

## Article

# Genomes of *Endotrypanum monterogei* from Panama and *Zelonia costaricensis* from Brazil: Expansion of Multigene Families in Leishmaniinae Parasites That Are Close Relatives of *Leishmania* spp.

Percy O. Tullume-Vergara <sup>1</sup>, Kelly Y. O. Caicedo <sup>1</sup>, Jose F. C. Tantalean <sup>1</sup>, Myrna G. Serrano <sup>2</sup>, Gregory A. Buck <sup>2</sup>, Marta M. G. Teixeira <sup>1</sup>, Jeffrey J. Shaw <sup>1</sup> and Joao M. P. Alves <sup>1,\*</sup>

<sup>1</sup> Department of Parasitology, Institute for Biomedical Sciences, University of Sao Paulo, Av. Prof. Lineu Prestes, 1374, Sao Paulo 05508-000, SP, Brazil; petullume@icb.usp.br (P.O.T.-V.); kellyolivos@usp.br (K.Y.O.C.); jcalderonta@ime.usp.br (J.F.C.T.); mmgteix@icb.usp.br (M.M.G.T.); jayusp@hotmail.com (J.J.S.)

<sup>2</sup> Department of Microbiology and Immunology, Virginia Commonwealth University School of Medicine, 1101 E Marshall St., Richmond, VA 23298, USA; myrna.serrano@vcuhealth.org (M.G.S.); gregory.buck@vcuhealth.org (G.A.B.)

\* Correspondence: jotaj@usp.br

**Abstract:** The Leishmaniinae subfamily of the Trypanosomatidae contains both genus *Zelonia* (monoxenous) and *Endotrypanum* (dixenous). They are amongst the nearest known relatives of *Leishmania*, which comprises many human pathogens widespread in the developing world. These closely related lineages are models for the genomic biology of monoxenous and dixenous parasites. Herein, we used comparative genomics to identify the orthologous groups (OGs) shared among 26 Leishmaniinae species to investigate gene family expansion/contraction and applied two phylogenomic approaches to confirm relationships within the subfamily. The *Endotrypanum monterogei* and *Zelonia costaricensis* genomes were assembled, with sizes of 29.9 Mb and 38.0 Mb and 9.711 and 12.201 predicted protein-coding genes, respectively. The genome of *E. monterogei* displayed a higher number of multicopy cell surface protein families, including glycoprotein 63 and glycoprotein 46, compared to *Leishmania* spp. The genome of *Z. costaricensis* presents expansions of BT1 and amino acid transporters and proteins containing leucine-rich repeat domains, as well as a loss of ABC-type transporters. In total, 415 and 85 lineage-specific OGs were identified in *Z. costaricensis* and *E. monterogei*. The evolutionary relationships within the subfamily were confirmed using the supermatrix (3384 protein-coding genes) and supertree methods. Overall, this study showed new expansions of multigene families in monoxenous and dixenous parasites of the subfamily Leishmaniinae.

**Keywords:** *Endotrypanum monterogei*; *Zelonia costaricensis*; sialidase; catalase; glycoprotein 63; phylogenomics; gene family expansion



**Citation:** Tullume-Vergara, P.O.; Caicedo, K.Y.O.; Tantalean, J.F.C.; Serrano, M.G.; Buck, G.A.; Teixeira, M.M.G.; Shaw, J.J.; Alves, J.M.P. Genomes of *Endotrypanum monterogei* from Panama and *Zelonia costaricensis* from Brazil: Expansion of Multigene Families in Leishmaniinae Parasites That Are Close Relatives of *Leishmania* spp.. *Pathogens* **2023**, *12*, 1409. <https://doi.org/10.3390/pathogens12121409>

Academic Editors: Nicola Carter and Hans-Peter Fuehrer

Received: 23 October 2023

Revised: 10 November 2023

Accepted: 24 November 2023

Published: 30 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Trypanosomatids are a well-known and widely distributed group of obligatory flagellate parasites that infect invertebrates, vertebrates, and plants. During their long evolutionary history, these eukaryotic parasites adopted two kinds of life cycles. While most are monoxenous, infecting only one invertebrate host, a few groups are dixenous, being found in vertebrates/plants and invertebrates that are their vectors [1,2].

Presently, the family Trypanosomatidae is divided into 24 genera [3]. Recently, there have been substantial increases in the identification of new taxa, as well as changes in the taxonomy and nomenclature of the subfamily Leishmaniinae, leading to the creation of new genera and subgenera [4]. The dixenous genera are *Leishmania*, *Porcisia*, and *Endotrypanum*, and the predominantly monoxenous ones are *Zelonia*, *Crithidia*, *Leptomonas*, *Novymonas*, *Lotmaria*, and *Borovskya*, which are parasites of Heteroptera, Diptera, and

Hymenoptera [4–6]. An earlier study had split the genus *Leishmania* into the following two informal sections: Euleishmania and Paraleishmania [7]. According to a recent classification based on molecular phylogeny, the latter section comprises the species *Endotrypanum* and its closest relative, *Porcisia*. The genus *Endotrypanum* currently harbors five species *Endotrypanum schaudinni*, *Endotrypanum monterogeei*, and the species *Endotrypanum colombiense*, *Endotrypanum equatorensis*, and *Endotrypanum herreri*, which were formerly classified as *Leishmania* [4].

According to the available data and the new classification of Leishmaniinae, human infection can be caused by *Leishmania* spp. and by *E. colombiense*, which has been reported to cause both cutaneous and visceral leishmaniasis [7–9]. In addition to the increasing number of dioxenous Leishmaniinae species infecting humans, molecular diagnosis has revealed concomitant infections of *Leishmania* species with *Leptomonas seymouri* and *Crithidia* spp. [10,11]. Evidence is accumulating that indicates that these monoxenous infections in man are related to a diminished immunity associated with other pathogenic parasites or an immunological deficiency. Therefore, the classification of trypanosomatids as monoxenous or heteroxenous must be viewed with caution.

The distinctive feature of *E. schaudinni* and *E. monterogeei* is that they are the only known intraerythrocytic trypanosomatids of mammals; however, *Leishmania* spp. of the subgenus *Sauroleishmania* are also reported within lizard erythrocytes [12,13]. The xenarthrans *Choloepus didactylus* and *Choloepus hoffmanni* are the natural vertebrate hosts, and Phlebotominae sand flies are the vectors of *Endotrypanum* spp. [14–16].

Today, genomic studies with close relatives of the *Leishmania* emphasize the necessity of comparative studies since they share the same habitats in their invertebrate vectors as *Endotrypanum* [17–19]. The genus *Zelonia* was created by Shaw, Camargo, and Teixeira in 2016 to accommodate the species named *Leptomonas costaricensis* [4,20]. *Z. costaricensis* (IRIC/CR/2003/15EC) was isolated from the predator hemipteran *Ricolla simillina* (Heteroptera, Reduviidae) in Costa Rica in 2003, whereas our isolate of *Z. costaricensis* (IZEL/BR/89/169E) was isolated from the predator hemipteran *Zellus* sp. in Brazilian Amazonia [4,20]. At first, this species was based on the promastigote morphology and named *Leptomonas costaricensis*. However, in all analyses using small subunit and large subunit rRNAs and heat shock protein genes, this species was separated from *Leptomonas pyrrocoris* and *Leptomonas seymouri* [20,21]. A previous phylogenetic study clustered *Z. costaricensis* with *Z. australiensis*, isolated in Australia from the black fly *Simulium (Morops) dycei* [21]. The phylogeny of the subfamily Leishmaniinae supports *Z. costaricensis* as a lineage basal to the clade harboring *Endotrypanum*, *Porcisia*, and *Leishmania* [4,21].

Cell surface proteins of trypanosomatid parasites have become important targets for comparative genomic studies. These protein families are considered to be of a cryptic nature due to their unknown origins in trypanosomatids [22–24]. The largest cell surface gene families have been identified in the genera *Trypanosoma*, *Leishmania*, *Crithidia*, and *Leptomonas* [25–28], including glycoprotein 63 (gp63), promastigote surface antigen (PSA, also known as gp46), amastin, tuzin, HASP (hydrophilic acylated surface protein), and SHERP (small hydrophilic ER-associated protein) [29–32].

The genome comparison of *Endotrypanum* and its closest known relative *Porcisia* (parasite of porcupines) revealed shared metabolic capacities with *Leishmania major* and a substantially reduced repertoire of surface proteins in both *Endotrypanum* and *Porcisia* compared to *Leishmania*, such as cell surface amastins required from the development of *Leishmania* amastigotes inside macrophages [33].

Comparative genomics approaches have been used to gain a better understanding of the genetic content in several trypanosomatids [22,24,34]. The focus of the present study is a genomic characterization of *E. monterogeei* from Panama (formerly named *E. schaudinni*) and *Z. costaricensis* from Brazil based on their whole-genome sequences and a comparative analysis with their close relatives, in particular those of the subfamily Leishmaniinae. This study also aims to confirm the phylogenetic position of *Z. costaricensis* and *E. monterogeei*

using robust phylogenomics methods (supermatrix and supertree), employing 3384 single-copy proteins from 27 trypanosomatid lineages.

## 2. Materials and Methods

### 2.1. Sample Origin, Genome Sequencing, and Assembly

High-quality genomic DNA was extracted, using the phenol–chloroform method, from *E. monterogeii* TCC224 (MCHO/PA/62/M907), which was originally isolated from sloth (*Choloepus hoffmanni*) in Colon Province, Panama, and from *Z. costaricensis* TCC169E (IZEL/BR/89/169E), which was isolated from a bug (*Zelus* sp.) in Amazonas, Brazil. These isolates are deposited in the Trypanosomatid Culture Collection of the University of Sao Paulo (TCC-USP).

Whole-genome sequencing was performed using standard pyrosequencing shotgun methodology on the Roche 454 platform, as recommended by the manufacturer. We estimated the quality of the sequences using FastQC tool v. 0.11.8 [35].

Genome assembly was performed using Roche’s Newbler software v. 2.3 (software distributed by the manufacturer). The pipeline Taxoblast v. 1.2 was used to identify and remove contigs from contaminants [36]. We calculated the N50 value using mfsizes v. 1.8.7 [37]. Additionally, to assess the completeness of the assemblies, Benchmarking Universal Single-Copy Ortholog (BUSCO) v. 5.2.2 [38] was run using the dataset Euglenozoa version 10 (n130 orthologous genes). Dot plot comparisons were run using the MUMmer tool v. 4.0 [39] to assess contig orientation and synteny.

### 2.2. Gene Models, Repetitive Elements, and RNAs Prediction

To predict protein-coding genes, we employed GeneMarkS software v. 4.28 [40], using self-training mode, considering ORFs of at least 200 bp, and using the less intron parameter. In addition, we also submitted the genomes to AUGUSTUS web server v. 2 [41], using gene models from *L. tarentolae*. Prediction and annotation of non-coding genes was performed with tRNAscan-SE v. 1.4 [42] for tRNA annotation; RNAmmer v. 1.2.1 [43] to identify 5S, SSU, and LSU rRNAs sequences; and the INFERNAL package v. 1.1 [44] against Rfam database v. 12.0 [45] to identify snoRNA, and microRNA genes. Additionally, some gene models of interest were manually inspected to generate a final annotation dataset.

For prediction of repetitive elements, RepeatModeler2 [46] pipeline was executed to construct a de novo repeat library for the *Z. costaricensis* and *E. monterogeii* genomes. The final library of each species was annotated with RepeatMasker software v. 4.1.2-p1 [47].

### 2.3. Functional Annotation of Genes

We used a combination of public databases to assign functions to predicted gene models (proteins) in the *E. monterogeii* and *Z. costaricensis* genomes. We searched protein sequences against the non-redundant database from NCBI (downloaded 6 September 2021) [48], where we used BLASTP v. 2.9.0+, and BLASTN v. 2.9.0+ [49] with default parameters (except for the E-value cutoff of  $1 \times 10^{-6}$ ). We also submitted predicted proteins to EggNOG v. 5.0 database’s eggNOG-mapper v. 2.0 tool [50]. The proteomes were characterized functionally according to the eukaryotic orthologous groups (KOG) classification scheme [51]. As a complementary analysis, we used InterProScan v. 5.30-69.0 [52] in standalone mode with default parameters. Pfam [53], Gene Ontology, and InterPro domains database were employed to assign putative functions to the proteomes. The proteomes were also submitted to the KAAS website tool v. 2.1 [54] in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

### 2.4. Orthologous Group Analysis

Orthologous groups (OGs) were inferred using the OrthoVenn2 online tool [55] with default settings. Three species that have never been found infecting humans (*Porcisia hertigi* TCC260, *E. monterogeii* LV88, and *Z. costaricensis*) and one that is pathogenic to humans (*L. major* Friedlin) were selected for comparison of their protein data. *L. (L.) major* and

*P. hertigi* were downloaded from TriTrypDB version 55 and 60 [56]. Also, specific-OGs for *Z. costaricensis* and *E. monterogei* were selected for further analysis.

### 2.5. Datasets for Phylogenomic Analysis

Besides the two proteomes from this study, we downloaded protein-coding gene sequences from 25 taxa from TriTrypDB version 55 and the Genbank database (accessed on 1 October 2022) [56] (Table S1). The criterion for including a species was its membership in the subfamily Leishmaniinae, with the exception of *Blechnomonas ayalai*, used as an outgroup. Single-copy orthologous groups as defined by Orthofinder2 v. 2.5.4 [57] were selected for the phylogenomic analysis. To evaluate the impact of missing data [58], we built two different datasets. Dataset 1 included all 3384 single-copy genes present in all 27 taxa, as defined above. Dataset 2 was built by selecting the 99 single-copy genes used by BUSCO, from all 27 species. Datasets were aligned by MUSCLE v. 3.8.31, [59], with subsequent removal of ambiguous aligned positions by Trimal v. 1.4.rev15, [60], using automated\_1 parameters, the optimal method for trimming in phylogenetic analysis. The resulting filtered alignments from each dataset were concatenated into a corresponding supermatrix with FASconCAT-G v. 1.0 [61].

### 2.6. Phylogenomic Analysis of the Subfamily Leishmaniinae

Supermatrix and supertree approaches were used to infer phylogenetic relationships within the subfamily Leishmaniinae. The maximum-likelihood (ML) method was used for phylogenomic analyses based on the supermatrix (concatenated proteins) approach. Dataset 1 was analyzed with IQ-TREE2 v. 2.0.4 [62] using partitions for the individual proteins. The best substitution models were selected automatically by ModelFinder (MFP option) [63]. Branch support was measured with the aLTR algorithm [64]. Dataset 2 was analyzed without partitions, using the JTT model (Jones–Taylor–Thornton) [65] +  $\Gamma$ 4 + F. To evaluate node support 1000 UFBS (Ultra-Fast Bootstrap) and 100 BS (non-parametric bootstrap) pseudoreplicates were run. Values higher than 95% and 75%, respectively, were considered robust.

In addition to the ML method, we applied Bayesian inference (BI) to dataset 2, the smaller of the two. Dataset 1 was not analyzed using BI due to the large size of the supermatrix (3384 OGs), which made runtime impractical. The Bayesian tree was inferred using MrBayes v. 3.2.6 [66] with two independent Monte Carlo Markov chain (MCMC) analyses; each chain started from a random tree and was run for 10,000,000 generations after convergence, with sampling every 1000 generations. Node support was considered robust for posterior probabilities (PP)  $\geq 0.95$ .

The supertree approach was carried out with dataset 1, to verify for inconsistent results at the individual gene level, which might be associated with incomplete lineage sorting (ILS) [67]. For this, we estimated an individual gene tree for each of the 3384 single-copy genes with IQ-TREE2 using the JTT +  $\Gamma$ 4 + F substitution model, and 1000 UFBS replicates to estimate node support. The supertree was inferred with ASTRAL-III v. 5.7.1 [68], and the option  $-t$  1 was chosen to estimate quartet support. Trees were drawn with iTOL v. 5.4 [69] and cosmetic adjustments were performed in the Inkscape vector editor ([www.inkscape.org](http://www.inkscape.org), accessed on 1 October 2022).

### 2.7. Characterization and Phylogenetic Analysis in gp63 Multigene Family

We selected gp63 (leishmanolysin), one of the gene families that have shown expansion, in *E. monterogei* and 26 other genomes (Table S1). The selected gp63 proteins were used to find the Orthofinder orthologous groups of interest, which were then aligned with MUSCLE with default settings. The gp63 phylogenetic tree was inferred using IQ-Tree2. The best tree was chosen using the JTT +  $\Gamma$ 4 model, and branch support was estimated by 100 bootstrap pseudoreplicates. Nodes were considered robust when displaying BS  $\geq 0.75\%$ .



### 2.8. Metabolic Pathway Prediction

We searched for potential biochemical pathways by deducing enzyme commission numbers and conducting comparative metabolic annotation analysis of four trypanosomatids (*L. (L.) major*, *E. monterogeei* LV88, *E. monterogeei* M907, and *Z. costaricensis*) using Asgard 1.6.5 [70], which uses Kyoto Encyclopedia of Genes and Genomes (KEGG) [71] and Uniref100 [72] databases as references. We then performed manual curation of the results of interest by thoroughly reading the relevant scientific literature [73,74]. To document significant enzymes in detail, we have used BRENDA [75]. Finally, the metabolic pathways of interest were then redrawn using Inkscape.

## 3. Results

### 3.1. Assembly and Genome Characteristics

The final genome assembly size of *E. monterogeei* M907 was 29.66 Mbp, split into 10,088 contigs, presenting the largest contig with 41,500 bp and an N50 contig length of 6.0 kb. *Z. costaricensis* TCC169E had a genome assembly of 39.80 Mbp, with 7896 contigs, N50 of 17.4 kb and largest contig with 126,447 bp. The mean GC content was 52.66% and 64.26%, respectively. Assembly statistics are summarized and compared in Table 1. Additionally, dot plots were generated and displayed a high degree of synteny with other species of trypanosomatids, which indicates that, overall, the *E. monterogeei* and *Z. costaricensis* genomes are probably assembled correctly (Figure S1A,B).

**Table 1.** Assembly statistics and completeness assessment for the *Endotrypanum monterogeei* M907 and *Zelonia costaricensis* TCC169E genomes.

Assembly Features	<i>E. monterogeei</i> M907	<i>Z. costaricensis</i> TCC169E
Number of contigs	10,088	7896
Longest contigs (bp)	41,459	126,447
N50 length (bp)	6007	17,448
L50	1445	614
GC%	52.66	64.26
Total length (Mb)	29.66	38.8
BUSCO v. 5.2.2		
Complete single-copy	125	129
Complete single duplicated	0	0
Fragmented	5	1
Missing	0	0

To assess the completeness of genome assembly and gene model predictions, we ran a BUSCO analysis employing the Euglenozoa dataset from OrthoDB v.10 (130 BUSCO orthologs). The *E. monterogeei* M907 genome presented slightly more fragmentation (8 fragmented orthologs) than *Z. costaricensis* (2 fragmented orthologs). The genomes of high quality, such as *L. major* Friedlin, *L. pyrrocoris* and *L. (V.) braziliensis*, all had no fragmented orthologs (Figure S2). To detect the presence of possible contaminants, we ran the taxoblast pipeline, detecting three prokaryote contigs on the draft assembly of *Z. costaricensis*, which were filtered out; the same analysis was performed in *E. monterogeei*, without any evidence of contaminants (bacteria).

### 3.2. Gene Prediction and Annotation

The best gene models were obtained using GeneMarkS. *E. monterogeei* and *Z. costaricensis* had 9711 and 12,201 putative protein-coding genes predicted, respectively, with a mean exon length of 1573 and 1580 bp, respectively. The content of repeat elements (RE) in the

genomes varied, reaching 4.47% in *E. monterogeii* and 13.16% in *Z. costaricensis*. Intriguingly, in this analysis, no transposable elements (TEs) were found in *E. monterogeii*. In *Z. costaricensis*, various repeating elements, including those in multigene families, account for 4.21% of the genes, and unknown repeats are 2.43%. Among Tes, LTR, LINE, and SINE represent 0.28%, 0.10%, and 0.23%, respectively (Table S2). We found non-autonomous Ingi elements within LINE; additionally, LTR was composed of TATE (telomere-associated transposable element) and VIPER (vestigial interposed retroelement) elements.

Functional annotation in both species was carried out using the following multiple public databases: NCBI's nr, KEGG, Pfam, EggNOG, and Interproscan (Table 2). Using BLASTP, 8834 *E. monterogeii* proteins had matches against the nr database (E-value cutoff of  $1 \times 10^{-10}$ ), which accounts for about 91% of the predicted proteins. Among those matches, 26% presented a high similarity to functionally annotated proteins, whereas 60.9% were assigned to hypothetical or unknown proteins. The remaining 13% of the proteome presented no hits against nr and were designated as unclassified or orphan. In addition, of the 8834 proteins, 8358 showed the best BLASTP hits with the genus *Leishmania* spp. and 387 with the genus *Leptomonas* spp.

**Table 2.** Comparative genomics metric for *Endotrypanum monterogeii* and *Zelonia costaricensis*.

Features	<i>E. monterogeii</i>	Percentage	<i>Z. costaricensis</i>	Percentage
Protein-coding genes	9711	NA	12,201	NA
G + C% of CDS	57.12	NA	68.39	NA
Mean exon length (bp)	1580	NA	1573	NA
Nr-DB *	8834	90.90%	11,380	93.30%
KAAS-KEGG *	2088	21.50%	2344	19.20%
Pfam-DB *	7241	74.60%	8975	73.50%
EggNOG-DB *	7738	79.60%	9173	75.20%
Interproscan *	8497	87.50%	11,171	95.50%

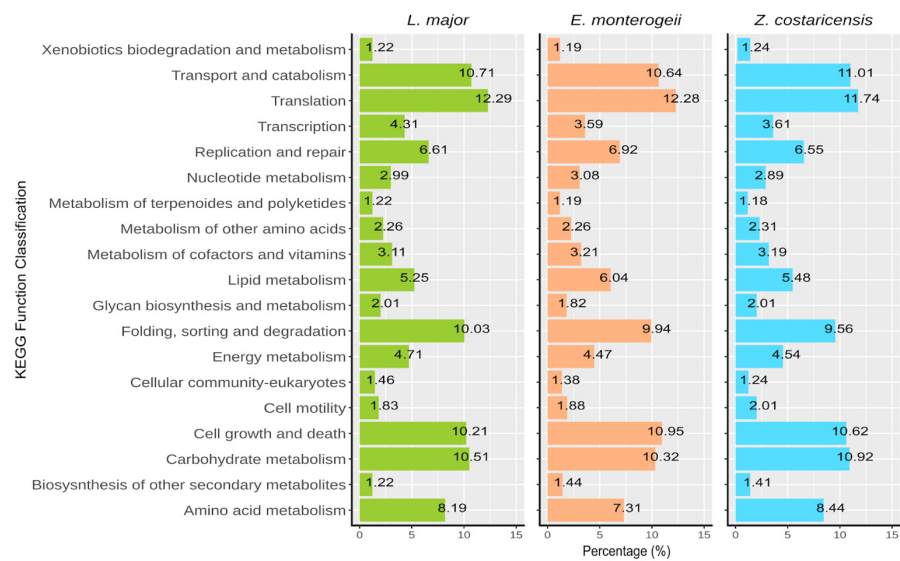
\*: indicates number of proteins and percentage of whole proteome matching the database; NA: not applicable.

KEGG mapping showed 2088 proteins (21.5%) in *E. monterogeii* and 2344 proteins (19.2%) in *Z. costaricensis* with matches against this database (Table 2).

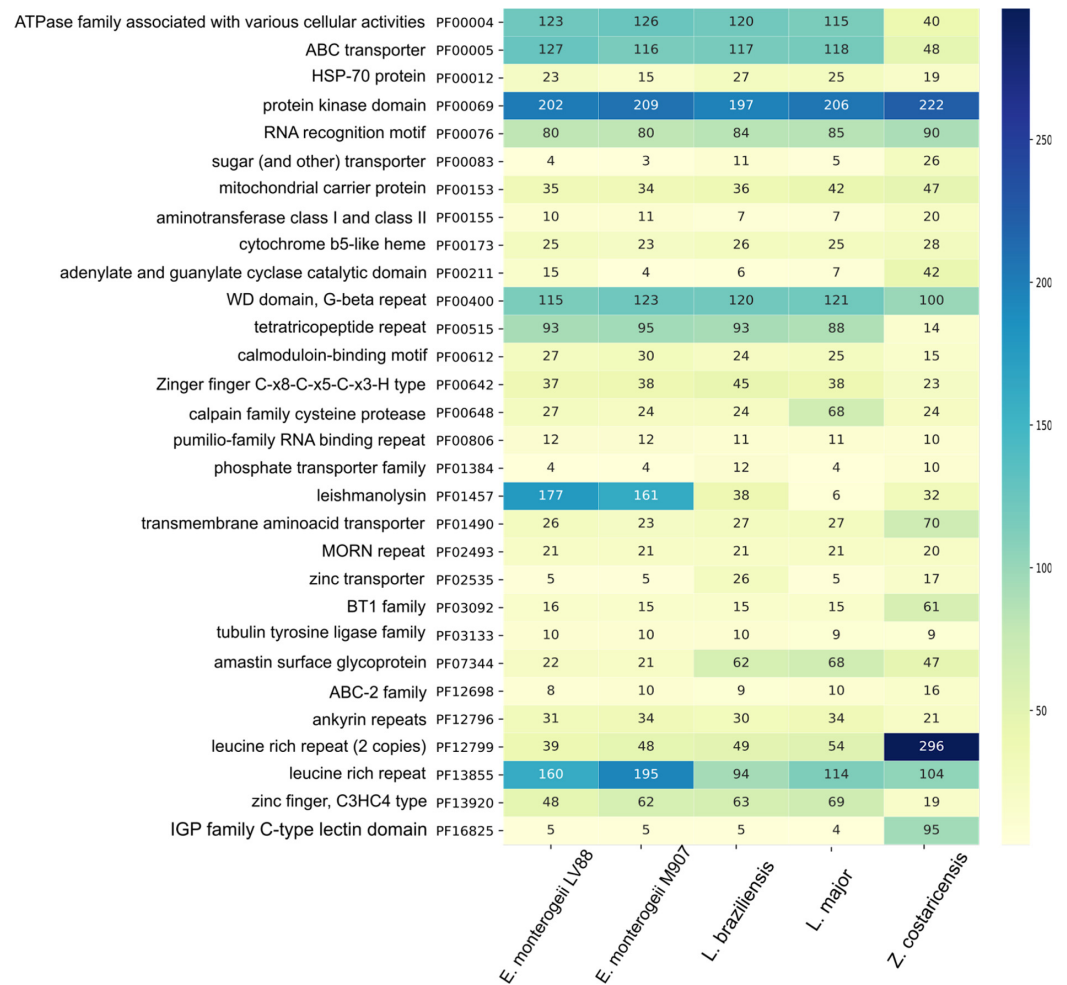
Comparative analysis using three specific categories (cellular processes, genetic information processing, and metabolic pathway) are shown in Figure 1. Most genes were associated with pathways related to transport and catabolism, translation, folding, sorting and degradation, cell growth and death, and carbohydrate metabolism.

Furthermore, KOG assignments were used to assess the complexity of cellular functions in both species (Figure S3). The categories that had the highest numbers of genes were "Post-Transcriptional Modification and Protein Turnover, Chaperones", "Signal Transduction Mechanisms", "Translation, Ribosomal Structure and Biogenesis", and "Carbohydrate Metabolism".

Functional annotation based on Pfam domain families showed identified domains in 7241 (74.6%) and 8975 (73.5%) proteins from *E. monterogeii* and *Z. costaricensis*, respectively (Table 2). Additionally, a comparative analysis of Pfam domains using five lineages was carried out to establish the abundance of the domains among them. From these, we identified eight families that are more abundant in *Z. costaricensis* (Figure 2). These families include leucine-rich repeat (2 copies) (PF12799), IGP family C-type lectin (PF16825), BT1 (PF03092), amino acid transporters (PF01490), and sugar transporters (PF00083). We also noticed a lower abundance of domains (48 genes) associated with ABC-type transporter (PF00005), and tetratricopeptide repeat (PF00515). Conversely, the most abundant families in *E. monterogeii* were protein kinase (PF00012), leishmanolysin (PF01457), and leucine-rich repeat (PF13855). Interestingly, in *Endotrypanum* spp., the amastin multigene family (PF07344) showed low abundance (17 copies).



**Figure 1.** KEGG function classification for enzymatic characterization based in cellular processes, genetic information processing, and metabolic pathways using the *Leishmania (Leishmania) major*, *Endotrypanum monterogeii* and *Zelonia costaricensis* genomes.



**Figure 2.** Abundance heatmap of 30 protein domain families in five Leishmaniinae species. The column on the left lists domain names along with Pfam code. The high number of domains is depicted in dark-blue and the low number of domains in light-green, following the color scale on the right.

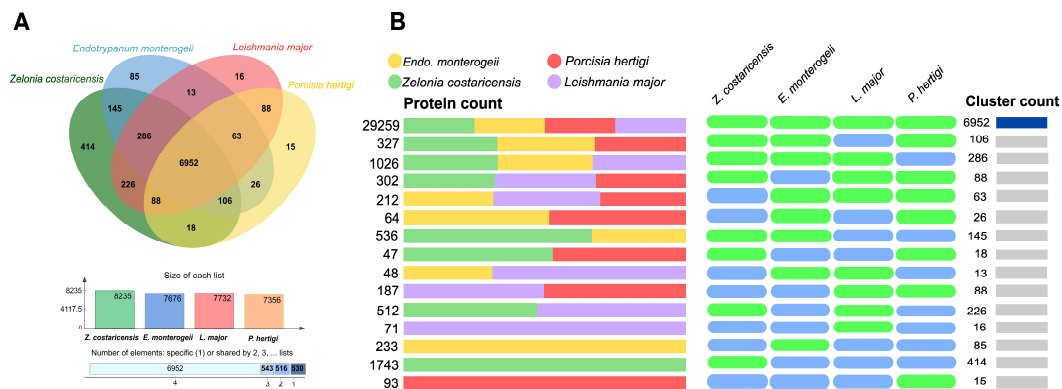
For the non-coding RNAs (Table 3), *Z. costaricensis* displayed 8 rRNAs, 201 tRNAs, 295 microRNAs, and 95 snoRNAs, while *E. monterogeei* had 6 rRNAs, 86 tRNAs, 307 microRNAs, and 213 snoRNAs.

**Table 3.** Non-coding RNA prediction summary.

Types	<i>Z. costaricensis</i>	<i>E. monterogeei</i>	Software
rRNA	8	6	RNAmmer
tRNA	201	295	tRNAscan
microRNA	295	307	Infernal
snoRNA	95	213	Infernal

### 3.3. Ortholog Clustering and Functional Analysis

The annotated protein sequences from *Z. costaricensis* and *E. monterogeei* were also compared with other close relative parasites, such as pathogenic *L. major* and non-pathogenic *P. hertigi*, by orthologous group analysis. Of the 8541 orthologous groups (Ogs), 6511 were single-copy Ogs. There is a core of 6952 Ogs (with a total of 7506 proteins in *Z. costaricensis*, 7131 in *E. monterogeei*, 7339 in *L. major*, and 7265 in *P. hertigi*) that are shared by all four species (Figure 3A,B). Furthermore, the highest number of Ogs shared was that between *Z. costaricensis* and *L. major*, with 7552 orthologs, whereas *E. monterogeei* and *Z. costaricensis* shared 7489 Ogs (Table S3). Another 145 OG were shared between *Z. costaricensis* and *E. monterogeei* but not found in *L. major*. In total, 414 and 85 lineage-specific Ogs were identified in *Z. costaricensis* and *E. monterogeei*, respectively, which is higher than the numbers in *L. major* (16) and *P. hertigi* (15). Lineage-specific ortholog groups in *Z. costaricensis* were involved in the plasma membrane, pathogenesis, intracellular signal transduction, glycerol transport, and cellulose catabolic process. In *E. monterogeei*, they were implicated in cell adhesion, phospholipid translocation, and intracellular signal transduction. In addition, both species showed several Ogs that were hypothetical proteins.



**Figure 3.** Comparative analysis of four species of Leishmaniinae. (A) Venn diagram of shared and unique orthologous group in four lineages. (B) Cluster count and proteins assigned to Ogs for *Leishmania* (*Leishmania*) *major*, *Endotrypanum monterogeei*, *Porcisia hertigi*, and *Zelonia costaricensis*. A core of 6952 orthologous groups was shared by all four lineages.

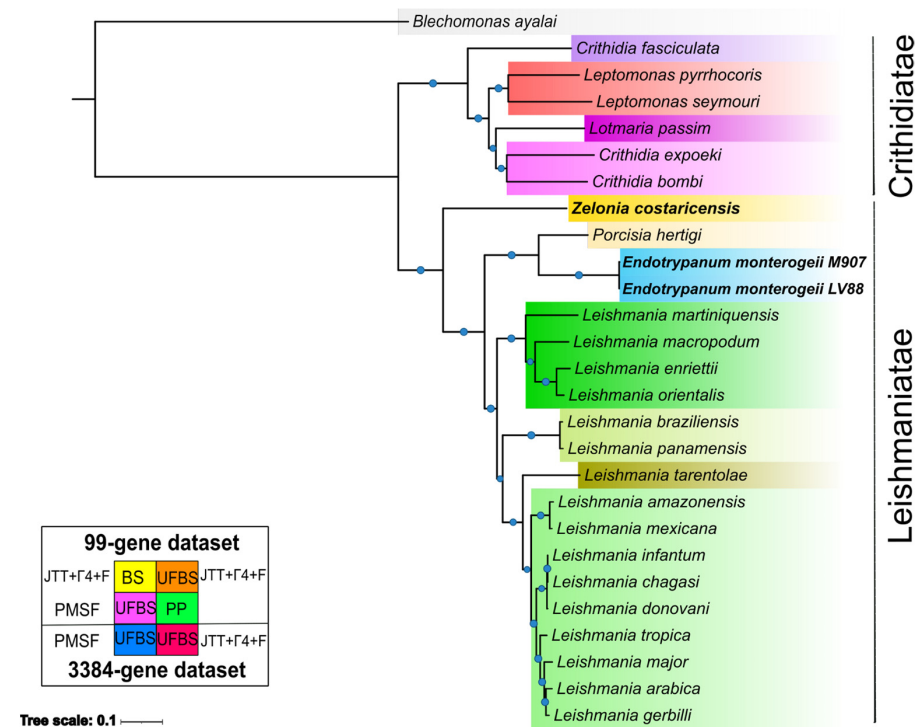
### 3.4. Phylogenomics Analysis into the Subfamily Leishmaniinae

The evolutionary and phylogenetic relationships of the *E. monterogeei* and *Z. costaricensis* species that belong to the subfamily Leishmaniinae were inferred by applying the supertree and supermatrix phylogenomic approaches. Two datasets were created based on single-copy genes. In the largest matrix, named Dataset.1 3384 single-copy genes were concatenated representing 50,950,701 amino acids, 1,198,087 indels, 49,752,083 parsimony-informative sites, and 531 missing data. Conversely, the smallest matrix, Dataset.2, was



obtained from 99 single-copy genes (conservative) from the BUSCO analysis, which represented 2,209,302 amino acids, 38,781 indels, and 2,170,511 parsimony-informative sites.

The phylogenomic tree was generated based on a group of 27 taxa. In brief, we applied several methods, statistical support metrics, and two datasets in order to find variation on the topological structure of species tree (Figures 4 and S4A–E, Table S4). All of the trees recovered showed that the subfamily Leishmaniinae was split in two known clades, the infrafamily Leishmaniatae, which is a group where almost all members are dioxenous, and the infrafamily Crithidiatae, which harbors monoxenous species. As expected, *Z. costaricensis* is placed in a basal branch into the Leishmaniatae, composed of *P. hertigi*, *Endotrypanum* spp., and *Leishmania* spp., with high support (BS: 100, PP: 1).



**Figure 4.** Maximum likelihood (ML) phylogenomic tree based on a concatenated dataset of 3384 single-copy genes. The tree was reconstructed using 27 species, and 5 different models using ML and 1 model using BI were tested. All analyses recovered the same topological structure. UFBS, BS, and PP values higher than 95%, 75%, and 1 are shown as light blue circles. The horizontal bar depicts 0.01 substitutions per site.

The monophyletic *Leishmania* clade was split in its expected subgenera, with *Mundinia* located basally, followed by *Viannia*, then *Sauroleishmania*, and *Leishmania*. The *Endotrypanum* + *Porcisia* clade was recovered as the outgroup for the genus *Leishmania*, and also showed strong support (BS:100, PP: 1). Additionally, *E. monterogeii* M907 and *E. monterogeii* LV88 are placed together and with negligible branch lengths, indicating the very low levels of sequence divergence between the two strains.

To confirm that topological variation among different gene phylogenies was not interfering with the overall phylogenetic signal that led to the supermatrix tree, we also ran a supertree analysis using ASTRAL to infer the separate history of the 3,384 single-copy genes. The phylogenomic tree obtained had the same topology as the supermatrix method, and high support for the nodes (LPP = 1) (Figure S5); moreover, we estimated the quartet score to be 53,851,596 and the normalized quartet score to be 0.919019. This means that 53,851,596 induced quartet trees from the gene trees are in the species tree, and these are 91.90% of all quartet trees that can be found in the species tree.

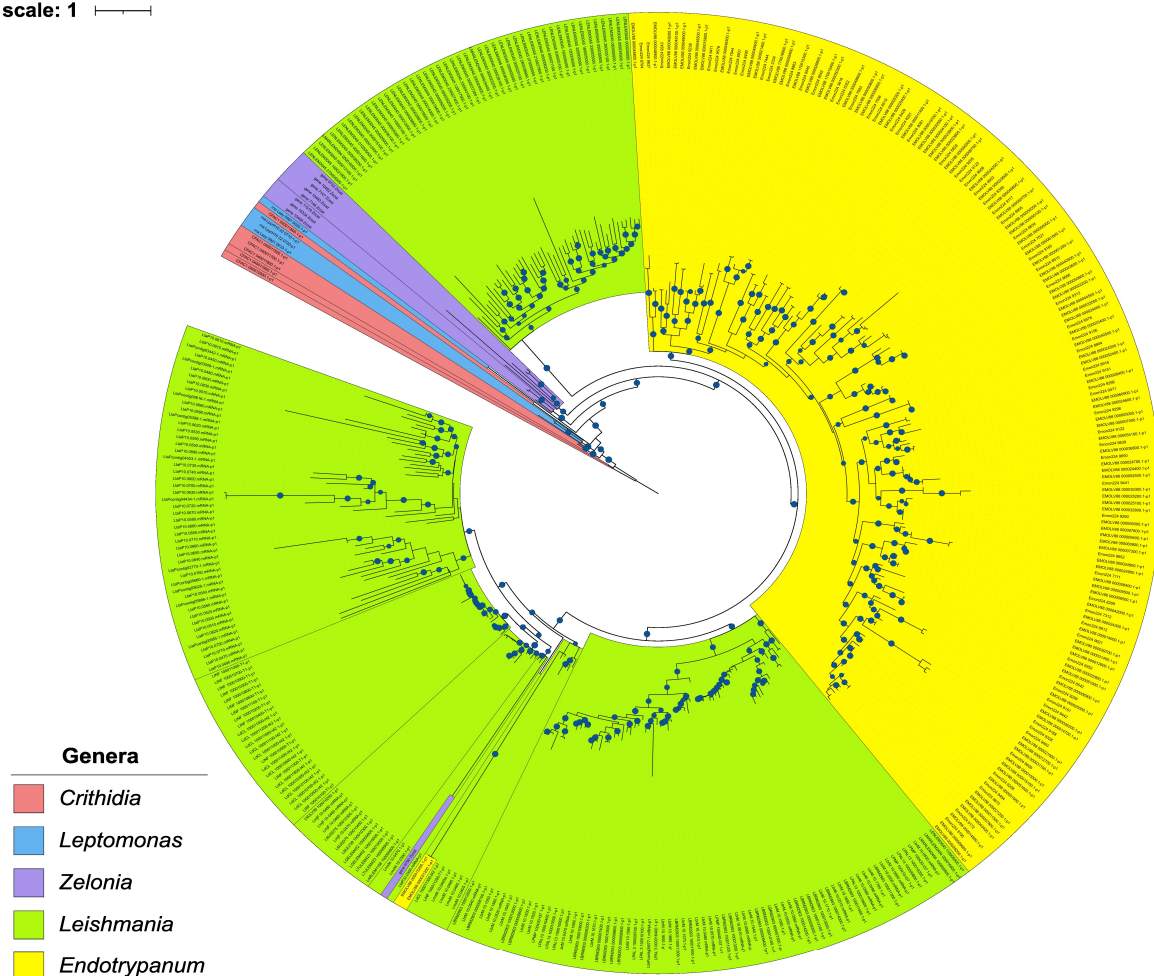
### 3.5. Phylogenetic Analysis of the GP63 Family

GP63 belongs to a multigene family and its functions have been extensively studied in *Leishmania* spp. [29]. For example, gp63 plays a role in the interaction and internalization with the macrophage [76,77]. Previous studies revealed a significant expansion of the gp63 family on chromosome 10 in *L. (V.) braziliensis* and *L. (Sauroleishmania) tarentolae* [78,79].

This expansion also happened in the genus *Endotrypanum* sp. (Table S5). We performed an evolutionary analysis of the gp63 family of *E. monterogeii* and *Z. costaricensis*. The 161 gp63 sequences identified and annotated in *E. monterogeii* were distributed in 28 ortholog groups, the largest being the following: OG00000 (with 69 gp63 genes) and OG00011 (with 52 gp63 genes) (Table S5). The length of the sequences ranged from 99 to 899 amino acids; the most frequent sequence length in *E. monterogeii* ranged from 100 to 199 aa (Figure S5). We also identified conserved motifs within gp63 sequences (Figure S6). For clarity, only the two most numerous groups (406 and 142 protein sequences) were considered when building these trees (Table S7).

The gp63 phylogenetic tree (Figure 5) mostly reflected species phylogeny. The basal clades belong to the genera *Crithidia* and *Leptomonas*, as expected. Most *Zelonia* gp63 sequences were placed external to the *Leishmania* and *Endotrypanum* genera. Even though *Endotrypanum* is a genus that diverged early in comparison with *Leishmania*, an interesting set of *Leishmania* spp. Gp63 sequences have possibly diverged in a former event.

Tree scale: 1 

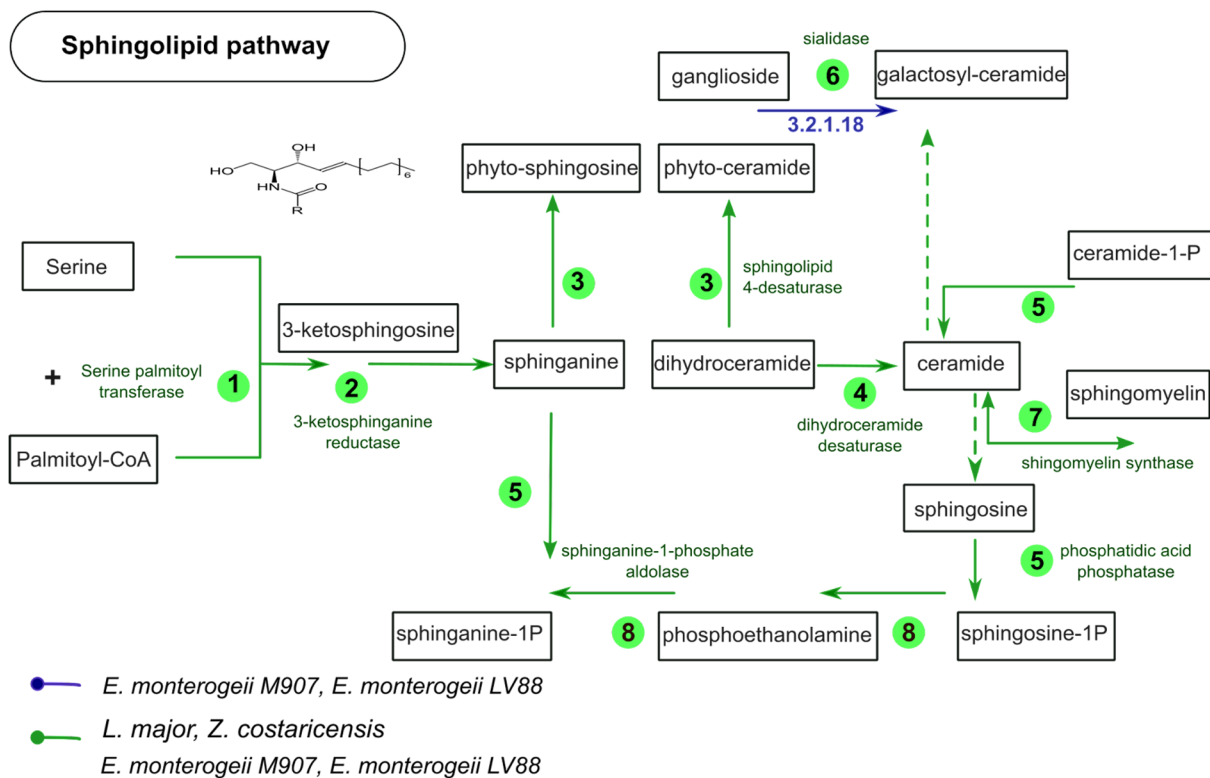


**Figure 5.** Phylogenetic relationships of glycoprotein 63 in 21 species using the ML method. The phylogram is represented by 406 gp63 protein sequences. The sequences from *Zelonia* and *Endotrypanum* genera are shown in red and yellow, respectively. Bootstrap exceeding 95% are shown in branches as blue circles.

### 3.6. Metabolic Pathway Comparison between *L. (L.) major*, *Z. costaricensis*, and *E. monterogeii*

By using comparative analysis with the Asgard tool, we screened 385 modules (metabolic pathways) from KEGG, searching for similarities and differences among all four taxa (*E. monterogeii* M907, *E. monterogeii* LV88, *Z. costaricensis*, and *L. major*). All pathways at the enzymatic level were almost identical for the three genera. We show some of the most distinctive enzymes differentially found in these three genera in Tables S6 and S7.

We found ten *Zelonia*-specific enzymes, which include catalase (E.C. 1.11.1.6), penicillin acylase (E.C. 3.5.1.11), and argininosuccinate lyase (E.C. 4.3.2.1). Likewise, we also looked for specific enzymes in *Endotrypanum*, finding only two, sialidase (E.C. 3.2.1.18) and phosphoglucosamine mutase (E.C. 5.4.2.10). Based on these, we inferred the sphingolipid pathway, as defined by KEGG, comparing the four species (Figure 6). Seven of the enzymes were conserved in all of the species, with sialidase as the only one exclusive to *E. monterogeii*.



**Figure 6.** Schematic representation of sphingolipid metabolism comparing four *Leishmaniinae* parasites. The numbers in green circles represent the number of predicted enzymes that are present in each pathway and in the four lineages. The blue arrow shows the sialidase enzyme present only in *Endotrypanum monterogeii* strains but absent in *Leishmania (Leishmania) major*.

A comparative enzymatic analysis between *Leishmania* and *Endotrypanum* revealed 485 KEGG ortholog enzymes. Furthermore, this analysis identified 13 specific enzymes of *Leishmania* that were absent in *Endotrypanum* spp.

Trypanosomatids are unable to synthesize heme, however, the last three enzymes of heme biosynthesis were acquired from bacteria by horizontal gene transfer (HGT) [80]. Alves et al. 2011 have identified these three genes in some *Leishmaniinae*, such as *Leishmania*, *Zelonia* (at the time, classified as *Leptomonas*), *Angomonas*, *Strigomonas*, *Crithidia*, and *Endotrypanum* [81]. Accordingly, we also screened the *Z. costaricensis* genome identifying these last three coding genes (PPOX, CPOX, and FeCH) (Table S7).

## 4. Discussion

Since the establishment of the subfamily *Leishmaniinae* by Jirků in 2012 [82], molecular phylogenetics studies have improved the current taxonomic organization of the *Leishmani-*

inae parasites. This includes the setting up of two new genera, *Zelonia* and *Porcisia*, and the new subgenus *Mundinia*; it was recognized that some species previously considered as belonging to the genus *Leishmania* were, in fact, species of *Endotrypanum* [4,8]. Expanding the genomic biology of the basal Leishmaniinae lineages [17] will potentially help in elucidating the monoxenous transition from a monoxenous to a dixenous life cycle. Until the present study, only a few genomes of monoxenous Leishmaniinae have been published [27,28,33,83]. In this study, we performed the first large-scale genomic analysis of a member of the genus *Zelonia*, a Brazilian isolate of *Z. costaricensis*, and compared it to that of the dixenous parasite *E. monterogei* and other closely related dixenous and monoxenous Leishmaniinae. The idea is that these comparisons will highlight the expansion or contraction of different gene families that contribute to different life cycles.

The *E. monterogei* M907 genome size (~29.9 Mb) was close to that of *E. monterogei* LV88 (30.4 Mb), and also similar in size to that of *P. hertigi* (~29.1 Mb) and *L. major* (32.8 Mb). [26,33] (Table 1). Interestingly, *Z. costaricensis* was assembled in ~38.8 Mb, a size that is close to that of *C. fasciculata* (~41.2 Mb). We postulate that this increase in genome size can be attributed to the occurrence of duplication events, explicitly, the expansion of multigene families, which comprise approximately 4.21% of the genome of *Z. costaricensis* (Table S2). In trypanosomatids, repetitive content can reach 50% in *Trypanosoma cruzi* and 10% in *L. major*, of which 29% and 4% are multigene families, respectively [22,26,84]. On the other hand, *Z. costaricensis* and *E. monterogei* have 12,201 and 9,711 predicted protein-coding genes, respectively, similar to other Leishmaniinae spp., such as *C. fasciculata* CfCl, *Lotmaria passim*, and *L. pyrrocoris* H10 [6,27,28].

Earlier phylogenies suggest that *Z. costaricensis* is placed in an early branch of the Leishmaniinae tree, basal to the *Endotrypanum*, *Porcisia*, and *Leishmania*. Given that these phylogenies were built using few genetic markers [4,21,85], this study employed a more robust phylogenomic analysis using whole-genome sequences to confirm the placement of *Z. costaricensis* TCC169E and *E. monterogei* M907 (previously referred to as *E. schaudinni* by Mesnil and Brimont, 1908). Despite the relatively fragmented nature of our genomes (Table 1), we obtained a very well supported and fully resolved phylogeny by adopting phylogenomic methods.

The phylogenomic tree was inferred with a variety of methods (ML and IB), statistical support metrics (BS = 100% and UFBS = 100%), and two different datasets (99 vs. 3384 protein-coding genes) in order to find variation in the topologies of the species tree. *B. ayalai*, a parasite of fleas (Siphonaptera) [86], was used as an outgroup. Additionally, due to the fact that spurious high support values (100% = UFBS, 100% = BS, and 1 = PP) can be seen in all nodes of phylogenomic trees using the supermatrix method, we also performed our inference using the supertree approach (Figure S5).

In all methods employed, we recovered *Z. costaricensis* as the sister group of *Porcisia*/*Endotrypanum* and *Leishmania*, in the infrafamily Leishmaniatae, in agreement with previous studies [33,85]. The genus *Endotrypanum*, along with the genus *Porcisia*, has previously been found to be an early branch compared to *Leishmania* [4,33]. In agreement with previous studies using different methods, our analyses confirmed, using ML, Bayesian inference, and quartet-based methods, the robust phylogenetic placement of the genus *Endotrypanum* spp. [4,7,33]. The two *E. monterogei* strains are located on the same branch and with branch lengths of nearly zero, confirming the findings of Espinosa et al. (2016) showing that the formerly named *E. schaudinni* M907 isolate is actually the same species as *E. monterogei* LV88 (MCHO/CR/62/A9).

The two phylogenomic trees obtained in this work fully agree with previous analyses (e.g., [4,5]) based on classical markers for trypanosomatids such as the SSU rRNA gene, the ITS rDNA, and the glycosomal glyceraldehyde 3-phosphate dehydrogenase (gGAPDH) gene. However, it is worth noting that these markers have different resolutions that are suitable for specific situations. Within the Leishmaniinae, primary sequence evolution is sometimes slow enough (e.g., within the *Viannia* subgenus of *Leishmania*) that only the ITS rDNA of the three classic markers has enough phylogenetic signal to separate closely



related species in the subfamily. Our phylogenomic analyses, employing 99 or 3384 genes displaying a broad range of evolutionary rates, confirm that the classic markers continue to be valuable tools in studies involving large numbers of isolates, for which genome sequencing is not efficient or not of interest. On the other hand, the use of many single-copy protein sequences yields phylogenies with much greater statistical support while also practically eliminating the risk of using markers whose gene trees differ from the species tree (even though that is an uncommon problem for the three classic markers used for trypanosomatid phylogeny, it remains a possibility). Therefore, marker selection should be performed on a case-by-case basis, depending on the specifics of each study.

Gene duplication is an essential mechanism that drives genome evolution, being a factor responsible for the expansion of gene families [87]. The trypanosomatid genomes are characterized by exhibiting a tandem duplication mechanism and, through this event, they have extended the number of multigene families [24]. For instance, cation transporter genes in *Trypanosoma* spp., and amino acids transporter (AAT), receptor-type adenylate cyclase, and gp63 genes in *Leishmania* spp. have been seen forming tandem gene arrays [24,88].

Gp63 proteins are zinc-dependent metalloproteases that form the structural basis of the parasite's cellular membrane. In *Leishmania*, gp63 plays important roles in host cell entry, immune modulation, intracellular survival, tissue invasion, and antigenic variation, which make it a key virulence factor in *Leishmania* infection. It inactivates the complement cascade, preventing damage to the parasite's membrane while allowing opsonization and phagocytosis. By facilitating parasite binding to macrophages through fibronectin receptors, gp63 reduces the production of TNF, IL-12, and nitric oxide, further promoting parasite survival. It can degrade host cell proteins, promoting the acquisition of nutrients from the host cell, and contribute to the establishment of the parasitophorous vacuole, the specialized compartment that guarantees the parasite's intracellular survival. Gp63 also activates a host tyrosine phosphatase, which aids in the parasite's entry into macrophages [29,76,77].

In this study, 161 gp63 genes were identified and annotated in the *E. monterogeei* M907 genome. This is a significant increase compared to *Z. costaricensis* with 30 gp63 genes. To verify this high number, we screened the *E. montegorei* LV88 genome from TriTrypDB (Table S1), and also found 171 sequences annotated as gp63 or leishmanolysin (Figure 3). Additionally, our phylogenetic analysis of gp63 (Figure 5) suggests that the expansion of this multigene family seen in *Endotrypanum* spp. may have happened after its split from the *Leishmania* genus.

Like *Endotrypanum*, *T. cruzi* possesses a large gp63 family (Tcgp63), with 60 proteins split into subfamilies; the Tcgp63-I and II groups have shown features of metalloproteases, while Tcgp63-III may contain some pseudogenes [89]. In *L. (V.) braziliensis*, 38 gp63 genes have been identified, of which the ones on chromosome 10 genes are associated with parasite interaction with the mammalian host, while another gp63 group is related to adhesion in the insect gut epithelium [78,90]. Pereira et al. (2009) reported that gp63 genes were involved in the interaction of *Leptomonas* spp. with their insect vector [91].

The large gp63 repertoire found in *Endotrypanum* opens a new focus of investigations on the evolution of the digenetic life mode. Its relatively low genetic expression in the monogenetic *Zelonia* suggests that increasing the number of gp63 genes could have been a crucial evolutionary step in the transition from the monogenetic to the digenetic lifestyle. We hypothesize that the gp63 multigene family may be intimately linked to novel features reflecting the adaptation to the digenetic life cycle within the Leishmaniinae.

Also, the expansion of multigene families has been a fundamental step for parasites to adapt and survive inside their host [24]. In this study, we identified 70 amino acid transporter (AAT) genes in *Z. costaricensis*, higher than in *Leishmania* spp., as shown in Figure 2. The AAT gene family has been biochemically well-characterized in a few trypanosomatids. Some AATs perform the uptake of L-arginine in *L. (Leishmania) amazonensis* and *L. (L.) donovani*, proline in *L. (L.) donovani* and *T. cruzi*, and glutamate in *T. cruzi* [92–94]. Mazared et al. (1999) identified that the activity of proline transporter has been active in the promastigote stage (insect host) in *L. (L.) donovani* [95]. Moreover, the duplication events are

induced and presumably maintained to meet the requirements of specific life cycle stages or to regulate amino acid uptake, as occurred in *T. cruzi*, *T. brucei*, and *L. (L.) major* [88].

Gene repertoires could be reduced due to adaptation of parasitic lifestyle into vertebrates or plants, as happened in *Leishmania* and *Phytomonas* [73,96]. Thus, we can explain that this large number of ATT genes is essential to adapt to its way of life inside the insect host, in which the parasite captures amino acids as their main source of energy. Indeed, the parasites inside the insect have access to a number of amino acid sources obtained from the insect's hemolymph, as well as amino acids from bacterial sources [97].

Conversely, ATP-binding cassette (ABC)-type transporters are another multigene family that displayed a decrease in the number of copies in *Z. costaricensis*. ABC comprises one of the largest protein families in living organisms, having conserved ABC domains that bind to and hydrolyze ATP [98]. In *Leishmania* spp., ABC transporters have been well-characterized into eight subfamilies (42 genes) and were sometimes implicated in drug resistance mechanisms [99]. For example, there is evidence that the ABC transporter MRPA subfamily confers antimony resistance [100]. Another, ABCC7 transporter, was a candidate to confer resistance to pentamidine in the amastigote and promastigote stages in *L. major* and *L. (L.) amazonensis* [101].

In our analyses of *Zelonia*, we identified 40 ABC-type transporter proteins, which is a low number when compared to *L. (L.) major* and *E. monterogeei*, with 120 and 126 proteins, respectively (Figure 2). Probably, this monoxenous organism did not need to increase its arsenal of ABC transporters, as probably occurred in *Leishmania* spp. In agreement with our data, this pattern of reduction in ABC-type transporter genes has already been reported in the genome of *L. pyrrocoris* [28], suggesting that the low repertory of such genes might be a characteristic for its lifestyle in the insect.

Trypanosomatid metabolism has evolved to adapt to new environments and food supplies in vertebrate and invertebrate hosts [102]. The evolutionary events of gain and loss of metabolism-related genes can be used to infer switches in the lifestyles of monoxenous and dixenous parasites [97]. Interestingly, Leishmaniinae is the only trypanosomatid subfamily with a notable gain of metabolic genes (23 enzymes), possibly of a bacterial origin [3]. *Leishmania* spp. contains within its genome a complex metabolic machinery, which allows alternations between insect vectors and vertebrate hosts [103,104]. We screened for enzyme-coding genes by comparing four lineages of Leishmaniinae and have seen that most enzymes are present in all four with a few exceptions (Table S7). Then, we focused on two interesting genes, one specific to *Endotrypanum* spp. and the other to *Z. costaricensis*.

Trans-sialidase is a large multigene family characterized in both *T. cruzi* and *T. brucei* [22,25,105]. Over ~1400 genes have been arranged in eight subfamilies, with special interest in the type I subfamily, which is associated with trans-sialidase/neuraminidase enzymatic activities [106]. Trans-sialidase has been localized in the cell body, flagellum, and flagellar pocket [107,108]. *Endotrypanum*, a lineage distantly related to *Trypanosoma*, has also shown high sialidase activity [109]. Herein, we confirmed an *Endotrypanum*-specific gene, which might be used in the sphingolipid pathway (Figure 6 and Table S7). Interestingly, this sialidase shares orthologs in both *T. cruzi* and *T. brucei*, suggesting that it may be a vestigial gene that still plays an important role in *Endotrypanum* spp.

The main function of sialidase is to transfer sialic acid from one glycoconjugate to another [110]. Another enzymatic activity of sialidase observed during the trypomastigote stage of *T. cruzi* is that, when released into the extracellular environment, it can be dispersed into the blood [111]. *T. cruzi* is unable to synthesize sialic acid, but it has a sialidase receptor to acquire sialic acid units from mammalian host glycoconjugates [112,113]. That way, at the trypomastigote stage, *T. cruzi* has a molecular tool to protect itself from host complement, enabling its survival in the bloodstream [114,115]. Given this ability of *T. cruzi*, it is possible that the sialidase enzyme of *Endotrypanum* spp. exerts essential enzymatic activity to evade attack via the vertebrate host's immune system, more specifically, attack via the complement system. Furthermore, we believe that sialidase can be used as an *Endotrypanum*-specific

marker enzyme to aid in epidemiological identification, especially when doing insect host surveys [109].

Previous studies have emphasized the critical role of the catalase enzyme, deeming it essential for the monoxenous lifestyle; accordingly, it has been identified in *Novymonas*, *Leptomonas*, and *Crithidia* genera, but it has been lost by *Leishmania* spp., presumably due to its adaptation to a dixenous lifestyle [116,117]. The enzyme catalase has been found only in the subfamily Leishmaniinae, and it was acquired through a recent HGT event [117]. *Leptomonas*, *Crithidia*, and *Novymonas* probably acquired their catalase enzyme from a *Brachyspira* spp. (Spirochaetes) bacterium, while *Blastocrithidia* spp. might have acquired its catalase gene from a bacterium related to *Snodgrassella alvi* (Betaproteobacteria). The catalase enzyme identified in *Zelonia* spp. (Table S7), has displayed 85% and 75% identities with the orthologs from *Novymonas esmeralda* and *B. alvinipulli*, respectively (data not shown). This result suggests that the *Zelonia* catalase is derived from a common ancestor with *N. esmeralda*, which has been strongly supported by previous studies [117,118].

Host-parasite interaction mechanisms appear to be conserved in trypanosomatid parasites, although there are some variations depending on the proteomic machinery of the organism [119]. Interestingly, the host immune system became the primary driving force behind trypanosomatid evolution, constantly placing the parasite's surface under selective pressure [120]. Consequently, surface protein families have evolved to enable trypanosomatid parasites to cope with environmental stresses encountered during their extracellular stage in the invertebrate host or intracellular stage in the vertebrate host [121]. Previous phylogenetic studies have demonstrated the relationship between subfamilies of surface proteins and their association with monoxenous or dixenous life cycles [121]. For instance, amastin has been categorized into four distinct subfamilies ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ) [96]. The position of basal amastins from the free-living kinetoplastid *Bodo saltans* is closely related to the  $\alpha$ -amastin subfamily. Conversely, in the dixenous *Leishmania*, an expansion of the  $\delta$ -amastin subfamily is associated with macrophage-parasite interaction. This evolutionary event coincides with the transition to the intracellular stage.

The comparative approach used in this study clearly suggests the existence of expanded multigene families that might confer adaptations to their life cycle, which alternates between insect and vertebrate host. Lastly, although short read-based sequencing technology produced fragmented genome assemblies, this did not affect the analysis and annotation of the whole genome sequences of *Z. costaricensis* and *E. monterogeii* (Figure S2 and Table 1). Nevertheless, based on the new findings about repetitive gene families uncovered in this study, we strongly believe that these genomes should be newly sequenced using third-generation sequencing [23], which should avoid assembly artifacts that collapse such repetitive families and thus more confidently characterize the significant expansion of multigene families in both genera.

## 5. Conclusions

The present comparative genomics study provides new insights into the possible mechanisms of host adaptation of *Zelonia costaricensis* and *Endotrypanum monterogeii*. It highlights the lineage-specific expansion in *E. monterogeii* of important gene families that interact with their host, such as gp63 and gp46, which are important modulators of *Leishmania* species. In *Z. costaricensis*, a large number of multigene families are involved in biopterin, amino acid transporters, adenylate cyclase, and hypothetical proteins carrying leucine-rich repeats (2 copies) and C-type lectin domains of the IGP family. A phylogenomic analysis using robust methods (ML and IB) and approaches (supermatrix and supertree) of single-copy genes confirmed the position of the two genera within the Leishmaniinae. The presented data add to our understanding of the genome biology of *Zelonia* (monoxenous) and *Endotrypanum* (dixenous), close relatives of *Leishmania* species, and the expansion and retraction of multigene families in trypanosomatids. Future genomic studies of other monoxenous Leishmaniinae will help in understanding the evolutionary acquisition of the dixenous life cycle.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pathogens12121409/s1>, Figure S1: Dot-plots representing whole genome comparison.; Figure S2: Genomic quality assessment of the 27 trypanosomatids genomes used in this study.; Figure S3: Histogram of functional classification of the genomes of the three species Leishmaniinae according to KOG functional groups.; Figure S4: (a) Phylogenomics analysis of subfamily Leishmaniinae based on the supermatrix method using 99 BUSCO conservative genes (JTT +  $\Gamma$ 4 + F + BS). (b) Phylogenomics analysis of subfamily Leishmaniinae based on the supermatrix method using 99 BUSCO conservative genes (JTT +  $\Gamma$ 4 + F + UFBS). (c) Phylogenomics relationships of subfamily Leishmaniinae based on the supermatrix method using 99 BUSCO conservative genes (PMSF+UFBS). (d) Phylogenomics relationships of subfamily Leishmaniinae based on the supermatrix method using 99 BUSCO conservative genes (inference Bayesian). (e) Phylogenomics relationships of subfamily Leishmaniinae based on supermatrix analyses (JTT +  $\Gamma$ 4 + F + UFBS) on 3384 single-copy genes.; Figure S5: Phylogenomics relationships of subfamily Leishmaniinae based on supertree analyses.; Figure S6: A bar graph indicating the comparison of the number of sequences between *E. monterogeei* LV88 (from TritrypDB) and *E. monterogeei* M907.; Figure S7: GP63 conserved motif in *E. monterogeei* M907. [28]; Table S1: List of genomes sequenced in this work and several available 25 trypanosomatid genome assemblies. [6,27,28,33,34,83,121–133]; Table S2: Percentage of repetitive element categories representation in the genome of *Z. costaricensis* and *E. monterogeei*.; Table S3: Ortholog clusters analysis using OrthoVenn version 2 webtool.; Table S4: Methods and models used in this phylogenomics analysis.; Table S5: Number of gp63 proteins forming ortholog groups in *Endotrypanum monterogeei*.; Table S6: List of specific enzymes for *Leishmania major*; Table S7: Summary of specific enzymes for *Zelonia costaricensis*, *Endotrypanum monterogeei*, and specific between *Z. costaricensis* and *Leishmania major*.

**Author Contributions:** Conceptualization, P.O.T.-V. and J.M.P.A.; methodology, P.O.T.-V. and J.M.P.A.; formal analysis, P.O.T.-V. and J.M.P.A.; investigation, P.O.T.-V., K.Y.O.C., J.F.C.T. and M.G.S.; data curation, P.O.T.-V.; writing—original draft preparation, P.O.T.-V.; writing—review and editing, J.M.P.A., M.M.G.T. and J.J.S.; visualization, P.O.T.-V. and J.M.P.A.; supervision, G.A.B. and J.M.P.A.; funding acquisition, J.M.P.A. and G.A.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** P.O.T.-V. was the recipient of fellowship #2018/01284-6, Sao Paulo Research Foundation (FAPESP). This project was partially funded by grant #2013/14622-3, Sao Paulo Research Foundation (FAPESP) awarded to J.M.P.A and award #0830056, National Science Foundation’s Assembling the Tree of Life program, awarded to G.A.B.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The assembled genomes generated in this work are available from NCBI under accession numbers JAXHDX000000000 (Zelonia costaricensis IZEL/BR/89/169E) and JAXHDY000000000 (Endotrypanum monterogeei MCHO/PA/62/M907).

**Acknowledgments:** We would like to thank Marta Campaner and Carmen C. Takata for cloning and culturing the protozoa used in this study; Mariana Meidani Ripoli for manuscript editing.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Lukeš, J.; Butenko, A.; Hashimi, H.; Maslov, D.A.; Votýpka, J.; Yurchenko, V. Trypanosomatids Are Much More than Just Trypanosomes: Clues from the Expanded Family Tree. *Trends Parasitol.* **2018**, *34*, 466–480. [CrossRef]
2. Maslov, D.A.; Votýpka, J.; Yurchenko, V.; Lukeš, J. Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends Parasitol.* **2013**, *29*, 43–52. [CrossRef]
3. Maslov, D.A.; Opperdoes, F.R.; Kostygov, A.Y.; Hashimi, H.; Lukeš, J.; Yurchenko, V. Recent advances in trypanosomatid research: Genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* **2019**, *146*, 1–27. [CrossRef]
4. 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* **2016**, *145*, 430–442. [CrossRef]



5. Votýpka, J.; d'Avila-Levy, C.M.; Grellier, P.; Maslov, D.A.; Lukeš, J.; Yurchenko, V. New Approaches to Systematics of Trypanosomatidae: Criteria for Taxonomic (Re)description. *Trends Parasitol.* **2015**, *31*, 460–469. [[CrossRef](#)]
6. Runckel, C.; DeRisi, J.; Flenniken, M.L. A draft genome of the honey bee trypanosomatid parasite *Crithidia mellificae*. *PLoS ONE* **2014**, *9*, e95057. [[CrossRef](#)]
7. Cupolillo, E.; Medina-Acosta, E.; Noyes, H.; Momen, H.; Grimaldi, G., Jr. A revised classification for *Leishmania* and *Endotrypanum*. *Parasitol. Today* **2000**, *16*, 142–144. [[CrossRef](#)]
8. Kostygov, A.Y.; Yurchenko, V. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol. (Praha)* **2017**, *64*, 020. [[CrossRef](#)]
9. Rodriguez-Bonfante, C.; Bonfante-Garrido, R.; Grimaldi, G., Jr.; Momen, H.; Cupolillo, E. Genotypically distinct *Leishmania colombiensis* isolates from Venezuela cause both cutaneous and visceral leishmaniasis in humans. *Infect. Genet. Evol.* **2003**, *3*, 119–124. [[CrossRef](#)]
10. Ghosh, S.; Banerjee, P.; Sarkar, A.; Datta, S.; Chatterjee, M. Coinfection of *Leptomonas seymouri* and *Leishmania donovani* in Indian leishmaniasis. *J. Clin. Microbiol.* **2012**, *50*, 2774–2778. [[CrossRef](#)]
11. Rogerio, L.A.; Takahashi, T.Y.; Cardoso, L.; Takamiya, N.T.; de Melo, E.V.; de Jesus, A.R.; de Oliveira, F.A.; Forrester, S.; Jeffares, D.C.; da Silva, J.S.; et al. Co-infection of *Leishmania infantum* and a *Crithidia*-related species in a case of refractory relapsed visceral leishmaniasis with non-ulcerated cutaneous manifestation in Brazil. *Int. J. Infect. Dis.* **2023**, *133*, 85–88. [[CrossRef](#)]
12. Shaw, J.J. A possible vector of *Endotrypanum schaudinni* of the sloth *Choloepus hoffmanni*, in Panama. *Nature* **1964**, *201*, 417–418. [[CrossRef](#)]
13. Paperna, I.; Boulard, Y.; Hering-Hagenbeck, S.H.; Landau, I. Description and ultrastructure of *Leishmania zuckermanni* n. sp. amastigotes detected within the erythrocytes of the South African gecko *Pachydactylus turneri* Gray, 1864. *Parasite* **2001**, *8*, 349–353. [[CrossRef](#)]
14. Mesnil, F.; Brimont, E. Sur un hématozoaire nouveau (*Endotrypanum* n. gen.) d'un édenté de la Guyane. *C. R. Séances Soc. Biol. Ses. Fil.* **1908**, *65*, 581–583.
15. Shaw, J.J.; Bird, R.G. The endoerythrocytic habitat of a member of the Trypanosomatidae, *Endotrypanum schaudinni*, Mesnil and Brimont, 1908. *Z. Trop. Parasitol.* **1969**, *20*, 144–150.
16. Shaw, J.J. *The Haemoflagellates of Sloths*; H. K. Lewis: London, UK, 1969; p. 132.
17. Yurchenko, V.; Butenko, A.; Kostygov, A.Y. Genomics of Trypanosomatidae: Where We Stand and What Needs to Be Done? *Pathogens* **2021**, *10*, 1124. [[CrossRef](#)]
18. Franco, A.M.; Machado, G.M.; Naiff, R.D.; Moreira, C.F.; Grimaldi, G. Characterization of *Endotrypanum* parasites infecting the Neotropical tree Sloths (Edentata). *Mem. Inst. Oswaldo Cruz* **1999**, *94*, 261–268. [[CrossRef](#)]
19. Medina-Acosta, E.; Paul, S.; Tomlinson, S.; Pontes de Carvalho, L.C. Combined occurrence of trypanosomal sialidase/transsialidase activities and leishmanial metalloproteinase gene homologues in *Endotrypanum* sp. *Mol. Biochem. Parasitol.* **1994**, *64*, 273–282. [[CrossRef](#)]
20. Yurchenko, V.Y.; Lukes, J.; Jirku, M.; Zeledón, R.; Maslov, D.A. *Leptomonas costaricensis* sp. n. (Kinetoplastea: Trypanosomatidae), a member of the novel phylogenetic group of insect trypanosomatids closely related to the genus *Leishmania*. *Parasitology* **2006**, *133 Pt 5*, 537–546. [[CrossRef](#)]
21. 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 Dioxenous Parasitism in the Leishmaniinae. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005215. [[CrossRef](#)]
22. El-Sayed, N.M.; Myler, P.J.; Blandin, G.; Berriman, M.; Crabtree, J.; Aggarwal, G.; Caler, E.; Renauld, H.; Worthey, E.A.; Hertz-Fowler, C.; et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* **2005**, *309*, 404–409. [[CrossRef](#)]
23. Jackson, A.P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol. Biol. Evol.* **2010**, *27*, 33–45. [[CrossRef](#)]
24. Jackson, A.P. Genome evolution in trypanosomatid parasites. *Parasitology* **2015**, *142* (Suppl. S1), S40–S56. [[CrossRef](#)]
25. Berriman, M.; Ghedin, E.; Hertz-Fowler, C.; Blandin, G.; Renauld, H.; Bartholomeu, D.C.; Lennard, N.J.; Caler, E.; Hamlin, N.E.; Haas, B.; et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **2005**, *309*, 416–422. [[CrossRef](#)]
26. Ivens, A.C.; Peacock, C.S.; Worthey, E.A.; Murphy, L.; Aggarwal, G.; Berriman, M.; Sisk, E.; Rajandream, M.A.; Adlem, E.; Aert, R.; et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **2005**, *309*, 436–442. [[CrossRef](#)]
27. Schmid-Hempel, P.; Aebi, M.; Barribeau, S.; Kitajima, T.; du Plessis, L.; Schmid-Hempel, R.; Zoller, S. The genomes of *Crithidia bombi* and *C. expoeki*, common parasites of bumblebees. *PLoS ONE* **2018**, *13*, e0189738. [[CrossRef](#)]
28. Flegontov, P.; Butenko, A.; Firsov, S.; Kraeva, N.; Eliáš, M.; Field, M.C.; Filatov, D.; Flegontova, O.; Gerasimov, E.S.; Hlaváčová, J.; et al. Genome of *Leptomonas pyrhrhcoris*: A high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci. Rep.* **2016**, *6*, 23704. [[CrossRef](#)]
29. Yao, C.; Donelson, J.E.; Wilson, M.E. The major surface protease (MSP or GP63) of *Leishmania* sp. biosynthesis, regulation of expression, and function. *Mol. Biochem. Parasitol.* **2003**, *132*, 1–16. [[CrossRef](#)]
30. McMahan-Pratt, D.; Traub-Cseko, Y.; Lohman, K.L.; Rogers, D.D.; Beverley, S.M. Loss of the GP46/M-2 surface membrane glycoprotein gene family in the *Leishmania braziliensis* complex. *Mol. Biochem. Parasitol.* **1992**, *50*, 151–160. [[CrossRef](#)]
31. Teixeira, S.M.; Russell, D.G.; Kirchhoff, L.V.; Donelson, J.E. A differentially expressed gene family encoding “amastin,” a surface protein of *Trypanosoma cruzi* amastigotes. *J. Biol. Chem.* **1994**, *269*, 20509–20516. [[CrossRef](#)]

32. Rochette, A.; McNicoll, F.; Girard, J.; Breton, M.; Leblanc, E.; Bergeron, M.G.; Papadopoulou, B. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. *Mol. Biochem. Parasitol.* **2005**, *140*, 205–220. [[CrossRef](#)]
33. Albanaz, A.T.S.; Gerasimov, E.S.; Shaw, J.J.; Sádlová, J.; Lukeš, J.; Volf, P.; Opperdoes, F.R.; Kostygov, A.Y.; Butenko, A.; Yurchenko, V. Genome Analysis of *Endotrypanum* and *Porcisia* spp., Closest Phylogenetic Relatives of *Leishmania*, Highlights the Role of Amastins in Shaping Pathogenicity. *Genes* **2021**, *12*, 444. [[CrossRef](#)]
34. Butenko, A.; Kostygov, A.Y.; Sádlová, J.; Kleschenko, Y.; Bečvář, T.; Podešvová, L.; Macedo, D.H.; Žihala, D.; Lukeš, J.; Bates, P.A.; et al. Comparative genomics of *Leishmania (Mundinia)*. *BMC Genom.* **2019**, *20*, 726. [[CrossRef](#)]
35. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 8 July 2020).
36. Dittami, S.M.; Corre, E. Detection of bacterial contaminants and hybrid sequences in the genome of the kelp *Saccharina japonica* using Taxoblast. *PeerJ* **2017**, *5*, e4073. [[CrossRef](#)]
37. Alves, J.M.P. Mfsizes: Version 1.8.7. Available online: <https://sourceforge.net/projects/mfsizes/files/> (accessed on 18 October 2020).
38. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
39. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*, R12. [[CrossRef](#)]
40. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607–2618. [[CrossRef](#)]
41. Hoff, K.J.; Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr. Protoc. Bioinform.* **2019**, *65*, e57. [[CrossRef](#)]
42. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [[CrossRef](#)]
43. Lagesen, K.; Hallin, P.; Rødland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [[CrossRef](#)]
44. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)]
45. Nawrocki, E.P.; Burge, S.W.; Bateman, A.; Daub, J.; Eberhardt, R.Y.; Eddy, S.R.; Floden, E.W.; Gardner, P.P.; Jones, T.A.; Tate, J.; et al. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* **2015**, *43*, D130–D137. [[CrossRef](#)]
46. Flynn, J.M.; Hubble, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [[CrossRef](#)]
47. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **2009**, *4*, 4.10.1–4.10.14. [[CrossRef](#)]
48. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2011**, *39*, D38–D51. [[CrossRef](#)]
49. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
50. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [[CrossRef](#)]
51. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)]
52. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
53. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [[CrossRef](#)]
54. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)] [[PubMed](#)]
55. Xu, L.; Dong, Z.; Fang, L.; Luo, Y.; Wei, Z.; Guo, H.; Zhang, G.; Gu, Y.Q.; Coleman-Derr, D.; Xia, Q.; et al. OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **2019**, *47*, W52–W58. [[CrossRef](#)] [[PubMed](#)]
56. Aslett, M.; Aurrecochea, C.; Berriman, M.; Brestelli, J.; Brunk, B.P.; Carrington, M.; Depledge, D.P.; Fischer, S.; Gajria, B.; Gao, X.; et al. TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **2010**, *38*, D457–D462. [[CrossRef](#)]
57. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)] [[PubMed](#)]
58. Roure, B.; Baurain, D.; Philippe, H. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol. Biol. Evol.* **2013**, *30*, 197–214. [[CrossRef](#)]
59. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [[CrossRef](#)]

60. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
61. Kück, P.; Longo, G.C. FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* **2014**, *11*, 81. [[CrossRef](#)]
62. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)]
63. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jeremiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
64. Anisimova, M.; Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **2006**, *55*, 539–552. [[CrossRef](#)]
65. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **1992**, *8*, 275–282. [[CrossRef](#)]
66. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)] [[PubMed](#)]
67. Roch, S.; Steel, M. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor. Popul. Biol.* **2015**, *100C*, 56–62. [[CrossRef](#)]
68. Zhang, C.; Rabiee, M.; Sayyari, E.; Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **2018**, *19* (Suppl. S6), 153. [[CrossRef](#)] [[PubMed](#)]
69. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)] [[PubMed](#)]
70. Alves, J.M.; Buck, G.A. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. *Chem. Biodivers.* **2007**, *4*, 2593–2602. [[CrossRef](#)] [[PubMed](#)]
71. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
72. Suzek, B.E.; Huang, H.; McGarvey, P.; Mazumder, R.; Wu, C.H. UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **2007**, *23*, 1282–1288. [[CrossRef](#)]
73. Opperdoes, F.R.; Butenko, A.; Flegontov, P.; Yurchenko, V.; Lukeš, J. Comparative Metabolism of Free-living *Bodo saltans* and Parasitic Trypanosomatids. *J. Eukaryot. Microbiol.* **2016**, *63*, 657–678. [[CrossRef](#)]
74. Butenko, A.; Opperdoes, F.R.; Flegontova, O.; Horák, A.; Hampl, V.; Keeling, P.; Gawryluk, R.M.R.; Tikhonenkov, D.; Flegontov, P.; Lukeš, J. Evolution of metabolic capabilities and molecular features of diplomonids, kinetoplastids, and euglenids. *BMC Biol.* **2020**, *18*, 23. [[CrossRef](#)] [[PubMed](#)]
75. Chang, A.; Jeske, L.; Ulbrich, S.; Hofmann, J.; Koblit, J.; Schomburg, I.; Neumann-Schaal, M.; Jahn, D.; Schomburg, D. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **2021**, *49*, D498–D508. [[CrossRef](#)] [[PubMed](#)]
76. Yao, C. Major surface protease of trypanosomatids: One size fits all? *Infect. Immun.* **2010**, *78*, 22–31. [[CrossRef](#)] [[PubMed](#)]
77. Olivier, M.; Atayde, V.D.; Isnard, A.; Hassani, K.; Shio, M.T. *Leishmania* virulence factors: Focus on the metalloprotease GP63. *Microbes Infect.* **2012**, *14*, 1377–1389. [[CrossRef](#)] [[PubMed](#)]
78. Castro Neto, A.L.; Brito, A.N.A.L.M.; Rezende, A.M.; Magalhães, F.B.; de Melo Neto, O.P. In silico characterization of multiple genes encoding the GP63 virulence protein from *Leishmania braziliensis*: Identification of sources of variation and putative roles in immune evasion. *BMC Genom.* **2019**, *20*, 118. [[CrossRef](#)] [[PubMed](#)]
79. Raymond, F.; Boisvert, S.; Roy, G.; Ritt, J.F.; Légaré, D.; Isnard, A.; Stanke, M.; Olivier, M.; Tremblay, M.J.; Papadopoulou, B.; et al. Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res.* **2012**, *40*, 1131–1147. [[CrossRef](#)] [[PubMed](#)]
80. Korený, L.; Lukes, J.; Oborník, M. Evolution of the haem synthetic pathway in kinetoplastid flagellates: An essential pathway that is not essential after all? *Int. J. Parasitol.* **2010**, *40*, 149–156. [[CrossRef](#)]
81. Alves, J.M.; Voegtly, L.; Matveyev, A.V.; Lara, A.M.; da Silva, F.M.; Serrano, M.G.; Buck, G.A.; Teixeira, M.M.; Camargo, E.P. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS ONE* **2011**, *6*, e23518. [[CrossRef](#)]
82. Jirků, M.; Yurchenko, V.Y.; Lukeš, J.; Maslov, D.A. New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. *J. Eukaryot. Microbiol.* **2012**, *59*, 537–547. [[CrossRef](#)]
83. Kraeva, N.; Butenko, A.; Hlaváčová, J.; Kostygov, A.; Myškova, 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*, e1005127. [[CrossRef](#)]
84. Pita, S.; Díaz-Viraqué, F.; Iraola, G.; Robello, C. The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens. *Genome Biol. Evol.* **2019**, *11*, 546–551. [[CrossRef](#)] [[PubMed](#)]
85. 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. [[CrossRef](#)] [[PubMed](#)]



86. Votýpka, J.; Suková, E.; Kraeva, N.; Ishemgulova, A.; Duží, I.; Lukeš, J.; Yurchenko, V. Diversity of trypanosomatids (Kinetoplastea: Trypanosomatidae) parasitizing fleas (Insecta: Siphonaptera) and description of a new genus *Blechomonas* gen. n. *Protist* **2013**, *164*, 763–781. [[CrossRef](#)]
87. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [[CrossRef](#)] [[PubMed](#)]
88. Jackson, A.P. Origins of amino acid transporter loci in trypanosomatid parasites. *BMC Evol. Biol.* **2007**, *7*, 26. [[CrossRef](#)]
89. Cuevas, I.C.; Cazzulo, J.J.; Sánchez, D.O. gp63 homologues in *Trypanosoma cruzi*: Surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect. Immun.* **2003**, *71*, 5739–5749. [[CrossRef](#)]
90. Santos, A.L.; Branquinha, M.H.; D’Avila-Levy, C.M. The ubiquitous gp63-like metalloprotease from lower trypanosomatids: In the search for a function. *Acad. Bras. Cienc.* **2006**, *78*, 687–714. [[CrossRef](#)]
91. Pereira, F.M.; Bernardo, P.S.; Dias Junior, P.F.; Silva, B.A.; Romanos, M.T.; d’Avila-Levy, C.M.; Branquinha, M.H.; Santos, A.L. Differential influence of gp63-like molecules in three distinct *Leptomonas* species on the adhesion to insect cells. *Parasitol. Res.* **2009**, *104*, 347–353. [[CrossRef](#)]
92. Shaked-Mishan, P.; Suter-Grottemeyer, M.; Yoel-Almagor, T.; Holland, N.; Zilberstein, D.; Rentsch, D. A novel high-affinity arginine transporter for the human parasitic protozoan *Leishmania donovani*. *Mol. Microbiol.* **2006**, *60*, 30–38. [[CrossRef](#)]
93. Pereira, C.A.; Alonso, G.D.; Paveto, M.C.; Flawiá, M.M.; Torres, H.N. L-arginine uptake and L-phosphoarginine synthesis in *Trypanosoma cruzi*. *J. Eukaryot. Microbiol.* **1999**, *46*, 566–570. [[CrossRef](#)]
94. Geraldo, M.V.; Silber, A.M.; Pereira, C.A.; Uliana, S.R. Characterisation of a developmentally regulated amino acid transporter gene from *Leishmania amazonensis*. *FEMS Microbiol. Lett.* **2005**, *242*, 275–280. [[CrossRef](#)] [[PubMed](#)]
95. Mazareb, S.; Fu, Z.Y.; Zilberstein, D. Developmental regulation of proline transport in *Leishmania donovani*. *Exp. Parasitol.* **1999**, *91*, 341–348. [[CrossRef](#)]
96. Jackson, A.P.; Otto, T.D.; Aslett, M.; Armstrong, S.D.; Bringaud, F.; Schlacht, A.; Hartley, C.; Sanders, M.; Wastling, J.M.; Dacks, J.B.; et al. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr. Biol.* **2016**, *26*, 161–172. [[CrossRef](#)] [[PubMed](#)]
97. de Castro Neto, A.L.; da Silveira, J.F.; Mortara, R.A. Role of Virulence Factors of Trypanosomatids in the Insect Vector and Putative Genetic Events Involved in Surface Protein Diversity. *Front. Cell Infect. Microbiol.* **2022**, *12*, 807172. [[CrossRef](#)]
98. Dassa, E. Natural history of ABC systems: Not only transporters. *Essays Biochem.* **2011**, *50*, 19–42.
99. Da Costa, K.M.; Valente, R.D.C.; Fonseca, L.M.D.; Freire-de-Lima, L.; Previato, J.O.; Mendonça-Previato, L. The History of the ABC Proteins in Human Trypanosomiasis Pathogens. *Pathogens* **2022**, *11*, 988. [[CrossRef](#)]
100. El Fadili, K.; Messier, N.; Leprohon, P.; Roy, G.; Guimond, C.; Trudel, N.; Saravia, N.G.; Papadopolou, B.; Légaré, D.; Ouellette, M. Role of the ABC transporter MRPA (PGPA) in antimony resistance in *Leishmania infantum* axenic and intracellular amastigotes. *Antimicrob. Agents Chemother.* **2005**, *49*, 1988–1993. [[CrossRef](#)]
101. Coelho, A.C.; Messier, N.; Ouellette, M.; Cotrim, P.C. Role of the ABC transporter PRP1 (ABCC7) in pentamidine resistance in *Leishmania amastigotes*. *Antimicrob. Agents Chemother.* **2007**, *51*, 3030–3032. [[CrossRef](#)]
102. Bringaud, F.; Rivière, L.; Coustou, V. Energy metabolism of trypanosomatids: Adaptation to available carbon sources. *Mol. Biochem. Parasitol.* **2006**, *149*, 1–9. [[CrossRef](#)] [[PubMed](#)]
103. Oppenheimer, F.R.; Coombs, G.H. Metabolism of *Leishmania*: Proven and predicted. *Trends Parasitol.* **2007**, *23*, 149–158. [[CrossRef](#)] [[PubMed](#)]
104. Chavali, A.K.; Whittemore, J.D.; Eddy, J.A.; Williams, K.T.; Papin, J.A. Systems analysis of metabolism in the pathogenic trypanosomatid *Leishmania major*. *Mol. Syst. Biol.* **2008**, *4*, 177. [[CrossRef](#)]
105. El-Sayed, N.M.; Myler, P.J.; Bartholomeu, D.C.; Nilsson, D.; Aggarwal, G.; Tran, A.N.; Ghedin, E.; Worthey, E.A.; Delcher, A.L.; Blandin, G.; et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **2005**, *309*, 409–415. [[CrossRef](#)] [[PubMed](#)]
106. Herreros-Cabello, A.; Callejas-Hernández, F.; Gironès, N.; Fresno, M. *Trypanosoma cruzi* Genome: Organization, Multi-Gene Families, Transcription, and Biological Implications. *Genes* **2020**, *11*, 1196. [[CrossRef](#)] [[PubMed](#)]
107. Lantos, A.B.; Carlevaro, G.; Araoz, B.; Ruiz Diaz, P.; Camara Mde, L.; Buscaglia, C.A.; Bossi, M.; Yu, H.; Chen, X.; Bertozzi, C.R.; et al. Sialic Acid Glycobiology Unveils *Trypanosoma cruzi* Trypomastigote Membrane Physiology. *PLoS Pathog.* **2016**, *12*, e1005559. [[CrossRef](#)]
108. Pech-Canul, Á.C.; Monteón, V.; Solís-Oviedo, R.L. A Brief View of the Surface Membrane Proteins from *Trypanosoma cruzi*. *J. Parasitol. Res.* **2017**, *2017*, 3751403. [[CrossRef](#)]
109. Medina-Acosta, E.; Franco, A.M.; Jansen, A.M.; Sampol, M.; Nevés, N.; Pontes-de-Carvalho, L.; Grimaldi Júnior, G.; Nussenzweig, V. Trans-sialidase and sialidase activities discriminate between morphologically indistinguishable trypanosomatids. *Eur. J. Biochem.* **1994**, *225*, 333–339. [[CrossRef](#)] [[PubMed](#)]
110. Frasch, A.C. Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitol. Today* **2000**, *16*, 282–286. [[CrossRef](#)] [[PubMed](#)]
111. Agusti, R.; Couto, A.S.; Campetella, O.E.; Frasch, A.C.; de Lederkremer, R.M. The trans-sialidase of *Trypanosoma cruzi* is anchored by two different lipids. *Glycobiology* **1997**, *7*, 731–735. [[CrossRef](#)]
112. Pereira-Chioccola, V.L.; Schenkman, S. Biological role of *Trypanosoma cruzi* trans-sialidase. *Biochem. Soc. Trans.* **1999**, *27*, 516–518. [[CrossRef](#)]



113. Previato, J.O.; Andrade, A.F.; Pessolani, M.C.; Mendonca-Previato, L. Incorporation of sialic acid into *Trypanosoma cruzi* macromolecules. A proposal for a new metabolic route. *Mol. Biochem. Parasitol.* **1985**, *16*, 85–96. [[CrossRef](#)]
114. Tomlinson, S.; Raper, J. Natural human immunity to trypanosomes. *Parasitol. Today* **1998**, *14*, 354–359. [[CrossRef](#)] [[PubMed](#)]
115. Neres, J.; Bryce, R.A.; Douglas, K.T. Rational drug design in parasitology: Trans-sialidase as a case study for Chagas disease. *Drug Discov. Today* **2008**, *13*, 110–117. [[CrossRef](#)] [[PubMed](#)]
116. Zakharova, A.; Saura, A.; Butenko, A.; Podešvová, L.; Warmusová, S.; Kostygov, A.Y.; Nenarokova, A.; Lukeš, J.; Opperdoes, F.R.; Yurchenko, V. A New Model Trypanosomatid, *Novymonas esmeraldas*: Genomic Perception of Its “*Candidatus Pandoraea novymonadis*” Endosymbiont. *Mbio* **2021**, *12*, e0160621. [[CrossRef](#)] [[PubMed](#)]
117. Kraeva, N.; Horáková, E.; Kostygov, A.Y.; Kořený, L.; Butenko, A.; Yurchenko, V.; Lukeš, J. Catalase in Leishmaniinae: With me or against me? *Infect. Genet. Evol.* **2017**, *50*, 121–127. [[CrossRef](#)] [[PubMed](#)]
118. Bianchi, C.; Kostygov, A.Y.; Kraeva, N.; Záhonová, K.; Horáková, E.; Sobotka, R.; Lukeš, J.; Yurchenko, V. An enigmatic catalase of *Blastocrithidia*. *Mol. Biochem. Parasitol.* **2019**, *232*, 111199. [[CrossRef](#)] [[PubMed](#)]
119. de Castro Neto, A.L.; da Silveira, J.F.; Mortara, R.A. Comparative Analysis of Virulence Mechanisms of Trypanosomatids Pathogenic to Humans. *Front. Cell Infect. Microbiol.* **2021**, *11*, 669079. [[CrossRef](#)] [[PubMed](#)]
120. Jackson, A.P. Gene family phylogeny and the evolution of parasite cell surfaces. *Mol. Biochem. Parasitol.* **2016**, *209*, 64–75. [[CrossRef](#)]
121. Real, F.; Vidal, R.O.; Carazzolle, M.F.; Mondego, J.M.; Costa, G.G.; Herai, R.H.; Würtele, M.; de Carvalho, L.M.; Carmona e Ferreira, R.; Mortara, R.A.; et al. The genome sequence of *Leishmania (Leishmania) amazonensis*: Functional annotation and extended analysis of gene models. *DNA Res.* **2013**, *20*, 567–581. [[CrossRef](#)]
122. González-de la Fuente, S.; Peiró-Pastor, R.; Rastrojo, A.; Moreno, J.; Carrasco-Ramiro, F.; Requena, J.M.; Aguado, B. Resequencing of the *Leishmania infantum* (strain JPCM5) genome and de novo assembly into 36 contigs. *Sci. Rep.* **2017**, *7*, 18050. [[CrossRef](#)]
123. González-de la Fuente, S.; Camacho, E.; Peiró-Pastor, R.; Rastrojo, A.; Carrasco-Ramiro, F.; Aguado, B.; Requena, J.M. Complete and de novo assembly of the *Leishmania braziliensis* (M2904) genome. *Mem. Inst. Oswaldo Cruz* **2018**, *114*, e180438. [[CrossRef](#)]
124. Silveira, F.T.; Sousa Junior, E.C.; Silvestre, R.V.; Costa-Martins, A.G.; da Costa Pinheiro, K.; Sosa Ochoa, W.; Vasconcelos Dos Santos, T.; Ramos, P.K.; Casseb, S.; da Silva, S.P.; et al. Whole-Genome Sequencing of *Leishmania infantum* chagasi Isolates from Honduras and Brazil. *Microbiol. Resour. Announc.* **2021**, *10*, e0047121. [[CrossRef](#)]
125. Camacho, E.; González-de la Fuente, S.; Rastrojo, A.; Peiró-Pastor, R.; Solana, J.C.; Tabera, L.; Gamarro, F.; Carrasco-Ramiro, F.; Requena, J.M.; Aguado, B. Complete assembly of the *Leishmania donovani* (HU3 strain) genome and transcriptome annotation. *Sci. Rep.* **2019**, *9*, 6127. [[CrossRef](#)] [[PubMed](#)]
126. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) orientalis*, Isolate LSCM4, Strain LV768. *Microbiol. Resour. Announc.* **2021**, *10*, e0057421. [[CrossRef](#)] [[PubMed](#)]
127. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Thomaz-Soccol, V.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) enriettii*, Isolate CUR178, Strain LV763. *Microbiol. Resour. Announc.* **2021**, *10*, e0057521. [[CrossRef](#)] [[PubMed](#)]
128. Camacho, E.; González-de la Fuente, S.; Solana, J.C.; Rastrojo, A.; Carrasco-Ramiro, F.; Requena, J.M.; Aguado, B. Gene Annotation and Transcriptome Delineation on a De Novo Genome Assembly for the Reference *Leishmania major* Friedlin Strain. *Genes* **2021**, *12*, 1359. [[CrossRef](#)]
129. Almutairi, H.; Urbaniak, M.D.; Bates, M.D.; Jariyapan, N.; Al-Salem, W.S.; Dillon, R.J.; Bates, P.A.; Gatherer, D. Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania (Mundinia) martiniquensis*, Isolate LSCM1, Strain LV760. *Microbiol. Resour. Announc.* **2021**, *10*, e0005821. [[CrossRef](#)]
130. Batra, D.; Lin, W.; Narayanan, V.; Rowe, L.A.; Sheth, M.; Zheng, Y.; Loparev, V.; de Almeida, M. Draft Genome Sequences of *Leishmania (Leishmania) amazonensis*, *Leishmania (Leishmania) mexicana*, and *Leishmania (Leishmania) aethiopica*, Potential Etiological Agents of Diffuse Cutaneous Leishmaniasis. *Microbiol. Resour. Announc.* **2019**, *8*, e00269-19. [[CrossRef](#)]
131. Llanes, A.; Restrepo, C.M.; Del Vecchio, G.; Anguizola, F.J.; Leonart, R. The genome of *Leishmania panamensis*: Insights into genomics of the *L. (Viannia) subgenus*. *Sci. Rep.* **2015**, *5*, 8550. [[CrossRef](#)]
132. Goto, Y.; Kuroki, A.; Suzuki, K.; Yamagishi, J. Draft Genome Sequence of *Leishmania tarentolae* Parrot Tar II, Obtained by Single-Molecule Real-Time Sequencing. *Microbiol. Resour. Announc.* **2020**, *9*, e00050-20. [[CrossRef](#)]
133. Unoarumhi, Y.; Batra, D.; Sheth, M.; Narayanan, V.; Lin, W.; Zheng, Y.; Rowe, L.A.; Pohl, J.; de Almeida, M. Chromosome-Level Genome Sequence of *Leishmania (Leishmania) tropica* Strain CDC216-162, Isolated from an Afghanistan Clinical Case. *Microbiol. Resour. Announc.* **2021**, *10*, e00842-20. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.