of alignment-free sequence analyses offers an alternative approach for large-scale comparisons that may have the potential to address these limitations. One such alignment-free method introduced by Sims et al. [35] uses a measure based on the divergence between feature frequency profiles (FFPs), where the features (syn. word length) called k-mers are short nucleotide or amino-acid sequences of length k . For example, two sequences $x = \text{TTAAGG}$ and $y = \text{AAGGCC}$ and a feature length or k-mer size of three nucleotides produces $K_3^x = (\text{TTA}, \text{TAA}, \text{AAG}, \text{AGG})$ and $K_3^y = (\text{AAG}, \text{AGG}, \text{GGC}, \text{GCC})$ [33]. The frequency of these sub-sequences of a defined length or k-mers (k) are counted via a sliding frame implementation and used to calculate distance scores, which are subsequently used to generate a phylogenetic tree [35].

The aim of this paper is two-fold. First, the validation of the FFP protocol using the trypanosomatid family, in order to determine the limitations and variables in which the method provides reliable and robust results. Second, the application of this method to establish and summarise the current evolutionary relationships of the Trypanosomatidae. Rather than using data from the entire kinetoplast (minicircles and maxicircles), we focus on assembled genomes of the maxicircle only. In our previous study, we demonstrated that the maxicircle represents a superior, phylogenetically informative marker for studying the evolutionary relationships of the Leishmaniinae [19]. In this study, we expand on this concept and use the maxicircle kDNA to analyse not only the members of the Leishmaniinae, but representatives of the entire trypanosomatid family to provide an updated scheme for the taxonomic classification and resolution of the Trypanosomatidae.

In this study we applied a traditional MSA and AF method to the coding region of the maxicircle from a dataset of 46 trypanosomatid species. In addition to sequences already available from online databases and our previous study [19], twelve maxicircle genomes were assembled using both SR and SR/LR hybrid assemblies from raw Illumina and PacBio sequence data freely available from online databases (Sequence Read Archive). As a contribution to these efforts, this study contributes to our understanding of the phylogenomic relationships of the trypanosomatid family, demonstrating the power and robustness of the alignment-free analysis method based on the divergence between feature frequency profiles.

2. Materials and Methods

2.1. Samples

The trypanosomatid species used in this study are listed in Supplementary Materials S1 (S1 file).

2.2. Genome Assembly and Sequence Analysis

To obtain the complete maxicircle genome, processed reads were assembled from WGS data freely available through the Sequence Read Archive (SRA) on NCBI. Four paired-end Illumina and eight hybrid assemblies used a combination of long-read (PacBio) and short-read (Illumina) data to generate a hybrid contig using SPAdes version 3.12.0 [36]. For the hybrid assemblies, the datasets analysed were selected on the basis that both Illumina and PacBio sequence reads were available from the same author/provider and available for identical isolates/species. The maxicircle kDNA assembled from the Illumina assemblies (*Leishmania macropodum*, *Leishmania martiniquensis*, *Trypanosoma grayi* and *Phytomonas françai*) and hybrid assemblies (*Leishmania aethiopica*, *Leishmania amazonensis*, *Leishmania braziliensis*, *Leishmania guyanensis*, *Leishmania infantum*, *Leishmania mexicana*, *Leishmania tropica* and *Trypanosoma brucei rhodesiense*) were identified through BLAST analysis using

NCBI BLAST software [37]. Annotation, gene identification and sequence analysis of the maxicircle from whole-genome sequencing was completed using Geneious version 11.0.2 [38]. The GC% content and GC skew was visualised with the software DNAPlotter [39].

In a given DNA sequence, we measure skewed mononucleotide frequencies by:

$$\text{ATS} = \frac{f_A - f_T}{f_A + f_T} \quad \text{and} \quad \text{GCS} = \frac{f_G - f_C}{f_G + f_C}$$

where f_N denotes the observed frequency of nucleotides A, T, G and C [40].

2.3. Comparative Analysis of the Non-Coding Divergent Region of the Trypanosomatid Family

Self-dot plot analyses were generated in Geneious using the EMBOSS 6.5.7 software Dottup suite add-on [38]. The repetitive portion of the non-coding divergent region of 42 trypanosomatid species were visualised as a dense block of identity in a dot matrix homology search comparing each species against its own DNA sequence. The divergent region of *Endotrypanum herreri*, *Trypanosoma vivax* (Liem strain), *Trypanosoma copemani* and *Trypanosoma cruzi* (Silvio strain) were not available from the sequence reads for subsequent analyses. Additionally, REPuter was also used to identify repeat sequences including direct and palindromic repeats within the divergent region of these 42 trypanosomatid species [41]. A minimum repeat size of 60 bp and 15 bp was chosen respectively for each repeat identification. Tandem repeats in all five species were identified using Tandem Repeats Finder version 4.09 with default settings [42].

2.4. MSA Phylogenetic Analysis

A phylogenetic analysis was performed using the entire coding region from the maxicircle of 46 trypanosomatid species to investigate the evolutionary relationships between members of the Trypanosomatidae. All sequences were aligned using the MUSCLE algorithm implemented in the Seaview software package [43]. Phylogenetic relationships were inferred using the maximum likelihood optimality criterion using PhyML version 3.0. For ML trees, the best-fit model of evolution, GTR+I+G was selected using jModelTest 2.1 under the Bayesian information criterion [44]. Bootstrap support for clade topologies was estimated following the analysis of 1000 pseudo-replicate datasets using a heuristic tree search.

2.5. Alignment-Free FFP Analysis

The alignment-free FFP analysis was performed on the maxicircle coding region genome of 46 trypanosomatid species using the command line software FFP version 3.1.9 [35]. To determine the optimal length of k-mer, the FFPvprof utility was used to determine the average lower word length limit and the FFPvpreprof utility was used to determine the average upper word limit of all 46 trypanosomatid species [20]. The optimum k-mer length can be identified from the overlapping region of length ranges (i.e., average of the upper and lower length limit). To determine the lower k-mer length limit, FFPvprof calculates the number of k-mers (and length) that occur at least twice between the dataset. To determine the upper k-mer length limit, FFPpreprof calculates the relative entropy frequency (REF) between the expected and observed frequencies of specified k-mer lengths. Following optimisation, the UNIX-style command line pipeline was used to generate a distance matrix, which is used for the subsequent phylogenomic inference using the program Phylip version 3.697 [45]. Bootstrap support for clade topologies was estimated using the FFPboot utility following the analysis of 1000 pseudo-replicate datasets.

2.6. Estimating Divergence Time

Divergence dates of the trypanosomatid family were estimated using the Realtime method [1] and the General Time Reversible model [2] of the MEGA7 package. The maximum likelihood of this timetree was computed using one calibration constraint; the divergence of *Leishmania enriettii* and *Leishmania macropodum* approximately 40 MYA [46,47]. The maxicircle sequence of the monoxenous trypanosomatid *Paratrypanosoma confusum* served as an outgroup.

3. Results

3.1. Assembly of Data from Illumina and Hybrid Illumina/Pacbio Reads

The initial assemblies of Illumina paired-end reads using the SPAdes assembler resulted in the generation of twelve maxicircle genomes from the trypanosomatid family. Of these, we used the same Illumina dataset of eight species; *L. aethiopica*, *L. amazonensis*, *L. braziliensis*, *L. guyanensis*, *L. infantum*, *L. mexicana*, *L. tropica* and *T. brucei rhodesiense* in addition to their corresponding PacBio data for a hybrid assembly. The assemblies from both short-read and combination short- and long-read data resulted in a high-quality assembly of the maxicircle kDNA. A comparison of the final contig generated from paired-end reads to the hybrid revealed a 100% sequence identity between the coding region, reflecting the excellent accuracy of the maxicircle genome assembled from whole gene sequence generated by both SGS and TGS (Table 1). Comparison of the non-coding divergent region revealed a sequence identity >99% between the contigs generated from paired-end reads to the hybrid. The long-read lengths of the PacBio data were extremely valuable for the de novo assembly of the maxicircle, allowing us to overcome the problems caused by the repetitive nature of the divergent region. The difference in sequence identity of the divergent region is due to the increased resolution of the hybrid approach, yielding a more accurate estimate of the length of the divergent region than that of Illumina reads alone. Compared to that of the hybrid approach, SGS reads of *L. aethiopica*, *L. amazonensis*, *L. braziliensis*, *L. guyanensis*, *L. infantum*, *L. mexicana*, *L. tropica* and *T. brucei rhodesiense* assembled only 20%, 11%, 20%, 43%, 49%, 46%, 96% and 88% of the divergent region produced from TGS respectively.

Table 1. Comparison of the short-read Illumina and hybrid Illumina/PacBio assembly of various trypanosomatid species showing output length and sequence identity (%).

	Coding Region		Non-Coding Region		Overall	
	Size (bp)	% Identity	Size (bp)	% Identity	Size (bp)	% Identity
<i>L. aethiopica</i> (Illumina)	16 210	100	304	100	17 264	100
<i>L. aethiopica</i> (Hybrid)	16 210	100	1552	100	18 571	100
<i>L. amazonensis</i> (Illumina)	16 381	100	527	99.7	17 941	100
<i>L. amazonensis</i> (Hybrid)	16 381	100	4621	99.7	23 616	100
<i>L. braziliensis</i> (Illumina)	16 232	100	1121	98.5	18 118	99.9
<i>L. braziliensis</i> (Hybrid)	16 232	100	5745	98.5	23 012	99.9
<i>L. guyanensis</i> (Illumina)	16 235	100	1607	100	18 753	100
<i>L. guyanensis</i> (Hybrid)	16 235	100	3694	100	20 986	100
<i>L. infantum</i> (Illumina)	16 277	100	628	99.5	18 287	100
<i>L. infantum</i> (Hybrid)	16 277	100	1280	99.5	18 637	100
<i>L. mexicana</i> (Illumina)	16 472	100	736	99.7	17 946	100
<i>L. mexicana</i> (Hybrid)	16 472	100	1606	99.7	18 696	100
<i>L. tropica</i> (Illumina)	16 229	100	1582	97	18 800	99.5
<i>L. tropica</i> (Hybrid)	16 229	100	1644	97	19 020	99.5
<i>T. brucei rhodesiense</i> (Illumina)	14 905	100	2785	97	18 200	99.5
<i>T. brucei rhodesiense</i> (Hybrid)	14 905	100	3131	97	18 583	99.5

3.2. Genomic Organisation and Patterns of the Maxicircle

Sequence analysis of the maxicircle from representative species of the genera *Leishmania*, *Porcisia*, *Endotrypanum*, *Zelonina*, *Leptomonas*, *Crithidia*, *Herpetomonas*, *Angomonas*, *Blephomonas*, *Trypanosoma* and *Paratrypanosoma* revealed a conserved region typical for the maxicircle. The conserved region includes; 12S rRNA (large subunit), 9S rRNA (small subunit), seven subunits of NADH dehydrogenase, (ND8, ND9, ND7, ND1, ND3, ND5 and ND5), three subunits of cytochrome *c* oxidase (COI, COII and COIII), one subunit of the cytochrome *b*_{c1} complex (CYb) a single ribosomal protein (RPS12) and four open-reading frames whose role is unknown (MURF1, MURF2, MURF4 and MURF5). There was also two G-rich (G3 and G4) or C-rich (CR3 and CR4) pan-edited cryptogenes in *Leishmania* and *Trypanosoma* species respectively. Analysis of the *Phytomonas francai* maxicircle

revealed that the genes for subunits I, II and III (COI, COII and COIII) of cytochrome *c* oxidase and cytochrome *bc*₁ complex (CYb) were missing (Figure 1).

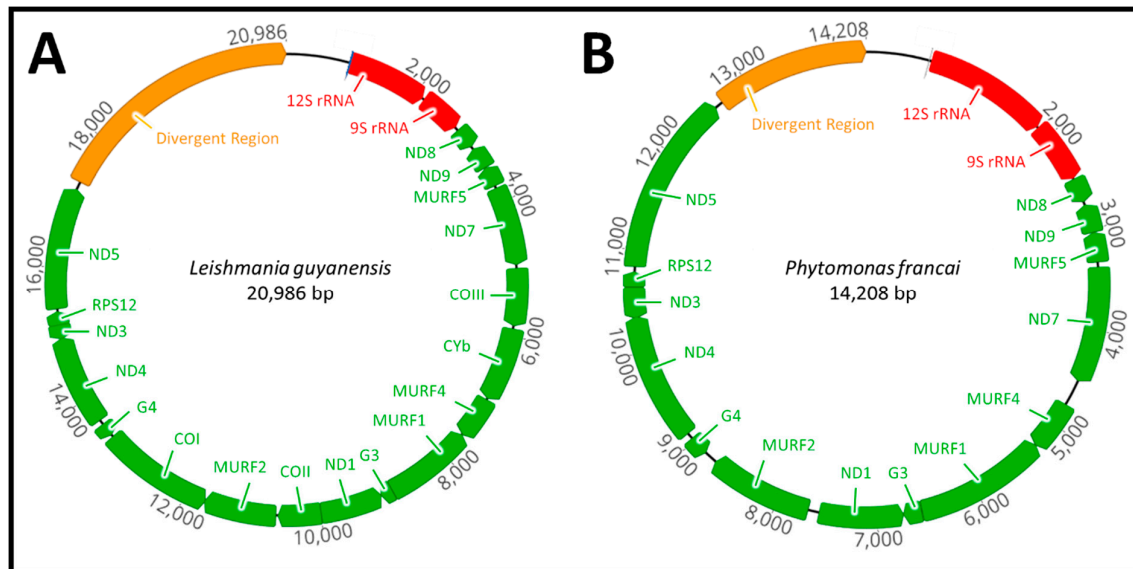


Figure 1. Graphical map of *Leishmania guyanensis* and *Phytomonas francai* maxicircle genomes assembled from Illumina and PacBio sequencing. Comparison of the maxicircle genome of (A) *Leishmania guyanensis* and (B) *Phytomonas francai*. *Phytomonas* spp. is characterised by the loss of the three subunits of the cytochrome *c* oxidase (COI, COII and COIII) and single subunit of cytochrome *b* (CYb).

The overall nucleotide frequency and skew throughout the maxicircle genome of all 46 trypanosomatid species is shown in Supplementary Materials S2. The overall AT-richness of the maxicircle ranged from 58%–82%, reflecting the AT-overabundance of the mitochondrial genome as a whole. Of the clinically important *L. braziliensis*, *T. brucei rhodesiense* and *T. cruzi* (CL strain), the GC skew and GC% content is represented in the top panel of Figure 2. From the GC content, it was observed that the AT-richness of *T. brucei rhodesiense* and *T. cruzi* is primarily due to the repeat region, where the GC% of the divergent region is predominately below average (dark blue lines). The AT-richness of *L. braziliensis* is still high (80%), although the GC plot demonstrates that the AT-richness in the divergent region is to a lesser extent less than that of *T. brucei rhodesiense* and *T. cruzi*, with the AT-abundance seen throughout the entire genome. The AT skew is shown in the bottom panel of Figure 2. The AT skew of all three species demonstrates the bias towards a T-rich sequence that is more pronounced in the coding region compared to that of the DR, which demonstrates a bias towards an A-rich nucleotide frequency.

A dot matrix search of the entire maxicircle sequence (left panel) and divergent region (right panel) of *L. braziliensis*, *T. brucei rhodesiense* and *T. cruzi* (CL strain), versus itself is shown in Figure 3. The comparative analysis for the remaining trypanosomatid species can be found in Supplementary Materials S3. Based on the left panel of each graph, we can see a clear distinction between the coding and non-coding region of the maxicircle sequence, which is highlighted in a black outline. The box structures located on and residing symmetrically around the line of identity are indicative of a clustered organisation of repetitive sequences. Multiple diagonal lines seen in all three panels are indicative of direct repeats. The length of the divergent region of the various trypanosomatid species is significantly variable and thus responsible for the size difference of the complete maxicircle genome of the kinetoplast.

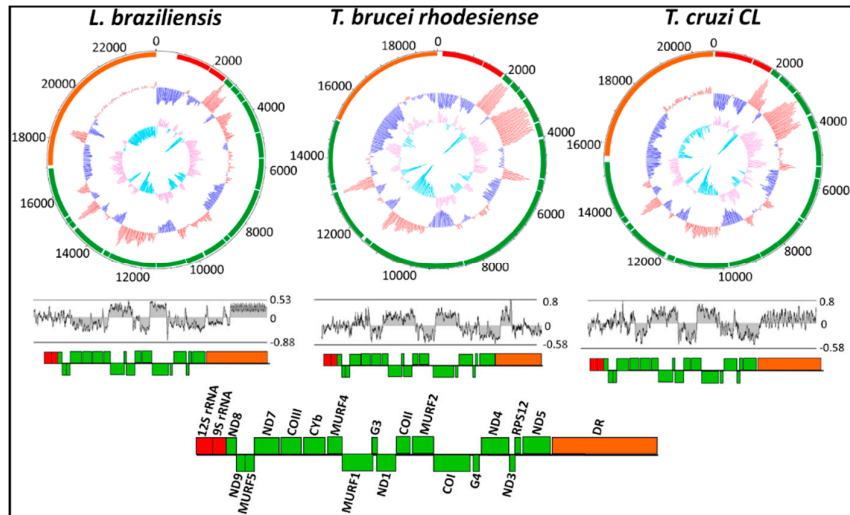


Figure 2. Circular confirmation of GC plot and GC skew of the maxicircle kinetoplast for *Leishmania braziliensis*, *Trypanosoma brucei rhodesiense* and *Trypanosoma cruzi*. The outer ring indicates gene arrangement and gene distribution. The middle circle represents the GC plot showing GC% content (dark blue for below-average and red for above-average) and the inner circle represents the GC skew (light blue for positive and orange for negative). The AT skew and corresponding gene arrangement and distribution of each species is shown below the circular plot.

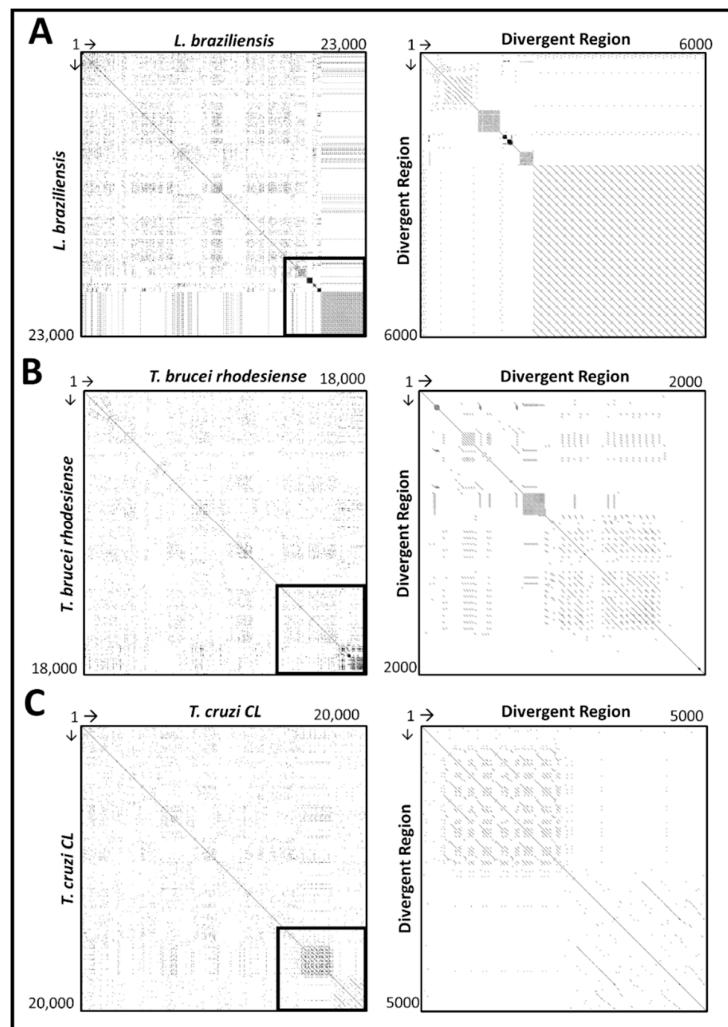


Figure 3. Self Dottup plot comparative analysis of the entire maxicircle genome (right panel) and divergent region (left panel) of *Leishmania braziliensis* (A), *Trypanosoma brucei rhodesiense* (B) and *Trypanosoma cruzi* (C) against their own sequences. The sequence of each species is placed on the axis and full identity over a 10 bp-long window is represented by a dot. The main diagonal line represents the sequence's alignment with itself and the lines about the main diagonal represent repetitive patterns within the maxicircle sequence.

The repeat analysis demonstrated that a different number of the various repeats were observed in the species compared (Figure 4). In these divergent region sequences, the repeat analysis detected tandem, palindromic, forward repeats. The various repeats found within the non-coding region differed significantly between species (Supplementary Materials S4). Following suit with the overall composition of the maxicircle, the forward, tandem and palindromic repeats all demonstrated a strong bias towards an AT-rich repeat in all trypanosomatid species compared. Tandem repeats identified were highly abundant and densely interspersed throughout the divergent region. The majority of tandem repeats are arranged in clusters composed of predominately short A_nT_n repeats.

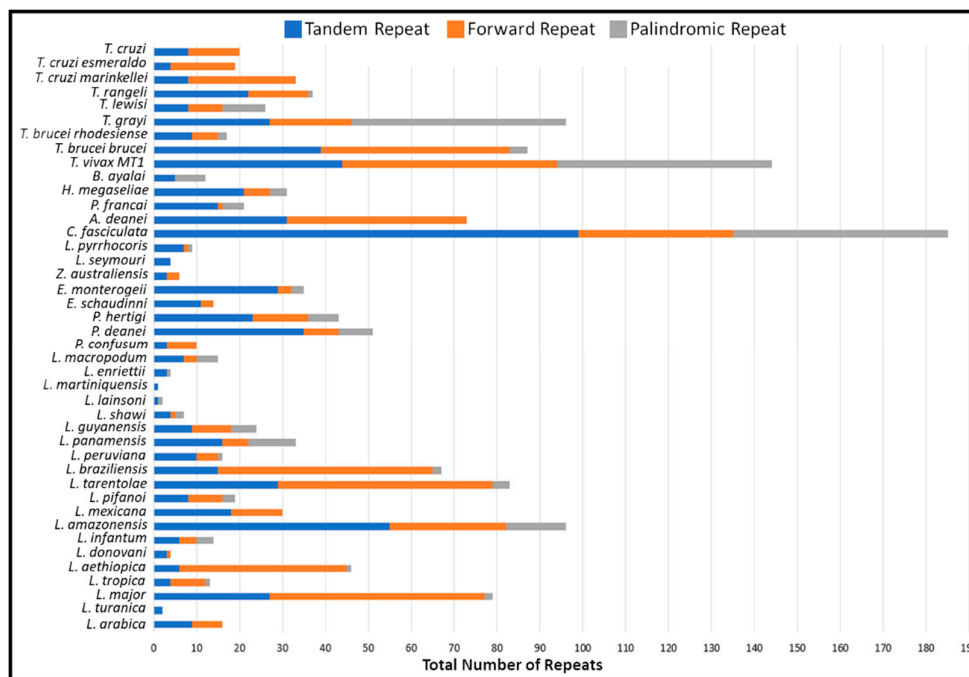


Figure 4. Analysis of total repeated sequences in the maxicircle divergent region of various trypanosomatid species. Totals of three repeat types; (Blue) Number of tandem repeats; (Orange) Number of forward repeats; (Grey) Number of palindromic repeats.

3.3. Phylogenetic Analyses

To assess how FFP-based phylogenies compares to the traditionally aligned-phylogenies, the phylogenetic relationships showing the genetic distance between members of the Trypanosomatidae based on the MSA- and AF-approach is shown in Figure 5A,B. The sequences of 46 trypanosomatid species were used to determine the optimal word length or k-mer for subsequent alignment-free analyses. From the FFPvprof and FFPpreprof analyses, the average lower word length (k-mer) limit of the 45 species was 14 ($k = 14$) and the average upper word length limit was 30 ($k = 30$) leading to the selection of k-mer lengths of 22 (Supplementary Materials S5).

In the inferred phylogenies, all *Leishmania* and *Trypanosoma* species formed two separate, strongly supported monophyletic clades. The FFP analysis showed the same subgrouping of the Trypanosomatidae, with all six subfamilies and eleven genera corresponding to the PhyML-generated phylogenetic tree. The subfamily Leishmaniinae unites the monoxenous protozoans of

insects (genera *Zelonia*, *Leptomonas* and *Crithidia*) and dixenous protozoans of insects and vertebrates (genera *Leishmania*, *Endotrypanum* and *Porcisia*). The monoxenous subfamilies Phytomonadinae, Strigomonadinae and Blechomonadinae include parasites of the genera *Herpetomonas*, *Angomonas* and *Blechomonas* respectively. In agreement with a recent study [11], despite *T. rangeli* and *T. lewisi* belonging to the same subgenus *Herpetosoma*, *T. rangeli* is more closely related to *T. cruzi*, clustering with the Schizotrypanum clade of trypanosomes.

The discrepancies observed were the branching of the Salivarian trypanosomes, where *T. brucei* *brucei* and *T. brucei* *rhodesiense* formed a clade basal to the *T. vivax* clade and between the monoxenous *A. deanei* and *P. francai*. The distance matrix generated from the alignment-free analysis is listed in Supplementary Materials S6.

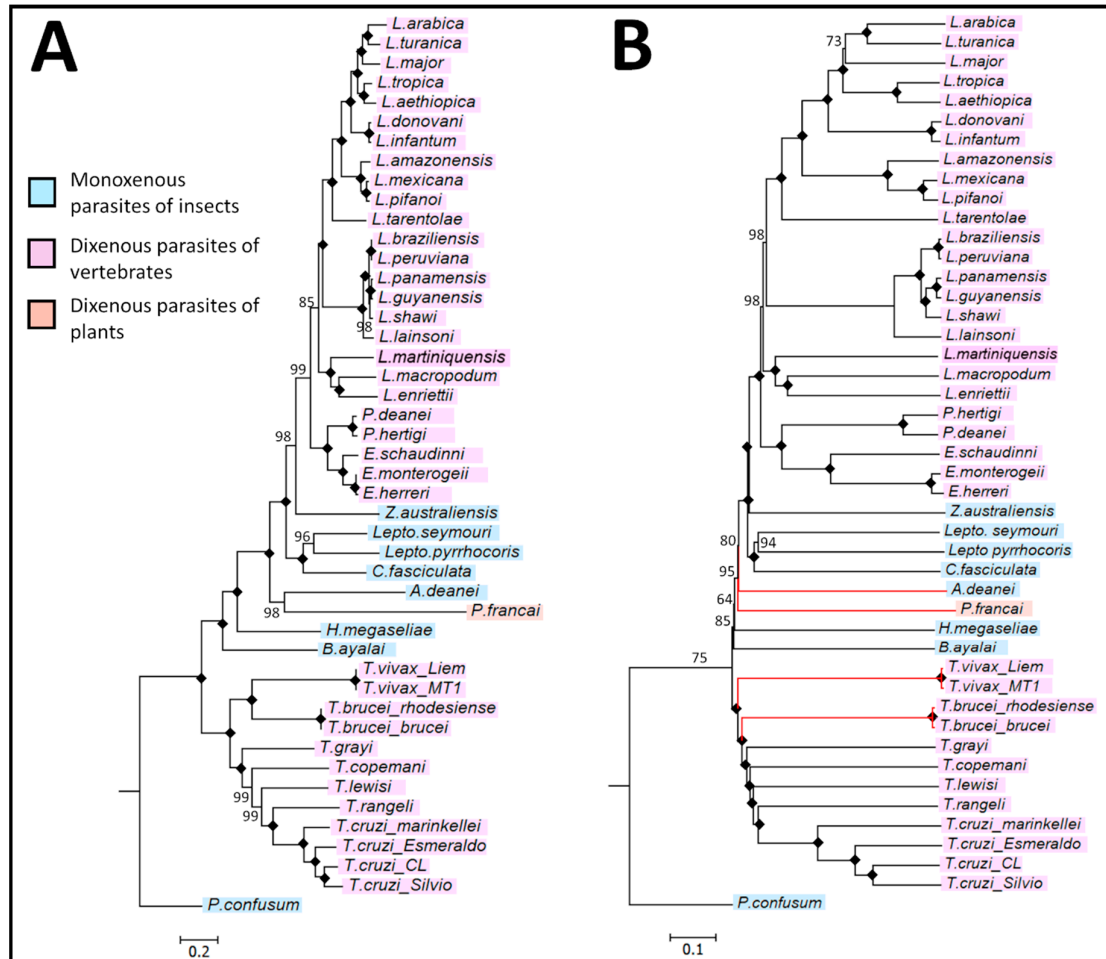


Figure 5. Inferred evolutionary relationships between species of the trypanosomatid family using aligned and alignment-free analysis of the maxicircle coding region. (A) PhyML-derived phylogeny showing the relationships between members of the trypanosomatid family using a multiple sequence alignment analysis. The maximum likelihood optimality criterion was used for the phylogenetic inference of the dataset with 1000 bootstrap samples. (B) Feature frequency profiles (FFP)-estimated phylogeny between species of the trypanosomatid family using an alignment-free analysis of the maxicircle coding region. Branches in red represent species whose clustering differs from those in (A). In both (A) and (B) a black diamond highlights a node that obtained a bootstrap value of 100% confidence and the scale bars depict the number of nucleotide substitutions per position.

3.4. Divergence Time Estimates of the Trypanosomatidae

The node depicting the separation of *L. macropodum* and *L. enriettii* was chosen as a calibration marker. The divergence date was established at an average of 40 MYA, which is the time period when

Australia and South America became separated, representing a minimum date of divergence [46,47]. From this calibration point, a common ancestor to the Trypanosomatinae subfamily and specifically the *T. cruzi* clade (*T. cruzi* and *T. rangeli*) was predicted to have appeared 150 MYA and 84 MYA respectively (Figure 6). The dixenous members of the Leishmaniinae was predicted to have first appeared 95 MYA, arising from a monoxenous ancestor.

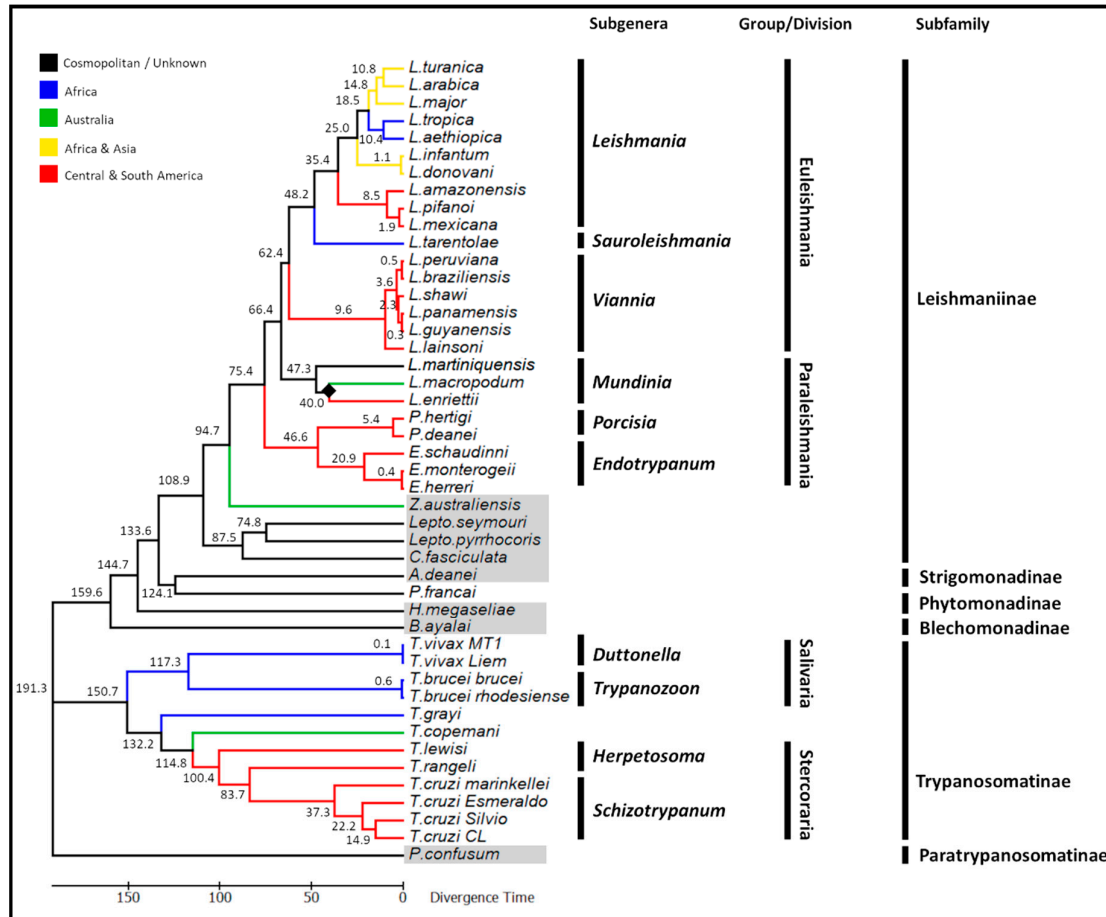


Figure 6. Phylogenetic time tree demonstrating the complex evolutionary relationships of the trypanosomatid family using the coding region of the maxicircle kDNA. The maximum likelihood of this tree was inferred using the GTR model. The time tree was calculated using a single calibration date, illustrated by a solid black diamond (the separation of *Leishmania macropodum* and *Leishmania enriettii* approximately 40 million years ago). Predicted divergence times of the various species are displayed on the nodes of the tree. Monoxenous trypanosomatid species are highlighted in grey.

4. Discussion

In this study we applied the alignment-free analysis of frequency feature profiles (FFP) for the phylogenomic analysis of the trypanosomatid family based on the entire maxicircle coding region genome. For closely related organisms i.e., within a single genus, alignment-based methods are preferred because in the most part, their genetic similarities allow a relatively straight-forward alignment of high-quality [33]. However, in the context of looking at an entire family of organisms (such as the Trypanosomatidae), the various members exhibit greater variation than species within a single genus. Knowledge of these evolutionary processes demonstrates that there is going to be genetic discrepancies (i.e., the loss of maxicircle genes in *Phytomonas* spp.) which is going to negatively impact on the quality of any sequence alignment. Due to these less conserved sequences and scale of the dataset (greater than 15 kb across 46 species), the alignment-free method is preferred

for the analysis of distant relatives, allowing the analysis of organisms which exhibit greater variation in the number and order of genetic elements.

In accordance with published standards in the assessment of robustness (i.e., bootstrapping), the percentile method validates the accuracy of a node, with a confidence value of >60% in support of the observed clade [48]. The MSA and AF trees showed extremely similar and robust topologies (Figure 5), demonstrating the suitability of the alignment-free FFP method for phylogenetic analyses of the Trypanosomatidae. All major clades (*Leishmania*, *Endotrypanum*, *Porcisia*, *Zelonia*, *Leptomonas*, *Crithidia*, *Herpetomonas*, *Phytomonas*, *Angomonas*, *Blechnomonas*, *Trypanosoma* and *Paratrypanosoma*, cluster with their respective subgenera within the trypanosomatid family with high bootstrap confidence. Analysing a large dataset from an entire family of parasites, a greater divergence is expected than a typical phylogenetic analysis of a few, closely related genera, which can impact on the accuracy of the sequence alignment. In addition, the overall running times of alignment-free phylogenetic inference were shorter (run time less than one day compared to >3 days for alignment and PhyML analysis), making it an attractive approach for large datasets >15 kb. However, important issues need to be taken into consideration for AF-approaches, namely that optimal k-mer lengths must be determined to establish accurate phylogenomic relationships between related organisms. Nevertheless, having the advantage of not requiring pre-alignment of a large collection of maxicircle genome sequences, FFP analysis allowed the rapid, accurate and robust phylogenetic analysis of a large group of diverse species within the Trypanosomatidae.

Kinetoplastid parasites of the trypanosomatid family possess a unique, single, densely-packed periflagellar network of DNA circles known as the kinetoplast [12,49]. In our analysis, both paired-end short reads and a hybrid protocol using short- and long-read data resulted in the high-quality assembly of the maxicircle genome (Table 1). The maxicircle sequences generated from SGS and hybrid assemblies resulted in 100% identity between the coding region and >99% identity between the divergent region (due to the difference in sequence length of the divergent region between Illumina and hybrid assemblies) of each respective species. The hybrid protocol offered an alternative approach, capable of overcoming the difficulties present in assembling complex genomic regions from short sequencing reads. Our hybrid assemblies resulted in the resolution of the divergent region (DR), generating highly contiguous, accurate assemblies of the DR, even when these regions contain a large number of near-identical repeats.

Analysis of the repetitive sequences demonstrated that the various trypanosomatid species have a variable number of repeat arrays throughout their divergent region (Figure 4 and Supplementary Materials S4). The non-coding divergent region of the maxicircle remains the most poorly studied region of kinetoplast DNA [16,50,51]. From our analysis, we can see that the overall DR structure is composed almost entirely of various repeat arrays, with a clear distinction between the coding-region of the maxicircle kDNA (Figures 3 and 4). The high repeat content of the divergent region posed a substantial obstacle to the assembly and output length of the maxicircle kDNA with Illumina short-reads alone. Compared to that of the hybrid approach, SGS reads of *L. aethiopica*, *L. amazonensis*, *L. braziliensis*, *L. guyanensis*, *L. infantum*, *L. mexicana*, *L. tropica* and *T. brucei rhodesiense* assembled only 20%, 11%, 20%, 43%, 49%, 46%, 96% and 88% of the divergent region produced from TGS respectively.

Protozoan mitochondrial genomes often display higher AT% than mitochondrial genomes from metazoans [12,52]. Thus, it is expected that in the mitochondrial homologue in trypanosomatids; the kinetoplast will have a low GC%. The extreme overabundance of AT% in the maxicircle genome of trypanosomatids is primarily due to the repetitive portion of the non-coding divergent region (Figure 2). The AT skew demonstrates the bias towards an AT-rich maxicircle genome of trypanosomatid species.

The vast majority of maxicircles in the trypanosomatid family contain the same set of genes and genomic organisation pattern. The notable exception to this is the plant parasite *Phytomonas françai*, which lacks the respiratory chain complexes III and IV, including the three subunits of the cytochrome *c* oxidase (COI, COII and COIII) and single cytochrome *b* (CYb), respectively (Figure 1). The respiratory chain complexes III and IV are required to maintain a complete electron transport

chain in trypanosomatid parasites, playing a critical role in the biochemical production and synthesis of adenosine triphosphate (ATP) [53,54]. It is speculated that the absence of these subunits is related to the adaptation of *Phytomonas* spp. to the carbohydrate-rich medium of the host plant environment [55]. In the presence of this carbohydrate-rich medium, the cytochrome-mediated respiration complexes of *Phytomonas* were lost after their function became nonessential and the presence of the glycolysis metabolic pathway in glycosomes alone was sufficient for ATP production [55,56]. The feeding behaviour of the insect host is also speculated to have played a role in the loss of cytochromes in *Phytomonas* [57]. Phytophagous hemipteran insects feed exclusively on the sap and juices of plants that are rich in carbohydrates. This suggests a passive, mechanical transmission from plant to insect host that does not require a metabolic shift from a carbohydrate to an amino acid metabolism [58]. Previous studies have also demonstrated the absence of these respective subunits in *Phytomonas serpens* and *Phytomonas* sp. HART1, demonstrating this respiratory deficiency is characteristic of these dixenous plant trypanosomatids [55,57,58].

A second notable exception is the mammalian infecting *Trypanosoma brucei evansi* and *Trypanosoma brucei equiperdum* [59]. Originally considered separate species due to their differences in host-range, transmission and pathogenicity, recent data demonstrates that *T. b. evansi* and *T. b. equiperdum* represent dyskinetoplastic subspecies of *T. brucei*, characterized by the complete or partial loss of their maxicircle kDNA, although their taxonomic status is often debated [60–62]. Recent studies have shown that mutations located in the nuclear-bound ATPase subunit γ of some *T. b. evansi* and *T. b. equiperdum* were found to compensate for the total or partial absence of maxicircle kDNA, demonstrating an important nuclear/kinetoplast interaction for the viability of these species [61].

The relationships between the monoxenous trypanosomatids and their dixenous relatives have been extensively debated over the last few decades [63,64]. It was speculated that dixenous parasitism of the Trypanosomatidae has independently evolved several times over the course of history from the monoxenous ancestors, giving rise to the *Trypanosoma*, *Leishmania*, *Endotrypanum*, *Porcisia* and *Phytomonas* lineages [63]. As such, it is impossible to answer questions surrounding the origins of the trypanosomatid family without studying the non-pathogenic, monoxenous relatives, which is reflected by the rising number of papers published in this field [46,65–69]. From the molecular data, several major clades can be identified including; *Leishmania*, *Endotrypanum*, *Porcisia*, *Zelonia*, *Leptomonas*, *Crithidia*, *Herpetomonas*, *Angomonas*, *Blechnomonas*, *Trypanosoma* and *Paratrypanosoma*. All clades generated support a recent proposal on the classification of the trypanosomatids, specifically the taxonomic validity of the *Endotrypanum*/*Porcisia* genera and the establishment of the subfamily Trypanosomatinae to encompass species of the trypanosomes [5,70]. Molecular clock analyses suggest that at least three lineages independently acquired the ability to infect two hosts (i.e., dixenous parasitism) including plants (*Phytomonas* spp.) and vertebrates (*Trypanosoma* and *Leishmania*/*Porcisia*/*Endotrypanum*) approximately 133, 190 and 95 MYA respectively. Thus, our analyses provide strong support for the multiple and independent origins of the dixenous life-style of the trypanosomatids.

The supercontinents origin of the dixenous Leishmaniinae, where the dixenous members first emerged from monoxenous ancestors during the continental separation of Gondwana is now widely accepted [46,47,63]. As suggested by the supercontinents hypothesis of dixenous parasitism, the earliest dixenous members of the Leishmaniinae first emerged in the late Cretaceous period between 77–140 MYA, during the predicted breakup of Gondwana [19]. Based on our phylogenetic analysis, the dixenous genera *Leishmania*, *Endotrypanum* and *Porcisia* emerged as distinct monophyletic lineages from a common monoxenous ancestor approximately 95 million years ago (Figure 6). Detailed extensively in our previous work [19], in summary the first emergence of dixenous parasitism within the Leishmaniinae subfamily coincides with when the radiation of mammals first began during the Cretaceous period [71].

The phylogenetic analysis of the trypanosomatid family supported the monophyletic lineage of the trypanosomes, having evolved from a monoxenous ancestor with high statistical support (Figure 5). Trypanosomes found in mammals (and humans) are grouped into two sections: Stercoraria (including members of the subgenera *Herpetosoma* and *Schizotrypanum*), which develops in the

posterior portion of the insect host digestive tract and *Salivaria* (including members of the subgenera *Duttonella* and *Trypanozoon*), which develops in the anterior region of the insect digestive tract [58]. Previous analyses proposed an early divergence of salivarian trypanosomes, which involved the ancient split of the trypanosomes into one clade containing all salivarian species and the other branch containing all non-salivarian lineages [72–74]. Based on the SSU rRNA, it was proposed that the emergence of the salivarian lineage first appeared approximately 300 MYA, although the authors were cautious about these date estimates due to the use of the SSU rRNA as the taxonomic marker [72]. Consistent with an ancient salivarian divergence, molecular clock analyses based on the maxicircle phylogenies suggest that the salivarian trypanosomes separated from other trypanosomes approximately 150 MYA (Figure 6). It is speculated that early African trypanosomes were most likely gut parasites or commensals of early insects, however the appearance of tsetse flies approximately 35 MYA facilitated the transmission to mammals by these blood-feeding insects [75]. In addition, the use of the maxicircle coding region as a taxonomic marker also provided strong support for the placement of *Trypanosoma vivax* with other African salivarian trypanosomes including *Trypanosoma brucei* [64].

The analysis of the coding region of the maxicircle also sheds light on the evolution of the *T. cruzi* clade, which encompasses most African, American and European trypanosomes from bats and terrestrial mammals [74]. In recent years, the southern super-continent trypanosome hypothesis has come under scrutiny, with some suggesting it can no longer be considered correct that *T. cruzi* first emerged in the New World and *T. brucei* in the Old World following the continental split of Africa and South America 100 MYA [47,76–78]. The new theory gaining the greatest support proposes that the *T. cruzi* clade (*T. cruzi* and *T. rangeli*) evolved more recently than originally thought within a broader clade of bat trypanosomes [73]. Our analysis suggests that the common ancestor of the *T. cruzi* clade first appeared approximately 84 million years ago. This timeframe provides strong additional support for the ‘bat-seeding’ hypothesis, suggesting that the common ancestor of the *T. cruzi* clade is likely to have evolved following the diversification of bats approximately 70–58 million years ago [79,80]. Using a geological time point i.e., the divergence of *L. enriettii* and *L. macropodum* as the calibration date facilitates suggestions on the emergence of the *T. cruzi* clade independently of the diversification of bats. This is an important advancement for the ‘bat-seeding’ hypothesis, previously not feasible in the absence of WGS data for *L. macropodum*, as it eliminates any bias or skew that may be introduced through using the estimated time of divergence of the host species (i.e., a secondary calibration) [80].

5. Conclusions

In conclusion, the use of the entire coding region of the maxicircle kinetoplast DNA in alignment-free analyses provided an exceptionally robust evolutionary insight into the relationships and patterns within the trypanosomatid family. We believe the present study demonstrates the applicability of AF-based approaches such as FFP to produce reliable and robust phylogenomic analyses to establish and summarise the current evolutionary relationships of the Trypanosomatidae. We suggest that future researchers aiming to analyse this diverse and widespread relationship should consider using this described approach to deal with the ever-increasing amount of sequence data that are becoming available. Ultimately, knowledge of these deep-rooted lineages is exceptionally useful for future analysis involving reconstruction of any evolutionary scenario involving these flagellated protozoans.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: List of trypanosomatid species used in this study. Table S2: The overall nucleotide frequency and skew throughout the maxicircle genome of various trypanosomatid species. Figure S3: Self Dottup plot comparative analysis of the entire maxicircle genome (right panel) and divergent region (left panel) of various trypanosomatid species. Figure S4: Analysis of repeated sequences in the maxicircle divergent region. Figure S5: Determination of optimal feature/k-mer length (k) for 46 trypanosomatid species. Table S6: Comparison of distance matrices showing genetic differences between maxicircle kDNA of various trypanosomatid species. Below diagonal:

genetic distance calculated from multiple sequence-alignment method, above diagonal: genetic distance calculated from alignment-free FFP method.

Author Contributions: A.K. conceived, wrote and edited all drafts. J.E. provided major suggestions on content and edits of the manuscript. D.S. provided additional feedback on edits and further suggestions on contents. All authors read and approved the final version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fraga, J.; Fernández-Calienes, A.; Montalvo, A.M.; Maes, I.; Deborggraeve, S.; Büscher, P.; Dujardin, J.C.; Van der Auwera, G. Phylogenetic analysis of the *Trypanosoma* genus based on the heat-shock protein 70 gene. *Infect. Genet. Evol.* **2016**, *43*, 165–172.
2. Kostygov, A.Y.; Yurchenko, V. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol.* **2017**, *64*, doi:10.14411/fp.2017.020.
3. Jackson, A.P. Genome evolution in trypanosomatid parasites. *Parasitology* **2015**, *142*, S40–S56.
4. Kaufer, A.; Ellis, J.; Stark, D.; Barratt, J. The evolution of trypanosomatid taxonomy. *Parasites Vectors* **2017**, *10*, 287.
5. Maslov, D.A.; Opperdoes, F.R.; Kostygov, A.Y.; Hashimi, H.; Lukes, J.; Yurchenko, V. Recent advances in trypanosomatid research: Genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* **2018**, *146*, 1–27.
6. Ghedin, E.; Bringaud, F.; Peterson, J.; Myler, P.; Berriman, M.; Ivins, A.; Andersson, B.; Bontempi, E.; Eisen, J.; Angiuoli, S.; et al. Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* **2004**, *134*, 183–191.
7. Hamilton, P.B.; Stevens, J.R.; Gaunt, M.W.; Gidley, J.; Gibson, W.C. Trypanosomes are monophyletic: Evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int. J. Parasitol.* **2004**, *34*, 1393–1404.
8. Lukeš, J.; Skalický, T.; Týč, J.; Votýpka, J.; Yurchenko, V. Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* **2014**, *195*, 115–122.
9. Deschamps, P.; Lara, E.; Marande, W.; López-García, P.; Ekelund, F.; Moreira, D. Phylogenomic Analysis of Kinetoplastids Supports That Trypanosomatids Arose from within Bodonids. *Mol. Biol. Evol.* **2011**, *28*, 53–58.
10. Hilu, K.W.; Black, C.M.; Oza, D. Impact of Gene Molecular Evolution on Phylogenetic Reconstruction: A Case Study in the Rosids (Superorder Rosanae, Angiosperms). *PLoS ONE* **2014**, *9*, e99725.
11. Lin RH, Lai DH, Zheng LL, Wu J, Lukes J, Hide G et al. Analysis of the mitochondrial maxicircle of *Trypanosoma lewisi*, a neglected human pathogen. *Parasites Vectors* **2015**, *8*, 665.
12. Lukeš, J.; Wheeler, R.; Jirsova, D.; David, V.; Archibald, J.M. Massive mitochondrial DNA content in diplomonid and kinetoplastid protists. *IUBMB Life* **2018**, *70*, 1267–1274.
13. Gerasimov, E.S.; Gasparyan, A.A.; Litus, I.A.; Logacheva, M.D.; Kolesnikov, A.A. Minicircle Kinetoplast Genome of Insect Trypanosomatid *Leptomonas pyrrocoris*. *Biochemistry* **2017**, *82*, 572–578.
14. Telleria, J.; Lafay, B.; Virreira, M.; Barnabe, C.; Tibayrenc, M.; Svoboda, M. *Trypanosoma cruzi*: Sequence analysis of the variable region of kinetoplast minicircles. *Exp. Parasitol.* **2006**, *114*, 279–288.
15. De Vries, B.F.; Mulder, E.; Brakenhoff, J.P.; Sloof, P.; Benne, R. The variable region of the *Trypanosoma brucei* kinetoplast maxicircle: Sequence and transcript analysis of a repetitive and a non-repetitive fragment. *Mol. Biochem. Parasitol.* **1988**, *27*, 71–82.
16. Flegontov, P.N.; Strelkova, M.V.; Kolesnikov, A.A. The *Leishmania major* maxicircle divergent region is variable in different isolates and cell types. *Mol. Biochem. Parasitol.* **2006**, *146*, 173–179.
17. Muhich, M.L.; Simpson, L.; Simpson, A.M. Comparison of maxicircle DNAs of *Leishmania tarentolae* and *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4060–4064.
18. Botero, A.; Kapeller, I.; Cooper, C.; Clode, P.L.; Shlomai, J.; Thompson, R.C.A. The kinetoplast DNA of the Australian trypanosome, *Trypanosoma copemani*, shares features with *Trypanosoma cruzi* and *Trypanosoma lewisi*. *Int. J. Parasitol.* **2018**, *48*, 691–700.
19. Kaufer, A.; Barratt, J.; Stark, D.; Ellis, J. The complete coding region of the maxicircle as a superior phylogenetic marker for exploring evolutionary relationships between members of the *Leishmaniinae*. *Infect. Genet. Evol.* **2019**, *70*, 90–100.

20. Van Vliet, A.H.M.; Kusters, J.G. Use of Alignment-Free Phylogenetics for Rapid Genome Sequence-Based Typing of *Helicobacter pylori* Virulence Markers and Antibiotic Susceptibility. *J. Clin. Microbiol.* **2015**, *53*, 2877–2888.
21. Fu, S.; Wang, A.; Au, K.F. A comparative evaluation of hybrid error correction methods for error-prone long reads. *Genome Biol.* **2019**, *20*, 26.
22. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289.
23. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics (Oxf. Engl.)* **2016**, *32*, 1009–1015.
24. Koren, S.; Schatz, M.C.; Walenz, B.P.; Martin, J.; Howard, J.T.; Ganapathy, G.; Wang, Z.; Rasko, D.A.; McCombie, W.R.; Jarvis, E.D.; et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **2012**, *30*, 693–700.
25. Patane, J.S.L.; Martins, J., Jr.; Setubal, J.C. Phylogenomics. *Methods Mol. Biol.* **2018**, *1704*, 103–187.
26. Amado Cattaneo, R.M.; Diambra, L.; McCarthy, A.N. Phylogenomics of tomato chloroplasts using assembly and alignment-free method. *Mitochondrial DNA Part A* **2018**, *29*, 1128–1138.
27. Zhang, Q.; Jun, S.R.; Leuze, M.; Ussery, D.; Nookaew, I. Viral Phylogenomics Using an Alignment-Free Method: A Three-Step Approach to Determine Optimal Length of k-mer. *Sci. Rep.* **2017**, *7*, 40712.
28. Bernard, G.; Chan, C.X.; Chan, Y.B.; Chua, X.Y.; Cong, Y.; Hogan, J.M.; Maetschke, S.R.; Ragan, M.A. Alignment-free inference of hierarchical and reticulate phylogenomic relationships. *Brief. Bioinform.* **2019**, *20*, 426–435.
29. Bernard, G.; Chan, C.X.; Ragan, M.A. Alignment-free microbial phylogenomics under scenarios of sequence divergence, genome rearrangement and lateral genetic transfer. *Sci. Rep.* **2016**, *6*, 28970.
30. Murray, K.D.; Webers, C.; Ong, C.S.; Borevitz, J.; Warthmann, N. kWIP: The k-mer weighted inner product, a de novo estimator of genetic similarity. *PLoS Comput. Biol.* **2017**, *13*, e1005727.
31. Fan, H.; Ives, A.R.; Surget-Groba, Y. Reconstructing phylogeny from reduced-representation genome sequencing data without assembly or alignment. *Mol. Ecol. Resour.* **2018**, *18*, 1482–1491.
32. Leimeister, C.A.; Schellhorn, J.; Dorrer, S.; Gerth, M.; Bleidorn, C.; Morgenstern, B. Prot-SpaM: Fast alignment-free phylogeny reconstruction based on whole-proteome sequences. *GigaScience* **2019**, *8*, giy148.
33. Zielezinski, A.; Vinga, S.; Almeida, J.; Karlowski, W.M. Alignment-free sequence comparison: Benefits, applications, and tools. *Genome Biol.* **2017**, *18*, 186–186.
34. Song, K.; Ren, J.; Reinert, G.; Deng, M.; Waterman, M.S.; Sun, F. New developments of alignment-free sequence comparison: Measures, statistics and next-generation sequencing. *Brief. Bioinform.* **2014**, *15*, 343–353.
35. Sims, G.E.; Jun, S.R.; Wu, G.A.; Kim, S.H. Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2677–2682.
36. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477.
37. NCBI. *BLAST® Command Line Applications User Manual US: Bethesda (MD)*; National Center for Biotechnology Information: Bethesda, MD, USA, 2008.
38. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649.
39. Carver, T.; Thomson, N.; Bleasby, A.; Berriman, M.; Parkhill, J. DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* **2009**, *25*, 119–120.
40. Shioiri, C.; Takahata, N. Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry. *J. Mol. Evol.* **2001**, *53*, 364–376.
41. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642.
42. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580.
43. Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221–224.
44. Posada, D. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **2008**, *25*, 1253–1256.
45. PHYLIP (Phylogeny Inference Package) Version 3.6. Available online: <http://evolution.genetics.washington.edu/phylip.html> (accessed on 10 September 2019).
46. Barratt, J.; Kaufer, A.; Peters, B.; Craig, D.; Lawrence, A.; Roberts, T.; Lee, R.; McAuliffe, G.; Stark, D.; Ellis, J. Isolation of Novel Trypanosomatid, *Zelonia australiensis* sp. nov. (Kinetoplastida: Trypanosomatidae)

- Provides Support for a Gondwanan Origin of Dixerous Parasitism in the Leishmaniinae. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005215.
47. Harkins, K.M.; Schwartz, R.S.; Cartwright, R.A.; Stone, A.C. Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infect. Genet. Evol.* **2016**, *38*, 101–109.
 48. Felsenstein, J. Confidence-limits on phylogenies—An approach using the bootstrap. *Evolution* **1985**, *39*, 783–791.
 49. Cavalcanti, D.P.; de Souza, W. The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy. *Scanning* **2018**, *2018*, 9603051.
 50. Flegontov, P.N.; Guo, Q.; Ren, L.; Strelkova, M.V.; Kolesnikov, A.A. Conserved repeats in the kinetoplast maxicircle divergent region of *Leishmania* sp. and *Leptomonas seymouri*. *Mol. Genet. Genom.* **2006**, *276*, 322–333.
 51. Westenberger, S.J.; Cerqueira, G.C.; El-Sayed, N.M.; Zingales, B.; Campbell, D.A.; Sturm, N.R. *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genom.* **2006**, *7*, 60.
 52. Millan, C.R.; Acosta-Reyes, F.J.; Lagartera, L.; Ebiloma, G.U.; Lemgruber, L.; Nué Martínez, J.J.; Saperas, N.; Dardonville, C.; De Koning, H.P.; Campos, J.L. Functional and structural analysis of AT-specific minor groove binders that disrupt DNA-protein interactions and cause disintegration of the *Trypanosoma brucei* kinetoplast. *Nucleic Acids Res.* **2017**, *45*, 8378–8391.
 53. Asato, Y.; Oshiro, M.; Myint, C.K.; Yamamoto, Y.I.; Kato, H.; Marco, J.D.; Mimori, T.; Gomez, E.A.; Hashiguchi, Y.; Uezato, H. Phylogenetic analysis of the genus *Leishmania* by cytochrome b gene sequencing. *Exp. Parasitol.* **2009**, *121*, 352–361.
 54. Horvath, A.; Kingan, T.G.; Maslov, D.A. Detection of the mitochondrially encoded cytochrome c oxidase subunit I in the trypanosomatid protozoan *Leishmania tarentolae*. Evidence for translation of unedited mRNA in the kinetoplast. *J. Biol. Chem.* **2000**, *275*, 17160–17165.
 55. Nawathean, P.; Maslov, D.A. The absence of genes for cytochrome c oxidase and reductase subunits in maxicircle kinetoplast DNA of the respiration-deficient plant trypanosomatid *Phytomonas serpens*. *Curr. Genet.* **2000**, *38*, 95–103.
 56. Maslov, D.A.; Nawathean, P.; Scheel, J. Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid *Phytomonas serpens*. *Mol. Biochem. Parasitol.* **1999**, *99*, 207–221.
 57. Jaskowska, E.; Butler, C.; Preston, G.; Kelly, S. *Phytomonas*: Trypanosomatids Adapted to Plant Environments. *PLoS Pathog.* **2015**, *11*, 17.
 58. Porcel, B.M.; Denoëud, F.; Opperdoes, F.; Noel, B.; Madoui, M.A.; Hammarton, T.C.; Field, M.C.; Da Silva, C.; Couloux, A.; Poulain, J.; et al. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet.* **2014**, *10*, e1004007.
 59. Desquesnes, M.; Holzmüller, P.; Lai, D.H.; Dargantes, A.; Lun, Z.R.; Jittaplapong, S. *Trypanosoma evansi* and surra: A review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects. *BioMed Res. Int.* **2013**, *2013*, 194176.
 60. Lai, D.H.; Hashimi, H.; Lun, Z.R.; Ayala, F.J.; Lukeš, J. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T-brucei*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1999–2004.
 61. Carnes, J.; Anupama, A.; Balmer, O.; Jackson, A.; Lewis, M.; Brown, R.; Cestari, I.; Desquesnes, M.; Gendrin, C.; Hertz-Fowler, C.; et al. Genome and Phylogenetic Analyses of *Trypanosoma evansi* Reveal Extensive Similarity to *T. brucei* and Multiple Independent Origins for Dyskinetoplasty. *PLoS Negl. Trop. Dis.* **2015**, *9*, e3404.
 62. Dewar, C.E.; MacGregor, P.; Cooper, S.; Gould, M.K.; Matthews, K.R.; Savill, N.J.; Schnauffer, A. Mitochondrial DNA is critical for longevity and metabolism of transmission stage *Trypanosoma brucei*. *PLoS Pathog.* **2018**, *14*, e1007195.
 63. Lukes, J.; Butenko, A.; Hashimi, H.; Maslov, D.A.; Votyčka, J.; Yurchenko, V. Trypanosomatids Are Much More than Just Trypanosomes: Clues from the Expanded Family Tree. *Trends Parasitol.* **2018**, *34*, 466–480.
 64. Simpson, A.G.B.; Stevens, J.R.; Lukes, J. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.* **2006**, *22*, 168–174.
 65. Maslov, D.A.; Votyčka, J.; Yurchenko, V.; Lukeš, J. Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends Parasitol.* **2013**, *29*, 43–52.
 66. Podlipaev, S. The more insect trypanosomatids under study—the more diverse Trypanosomatidae appears. *Int. J. Parasitol.* **2001**, *31*, 648–652.

67. Kostygov, A.Y.; Dobáková, E.; Grybchuk-Ieremenko, A.; Váhala, D.; Maslov, D.A.; Votýpka, J.; Lukeš, J.; Yurchenko, V. Trypanosomatid-Bacterium Association: Evolution of Endosymbiosis in Action. *MBio* **2016**, *7*, e01985-15.
68. Kraeva, N.; Butenko, A.; Hlaváčová, J.; Kostygov, A.; Myšková, J.; Grybchuk, D.; Leštinová, T.; Votýpka, J.; Volf, P.; Opperdoes, F.; et al. *Leptomonas seymouri*: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with *Leishmania donovani*. *PLoS Pathog.* **2015**, *11*, 23.
69. Podlipaev, S.A. Insect trypanosomatids: The need to know more. *Mem. Inst. Oswaldo Cruz* **2000**, *95*, 517–522.
70. Espinosa, O.A.; Serrano, M.G.; Camargo, E.P.; Teixeira, M.M.G.; Shaw, J.J. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. *Parasitology* **2018**, *145*, 430–442.
71. Cox, C.B. Plate tectonics, seaways and climate in the historical biogeography of mammals. *Mem. Inst. Oswaldo Cruz* **2000**, *95*, 509–516.
72. Haag, J.; O’Hugin, C.; Overath, P. The molecular phylogeny of trypanosomes: Evidence for an early divergence of the Salivaria. *Mol. Biochem. Parasitol.* **1998**, *91*, 37–49.
73. Hamilton, P.B.; Teixeira, M.M.; Stevens, J.R. The evolution of *Trypanosoma cruzi*: The ‘bat seeding’ hypothesis. *Trends Parasitol.* **2012**, *28*, 136–141.
74. Lima, L.; Espinosa-Álvarez, O.; Hamilton, P.B.; Neves, L.; Takata, C.S.; Campaner, M.; Attias, M.; de Souza, W.; Camargo, E.P.; Teixeira, M.M. *Trypanosoma livingstonei*: A new species from African bats supports the bat seeding hypothesis for the *Trypanosoma cruzi* clade. *Parasites Vectors* **2013**, *6*, 221.
75. Steverding, D. The history of African trypanosomiasis. *Parasites Vectors* **2008**, *1*, 3–3.
76. Lukeš, J.; Mauricio, I.L.; Schönián, G.; Dujardin, J.C.; Soteriadou, K.; Dedet, J.P.; Kuhls, K.; Tintaya, K.W.; Jirků, M.; Chocholová, E.; et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9375–9380.
77. Stevens, J.R.; Noyes, H.; Dover, G.A.; Gibson, W.C. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T-cruzi*. *Parasitology* **1999**, *118*, 107–116.
78. Stevens, J.R.; Noyes, H.A.; Schofield, C.J.; Gibson, W. The molecular evolution of Trypanosomatidae. In *Advances in Parasitology*; Elsevier: Amsterdam, The Netherlands, 2001; Volume, 48, pp. 1–56.
79. Espinosa-Álvarez, O.; Ortiz, P.A.; Lima, L.; Costa-Martins, A.G.; Serrano, M.G.; Herder, S.; Buck, G.A.; Camargo, E.P.; Hamilton, P.B.; Stevens, J.R.; et al. *Trypanosoma rangeli* is phylogenetically closer to Old World trypanosomes than to *Trypanosoma cruzi*. *Int. J. Parasitol.* **2018**, *48*, 569–584.
80. Sauquet, H.; Ho, S.Y.; Gandolfo, M.A.; Jordan, G.J.; Wilf, P.; Cantrill, D.J.; Bayly, M.J.; Bromham, L.; Brown, G.K.; Carpenter, R.J.; et al. Testing the Impact of Calibration on Molecular Divergence Times Using a Fossil-Rich Group: The Case of Nothofagus (Fagales). *Syst. Biol.* **2012**, *61*, 289–313.

