

## Article

# Genome Analysis of *Endotrypanum* and *Porcisia* spp., Closest Phylogenetic Relatives of *Leishmania*, Highlights the Role of Amastins in Shaping Pathogenicity

Amanda T. S. Albanaz <sup>1</sup>, Evgeny S. Gerasimov <sup>2</sup> , Jeffrey J. Shaw <sup>3</sup>, Jovana Sádlová <sup>4</sup>, Julius Lukeš <sup>5,6</sup> , Petr Volf <sup>4</sup> , Fred R. Opperdoes <sup>7</sup> , Alexei Y. Kostygov <sup>1,8</sup> , Anzhelika Butenko <sup>1,5,\*</sup> and Vyacheslav Yurchenko <sup>1,9,\*</sup> 

- <sup>1</sup> Life Science Research Centre, Faculty of Science, University of Ostrava, 71000 Ostrava, Czech Republic; amandatabitaalbanaz@gmail.com (A.T.S.A.); kostygov@gmail.com (A.Y.K.)  
<sup>2</sup> Faculty of Biology, M. V. Lomonosov Moscow State University, 119991 Moscow, Russia; jalgard@gmail.com  
<sup>3</sup> Biomedical Institute, São Paulo University, São Paulo 05508, Brazil; jayusp@hotmail.com  
<sup>4</sup> Department of Parasitology, Faculty of Science, Charles University, 12844 Prague, Czech Republic; JovanaS@seznam.cz (J.S.); volf@cesnet.cz (P.V.)  
<sup>5</sup> Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice, Czech Republic; jula@paru.cas.cz  
<sup>6</sup> Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czech Republic  
<sup>7</sup> de Duve Institute, Université Catholique de Louvain, 1200 Brussels, Belgium; Fred.Opperdoes@uclouvain.be  
<sup>8</sup> Zoological Institute of the Russian Academy of Sciences, 199034 St. Petersburg, Russia  
<sup>9</sup> Martsinovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov University, 119435 Moscow, Russia  
\* Correspondence: anzhelika.butenko@paru.cas.cz (A.B.); vyacheslav.yurchenko@osu.cz (V.Y.)



**Citation:** 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. <https://doi.org/10.3390/genes12030444>

Academic Editors: Jose M. Requena and Begoña Aguado

Received: 26 February 2021  
Accepted: 18 March 2021  
Published: 20 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Abstract:** While numerous genomes of *Leishmania* spp. have been sequenced and analyzed, an understanding of the evolutionary history of these organisms remains limited due to the unavailability of the sequence data for their closest known relatives, *Endotrypanum* and *Porcisia* spp., infecting sloths and porcupines. We have sequenced and analyzed genomes of three members of this clade in order to fill this gap. Their comparative analyses revealed only minute differences from *Leishmania major* genome in terms of metabolic capacities. We also documented that the number of genes under positive selection on the *Endotrypanum*/*Porcisia* branch is rather small, with the flagellum-related group of genes being over-represented. Most significantly, the analysis of gene family evolution revealed a substantially reduced repertoire of surface proteins, such as amastins and bioprotein transporters BT1 in the *Endotrypanum*/*Porcisia* species when compared to amastigote-dwelling *Leishmania*. This reduction was especially pronounced for  $\delta$ -amastins, a subfamily of cell surface proteins crucial in the propagation of *Leishmania* amastigotes inside vertebrate macrophages and, apparently, dispensable for *Endotrypanum*/*Porcisia*, which do not infect such cells.

**Keywords:** leishmaniinae; genome analysis; gene gain; gene loss

## 1. Introduction

Trypanosomatids (family Trypanosomatidae) is a diverse group of mono-flagellated kinetoplastids, which unites obligate parasites of invertebrates (monoxenous species, one-host developmental cycle) with those, shuttling between invertebrates and vertebrates or plants (dixenous species, two-host developmental cycle) [1,2]. The following five genera represent the latter group—*Trypanosoma*, *Leishmania*, *Phytomonas*, *Porcisia*, and *Endotrypanum*. Dixenous trypanosomatids evolved from monoxenous ones independently at least three times [3]. One such transition had happened within the subfamily Leishmaniinae [4,5], giving rise to the prominent genus *Leishmania*. Biology of its representatives has been extensively studied due to its medical importance, leading to the well-resolved

taxonomy of this genus and its closest relatives [6,7]. The genus *Leishmania* is currently subdivided into four subgenera—*Leishmania*, *Mundinia*, *Sauroleishmania*, and *Viannia*, which are well-defined based on their biology (host or vector specificity and clinical manifestations) and phylogeny [8]. Many of these parasites have been scrutinized using modern genomic methods and the comparative analyses have revealed their relationships and evolutionary history [9–13]. At the same time, the closest phylogenetic relatives of *Leishmania*, specifically the genera *Endotrypanum* and *Porcisia*, remained neglected and did not attract much of attention for the reasons that are discussed below.

Mesnil and Brimont described an enigmatic intra-erythrocytic flagellate in 1908 from a French Guianan two-toed sloth (*Choloepus didactylus*) and named *Endotrypanum schaudinni* [14]. Its intracellular localization was subsequently confirmed using electron microscopy [15]. This species turned out to be very unusual, as the intra-erythrocytic forms were represented by epimastigotes, while, in culture, only promastigotes, reminiscent of *Leishmania* spp., could be observed [16]. This led to a suggestion that the two morphotypes belong to distinct lineages, of which the intra-erythrocytic parasite represents an unidentified trypanosome, while the cultured forms are related to *Leishmania* [17]. It is currently accepted that both of the morphotypes belong to *Endotrypanum* [8,18]. Of note, *E. colombiensis* [19] and *E. herreri* [20] parasitize different white cells in sloth blood. They also produce amastigotes in tissue culture, but, unlike *Leishmania* spp., those of *E. colombiensis* die out [19]. Neither species can infect hamsters *in vitro*, but *E. colombiensis* was associated with cutaneous and visceral leishmaniases in men and dogs in Colombia and Venezuela [21,22].

Besides *Endotrypanum*, two other somewhat mysterious parasites, named *Leishmania hertigi* and *L. deanei*, were described from the American tropical porcupines [23,24]. Biochemical and molecular studies have shown that these flagellates are related, yet distinct from *Leishmania* spp. [17,25], which leads to the erection of a new genus *Porcisia* to accommodate them [8]. Another name, *Paraleishmania*, was proposed for this taxon [4], but it did not become formally available according to the article 16.1 of the International Code of Zoological Nomenclature. Although only two species of this genus have been described so far; it is conceivable that others will be discovered in the future, given that there are 17 porcupine species present in the Americas. Both flagellate species can be found in the upper dermis of the skin, and in the liver and spleen of their vertebrate hosts [23,26,27]. They cause no apparent pathology, except for the vacuolization of the host cell's cytoplasm. Some of the flagellates even appear to be extracellular [24]. In culture, these parasites proliferate as long aciculate nectomonad-like promastigotes, morphologically resembling those of *L. (Mundinia)* spp. [28]. No lesions were observed in experimental infections of hamsters using intradermal inoculation of culture. Parasites from the inoculation site could be introduced into culture for up to a year, although amastigotes could only be microscopically detected for a few weeks [23]. The absence of any pathology in the natural host and their long-term survival in experimental animals indicates that these parasites have developed sophisticated mechanisms of evading the host's immune system, which are probably responsible for their high infection rates seen in natural populations.

There is good evidence from the experimental [29] and natural [30] infections that *Endotrypanum* spp. use phlebotomine sand flies as vectors and develop in their hindgut and pylorus, as do *L. (Viannia)* spp. This makes it difficult to distinguish the two groups of parasites in the wild sand flies. However, unlike *L. (Viannia)* spp., *Endotrypanum* can also be detected in Malpighian tubules [31,32]. *Endotrypanum schaudinni* has been documented in six Brazilian sand fly species [30], whereas *E. colombiensis* and *E. equatorensis* only appear to be vectored by the Panamanian and Ecuadorian *Lutzomyia hartmanni* [19,33]. Overall, there is evidence of vector specificity, with infection rates varying between species. For example, the rate of *E. schaudinni* infection of *Lu. gomezi* was significantly higher than that of *Lu. sanguinaria*, which suggests that the former is a more susceptible natural vector [34]. So far, there are no clear leads as to the transmission of *Porcisia*, but a recent study has identified *Lu. antunesi* as a potential vector for *P. hertigi* [35].

According to phylogenetic inferences, the *Endotrypanum-Porcisia* clade separated from *Leishmania* 70–120 MYA, in the Cretaceous period [36,37], when placental mammals (that emerged ~66 MYA, after the Cretaceous-Paleogene boundary) did not yet exist [38]. Xenarthrans, one of the most ancient groups of placental mammals in South America, are hosts for *Endotrypanum* spp. It seems plausible that the parasite clade under study, to which *Endotrypanum* belongs, has originated in this mammalian lineage. However, *Porcisia* spp. have switched to other suitable hosts, including the ancestral American porcupines (Erethizontidae).

In this work, we sequenced the genomes of three species of the *Endotrypanum*—*Porcisia* clade and performed their comparative analyses, demonstrating correlations between their genomic content and biological peculiarities.

## 2. Materials and Methods

### 2.1. Cultivation, DNA Isolation and Species Verification

The strains that were studied in this work were *Porcisia deanei* TCC258 (MCOE/BR/91/-M13451), which were isolated from *Coendou* sp. in Brazil in 1991, *P. hertigi* TCC260 (MCOE/PA/80/C8), isolated from *Coendou rothschildi* in Panama in 1980, and *Endotrypanum* sp. ATCC 30507 (MCHO/PA/72/3130) isolated from the sloth's (*Choloepus* sp.) blood in Panama in 1972 and representing the *E. monterogeii* group B in [39]. Promastigotes were cultivated in M199 medium (Sigma–Aldrich, St. Louis, MO, USA) supplemented with 10% heat-inactivated fetal bovine calf serum (Thermo Fisher Scientific, Waltham, MA, USA), 1% Basal Medium Eagle vitamins (Sigma–Aldrich, St. Louis, MO, USA), 2% sterile urine, and 250 µg/mL of amikacin (Bristol-Myers Squibb, New York, NY, USA). The total genomic DNA was isolated from 10 mL of trypanosomatid cultures with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. 18S rRNA and gGAPDH genes were amplified while using primers S762 and S763 [40] and M200 and M201 [41], respectively, following the previously described protocol [42]. PCR amplicons were directly sequenced at MacroGen Europe (Amsterdam, The Netherlands) using internal primers, as described previously [43,44]. The obtained nucleotide sequences were deposited to GenBank under the accession numbers MT862138–MT862140 (18S rRNA) and MT887294–MT887296 (gGAPDH). BLAST analysis confirmed the identity of species under study [45].

### 2.2. Whole-Genome and Transcriptome Sequencing and Annotation

The whole genomes and transcriptomes of *Endotrypanum monterogeii* ATCC 30507, *Porcisia deanei* TCC258 and *P. hertigi* TCC260 were sequenced, as described previously [9] while using the Illumina HiSeq platform. On average, 63 and 46 million 100 nt long paired-end raw reads were produced for genomes and transcriptomes, respectively. The raw reads were trimmed using Trimmomatic v.0.39 [46] with the following settings: ILLUMINACLIP:TruSeq3-PE-2.fa:2:20:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75. The read quality before and after the trimming was checked with FastQC v.0.11.8 [47].

The trimmed genomic reads were assembled *de novo* using SPAdes Genome assembler v.3.13.0 [48] with default settings and automatic k-mer selection (k-mers of lengths 21, 33, and 55 nt were used). The resulting scaffolds were checked for potential contamination with BlobTools v.1.1 [49] and those shorter than 500 nucleotides or showing high-quality BLAST hits at the nucleotide level (identity > 95% and coverage > 85%) to sequences outside Euglenozoa in NCBI database were discarded. Using these criteria, 1008 (348,597 bp), 756 (255,773 bp), and 1987 (673,383 bp) sequences for *E. monterogeii*, *P. deanei* and *P. hertigi*, respectively, were identified as contamination (Figure S1). The quality of the resulting assemblies was assessed using QUAST v.5.0.2 [50]. The genome and transcriptome read mapping was performed with Bowtie2 v.2.3.5.1 using “–end-to-end” and “–very-sensitive” options [51] and HISAT2 v.2.1.0 with “–dta-cufflinks” option [52], respectively. The raw reads and assembled genome sequences were deposited to NCBI database under BioProject

accession numbers PRJNA680236, PRJNA680237, and PRJNA680239 for *E. monterogei* ATCC 30507, *P. deanei* TCC258, and *P. hertigi* TCC260, respectively.

Genome annotation using transcriptome evidence was performed in the web-based program Companion with default options [53], using *L. major* Friedlin as the most closely related available reference. The pseudo-chromosome level sequences produced with Companion software were only used for the purpose of synteny analysis, in all other cases scaffold-level sequences produced by Spades assembler were analyzed. The genome completeness and annotation quality were assessed with BUSCO v.3 using the eukaryota\_odb9 reference database [54].

### 2.3. Repeats Identification and Synteny Analysis

The *de novo* repeat identification was performed using RepeatModeler v.2.0.1 and RepeatMasker v.4.1.0 [55] with the option ‘-species’ set to Euglenozoa. The repeats families were annotated using BLASTX and BLASTN with the *e*-value set to 0.01.

Synteny analysis was completed using SyMAP v.5.0.5 [56] with the following settings: minimum size of sequence to load, 500 bp; minimum number of anchors required to define a synteny block, 7; synteny blocks merged in case of overlaps; and, only the larger block kept if two synteny blocks overlapped on a chromosome. For synteny inferences, the pseudo-chromosomes that were produced by Companion were used with the sequences of *L. major* Friedlin as a reference. The cross-mapping of pseudo-chromosomes was visualized using Chromosomer v.0.1.4 [57].

### 2.4. Genome Coverage Analysis and Ploidy Estimation

The trimmed genomic reads were mapped onto the genome assembly with Bowtie2 [51] using “-end-to-end” and “-very-sensitive” options. The GenomeCov tool from the BEDTools v.2.28.0–33 package [58] was used to calculate the per-base read coverage for the 50 longest scaffolds. The median genome coverage (represented by the 50 longest scaffolds) was calculated using the dplyr package in R v. 3.6.3 [59]. For ploidy estimation, the relative coverage values were obtained by dividing the average coverage of each of the 50 longest scaffold sequences by the average genome coverage. The ploidy was inferred assuming that the majority of the chromosomes are diploid. The coverage plots were visualized using the R v. 3.6.3 packages ggplot2 [60] and tidyverse [61]. WeeSAM v.1.5 [62] was used to obtain the multiple genome coverage statistics that are represented in Table S2.

### 2.5. Variant Calling

After the genomic reads were mapped, as described above, and, prior to variant calling, the read duplicates were removed and the reads were locally realigned using the MarkedDuplicates and IndelRealigner tools of GenomeAnalysisTK v.4.1.4.0 [63] with default settings, except for REMOVE\_DUPLICATES = true. The variant calling was performed using Platypus v.0.8.1 [64] with default settings.

### 2.6. Orthology and Phylogenomic Analyses

The OrthoFinder v.2.3.8 [65] with default settings was used on a dataset of 44 trypanosomatid species with the eubodoniid *Bodo saltans* representing an outgroup in order to infer protein orthology. Out of a total 14,511 orthologous groups (OGs), 522 contained proteins that were encoded by single-copy genes. Out of these, 410 OGs with the average percent identity within the group  $\geq 60\%$  were selected for phylogenomic inferences. The amino acid sequences in each OG were aligned using the L-INS-i algorithm in MAFFT v.7.453 with default settings [66] and trimmed using TrimAl v.1.4 [67] with “-strict”, “-sident”, “-sgc”, and “-sgt” options, and then concatenated. The average protein identity within OGs was assessed using the esl-alistat script v.0.46 from HMMER package [68].

The maximum likelihood phylogenetic tree was inferred in IQ-TREE v.1.6.12 [69] with JTT + F + I + G4 being automatically selected as the best fit model and branch support estimated using 1000 standard bootstrap replicates. For the Bayesian inference,

two independent chains were run in PhyloBayes-MPI [70] for ~16,000 iterations under the JTT + CAT + G model with the removal of invariant sites. The absolute topological convergence was achieved after 300 iterations. For all run parameters at the end of the analysis, the relative differences were below 0.1 and effective sample sizes  $\geq 596$ . The final tree was visualized using the dplyr, ggplot2, and ggtree packages in R v. 3.6.3 [71].

The gene family gains, losses, expansions, and contractions were analyzed with Dollo's and Wagner's (gain penalty set to 3) parsimony algorithms implemented in the COUNT software [72], as described previously [9]. KEGG and Interpro IDs were assigned to the annotated proteins with BlastKOALA [73] and locally installed InterproScan v.5.45-80 with "-dp", "-goterms", and "-pathways" settings [74], respectively. OG intersections were inferred and visualized with UpSetR package in R v. 3.6.3 [75].

Metabolic pathways were analyzed using "all against all" BLASTP searches with an *e*-value cut-off of  $1e^{-50}$ , as described previously [76]. This rather strict *e*-value was chosen in order to distinguish between true orthologous proteins and more distant homologues, which are not necessarily functional orthologues.

### 2.7. Gene Ontology Analysis and Functional Annotation

Gene ontology (GO) identifiers and related GO terms were assigned to the annotated proteins using the InterproScan v. 5.45-80 and QuickGo web server, respectively [77]. When possible, *L. major* proteins were used as representative sequences in these analyses.

### 2.8. Analyses of Amastin Surface Proteins and Biopterin Transporters

Taking the large amastin protein family size in *Leishmania* spp. into account, we restricted our analyses to a subset of ten selected trypanosomatid species/strains: *Crithidia fasciculata*, *E. monterogei* ATCC30507 and LV88, *Leishmania braziliensis*, *Leishmania major*, *Leptomonas pyrrhocoris*, *P. deanei*, *P. hertigi*, *Trypanosoma brucei brucei*, and *Trypanosoma cruzi*. For studying the biopterin BT1 protein family, the whole protein set of 44 trypanosomatids and the outgroup *B. saltans* was used.

We performed the HMM search (HMMER v.3.3.1, [68]) using the amastin (PF07344) and BT1 (PF03092) HMM profiles from Pfam database [78], along with the respective datasets described above. Only hits with *e*-values below  $1e^{-10}$  were kept for further steps. The pairwise identity of the hits was assessed using Clustal Omega 2.1 [79]. For amastins, only the sequences having more than 20% identity to the  $\alpha$ -amastin LmjF.28.1400 of *L. major* Friedlin were kept. Of note, in this filtering step, the proto- $\delta$ -amastin LmjF.34.0970 was formally excluded and, therefore, it is not present on the tree. The same criteria were used to filter BT1 sequences, with the protein identity of the hits being compared to the BT1 of *L. major* Friedlin (LmjF.35.5150). Finally, for both amastins and BT1 transporters, the remaining hits were aligned using the L-INS-i algorithm in MAFFT v.7.453 [66] and only sequences with  $<90\%$  of gaps were kept. These sequences were then re-aligned and trimmed with TrimAl v.1.4 [67] using the option "-gappyout". In total, we identified 239 amastin sequences in our dataset and, after applying the filtering criteria mentioned above, 188 sequences were retained and used in phylogenetic analysis. For putative BT1 transporters, 320 sequences were retained out of 544 initial hits.

The ML trees were built using IQ-TREE v.1.6.12 [69], with 1000 bootstrap replicates, and the best-fit models for amastins and BT1 being WAG + F + G4 and JTT + F + I + G4, respectively. The trees were visualized in FigTree v.1.4.4 [80]. For predicting transmembrane domains (TMD), the protein sequences that were presented in the trees were submitted to the TMHMM Server v. 2.0 [81] with the default settings.

We performed a reconstruction of the sequence similarity-based protein network in order to gain some insight into affiliation of the amastins excluded from the phylogenetic analysis according to the filtering criteria mentioned above. In the case of phylogenetic analysis after application of the abovementioned thresholds, 10 amastins out of 15 were retained for *E. monterogei*, seven out of 10 for *P. deanei*, and five out of eight for *P. hertigi*. The amastin protein network was inferred from a dataset of 237 protein sequences longer

than 100 amino acids using EFI-EST [82] with a BLAST  $e$ -value threshold of  $1e^{-10}$  and a minimum alignment score (roughly corresponding to sequence similarity) set to 30. The result was visualized in Cytoscape v.3.8.0 [83]. In this analysis, only two short sequences were discarded from the original dataset containing 239 HMMER hits, being identified with an  $e$ -value lower than  $1e^{-10}$ . Putative annotations were assigned to the inferred protein clusters based on the results of phylogenetic analysis. Sequences, which were excluded from the phylogenetic analysis by filtering criteria, were annotated based on previously published results [84,85].

### 2.9. Selection Analysis

A subset of six species that includes three investigated *Endotrypanum/Porcisia* spp., *Leishmania major*, *L. tarentolae*, and *Leptomonas seymouri* was used for positive selection analysis. From all OGs, we only selected those that contained sequences of all six species. Tuples of orthologous protein sequences were aligned with MAFFT v.7.453 and multiple alignments were converted into codon alignments using a custom Python script. In order to identify genes under positive selection, a branch-site model A [86] was used for *Endotrypanum/Porcisia* and *Leishmania* branches (two independent tests), while other branches were set as a background. The LRT was used to evaluate whether branch-site model A had a significantly better fit for the codon site with  $\omega > 1$  in comparison with the branch-site model A1, which fixes  $\omega$  to 1.0 on the branches of interest. The analysis was carried out using the ETE3 framework [87]. If positive selection was detected within an OG, a gene of *L. major* was used as a representative sequence for the group. Genes that were under positive selection on the *Endotrypanum/Porcisia* and *Leishmania* branches were subjected to GO enrichment analysis in the top.GO R package [88].

## 3. Results

### 3.1. *Endotrypanum* Sp. ATCC 30507 (MCHO/PA/72/3130) Is *E. monterogeei*

Confirming previous results (Table S2 in [39]), 18S rRNA sequence analysis established the identity of *Endotrypanum* sp. ATCC 30507 (MCHO/PA/72/3130) as bona fide *E. monterogeei*. This name is used hereafter.

### 3.2. General Features of *Endotrypanum* and *Porcisia* Genomes

The three genome assemblies that were obtained herein (Figure S1) had total length and N50 values (given in parentheses) of 30.4 Mb (57 kb) for *E. monterogeei*, 29.5 Mb (30.35 kb) for *P. deanei*, and 29.1 Mb (29.55 kb) for *P. hertigi* (Table S1). They were conspicuously shorter than the reference ~32 Mb genome of *L. major* Friedlin. This can be explained by multiple factors, including differences in real gene content and assembly procedures (different scaffolding methods may result in discordant gap content, some repeats may not completely resolved in the absence of long reads, etc.).

The genomes that were sequenced here were predicted to encode about 7600 proteins on average, which is significantly less than in *L. major* Friedlin (8519), but correlates well with the estimated genome sizes of these species (Table S1). The percentages of missing benchmarking universal single-copy and duplicated orthologs (BUSCOs), which are used to estimate completeness of the assembly and annotation quality, are as low as ~20% and 5% for each of the sequenced genomes, similarly to the respective estimates for the high-quality reference genome of *L. major* Friedlin (19.8% and 6.3%, respectively). Along with the results of the coverage homogeneity analysis (described below), this suggests that most of the repeated regions were properly resolved. A very low proportion of homozygous single nucleotide polymorphisms (SNPs) (around 1%) indicates a minimal number of genome assembly errors (Table S1). The variant calling procedure led to the identification of the highest total SNP number (74,038) in the genome of *P. deanei*, while those of *P. hertigi* and *E. monterogeei* displayed less variation with 58,586 and 40,923 SNPs, respectively (Table S1).

### 3.3. Genome Coverage Analysis, Ploidy Estimation and Synteny Analysis

For the analysis of genome assembly coverage and ploidy estimation, genomic reads were mapped back onto the scaffolds (see Materials and Methods). The coverage is uniform across all three analyzed genome assemblies, with the median numbers being 152, 108, and 112 for *E. monterogeii*, *P. deanei*, and *P. hertigi*, respectively (Figure S2). The per-scaffold average proportion of low-coverage sites (the percentage of sites with coverage  $\leq 0.2$  of the average depth) is small for all three genome assemblies: 1.67% (*E. monterogeii*), 2.42% (*P. deanei*), and 2.82% (*P. hertigi*) (Figure S2, Table S2). In all three species, most of the scaffolds (~96%) have homogeneous coverage with the coefficient of coverage variation below 1 (Table S2). Most of the 50 longest scaffolds in the obtained assemblies are diploid (2n), and just a few have other ploidy levels (3–4n) (Figure S3, Table S2). The highest rate of aneuploidy was detected in *E. monterogeii*, with five out of 50 largest scaffolds demonstrating estimated ploidy over 2n (Figure S3).

We documented variable levels of gene order conservation among analyzed trypanosomatid genomes, with 41 to 85% of genes located within synteny blocks in the various intra- and interspecies comparisons (Figure S4, Table S3). These numbers are similar to the estimates for other Leishmaniinae, and they are consistent with the majority of trypanosomatid genes located within relatively well-conserved polycistronic transcription units [9,89,90].

### 3.4. Analysis of Repetitive Sequences

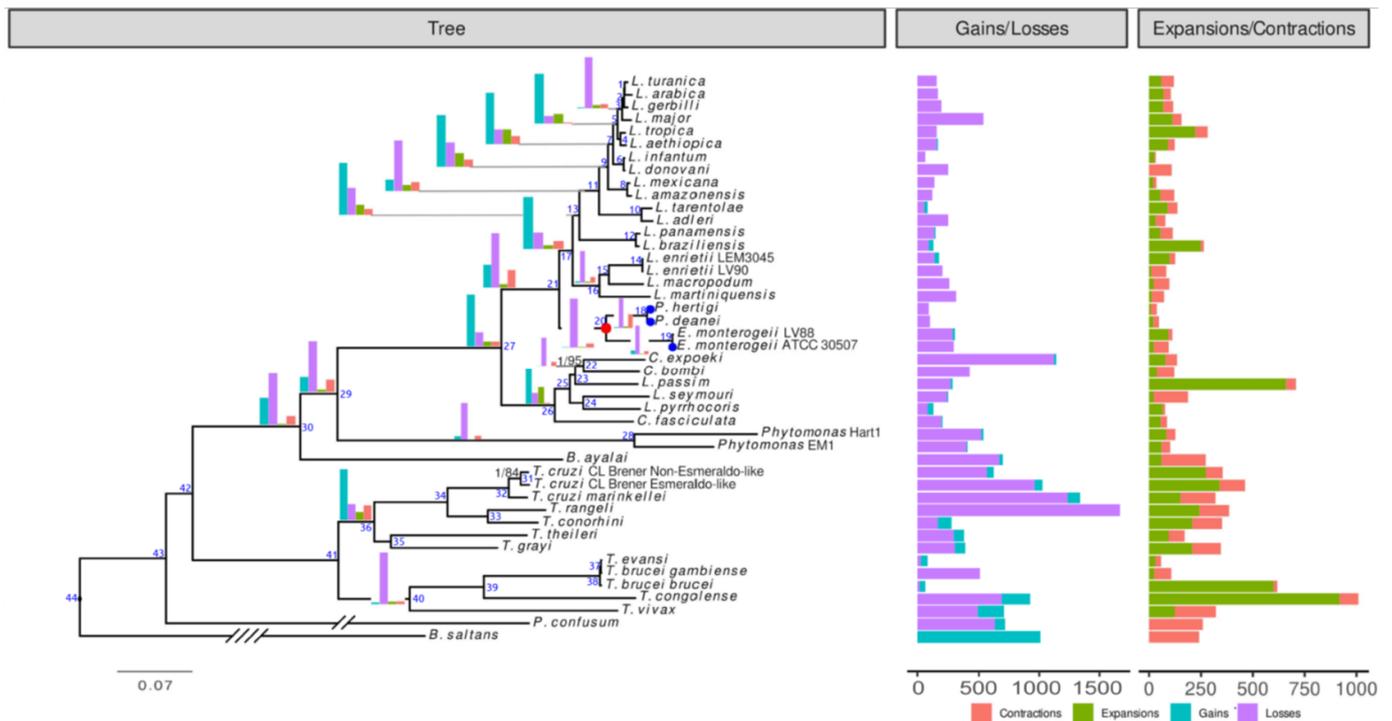
Twenty seven families of repeats spanning 3.66% (~1.1 Mb) of the genome assembly were identified in the *E. monterogeii* genome. Out of these, 0.3% are low complexity repeats. *Porcisia deanei* and *P. hertigi* have 40 and 45 families of repeats, covering 4.22% and 4.49% of their genomes (Table S4), with 0.52% and 0.58% of low complexity repeats, respectively. Even though *L. major* Friedlin has a higher number of identified repeat families (321), the genomic spanning of these repeats is also comparable 3.66%, from which 0.38% are low complexity repeats. For most of the identified repetitive sequences (including species-specific groups of repeats), no functional annotation could be inferred (Table S4). Among the annotated families of repetitive sequences, the majority contain surface antigens (leishmanolysin GP63 and protease GP46, GP stands for a glycoprotein), as well as serine/threonine-protein phosphatases, which possibly play a role in cell division and the modulation of host immune response [91–95].

### 3.5. Gene Family Sharing Analysis

Annotated proteins of 44 trypanosomatids and *B. saltans* (Table S5) cluster into 14,511 orthogroups (OGs) that contain at least two sequences. OG sharing analysis (group composition is presented in Table S6) shows that 1650 OGs (11.4% of the total OG number), incorporating mostly housekeeping genes, are shared among all kinetoplastid groups in our dataset (Figure S5, Table S7). The analysis of OGs that are uniquely shared among various representatives of the *Endotrypanum/Porcisia* clade revealed several dozen of OGs incorporating proteins, to which function could not be assigned with confidence (Table S7). The same analysis performed only for the representatives of the *Endotrypanum/Porcisia* clade led to the identification of a large set of 6764 OGs that were shared by all four species/strains (Figure S5, Table S7), which is in agreement with the high synteny levels for genomes of these species.

### 3.6. Phylogenomic Analysis

The maximum likelihood and Bayesian phylogenomic trees inferred using the supermatrix of 410 proteins encoded by single-copy genes have the same topology and demonstrate maximal supports for almost all branches (Figure 1). This topology is compatible with those inferred previously [2,9,96,97], which confirms the position of the genera *Endotrypanum* and *Porcisia* as the closest known relatives of the genus *Leishmania*.



**Figure 1.** Phylogenomic tree based on 410 proteins encoded by single-copy genes from 44 trypanosomatids and the eubodid *Bodo saltans*. Posterior probabilities and bootstrap supports are shown (in black) only if the latter is <100%. The scale bar represents substitutions per site. The numbers of orthologous groups (OG) gained/lost/expanded/contracted at certain nodes and leaves (species) are depicted using bar plots placed at the nodes and on the right of the tree, respectively (see Table S8 for exact counts; node numbers indicated in blue correspond to those in the Table S8). The *Endotrypanum/Porcisia* node (node 20) and the isolates sequenced in this study are marked with red and blue circles, respectively. The length of *B. saltans* and *P. confusum* branches was reduced four- and two-fold, respectively, for visualization purposes.

### 3.7. Evolution of Gene Families

Aiming at elucidating gene content differences between *Endotrypanum/Porcisia* and other trypanosomatids, we performed a genome-wide analysis of gene content with the emphasis on genes and genes families gained/lost/expanded/contracted at the *Endotrypanum/Porcisia* branch (node 20 in the Figure 1), revealing evolutionary changes on this branch as compared to other Leishmaniinae. In addition, we systematically examined the differences in metabolic pathways between *Endotrypanum/Porcisia* and *L. major* (below). The *Endotrypanum/Porcisia* node is characterized by the prevalence of gene family losses and contractions over gains and expansions (node 20 in the Figure 1, Tables S8–S10), 150 and seven-fold, respectively. This is reminiscent of the situation that was inferred for the subgenus *Leishmania (Mundinia)* [9]. No functional annotation could be confidently assigned to the OGs gained and expanded at the ancestral *Endotrypanum/Porcisia* node (Tables S9 and S10). Among OGs lost and contracted at this node, ~70% and 37% of families, respectively, are represented by hypothetical proteins. We also analyzed the genus-specific changes in the gene family repertoire, focusing on the evolutionary events at the *Porcisia* and *Endotrypanum* nodes (Figure 1, nodes 18 and 19, respectively). Similar to the ancestral *Endotrypanum/Porcisia* node, both of the nodes are dominated by gene family losses,

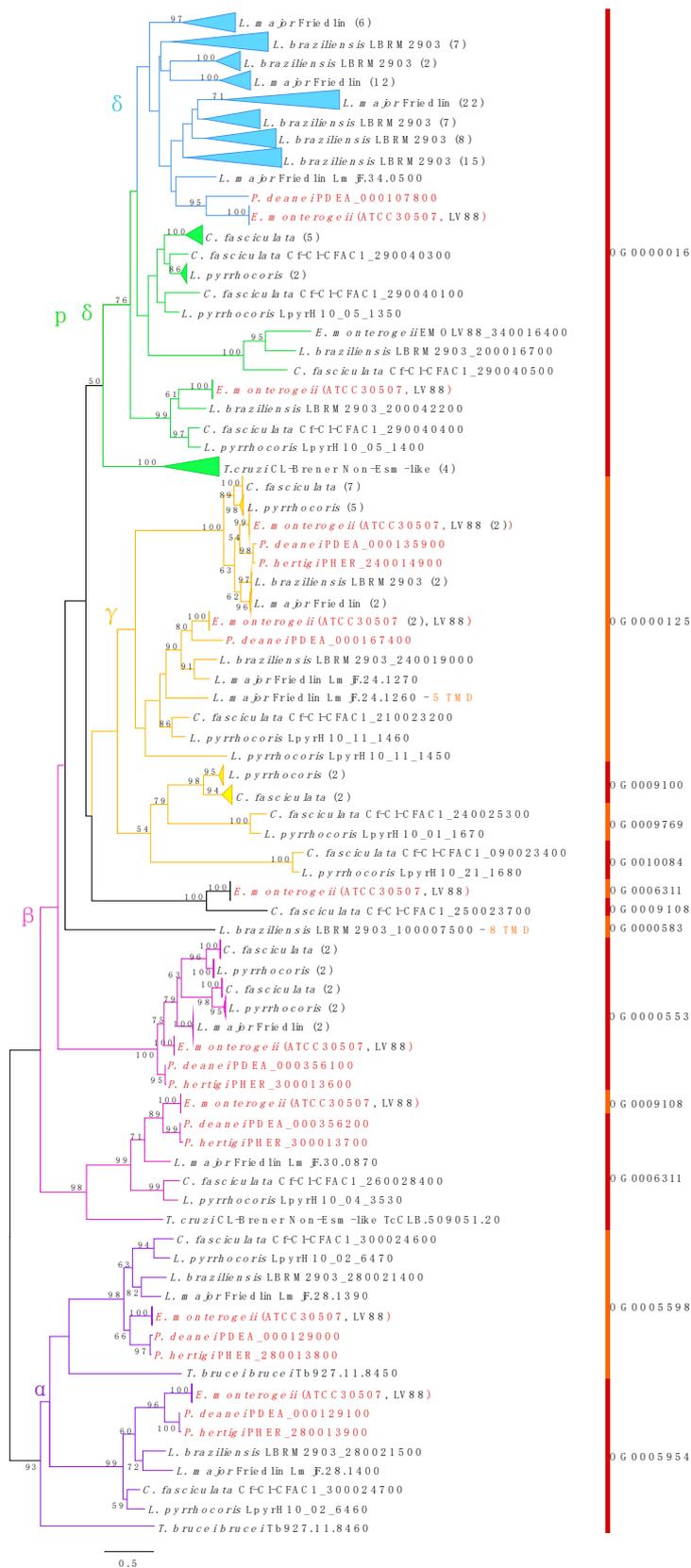
with the majority of proteins having no functional annotation. Several protein families at this node have undergone noticeable evolutionary changes in their composition and size, including membrane proteins (transporters, cell surface proteins), proteins involved in cell signaling (kinases, phosphatases, GTPases, adenylate cyclase-like proteins), subtilisins and peptidases, families of housekeeping genes encoding motor proteins (actin, dynein, myosin, and kinesin), as well as ribosomal and DNA repair proteins (Table S9). Out of these, we analyzed, in detail, amastins and biopterin transporter family BT1, the two protein families displaying the most significant changes and, at the same time, playing a key role in host-parasite interactions and triggering the host immune system response [98]. Changes in the repertoire of these proteins may represent an adaptive mechanism for the successful evasion of the host immune system and be associated with lower pathogenicity.

### 3.8. Amastins

Amastins are a large family of transmembrane glycoproteins (GPs) that are widely conserved across trypanosomatids and expressed mainly during the amastigote stage of their life cycle [84,85,99]. These GPs are among the most immunogenic surface antigens in *Leishmania*, enabling parasites to invade host cells and provide other advantages, such as fast and efficient response to the changes of physiological conditions inside macrophages [100]. The number of genes encoding putative amastins vary across *Leishmania* spp., with the highest counts being documented for the representatives of *Leishmania* and *Viannia* subgenera, such as *L. infantum* (68 proteins), *L. major* (63 proteins), and *L. braziliensis* (66 proteins) (Tables S11 and S12). In the representatives of the *Endotrypanum/Porcisia* clade, there are from 8 (*P. hertigi*) to 15 (*E. monterogeii*) amastin domain-containing proteins, which is even less than in *Leishmania (Mundinia)* spp. [9]. Of note, in all of these proteins, three to four transmembrane domains (TMDs) were identified (Table S12).

The amastins repertoire also varies across Leishmaniinae (Figure 2). Based on phylogeny, expression pattern, and secondary structure, these proteins are classified into four subfamilies— $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -amastins (including proto- $\delta$ -amastins) [85]. While the repertoires of  $\alpha$ - and  $\beta$ -amastins are highly conserved across *Leishmania*, *Endotrypanum*, *Porcisia*, and even monoxenous representatives of the subfamily Leishmaniinae, *P. hertigi* contains a slightly reduced set of  $\gamma$ -amastins, and lacks detectable homologues of proto- $\delta$  and  $\delta$ -amastins (Figure 2).

The repertoire of  $\delta$ -amastins is substantially expanded in *Leishmania* as compared to *Endotrypanum* and *Porcisia*. Because some amastin domain-containing proteins that were initially identified by homology-based searches were discarded from phylogenetic analysis based on set threshold (see Materials and Methods for details), we estimated their affinity to known amastin subfamilies by a similarity-based sequence clustering approach using the unfiltered dataset (Figure S6). The composition of the inferred protein clusters strongly corresponds to that of the clades on the amastin phylogenetic tree (Figure 2). Almost all amastin domain-containing proteins of *Endotrypanum/Porcisia* that were excluded from the phylogenetic analysis cluster with divergent sequences, which were previously annotated in other Leishmaniinae as putative  $\beta$ -amastins (Figure S6). The exceptions are the two amastin domain-containing proteins of *E. monterogeii* ATCC30507: one is a putative divergent proto- $\delta$  amastin (EMON\_000317000.1), while, for the other, no affiliation could be established (EMON\_000357800.1), similarly to the one of the *P. hertigi* sequences (PHER\_000076200.1).



**Figure 2.** Maximum-likelihood phylogenetic tree of 188 kinetoplastid amastins. Only bootstrap supports over 50% are shown. The sequences obtained in this study are shown in red with the respective OG IDs. The five classes of amastins are highlighted in different colors. Most analyzed proteins have four transmembrane domains (TMDs), with a few exceptions indicated in the tree and Table S12. Numbers of sequences within collapsed clades are shown in brackets.

### 3.9. Biopterin Transporter BT1

The biopterin transporters (BT) are integral membrane proteins [101] of the major facilitator superfamily [102]. The BT1 is a high-affinity biopterin transporter and a low-affinity folate transporter. It is the only non-conjugated pterin transporter and the main biopterin transporter in *Leishmania* [103]. All of the trypanosomatid species with studied metabolism are biopterin auxotrophs [102,104]. Biopterin is a co-factor of endogenous enzymes known to play a role in the parasite's differentiation and growth. All of the studied *Leishmania* spp. not only possess pterin salvage pathways and pterin transporters, but also the highest number of BT1 family members among trypanosomatids [104]. The BT1 proteins are known to contain 10 to 12 putative TMDs, which are predicted to form amphiphilic  $\alpha$ -helices and  $\beta$ -strands, involved in the formation of aqueous channels across the lipid membrane [105]. All of the kinetoplastid BT1 transporters analyzed here encode six to 24 TMDs, with the majority possessing 12 TMDs (Figure S7, Table S13). Of note, the proteins having similar number of TMDs tend to cluster together, which likely reflecting their shared evolutionary history.

The OG gain/loss analysis showed moderate BT1 repertoire changes in the *Endotrypanum/Porcisia* clade (Table S9). According to the phylogenetic analysis, some BT1 orthologues of *L. (Leishmania)* spp. proteins are absent in the *Endotrypanum/Porcisia* spp. (Figure S7). The distribution of the proteins of this family among kinetoplastids suggests that their diversification happened several times during the trypanosomatid evolution (e.g., in the common ancestor of trypanosomes, in that of *Crithidia/Leptomonas* clade, *Phytomonas*, and *Crithidia* spp.). In addition, such events also happened in the common ancestor of Leishmaniinae, as judged by the presence of multiple BT1-encoding genes in the representatives of this subfamily and the absence of closely related sequences in trypanosomes. The members of the *Endotrypanum/Porcisia* clade apparently secondarily lost some of the BT1 homologues, which are present in other Leishmaniinae (Figure S7). Functional studies are needed to shed light on the role of the reduced BT1 repertoire in *Endotrypanum/Porcisia* spp. and whether it plays a role in their pathogenicity.

### 3.10. Notes on Metabolism of *Endotrypanum* and *Porcisia*

The metabolic capacities of *Leishmania* spp. have been reviewed elsewhere [76,106–108], and they may serve as a reference for the interpretation of the *Endotrypanum/Porcisia* proteomes. The major differences between *L. major* and species under study rest in the absence of many amastin-like genes (discussed above) and hypothetical proteins. Apart from the fact that complete gene families have been missing or reduced to a few gene copies in the *Endotrypanum/Porcisia* clade, the metabolic arsenal of these flagellates is generally similar to that of *L. major*. The main differences include the prominent absence of genes for methionine synthase, methionine synthase reductase, methylmalonyl-CoA epimerase, and methylmalonyl-CoA mutase from the genomes of all three analyzed species. This suggests that *Endotrypanum* and *Porcisia* cannot use the two branched amino acids, Ile and Val, as well as Met, for energy production and gluconeogenesis, because their common degradative intermediate, propionyl-CoA, cannot be converted to succinyl-CoA. On the other hand, the absence of a gene for methionine synthase does not mean that these species are auxotrophic for Met, since all Leishmaniinae possess a second methionine synthase isofunctional enzyme [107]. The enzymes of the methionine salvage pathway are present in all trypanosomatids [76]. The genomes of *Endotrypanum*, *Leishmania*, and *Porcisia* encode a methylthioadenosine phosphorylase that compensates for the loss of a 5-methylthioribose kinase, while monoxenous Leishmaniinae (*C. fasciculata* and *L. pyrrocoris*) have genes for both of the enzymes.

Leishmaniinae (*L. major*, *C. fasciculata*, *L. pyrrocoris*) possess two asparaginase isoenzyme genes (orthologues of *LmjF.15.0390* and *LmjF.36.4430*), allowing them to utilize Asn as an energy source. Both isoenzymes usually carry the peroxisomal C-terminal targeting signal (PTS1). Interestingly, the analyzed *Endotrypanum/Porcisia* spp. have lost one of the two genes (orthologue of *LmjF.15.0390*), while the remaining one now also lacks the PTS1.

Within Leishmaniinae, only *C. fasciculata* and *L. pyrrocoris* have the capacity to transform the bacterial amino acid diaminopimelate into Lys, owing to the acquisition of diaminopimelate decarboxylase and diaminopimelate epimerase genes [89]. The absence of these two genes that were previously reported for *Leishmania* spp. is now also extended for *Endotrypanum/Porcisia* spp. As in all *Leishmania* [109], the gene encoding catalase was not retained in the *Endotrypanum/Porcisia* genomes, further supporting the hypothesis of its incompatibility with the dixenous life cycle [110].

No differences between *Leishmania* and *Endotrypanum/Porcisia* with respect to their capacity to synthesize sugar nucleotides were detected, and pools of GDP-Ara, UDP-Fuc, UDP-GlcNAc, GDP-Man, UDP-Glc, UDP-Galp, and UDP-Galf are predicted to be available for the incorporation of the respective sugar residues into glycoproteins on the surface of these flagellates [111]. However, the absence of  $\beta$ -galactofuranosyl/glycosyltransferase in *P. hertigi* and UDP-glucuronosyl and UDP-glucosyl transferase from genomes of all three analyzed species indicates differences in the surface glycoprotein composition between members of this clade and *L. major*.

### 3.11. Selection Analysis

In total, 5901 OGs were found to contain genes from a selected subset of species. Applying a branch-site model A, we identified 280 and 169 genes with positively selected sites ( $p$ -value below 0.05) on the branches, leading to the *Endotrypanum/Porcisia* and *Leishmania* clades, respectively. Only 19 genes were in the intersection of these two sets. The most commonly found annotations in these lists were ‘hypothetical protein’ and ‘protein containing domain with unknown function’. However, some well-described proteins were also present, e.g., ornithine decarboxylase implicated in the survival of *Leishmania* amastigotes [112] and ascorbate peroxidase, which is essential for the defense against oxidative stress [113]. No considerable GO enrichment for the positively selected genes was documented in the categories ‘biological process’ or ‘molecular function’ on any of the two branches (Figure S8), which indicated that positive selection pressure affects functionally diverse genes. However, the enrichment was revealed in the ‘cellular component’ category for proteins with ‘cilium’ and ‘cell projection’ annotations. These are membrane-bound, intermembrane, and excreted proteins that are commonly involved in host-parasite interactions [114].

## 4. Discussion

The genome analysis of *Endotrypanum* and *Porcisia* performed here revealed that these parasites and their morphologically indistinguishable closest phylogenetic relatives, *Leishmania* spp., followed different evolutionary paths, resulting in distinct biology. Although we identified specific sets of genes under positive selection in these two lineages, possibly reflecting their adaptation to different hosts, the number of such genes is rather small. Meanwhile, gene gains and losses, as well as gene family expansions and contractions, show stronger signals in both lineages, indicating that these were the main mode of genome evolution in dixenous Leishmaniinae.

The differences in the metabolic capabilities of *Endotrypanum/Porcisia* spp. and those of *L. major*, as a representative of the genus *Leishmania*, are rather subtle and consist mainly of the repertoire of enzymes participating in the metabolism of amino acids and biosynthesis of surface glycoproteins. More substantial changes can be seen in gene families incorporating hypothetical proteins, about which no definite conclusions can be drawn, as well as in those containing genes encoding members of large gene families, such as membrane proteins, proteins that are involved in cell signaling, parasite–host interaction, and even several families of housekeeping genes (e.g., those encoding motor and DNA repair proteins). Out of these, we identified two protein families, amastins and bioplerin transporters BT1, which we find to be at least partially responsible for the differences in pathogenicity between *Endotrypanum/Porcisia* and *Leishmania* spp. based on the drastically different evolutionary patterns of these proteins in the two lineages.

Amastins are transmembrane glycoproteins present on the cell surfaces of all trypanosomatids. In *Leishmania*, with up to ~70 members, the amastins represent the largest developmentally regulated gene family reported so far [100]. These proteins were first identified in *T. cruzi* [115], and they all share similar structural organization with an extracellular domain, several transmembrane segments and an amastin domain. For the majority of amastins, the expression is amastigote-specific and strictly dependent on acidic pH [116,117]. These proteins serve as membrane transporters that are essential for the survival inside the vertebrate cell or as signal transducers allowing for sensing the lysosomal acidic milieu. Amastins are among the most immunogenic leishmanial surface antigens for mice [118] and solicit strong immune responses in humans, which makes these proteins promising vaccine candidates [119]. The amastin repertoire is expanded in *Leishmania* spp. relative to that in other trypanosomatids. The proteins are encoded by a diverse gene family, including four subfamilies ( $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -amastins), which have distinct genomic positions and diverged already in an ancestral trypanosomatid [85,120]. In *Leishmania* spp., the group of  $\delta$ -amastins rapidly diversified even further, while such diversification never happened in other trypanosomatids, including the *Endotrypanum/Porcisia* lineage. As a case of extreme reduction, *P. hertigi* lacks detectable homologues of  $\delta$ -amastins (including divergent sequences of the proto- $\delta$  group). Unfortunately, we could not obtain high supports at many important branches of the phylogenetic tree due to a short length of the amastin sequences (most are only 180–210 amino acids long). However, given the amastin distribution patterns that are observed here and in previous studies [84,85,121], we can make certain assumptions regarding the evolutionary history of this protein family in Euglenozoa. The amastin domain-containing proteins with unknown function were already present in the common ancestor of Euglenozoa, as suggested by the presence of the respective homologues in the representatives of all main euglenozoan groups that were analyzed in this respect, except *Perkinsella*, an extremely reduced symbiont of amoebae [84]. As mentioned above, the phylogenetic analysis of amastins incorporating the data from all sequenced genomes is a challenging task, due to the large size of this family and a relatively short protein length. Still, the available data suggest that diversification into  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and proto- $\delta$ -amastins had not happened later than in the common ancestor of the human pathogenic genera *Leishmania* and *Trypanosoma* (node 42 in Figure 1). The  $\delta$ -amastin subfamily is apparently *Leishmania*-specific, since no obvious homologues of these sequences were identified with confidence in other trypanosomatids, neither by phylogenetic analysis, nor by using similarity-based protein clustering approach (Figure 2, Figure S6). These genes developed from ancestral proto- $\delta$ -amastins and they significantly diversified in the common ancestor of *Leishmania*, while, in *Endotrypanum/Porcisia*, they remained scarce [121]. The evolution of the subgenera *L. (Leishmania)* and *L. (Viannia)* was accompanied by a further diversification of  $\delta$ -amastins, as judged by the presence of specific clades on the phylogenetic tree of amastins for *L. major* and *L. braziliensis*. In *L. (Sauroleishmania)*, the repertoire of  $\delta$ -amastins was secondarily reduced to only two genes [122]. These parasites reside in the bloodstream. Amastigotes (either free or inside monocytes or erythrocytes) are rarely observed and infections are principally detected by culture [123]. To date, there is no evidence that the flagellates seen in the intestine and cloaca of some lizards are *L. (Sauroleishmania)* [124]. The view that the expansion of  $\delta$ -amastin in *Leishmania* was associated with adaptation of the amastigote to the life in vertebrate macrophages [85] is now further supported, since not only *L. (Sauroleishmania)*, but also *Endotrypanum/Porcisia*, which do not infect macrophages, possess a very limited diversity of  $\delta$ -amastins. Thus, a limited repertoire of  $\delta$ -amastins in both *L. (Sauroleishmania)* and *Endotrypanum/Porcisia* is connected to the inability of these pathogens to infect host macrophages. However, while in the ancestor of *Endotrypanum/Porcisia*, the  $\delta$ -amastin family was never expanded, a rather limited set of these proteins in *L. (Sauroleishmania)* is likely a result of secondary losses. The results of our phylogenetic analysis are in agreement with the experimental data showing that the knockdown of  $\delta$ -amastin in *L. braziliensis* affects the parasite-macrophage interaction and results in impaired viability of intracel-

lular amastigotes, which certifies this protein as a virulence factor [125]. Species-specific differences among macrophage-infecting species may be explained by multiple factors, such as the vertebrate host species, the infected macrophage type, which can, in turn, even be affected by the composition of the insect vector's saliva [126,127]. The phylogenetic distribution of the representatives of other amastin subfamilies suggests that the respective proteins might be functionally significant in the vector, or both vector and host.

The repertoire of biopterin transporters is also narrower in the *Endotrypanum/Porcisia* clade as compared to *L. (Leishmania)*, but experimental approaches have to address the potential contribution of this feature to the reduced pathogenicity of these parasite. We speculate that, since these proteins are associated with cell differentiation, *Endotrypanum* and *Porcisia* were not forced to develop very precise and diverse mechanisms for this process, as were *Leishmania* spp., which have one of their life cycle stages confined to host macrophages and they demonstrate pronounced antagonistic relationship with the host. While the details of the life cycles and the respective roles of BT1 transporters at its different stages remain to be elucidated for *Endotrypanum* and *Porcisia*, the role of BT1 transporters in several *Leishmania* spp. was clearly connected to survival and growth inside the host macrophages [128]. *L. donovani* cells overexpressing the BT1 gene demonstrated increased infectivity and survival in the macrophages, with the opposite effect being observed in the knock-out cell line [128]. We suggest that, similar to the situation observed for amastins, *Endotrypanum* and *Porcisia* spp. do not require an elaborate repertoire of BT1 transporters, as do macrophage-dwelling *Leishmania*.

In sum, our genomic analysis of *Endotrypanum* and *Porcisia* spp. allows for a better understanding of the evolutionary trajectories within the dixenous Leishmaniinae and the potentially critical role of the two protein families, amastins and biopterin transporters BT1, in the biology of trypanosomatids.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2073-4425/12/3/444/s1>, Figure S1: BlobTools statistics for *Endotrypanum monterogeei* ATCC 30507, *Porcisia deanei* TCC258, and *P. hertigi* TCC260 before and after filtering. Figure S2: Distribution of genomic read coverage for the genome assemblies of *E. monterogeei* ATCC30507 (red), *P. deanei* (blue) and *P. hertigi* (green). Figure S3: The distribution of 50 longest scaffolds of *E. monterogeei* ATCC30507, *P. deanei*, and *P. hertigi* according to the estimated ploidy levels. Figure S4: (A)–(J): Schematic representation of the two-way synteny between the genomes of *Endotrypanum monterogeei* ATCC305007, *Porcisia deanei* TCC258 and *P. hertigi* TCC260 and the reference genome of *L. major* Friedlin. Inverted synteny blocks are in green, direct ones are in red. Only pseudo-chromosome level scaffolds (produced using Companion software) carrying regions of synteny to the respective regions of the reference genomes are shown. (K) The table summarizing the synteny statistics for each pairwise comparison. Figure S5: OGs sharing among kinetoplastids. (A) UpSet plot for the whole dataset of 44 trypanosomatids and the eubodonid *B. saltans* (the species were grouped in accordance with their phylogenomic position, see Table S06 for details). The Y-axis represents the intersection size (the number of shared OGs) and the X-axis shows the groups/species being compared. The clade *Endotrypanum/Porcisia* is highlighted in pink and the species sequenced in this study are in bold. Red dots indicate OGs uniquely shared among *Endotrypanum* and/or *Porcisia* and unique OGs to each species of this clade. (B) Results of the OG sharing analysis involving only on the *Endotrypanum/Porcisia* group. Figure S6: A sequence similarity network of amastin surface proteins. The network was inferred from a dataset of 237 sequences longer than 100 amino acids using EFI-EST [82] with a BLAST *e*-value threshold of  $10^{-10}$  and a minimum alignment score set to 30. Nodes are color-coded according to the amastin type as follows: putative  $\alpha$  amastins are in violet,  $\beta$  in magenta, proto- $\delta$  in green and  $\delta$  in blue. Sequences not included in phylogenetic analysis due to exclusion criteria (see Materials and Methods), but present in the network, are in pink. When possible, they are putatively annotated based on the phylogenetic analyses (Figure 2 and [121]). Species abbreviations before the protein IDs are as follows: *Crithidia fasciculata* (CFAC1), *E. monterogeei* ATCC30507 (EMON), *E. monterogeei* LV88 (EMOLV88), *Leishmania braziliensis* LBRM2903 (LBRM2903), *L. major* Friedlin (LmjF), *Leptomonas pyrrocoris* (LpyrH10), *P. deanei* (PDEA), *P. hertigi* (PHER), *Trypanosoma brucei brucei* (Tb927) and *T. cruzi* CL-Brener Non-Esmeraldo-like (TcCLB). Figure S7: Maximum-likelihood phylogenetic tree of 320 trypanosomatid BT1 proteins. Only bootstrap supports over 50% are shown. The isolates sequenced and analyzed

in this study are shown in red. The numbers of predicted transmembrane domains (TMDs) are indicated beside the protein IDs using different colors (see also Table S13). Figure S8: GO enrichment of positively selected genes. Bar plot showing the GO enrichment of positively selected genes on *Leishmania* (169 genes) and *Endotrypanum/Porcisia* (280 genes) branches. The number of annotated sequences was log-transformed in order to optimize visualization, thus, all the GOs with zero sequence in this plot, contain one sequence. Only GO enrichment with a Fisher  $p$ -value below 0.05 are shown. Table S1: Statistics of the genome and whole-transcriptome sequencing for *Endotrypanum monterogeii* ATCC 30507 (this work), *Porcisia deanei* TCC258 (this work), *P. hertigi* TCC260 (this work), and *Leishmania major* Friedlin (TriTrypDB). Table S2: Statistics of genome assembly coverage. Table S3: (A) Statistics of cross-mapping of scaffolds of *Endotrypanum monterogeii* ATCC 30507 (this work), *Porcisia deanei* TCC258 (this work), *P. hertigi* TCC260 (this work), and *Leishmania major* Friedlin (TriTrypDB); (B) Statistics of synteny analysis based on the pseudo-chromosome-level scaffolds of *E. monterogeii*, *P. deanei* and *P. hertigi* produced using Companion software. Genome sequence of *L. major* Friedlin was used as a reference. Table S4: Repeated sequences in the analyzed genome assemblies. The upper table contains overall statistics, the bottom one lists identified repeat families. Table S5: Dataset used in the phylogenomic analysis. Species, whose genomes and transcriptomes were sequenced in this work, are in bold. Table S6: Grouping criteria used for OG sharing analysis. Table S7: Functional annotation of proteins belonging to OGs shared among various trypanosomatids groups. Table S8: Counts for gene family gains/losses/expansions/contractions for the branches and nodes of the phylogenomic tree. Table S9: Gene family gains and losses for the species under study. Table S10: Gene family expansions and contractions for the species under study. Table S11: Pairwise identity matrix of 239 amastin domain-containing proteins in trypanosomatids. Sequences removed from further analysis (see Materials and Methods for exclusion thresholds) are highlighted in red. Table S12: Composition of collapsed clades in the amastin phylogenetic tree. The sequences are listed in the same order as in the tree. Table S13: Composition of collapsed clades on the BT1 tree. The sequences are listed in the same order as in the tree. The lower table contains a list of species abbreviation used on the BT1 tree.

**Author Contributions:** Conceptualization: V.Y., A.B.; Data curation: A.B.; Formal analysis: A.T.S.A., E.S.G., F.R.O.; Funding acquisition: V.Y., J.L., P.V.; Investigation: A.T.S.A., F.R.O., E.S.G., A.B.; Methodology: A.B., E.S.G.; Project administration: V.Y., A.B.; Resources: J.J.S., J.S., P.V.; Supervision: V.Y., P.V., J.L., A.B., E.S.G.; Validation: A.B., E.S.G., A.Y.K.; Visualization: A.T.S.A., E.S.G., A.B.; Writing—original draft: V.Y., A.B., A.T.S.A., J.L., P.V.; Writing—review & editing: all co-authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the European Regional Funds (project “Centre for Research of Pathogenicity and Virulence of Parasites” CZ.02.1.01/16\_019/0000759 to A.Y.K., J.L., J.S., and P.V.), the Grant Agency of the Czech Republic (grants 20-07186S and 21-09283S to J.L. and V.Y.), the Russian Science Foundation (grant 19-15-00054 to V.Y. and E.S.G., analysis of amastins and phylogenomics), ERC CZ (grant LL1601 to J.L.), the University of Ostrava (grant SGS/PrF/2021 to A.T.S.A.), Moravskoslezský kraj research initiative (grant RRC/02/2020 to A.T.S.A.), the State assignment for the Zoological Institute AAAA-A19-119031200042-9 to A.Y.K., and the de Duve Institute (to F.R.O.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Raw reads and assembled genome sequences were deposited to NCBI database under BioProject accession numbers PRJNA680236, PRJNA680237, and PRJNA680239 for *E. monterogeii* ATCC 30507, *P. deanei* TCC258 and *P. hertigi* TCC260, respectively.

**Acknowledgments:** We thank members of our laboratories for stimulating discussions.

**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. 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](#)] [[PubMed](#)]
2. 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](#)] [[PubMed](#)]
3. Lukeš, J.; Skalický, T.; Týč, J.; Votýpka, J.; Yurchenko, V. Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* **2014**, *195*, 115–122. [[CrossRef](#)] [[PubMed](#)]
4. Kostygov, A.Y.; Yurchenko, V. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol.* **2017**, *64*, 020. [[CrossRef](#)] [[PubMed](#)]
5. Jirků, M.; Yurchenko, V.; 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](#)] [[PubMed](#)]
6. Bruschi, F.; Gradoni, L. *The Leishmaniasis: Old Neglected Tropical Diseases*; Springer: Cham, Switzerland, 2018; p. 245. [[CrossRef](#)]
7. Akhoundi, M.; Downing, T.; Votýpka, J.; Kuhls, K.; Lukeš, J.; Cannet, A.; Ravel, C.; Marty, P.; Delaunay, P.; Kasbari, M.; et al. *Leishmania* infections: Molecular targets and diagnosis. *Mol. Asp. Med.* **2017**, *57*, 1–29. [[CrossRef](#)]
8. Espinosa, O.A.; Serrano, M.G.; Camargo, E.P.; Teixeira, M.M.; Shaw, J.J. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. *Parasitology* **2018**, *145*, 430–442. [[CrossRef](#)]
9. 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](#)]
10. Coughlan, S.; Taylor, A.S.; Feane, E.; Sanders, M.; Schonian, G.; Cotton, J.A.; Downing, T. *Leishmania naiffi* and *Leishmania guyanensis* reference genomes highlight genome structure and gene evolution in the *Viannia* subgenus. *R. Soc. Open Sci.* **2018**, *5*, 172212. [[CrossRef](#)]
11. Coughlan, S.; Mulhair, P.; Sanders, M.; Schonian, G.; Cotton, J.A.; Downing, T. The genome of *Leishmania adleri* from a mammalian host highlights chromosome fission in *Sauroleishmania*. *Sci. Rep.* **2017**, *7*, 43747. [[CrossRef](#)] [[PubMed](#)]
12. Valdivia, H.O.; Reis-Cunha, J.L.; Rodrigues-Luiz, G.F.; Baptista, R.P.; Baldeviano, G.C.; Gerbasi, R.V.; Dobson, D.E.; Pratloug, F.; Bastien, P.; Lescano, A.G.; et al. Comparative genomic analysis of *Leishmania* (*Viannia*) *peruviana* and *Leishmania* (*Viannia*) *braziliensis*. *BMC Genom.* **2015**, *16*, 715. [[CrossRef](#)]
13. Peacock, C.S.; Seeger, K.; Harris, D.; Murphy, L.; Ruiz, J.C.; Quail, M.A.; Peters, N.; Adlem, E.; Tivey, A.; Aslett, M.; et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.* **2007**, *39*, 839–847. [[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. Cunha, A.M.; Muniz, J. Pesquisas sobre o *Endotrypanum schaudinni* Mesnil e Brimont, 1908, parasita do *Choloepeus didactylus* (L.). *Mem. Do Inst. Oswaldo Cruz* **1944**, *41*, 179–193. [[CrossRef](#)]
17. 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](#)]
18. Shaw, J.J. *The Haemoflagellates of Sloths*; H. K. Lewis: London, UK, 1969; p. 132.
19. Kreutzer, R.D.; Corredor, A.; Grimaldi, G., Jr.; Grogl, M.; Rowton, E.D.; Young, D.G.; Morales, A.; McMahon-Pratt, D.; Guzman, H.; Tesh, R.B. Characterization of *Leishmania colombiensis* sp. n. (Kinetoplastida: Trypanosomatidae), a new parasite infecting humans, animals, and phlebotomine sand flies in Colombia and Panama. *Am. J. Trop. Med. Hyg.* **1991**, *44*, 662–675. [[CrossRef](#)]
20. Zeledón, R.; Ponce, C.; Murillo, J. *Leishmania herreri* sp. n. from sloths and sandflies of Costa Rica. *J. Parasitol.* **1979**, *65*, 275–279. [[CrossRef](#)] [[PubMed](#)]
21. Delgado, O.; Castes, M.; White, A.C., Jr.; Kreutzer, R.D. *Leishmania colombiensis* in Venezuela. *Am. J. Trop. Med. Hyg.* **1993**, *48*, 145–147. [[CrossRef](#)] [[PubMed](#)]
22. 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. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2003**, *3*, 119–124. [[CrossRef](#)]
23. Herrer, A. *Leishmania hertigi* sp. n., from the tropical porcupine, *Coendou rothschildi* Thomas. *J. Parasitol.* **1971**, *57*, 626–629. [[CrossRef](#)] [[PubMed](#)]
24. Lainson, R.; Shaw, J.J. Leishmaniasis of neotropical porcupines: *Leishmania hertigi deanei* nov. subsp. *Acta Amaz.* **1977**, *7*, 51–57. [[CrossRef](#)]
25. Gardener, P.J.; Chance, M.L.; Peters, W. Biochemical taxonomy of *Leishmania*. II: Electrophoretic variation of malate dehydrogenase. *Ann. Trop. Med. Parasitol.* **1974**, *68*, 317–325. [[CrossRef](#)] [[PubMed](#)]
26. da Silva, D.A.; Madeira Mde, F.; Barbosa Filho, C.J.; Schubach, E.Y.; Barros, J.H.; Figueiredo, F.B. *Leishmania* (*Leishmania*) *hertigi* in a porcupine (*Coendou* sp.) found in Brasilia, Federal District, Brazil. *Rev. Bras. Parasitol. Vet.* **2013**, *22*, 297–299. [[CrossRef](#)]
27. Deane, L.M.; da Silva, J.E.; de Figueiredo, P.Z. Leishmaniasis in the viscera of porcupines from the state of Piauí, Brazil. *Rev. Do Inst. De Med. Trop. De Sao Paulo* **1974**, *16*, 68–69.

28. Pothirat, T.; Tantiworawit, A.; Chaiwarith, R.; Jariyapan, N.; Wannasan, A.; Siriyasatien, P.; Supparatpinyo, K.; Bates, M.D.; Kwakye-Nuako, G.; Bates, P.A. First isolation of *Leishmania* from Northern Thailand: Case report, identification as *Leishmania martiniquensis* and phylogenetic position within the *Leishmania enriettii* complex. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3339. [[CrossRef](#)]
29. Shaw, J.J. A possible vector of *Endotrypanum schaudinni* of the sloth *Choloepus hoffmanni*, in Panama. *Nature* **1964**, *201*, 417–418. [[CrossRef](#)]
30. Shaw, J.J.; de Rosa, A.T.; Cruz, A.C.R.; Vasconcelos, P.F.C. Brazilian phlebotomines as hosts and vectors of viruses, bacteria, fungi, protozoa (excluding those belonging to the genus *Leishmania*) and nematodes. In *Brazilian Sand Flies*; Rangel, E.F., Shaw, J.J., Eds.; Springer International Publishing AG: Basel, Switzerland, 2018; pp. 417–441.
31. Shaw, J.J. The behaviour of *Endotrypanum schaudinni* (Kinetoplastida: Trypanosomatidae) in three species of laboratory-bred neotropical sandflies (Diptera: Psychodidae) and its influence on the classification of the genus *Leishmania*. In *Parasitological Topics. A Presentation Volume to P. C. C. Garnham, F. R. S., on the Occasion of His 80th Birthday*; Canning, E.U., Ed.; Allen Press: Lawrence, KS, USA, 1981; pp. 232–241.
32. Franco, A.M.; Tesh, R.B.; Guzman, H.; Deane, M.P.; Grimaldi Junior, G. Development of *Endotrypanum* (Kinetoplastida: Trypanosomatidae) in experimentally infected phlebotomine sand flies (Diptera: Psychodidae). *J. Med. Entomol.* **1997**, *34*, 189–192. [[CrossRef](#)] [[PubMed](#)]
33. Katakura, K.; Mimori, T.; Furuya, M.; Uezato, H.; Nonaka, S.; Okamoto, M.; Gomez, L.E.; Hashiguchi, Y. Identification of *Endotrypanum* species from a sloth, a squirrel and *Lutzomyia* sandflies in Ecuador by PCR amplification and sequencing of the mini-exon gene. *J. Vet. Med. Sci.* **2003**, *65*, 649–653. [[CrossRef](#)] [[PubMed](#)]
34. Christensen, H.A.; Herrer, A. Neotropical sand flies (Diptera: Psychodidae), invertebrate hosts of *Endotrypanum schaudinni* (Kinetoplastida: Trypanosomatidae). *J. Med. Entomol.* **1976**, *13*, 299–303. [[CrossRef](#)]
35. Thies, S.F.; Bronzoni, R.V.M.; Michalsky, E.M.; Santos, E.S.D.; Silva, D.; Dias, E.S.; Damazo, A.S. Aspects on the ecology of phlebotomine sand flies and natural infection by *Leishmania hertigi* in the Southeastern Amazon Basin of Brazil. *Acta Trop.* **2018**, *177*, 37–43. [[CrossRef](#)]
36. 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](#)]
37. Harkins, K.M.; Schwartz, R.S.; Cartwright, R.A.; Stone, A.C. Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2016**, *38*, 101–109. [[CrossRef](#)] [[PubMed](#)]
38. O’Leary, M.A.; Bloch, J.I.; Flynn, J.J.; Gaudin, T.J.; Giallombardo, A.; Giannini, N.P.; Goldberg, S.L.; Kraatz, B.P.; Luo, Z.X.; Meng, J.; et al. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **2013**, *339*, 662–667. [[CrossRef](#)]
39. Lopes, A.H.; Iovannisci, D.; Petrillo-Peixoto, M.; McMahon-Pratt, D.; Beverley, S.M. Evolution of nuclear DNA and the occurrence of sequences related to new small chromosomal DNAs in the trypanosomatid genus *Endotrypanum*. *Mol. Biochem. Parasitol.* **1990**, *40*, 151–161. [[CrossRef](#)]
40. Maslov, D.A.; Lukeš, J.; Jirků, M.; Simpson, L. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. *Mol. Biochem. Parasitol.* **1996**, *75*, 197–205. [[CrossRef](#)]
41. Maslov, D.A.; Yurchenko, V.Y.; Jirků, M.; Lukeš, J. Two new species of trypanosomatid parasites isolated from Heteroptera in Costa Rica. *J. Eukaryot. Microbiol.* **2010**, *57*, 177–188. [[CrossRef](#)] [[PubMed](#)]
42. Yurchenko, V.; Votýpka, J.; Tesařová, M.; Klepetková, H.; Kraeva, N.; Jirků, M.; Lukeš, J. Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. *Folia Parasitol.* **2014**, *61*, 97–112. [[CrossRef](#)]
43. Losev, A.; Grybchuk-Ieremenko, A.; Kostygov, A.Y.; Lukes, J.; Yurchenko, V. Host specificity, pathogenicity, and mixed infections of trypanoplasms from freshwater fishes. *Parasitol. Res.* **2015**, *114*, 1071–1078. [[CrossRef](#)] [[PubMed](#)]
44. Kostygov, A.Y.; Grybchuk-Ieremenko, A.; Malysheva, M.N.; Frolov, A.O.; Yurchenko, V. Molecular revision of the genus *Wallaceina*. *Protist* **2014**, *165*, 594–604. [[CrossRef](#)]
45. 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](#)]
46. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
47. 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 March 2021).
48. 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. J. Comput. Mol. Cell Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
49. Laetsch, D.R.; Blaxter, M.L. BlobTools: Interrogation of genome assemblies. *F1000Research* **2017**, *6*, 1287. [[CrossRef](#)]
50. Mikheenko, A.; Prjibelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150. [[CrossRef](#)]
51. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
52. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]

53. Steinbiss, S.; Silva-Franco, F.; Brunk, B.; Foth, B.; Hertz-Fowler, C.; Berriman, M.; Otto, T.D. Companion: A web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* **2016**, *44*, W29–W34. [CrossRef] [PubMed]
54. Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [CrossRef] [PubMed]
55. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. Available online: <http://www.repeatmasker.org> (accessed on 19 March 2021).
56. Soderlund, C.; Nelson, W.; Shoemaker, A.; Paterson, A. SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* **2006**, *16*, 1159–1168. [CrossRef]
57. Tamazian, G.; Dobrynin, P.; Krashennikova, K.; Komissarov, A.; Koepfli, K.P.; O'Brien, S.J. Chromosomer: A reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* **2016**, *5*, 38. [CrossRef]
58. Quinlan, A.R. BEDTools: The swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11–12. [CrossRef] [PubMed]
59. Wickham, H.; François, R.; Henry, L.; Müller, K. *Dplyr: A Grammar of Data Manipulation. R Package Version 1.0.2*; 2020.
60. Ginestet, C. ggplot2: Elegant graphics for data analysis. *J. R. Stat. Soc.* **2011**, *174*, 245. [CrossRef]
61. Lipovetsky, S. Statistical inference via data science: A modern dive into R and the tidyverse. *Technometrics* **2020**, *62*, 283. [CrossRef]
62. Boyd, Z.; Hughes, J. *WeeSAM: Script for Parsing SAM/BAM Files for Coverage Statistics*; 2018.
63. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]
64. Rimmer, A.; Phan, H.; Mathieson, I.; Iqbal, Z.; Twigg, S.R.; Wilkie, A.O.; McVean, G.; Lunter, G. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **2014**, *46*, 912–918. [CrossRef]
65. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef] [PubMed]
66. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
67. Capella-Gutiérrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
68. Eddy, S.R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [CrossRef] [PubMed]
69. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
70. Lartillot, N.; Rodrigue, N.; Stubbs, D.; Richer, J. PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **2013**, *62*, 611–615. [CrossRef]
71. Yu, G.C.; Smith, D.K.; Zhu, H.C.; Guan, Y.; Lam, T.T.Y. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. [CrossRef]
72. Csűrös, M. Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **2010**, *26*, 1910–1912. [CrossRef]
73. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef] [PubMed]
74. 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]
75. Lex, A.; Gehlenborg, N.; Strobel, H.; Vuilleumot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1983–1992. [CrossRef]
76. 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]
77. Binns, D.; Dimmer, E.; Huntley, R.; Barrell, D.; O'Donovan, C.; Apweiler, R. QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **2009**, *25*, 3045–3046. [CrossRef]
78. 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]
79. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]
80. Rambaut, A. FigTree v.1.4.4. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 19 March 2021).
81. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef]
82. Gerlt, J.A.; Bouvier, J.T.; Davidson, D.B.; Imker, H.J.; Sadkhin, B.; Slater, D.R.; Whalen, K.L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019–1037. [CrossRef] [PubMed]

83. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
84. Butenko, A.; Hammond, M.; Field, M.C.; Ginger, M.L.; Yurchenko, V.; Lukeš, J. Reductionist pathways for parasitism in euglenozoans? Expanded datasets provide new insights. *Trends Parasitol.* **2021**, *37*, 100–116. [[CrossRef](#)]
85. Jackson, A.P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol. Biol. Evol.* **2010**, *27*, 33–45. [[CrossRef](#)]
86. Zhang, J.; Nielsen, R.; Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **2005**, *22*, 2472–2479. [[CrossRef](#)] [[PubMed](#)]
87. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [[CrossRef](#)]
88. Alexa, A.; Rahnenfuhrer, J.; Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **2006**, *22*, 1600–1607. [[CrossRef](#)]
89. 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 pyrrhocoris*: A high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci. Rep.* **2016**, *6*, 23704. [[CrossRef](#)]
90. Sloan, M.A.; Brooks, K.; Otto, T.D.; Sanders, M.J.; Cotton, J.A.; Ligoxygakis, P. Transcriptional and genomic parallels between the monoxenous parasite *Herpetomonas muscarum* and *Leishmania*. *PLoS Genet.* **2019**, *15*, e1008452. [[CrossRef](#)] [[PubMed](#)]
91. Szöör, B. Trypanosomatid protein phosphatases. *Mol. Biochem. Parasitol.* **2010**, *173*, 53–63. [[CrossRef](#)]
92. Soulat, D.; Bogdan, C. Function of macrophage and parasite phosphatases in leishmaniasis. *Front. Immunol.* **2017**, *8*, 1838. [[CrossRef](#)]
93. Orr, G.A.; Werner, C.; Xu, J.; Bennett, M.; Weiss, L.M.; Takvorkan, P.; Tanowitz, H.B.; Wittner, M. Identification of novel serine/threonine protein phosphatases in *Trypanosoma cruzi*: A potential role in control of cytokinesis and morphology. *Infect. Immun.* **2000**, *68*, 1350–1358. [[CrossRef](#)]
94. Kraeva, N.; Leštinová, T.; Ishemgulova, A.; Majerová, K.; Butenko, A.; Vaselek, S.; Bespyatykh, J.; Charyyeva, A.; Spitzová, T.; Kostygov, A.Y.; et al. *LmxM.22.0250*-encoded dual specificity protein/lipid phosphatase impairs *Leishmania mexicana* virulence in vitro. *Pathogens* **2019**, *8*, 241. [[CrossRef](#)]
95. Qureshi, R.; Jakkula, P.; Sagurthi, S.R.; Qureshi, I.A. Protein phosphatase 1 of *Leishmania donovani* exhibits conserved catalytic residues and pro-inflammatory response. *Biochem. Biophys. Res. Commun.* **2019**, *516*, 770–776. [[CrossRef](#)]
96. Kaufer, A.; Stark, D.; Ellis, J. Evolutionary insight into the Trypanosomatidae using alignment-free phylogenomics of the kinetoplast. *Pathogens* **2019**, *8*, 157. [[CrossRef](#)] [[PubMed](#)]
97. Ludwig, A.; Krieger, M.A. Genomic and phylogenetic evidence of VIPER retrotransposon domestication in trypanosomatids. *Mem. Do Inst. Oswaldo Cruz* **2016**, *111*, 765–769. [[CrossRef](#)]
98. Kelly, F.D.; Sanchez, M.A.; Landfear, S.M. Touching the surface: Diverse roles for the flagellar membrane in kinetoplastid parasites. *Microbiol. Mol. Biol. Rev. Mmbr.* **2020**, *84*, e00079-19. [[CrossRef](#)] [[PubMed](#)]
99. Liu, Q.; Lei, J.; Darby, A.C.; Kadowaki, T. Trypanosomatid parasite dynamically changes the transcriptome during infection and modifies honey bee physiology. *Commun. Biol.* **2020**, *3*, 51. [[CrossRef](#)]
100. 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](#)]
101. Myler, P.J.; Lodes, M.J.; Merlin, G.; de Vos, T.; Stuart, K.D. An amplified DNA element in *Leishmania* encodes potential integral membrane and nucleotide-binding proteins. *Mol. Biochem. Parasitol.* **1994**, *66*, 11–20. [[CrossRef](#)]
102. Ravooru, N.; Paul, O.S.; Nagendra, H.G.; Sathyanarayanan, N. Data enabled prediction analysis assigns folate/biopterin transporter (BT1) family to 36 hypothetical membrane proteins in *Leishmania donovani*. *Bioinformatics* **2019**, *15*, 697–708. [[CrossRef](#)]
103. Ouellette, M.; Drummel-Smith, J.; El-Fadili, A.; Kundig, C.; Richard, D.; Roy, G. Pterin transport and metabolism in *Leishmania* and related trypanosomatid parasites. *Int. J. Parasitol.* **2002**, *32*, 385–398. [[CrossRef](#)]
104. Ouameur, A.A.; Girard, I.; Legare, D.; Ouellette, M. Functional analysis and complex gene rearrangements of the folate/biopterin transporter (FBT) gene family in the protozoan parasite *Leishmania*. *Mol. Biochem. Parasitol.* **2008**, *162*, 155–164. [[CrossRef](#)]
105. Singer, S.J. The structure and insertion of integral proteins in membranes. *Annu. Rev. Cell Biol.* **1990**, *6*, 247–296. [[CrossRef](#)]
106. 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](#)] [[PubMed](#)]
107. Opperdoes, F.R.; Coombs, G.H. Metabolism of *Leishmania*: Proven and predicted. *Trends Parasitol.* **2007**, *23*, 149–158. [[CrossRef](#)]
108. Opperdoes, F.; Michels, P.A. The metabolic repertoire of *Leishmania* and implications for drug discovery. In *Leishmania: After the Genome*; Myler, P., Fasel, N., Eds.; Caister Academic Press: Norfolk, UK, 2008; pp. 123–158.
109. Škodová-Sveráková, I.; Záhonová, K.; Bučková, B.; Füssy, Z.; Yurchenko, V.; Lukeš, J. Catalase and ascorbate peroxidase in euglenozoan protists. *Pathogens* **2020**, *9*, 317. [[CrossRef](#)]
110. Kraeva, N.; Horáková, E.; Kostygov, A.; Kořený, L.; Butenko, A.; Yurchenko, V.; Lukeš, J. Catalase in Leishmaniinae: With me or against me? *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2017**, *50*, 121–127. [[CrossRef](#)]

111. Turnock, D.C.; Ferguson, M.A. Sugar nucleotide pools of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. *Eukaryot. Cell* **2007**, *6*, 1450–1463. [[CrossRef](#)]
112. Boitz, J.M.; Yates, P.A.; Kline, C.; Gaur, U.; Wilson, M.E.; Ullman, B.; Roberts, S.C. *Leishmania donovani* ornithine decarboxylase is indispensable for parasite survival in the mammalian host. *Infect. Immun.* **2009**, *77*, 756–763. [[CrossRef](#)] [[PubMed](#)]
113. Moreira, D.S.; Xavier, M.V.; Murta, S.M.F. Ascorbate peroxidase overexpression protects *Leishmania braziliensis* against trivalent antimony effects. *Mem. Do Inst. Oswaldo Cruz* **2018**, *113*, e180377. [[CrossRef](#)] [[PubMed](#)]
114. Sunter, J.D.; Yanase, R.; Wang, Z.; Catta-Preta, C.M.C.; Moreira-Leite, F.; Myšková, J.; Pružinová, K.; Volf, P.; Mottram, J.C.; Gull, K. *Leishmania* flagellum attachment zone is critical for flagellar pocket shape, development in the sand fly, and pathogenicity in the host. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6351–6360. [[CrossRef](#)] [[PubMed](#)]
115. 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](#)]
116. Wu, Y.; El Fakhry, Y.; Sereno, D.; Tamar, S.; Papadopoulou, B. A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins. *Mol. Biochem. Parasitol.* **2000**, *110*, 345–357. [[CrossRef](#)]
117. Coughlin, B.C.; Teixeira, S.M.; Kirchhoff, L.V.; Donelson, J.E. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3′-untranslated region position-dependent *cis*-element and an untranslated region-binding protein. *J. Biol. Chem.* **2000**, *275*, 12051–12060. [[CrossRef](#)]
118. Stober, C.B.; Lange, U.G.; Roberts, M.T.; Alcamí, A.; Blackwell, J.M. IL-10 from regulatory T cells determines vaccine efficacy in murine *Leishmania major* infection. *J. Immunol.* **2005**, *175*, 2517–2524. [[CrossRef](#)]
119. Ribeiro, P.A.F.; Vale, D.L.; Dias, D.S.; Lage, D.P.; Mendonça, D.V.C.; Ramos, F.F.; Carvalho, L.M.; Carvalho, A.; Steiner, B.T.; Roque, M.C.; et al. *Leishmania infantum* amastin protein incorporated in distinct adjuvant systems induces protection against visceral leishmaniasis. *Cytokine* **2020**, *129*, 155031. [[CrossRef](#)]
120. Pérez-Díaz, L.; Silva, T.C.; Teixeira, S.M. Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **2017**, *211*, 1–8. [[CrossRef](#)]
121. Butenko, A.; Opperdoes, F.R.; Flegontova, O.; Horak, A.; Hampl, V.; Keeling, P.; Gawryluk, R.M.R.; Tikhonenkov, D.; Flegontov, P.; Lukeš, J. Evolution of metabolic capabilities and molecular features of diplomonads, kinetoplastids, and euglenids. *BMC Biol.* **2020**, *18*, 23. [[CrossRef](#)]
122. Raymond, F.; Boisvert, S.; Roy, G.; Ritt, J.F.; Legare, 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](#)]
123. Wilson, V.; Southgate, B. Lizard *Leishmania*. In *Biology of Kinetoplastida*; Lumsden, W., Evans, D.A., Eds.; Academic Press: New York, NY, USA, 1979; pp. 242–268.
124. Ovezmukhammedov, A.; Saf’ianova, V.M. Taxonomic problems of the *Leishmania* of reptiles. *Parazitologiya* **1989**, *23*, 334–343. (In Russian)
125. de Paiva, R.M.; Grazielle-Silva, V.; Cardoso, M.S.; Nakagaki, B.N.; Mendonça-Neto, R.P.; Canavaci, A.M.; Souza Melo, N.; Martinelli, P.M.; Fernandes, A.P.; da Rocha, W.D.; et al. Amastin knockdown in *Leishmania braziliensis* affects parasite-macrophage interaction and results in impaired viability of intracellular amastigotes. *PLoS Pathog.* **2015**, *11*, e1005296. [[CrossRef](#)] [[PubMed](#)]
126. de Menezes, J.P.; Saraiva, E.M.; da Rocha-Azevedo, B. The site of the bite: *Leishmania* interaction with macrophages, neutrophils and the extracellular matrix in the dermis. *Parasites Vectors* **2016**, *9*, 264. [[CrossRef](#)] [[PubMed](#)]
127. Pinheiro, L.J.; Paranaíba, L.F.; Alves, A.F.; Parreiras, P.M.; Gontijo, N.F.; Soares, R.P.; Tafuri, W.L. Salivary gland extract modulates the infection of two *Leishmania enriettii* strains by interfering with macrophage differentiation in the model of *Cavia porcellus*. *Front. Microbiol.* **2018**, *9*, 969. [[CrossRef](#)] [[PubMed](#)]
128. Jain, M.; Dole, V.S.; Myler, P.J.; Stuart, K.D.; Madhubala, R. Role of *biopterin transporter* (BT1) gene on growth and infectivity of *Leishmania*. *Am. J. Biochem. Biotechnol.* **2007**, *3*, 199–206. [[CrossRef](#)]