Genomics 115 (2023) 110661

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Complete assembly, annotation of virulence genes and CRISPR editing of the genome of *Leishmania amazonensis* PH8 strain

Wanessa Moreira Goes [a], Carlos Rodolpho Ferreira Brasil [b], João Luis Reis-Cunha [c,d], Anderson Coqueiro-dos-Santos [d], Viviane Grazielle-Silva [a], Júlia de Souza Reis [b], Tatiane Cristina Souto [b], Maria Fernanda Laranjeira-Silva [e], Daniella Castanheira Bartholomeu [d], Ana Paula Fernandes [b,f], Santuza Maria Ribeiro Teixeira [a,f,*]

[a] Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG CEP 31.270-901, Brazil
[b] Departamento de Análises Clínicas e Toxicológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG CEP 31.270-901, Brazil
[c] Departamento de Veterinária Preventiva, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG CEP 31.270-901, Brazil
[d] Departamento de Parasitologia, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG CEP 31.270-901, Brazil
[e] Departamento de Fisiologia, Universidade de São Paulo, Rua do Matão 101, Cidade Universitária, São Paulo, SP CEP 05508-900, Brazil
[f] Centro de Tecnologia de Vacinas, Universidade Federal de Minas Gerais, Rua Professor José Vieira de Mendonça 770, Belo Horizonte, MG, CEP 31.210-360, Brazil

ARTICLE INFO

Keywords:
Leishmania amazonensis
Genome
Multigene families
Virulence factors
CRISPR editing

ABSTRACT

We report the sequencing and assembly of the PH8 strain of *Leishmania amazonensis* one of the etiological agents of leishmaniasis. After combining data from long Pacbio reads, short Illumina reads and synteny with the *Leishmania mexicana* genome, the sequence of 34 chromosomes with 8317 annotated genes was generated. Multigene families encoding three virulence factors, A2, amastins and the GP63 metalloproteases, were identified and compared to their annotation in other *Leishmania* species. As they have been recently recognized as virulence factors essential for disease establishment and progression of the infection, we also identified 14 genes encoding proteins involved in parasite iron and heme metabolism and compared to genes from other Trypanosomatids. To follow these studies with a genetic approach to address the role of virulence factors, we tested two CRISPR-Cas9 protocols to generate L. *amazonensis* knockout cell lines, using the Miltefosine transporter gene as a proof of concept.

1. Introduction

Leishmaniasis is considered a neglected tropical disease caused by several species of *Leishmania* spp. Transmitted to mammalian hosts by infected female sand flies, the disease is widespread in the tropical and subtropical areas of the world [1]. Due to the large heterogeneity of *Leishmania* parasites, different clinical and pathological forms of tegumentary leishmaniasis (TL) and visceral leishmaniasis (VL) are observed [2]. In addition to the infecting parasite species and their tissues' tropism and virulence factors, development of each clinical manifestation depends on other factors that include the biology of the vector, the genetics of the host and its immune system. The combination of these elements results in the wide spectrum of clinical manifestations of TL, diffuse cutaneous leishmaniasis (DCL), localized cutaneous leishmaniasis (LCL) and mucosal leishmaniasis (ML) [2]. Endemic in 92 countries, 700,000 to 1 million new cases of leishmaniasis occur annually, resulting in 20,000 to 30,000 deaths per year, mainly associated with VL. Still, around 95% of cases of cutaneous leishmaniasis (CL) occur in Latin America, the Mediterranean basin and West Asia [1]. In the last 20 years, 1,067,759 cases of CL were reported to the Pan American Health Organization (PAHO), with 16,000 new cases reported in Brazil alone in 2020 [3].

Among twenty *Leishmania* species that are pathogenic to humans, *Leishmania amazonensis* stands out because it can cause virtually all the clinical forms of TL and also VL in humans [4] and dogs [5,6]. Being a New World autochthonous species, the fact that L. *amazonensis* can cause VL infection is intriguing because VL is caused mainly by Old World species of the *Leishmania donovani complex and Leishmania*

* Corresponding author at: Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG CEP 31.270-901, Brazil.
E-mail address: santuzat@ufmg.br (S.M.R. Teixeira).

https://doi.org/10.1016/j.ygeno.2023.110661
Received 22 November 2022; Received in revised form 4 May 2023; Accepted 27 May 2023
Available online 30 May 2023

infantum chagasi [4]. Moreover, while LCL or ML caused by *Leishmania braziliensis* are characterized by exacerbated T cell responses to the parasite and early production of excessive amounts of pro-inflammatory cytokines (eg, IFN-γ, TNF-α and IL-6), which are associated with exacerbated tissue damage, DCL caused by L. *amazonensis* correlates with anergy of T cell responses and large amounts of parasites in lesions [7]. It is also noteworthy the fact that all mice strains are susceptible to L. *amazonensis,* while some strains are resistant to *Leishmania major* infections and all mice strains are usually resistant to *L. braziliensis* infections [8,9]. All these different patterns of interactions with the host's immune system suggest that L. *amazonensis* has distinct virulence factors or may express them differently as compared to other species. To better understand how L. *amazonensis* evolved and developed so many strategies to overcome host immune responses and cause the whole spectrum of disease forms, the availability of complete genome sequences of reference strains and isolates is very important. However, until recently very few L. *amazonensis* genomes were available, some have important gaps and just one was partially annotated. Although the PH8 strain has been employed in several studies aimed at investigating basic aspects of the infection as well as studies towards vaccine development, the genome of this strain has not yet been fully sequenced.

Together with experimental studies, genome sequencing became a powerful tool that largely contributes to a comprehensive understanding of the biology of *Leishmania* as well as other protozoan parasites. Moreover, genome studies allow the identification of parasite factors involved with interactions with the insect and mammalian hosts and the pathogenesis of different diseases they cause. *Leishmania* is a diploid organism, but several studies have reported the occurrence of aneuploidy in many chromosomes [10–14], depending on the species and strains analysed [15–17]. Two attempts at sequencing and assembling the genome of L. *amazonensis* (MHOM/BR/71973/M2269), which belongs to the *L. mexicana* complex, resulted in the generation of highly fragmented genomes [18,19]. In 2019, sequencing and assembly of the genomes of two additional L. *amazonensis* strains, CDC210-L1346 and UA301 were also described, with the UA301 strain being the only published study that reached the number of chromosomes predicted for the species [20]. Guided by the reference genome of L. *mexicana*, the assembly of the L. *amazonensis* UA301 strain showed that 24 of the 34 chromosomes were disomic, 9 trisomic and 1 tetrasomic. Consistent with this and other studies, genome analyses of two L. *amazonensis* isolates obtained from lymph node aspirates from dogs with clinical manifestation of VL confirmed that, while most chromosomes are disomic, chromosome 30 is polysomic [21]. In addition, copy number variation has been also recognized as an important factor affecting gene expression levels [22]. It has been shown that variations in the number of gene copies among multigene families involved with parasite virulence are highly frequent in different *Leishmania* species [12,13,18,20,23,24]. However, for a reliable evaluation of gene copy number variation, it is necessary to have well-assembled and annotated reference genomes not only for various *Leishmania* species but also for different strains and isolates from the same species.

Once a high-quality reference genome is available, efficient genome manipulation protocols to allow gene knockout of single copy genes as well as multigene families are essential tools for post-genomic studies aimed at characterizing parasite molecules that act as virulence factors. Besides describing the genome of the PH8 strain, we have also described two protocols that allowed a rapid and efficient generation of L. *amazonensis* knockout mutants using the CRISPR/Cas9 technology. By expanding the knowledge about the molecular diversity of this species, the current work may contribute to a better understanding of the factors responsible for the broad range of clinical manifestations of leishmaniasis.

2. Materials and methods

2.1. Parasites, DNA isolation and genome sequencing

Genomic DNA (gDNA) was obtained from promastigote cultures of L. *amazonensis* (strain IFLA/BR/67/PH8), isolated from *Lutzomyia flaviscutellata* in the city of Belém, Brazil. The extraction and purification of gDNA were performed using Life Technologies DNA extraction kit and sequenced using TruSeq™ RNA and DNA Library Preparation Kit v2 to produce 350 bp-size paired-end reads on the Illumina Hiseq 2000 platform. A second sample of L. *amazonensis* was obtained from gDNA extraction and purification of 6×10^8 promastigotes using the Wizard Genomic DNA Purification kit (Promega). The sample was sequenced using the Pacbio RS II platform and the DNA Sequencing Reagent kit 4.0 v2.

2.2. Nuclear and maxicircle genome assembly and annotation

The quality of Illumina raw reads was verified using FastQC v0.11.3 [25]. Adapter sequences and poor-quality regions were removed using Trimmomatic v.0.39 [26] and the following trimming parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:20:10 LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 3:15, MILEN: 75. The complete genome sequence of L. *amazonensis* PH8 strain was obtained through a de novo assembly approach and correction using short reads. Approximately 86× genome coverage long Pacbio reads were combined with ~43× genome coverage short Illumina reads. Canu v.1.5 assembler [27] was used to assemble initial contigs, considering only Pacbio reads. Later, the pipeline IPA (https://github.com/ThomasDOtto/IPA) [28] was used to remove smaller (< 5 kb) and redundant contigs. The scaffolding step was performed using SSPACE v.3.0 [29] and gap filling with GapFiller v.1.10 [30], based on Illumina reads. Finally, the assembly was polished using Pilon tool v.1.22 [31], and again Illumina reads were aligned to the pre-assembled genome. The scaffolds were ordered based on L. *mexicana* chromosomes using Abacas v.1.3.1 [32], where a minimum coverage equal to 20 was considered.

To verify the genome completeness, the Benchmarking Universal Single-Copy Orthologs (BUSCO, v5.1.2) metric [33] was employed, in which the single-copy ortholog genes database euglenozoa_odb10 was selected for the search performed against the PH8 genome. Lastly, genome coverage was calculated from the mapping of Illumina and Pacbio read libraries to the PH8 assembly. The BWA-MEM algorithm [34] was run with standard parameters and the count of reads mapped to each position in the genome (-d flag) was performed using BED tools [35].

The automatic annotation of the PH8 genome was performed by transferring the available annotation of the nuclear genome of the L. *mexicana* U1103, using the Rapid Annotation Transfer Tool (RATT) [36]. In parallel, a de novo annotation step was performed using the Companion software [37], which runs the AUGUSTUS algorithm, trained for L. *mexicana*. On the other hand, to annotate the maxicircle genome, only RATT tool was used, however, the reference genome used was L. *tarentolae* (strain UC), as described by Camacho and co-workers (2019) [38].

RepeatModeler v2.0.2 (http://www.repeatmasker.org/RepeatModeler) [39] was used to build a de novo repeats library using *Leishmania* spp. database (TritrypDB Release 52), and then used RepeatMasker v.4.1.2-p1 [40] (http://www.repeatmasker.org) to annotate all the repetitive sequences.

2.3. Chromosome and gene copy number analyses

Ploidy estimation was done using CADIn (Coqueiro-dos-Santos et al., submitted), based on the allele frequency of heterozygous positions and Read Deep Coverage (RDC) variations. Briefly, PH8 Illumina libraries were mapped in the *L. mexicana* reference chromosomes using BWA-

mem v0.7.12 [41]. Reads that mapped to the reference genome were then filtered by mapping quality 30 using SAMtools v1.1 [42] and the RDC of each position in each gene in all chromosomes was calculated using BEDtools genomecov v2.16.2 [35]. Genes with <50% of their length covered or with outlier coverage compared to other genes in the chromosomes were excluded. The somy (i.e., chromosome copy number variation) of each chromosome was calculated by the ratio of the median coverage of genes in each chromosome and the median genome RDC. The values of the chromosomal somy were submitted to Mann-Whitney-Wilcoxon tests and the significance of $p < 0.05$, as previously described [43]. To differentiate single-copy and expanded genes, we established a cut-off of 1.80× the haploid genome copy.

2.4. In silico characterization of genes encoding virulence factors

Members of multigene families encoding virulence factors such as A2 proteins, amastins and GP63 metalloprotease were identified using an automated pipeline developed by Wei-Wang and co-workers (2021) [44]. Candidate genes with at least 150 bp are predicted by searching for annotated members of each of the families in other *Leishmania* species, *T. cruzi* and *T. brucei* (TriTrypDB release 52). An additional validation step of the predicted candidate genes was performed using InterProScan [45].

Alpha, beta, gamma and delta-amastins signatures were identified according to Jackson (2010) [46]. Alignments between the amastin signatures and sequences present in the PH8 genome were performed using Blastp with an e-value set to 10^{-10}. A2 proteins, amastins, GP63 and heme and iron amino acid sequences were aligned using the MAFFT v7 [47] and a search for conserved domains was performed manually for all families. To analyze the clusters formed between amastin subfamilies and the evolutionary distances between them, a maximum-likelihood (ML) phylogenetic reconstruction using PhyML v3.0 [48] was performed. The best substitution model was set to LG + G + I, previously defined by ProtTest 3.4.2 software [49].

Protein coding sequences related to heme and iron metabolism of L. *braziliensis* M2904, *L. donovani* BPK282A1, *L. infantum* JPCM5, *L. major* Friedlin, *L. mexicana* U1103, *Trypanosoma brucei* TREU927, *Trypanosoma cruzi* CL Brener and L. *amazonensis* PH8 were downloaded from the TriTrypDB (release 52) and were aligned M-Coffee v13.45 [50]. ML tree were reconstructed using again PhyML (48) v3.0 for the alignment of 107 sequences and the best substitution model was set to JTT + I + G [51] determined using ProtTest 3.4.2 [49] according to the agreement between the Akaike information criterion and Bayesian information criterion.

2.5. Parasite culture, plasmids and in vitro transcription of sgRNAs for CRISPR gene editing

Promastigote forms of L. *amazonensis* (strain PH8) were grown at 26 °C using medium 199 (Gibco) as previously described [52]. The pLDCN plasmid, containing the Neomycin resistance gene, was used to express *Streptococcus pyogenes* nuclease Cas9 under the control of the L. *donovani* ribosome RNA promoter. Two sgRNA sequences containing sequences of the Miltefosine Transporter (MT) gene were designed with the aid of the LeishGEdit tool. For the donor sequence, an oligonucleotide containing sequences of the MT gene flanking the Cas9 endonuclease cleavage point followed by three tandem stop codons covering all three possible reading frames was synthetized. Two 30 bp homology arms corresponding to part of the MT gene was also part of the oligonucleotide donor sequence. The oligonucleotide also contains restriction sites for *Xho*I and *Hin*dIII endonucleases, which allowed after PCR amplification of the edited gene, a confirmation of the disruption of the gene by enzymatic digestion (Table S1).

In vitro transcription reactions were used to synthesize the sgRNAs targeting the MT gene, as previously described by Burle-Caldas et al. (2018) [53]. A PCR fragment generated with a forward primer containing the sequence corresponding to the T7 RNA polymerase promoter, a 20-nucleotide specific sequence complementary to the MT gene target site and a portion of the Cas9 scaffold and a reverse primer containing the sequence to amplify the remaining sgRNA Cas9 scaffold (Table S1) were used as a template. Following transcription with the MEGAshortscript™ T7 Kit (Thermo Fischer Scientific – USA), the products were purified using MEGAclear Kit (Thermo Fischer Scientific – USA), quantified, analysed in agarose gels and used for transfections.

2.6. Parasite transfection and CRISPR gene editing

A culture with 8×10^7 promastigotes of L. *amazonensis* was centrifuged at 1500 xg, for 10 min at 4 °C. The cell pellet was resuspended with 160 μL of Tb-BSF (90 mM $NaPO_4$, 0.15 mM $CaCl_2$, 5 mM KCl, 50 mM Hepes, pH 7.3), as described by Burle-Caldas et al. (2018) [53]. Ten μg of pLDCN plasmid containing the Cas9 gene were mixed with the same buffer, to a final volume of 20 μL. After adding 80 μL of the cell suspension to the DNA mixture the mix was pipetted into an electroporation cuvette (Gene Pulser 0.2 cm, BioRad). Samples were submitted to the X-001 program, one pulse, of the Amaxa Nucleofector II (Lonza Biotechnologies). The contents of the cuvette were transfered to a flask with 5 mL of M199 medium and 48 h after transfection, 20 μg/mL G418 was added to select for SpCas9 expressing cell lines.

After the selection of L. *amazonensis* expressing SpCas9, 4×10^7 promastigote forms were centrifuged as described above. The cell pellet was resuspended in 80 μL of Tb-BSF. Twenty μg of the in vitro transcribed RNA was added to 1 mM of the oligonucleotide donor sequence in a final volume of 20 μL of Tb-BSF. This mix was added to the cell suspension, transferred to the electroporation cuvette and submitted again to the Amaxa Nucleofector II. After the procedure, cells were transferred to a flask containing M199 medium and 48 h later, 42.5 μM of Miltefosine was added to select resistant parasites.

For parasite transfection with the ribonucleoprotein complex, 4×10^6 *L. amazonensis* promastigote forms were centrifuged as described. The cell pellet was resuspended using 70 μL of Tb-BSF. Fifteen μg of recombinant SaCas9 and 20 μg of the in vitro transcribed sgRNA were incubated in 10 μL of Tb-BSF 3× for 10 min at room temperature followed by another 10 min at 37 °C, to form the ribonucleoprotein complex. After adding 1mMol of the donor oligonucleotide to the mixture in a volume of 30 μL, the parasite suspension together with the ribonucleoprotein complex mixture were transferred to the electroporation cuvette and submitted to the Amaxa nucleofector II. The mixture was transferred to a flask with 5 mL of M199 medium and 48 h after transfection, 42.5 μM of Miltefosine were added to select resistant parasites. The recombinant Cas9 nuclease from *Staphylococcus aureus* (SaCas) was obtained from *E. coli* and used in an in vitro activity assay to test the nuclease activity as previously described [53].

After selecting MT resistant parasites, DNA was extracted using the Life Technologies DNA extraction kit as previously described [53] and PCR amplification of part of the MT gene (3.2Kbp) was carried with primers described in Table S1. The PCR products were submitted to enzymatic digestion with *Hin*dIII.or *Xho*I and analysed by electrophoresis in agarose gels.

3. Results and discussion

3.1. Illumina and Pacbio sequencing and genome assembly

Two libraries, prepared from DNA extracted from promastigote cultures of L. *amazonensis* PH8 strain were sequenced using Illumina and Pacific Biosciences (Pacbio) platforms. The first library, consisting of paired-end Illumina sequenced fragments of 350 bp medium-sized, yielded a total of 15,869,052 reads with 100 bp. After quality analysis and removal of low-quality regions and adapters, the total number of reads was reduced to 12,286,419 (~77.5% of the total obtained). As shown in the Phred score distribution profile generated with all reads

from the Illumina platform, after removing low-quality sequences, only reads with Phred >30 were used (Fig. S1A). The second library, sequenced with the Pacbio platform, resulted in the generation of 560,797 reads with an average size of 8056 bp and N50 equal to 10,196 (Fig. S1B). After combining these reads and removing redundancy, we obtained a total of 349,217 Pacbio reads with an average size of 12,980 bp, N50 of 17,985 and average quality equal to 0.83.

To assemble the L. *amazonensis* PH8 genome, first, Pacbio reads were

Fig. 1. The 34 chromosomes of L. *amazonensis* PH8 strain. Genes are depicted as rectangles drawn as proportional to their length, and their coding strand is indicated by their position above (top strand) or below (bottom strand) the central line. Coloured boxes represent gene families, whereas black and grey rectangles represent hypothetical and housekeeping genes, respectively. Chromosome sizes are indicated in kb or Mb.

processed using the Canu assembler, resulting in 380 contigs, followed by "clean up" using the IPA pipeline, resulting in 329 contigs. Next, scaffolding with SSPACE and gap filling with GapFiller resulted in 314 scaffolds with N50 equal to 188,152. Finally, the Pilon tool was used to "polish" the assembled PH8 genome, to eliminate inconsistencies between Illumina reads that were mapped to this assembly. We performed indel corrections (insertion and deletions: 29.76%), single base corrections (20.23%), block replacement events (20.23%), gap filling and identified local missassemblies, which together represent the 29.77% of all discrepancies between Illumina and Pacbio reads identified in the assembled sequences.

To better understand the difficulty in achieving assemblies with levels of contiguity that correspond to the exact number of chromosomes predicted for each species, without any gap, it is necessary to pay attention to the main intrinsic characteristics of the platforms used in sequencing, software used in assemblies as well as to the characteristics of the genome that is being analysed. Reconstruction of highly homologous sequences and structural variants such as deletions, duplications, insertions, inversions and translocations is a difficult task when using only short reading sequencing approaches. [54]. In fact, when analysing the complexity of different genomes of Trypanosomatids, it became clear the need to use sequencing platforms capable of covering ever larger regions, such as the Pacbio and Nanopore sequencing platforms. Another aspect to be considered is the accuracy of the base call, which must minimize errors produced during sequencing. Likewise, the number of reads generated per run is also highly important, in order to obtain satisfactory coverage and depth of the genome in question. Parallel to the evolution of sequencing techniques, genome assembly algorithms have also undergone major changes, due to the creation of new types of files generated by sequencing platforms and new strategies for obtaining complete genomes, as is the case of hybrid approaches, which considers more than one type of read [55]. Finally, the combination of all these factors together with gap-directed sequencing, may lead to the successful assembly of a genome, increasing its completeness to 100%.

The final assembly of the PH8 genome was built with 270 scaffolds syntenic to the 34 chromosomes of the L. *mexicana* reference genome (U1103 strain) (Fig. 1, Fig. S2). One scaffold, with 41,489 bp, contains the PH8 maxicircle (Fig. 2) and 42 scaffolds represent short sequences ranging from 1 to 35 kb that have low homology to the L. *mexicana* genome and, therefore, were not incorporated into the assembly. As indicated in the following sections, some of the small scaffolds contain repetitive sequences encoding amastins and A2 proteins. The 34 *L. amazonensis* PH8 chromosomes have >88% of their sequences aligned to the reference genome of L. *mexicana* U1103 with ambiguous base (N) ratio ranging from 0.001 to 0.057. Telomeric repeats [(TTAGGG)n] were identified at the 3' ends of scaffolds 6, 17 and 32, and complement reverse [(CCCTAA)n] at the 5' ends of scaffolds 3, 7 and 19. Based on the estimated size of 31.9 Mb for the L. *amazonensis* genome [16,18,20], the average sequencing coverage was estimated at 43× for Illumina reads and 86× for Pacbio reads (Fig. S3). Analysis to evaluate the completeness of PH8 assembly in comparison to a dataset of 130 markers of single-copy ortholog genes [33] that are expected to be present in 31 protozoan species, including L. *mexicana* U1103 strain [13], indicated that the PH8 strain assembly is 95.4% complete. A comparison with last two published assemblies from different L. *amazonensis* strains available in public databases (RZOD01 and UA301) showed similar values for

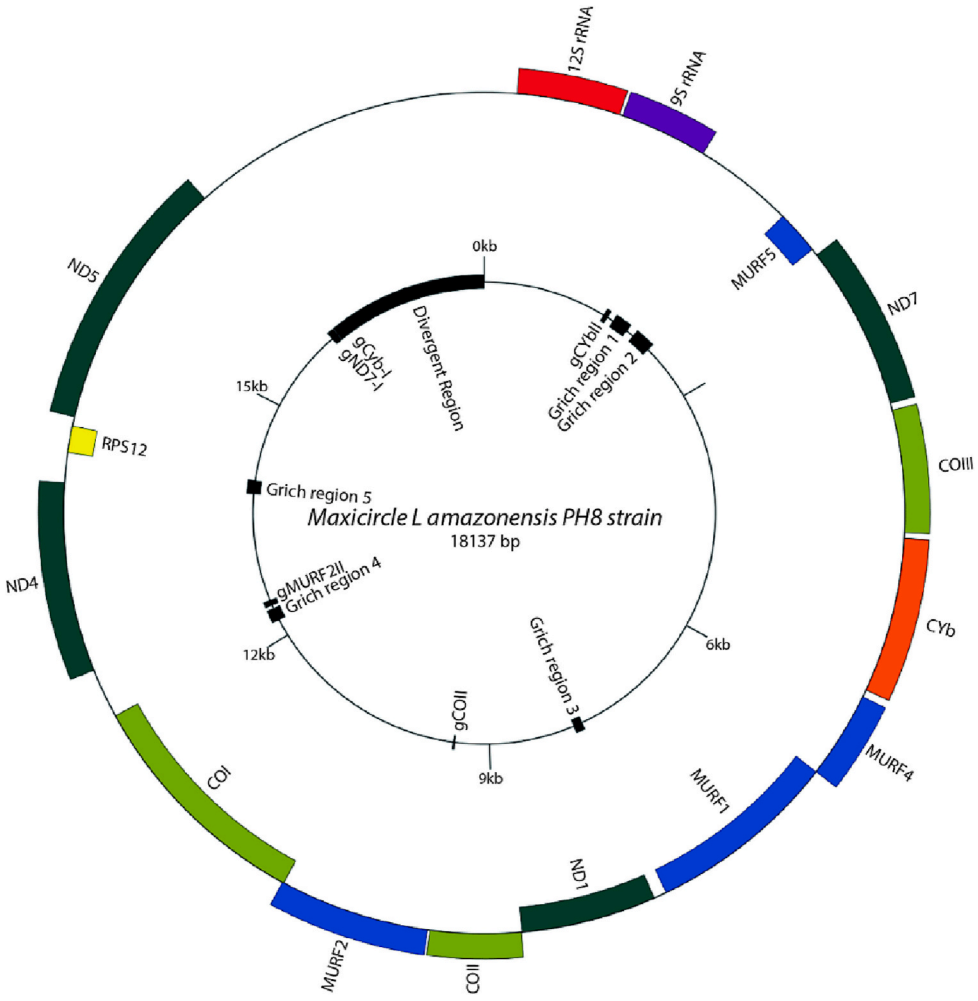

Fig. 2. The maxicircle mitochondrial genome of L. *amazonensis*. Coloured rectangles along the outer circle (strand) represent the location of genes and cryptogens. Upward-facing genes indicate the location on the sense strand and downward-facing genes on the antisense strand. Group of genes are indicated by similar colours. The rectangles in the inner circle represent miscellaneous features, including regions that encode guide RNAs (gRNAs), guanine-rich (G-rich) regions, and a ~ 9 kb repeat region.

most parameters evaluated, except contiguity levels (Table 1). The assembly of the PH8 genome is significantly improved compared to the genome of the M2269 strain, which is the only genome currently available and annotated in TriTrypDB. Compared to the UA301 strain, the results are comparable, since both assemblies reached chromosomal levels, however, no annotation was described for the genes present in the UA301 genome.

3.2. Nuclear and maxicircle genome content

Although the first description of a sequenced genome of L. *amazonensis* was published in 2013[18] and other sequences were made available later [5,16,19,20], these genome studies resulted in draft genome assemblies that are highly fragmented or genomes with limited gene annotation. In contrast, the L. *mexicana* genome, consisting of 554 contigs, has a total of 32 Mb of sequences ordered and scaffolded in 34 chromosomes. Since the PH8 genome described here is highly syntenic with the 34 chromosomes of L. *mexicana* (Fig. S2), we transferred the annotation of all 8225 genes from L. *mexicana* to the L. *amazonensis* genes. As shown in Table 2 a total of 7437 protein-coding genes with an average coding region of 1828 bp was identified. After performing an ab initio prediction of the gene models, 8317 genes were obtained, 7999 of them being protein-coding genes, 15 rRNA, 92 tRNA as well as 138 sequences encoding non-coding RNAs. This result means that 190 new annotated L. *amazonensis* genes were found, which were not described in the previous annotation of any other L. *amazonensis* genome[18].

Comparisons between the genetic content of different strains of L. *amazonensis*, considering mainly 12 universal orthologous genes that are shared among all *Leishmania* species, as described by Patino et al., 2020 [16], showed a clear evolutionary similarity of the PH8 strain with the genome of UA301 strain and RZOD01 (Fig. S4). Furthermore, the evolutionary proximity with L. *mexicana* was also confirmed, presenting the same evolutionary origin (emerging from the same node in the tree). The species classified as Old World *Leishmania* formed a distinct cluster, evidenced by the root configured in the orthologous genes of *T. cruzi*, used as an outgroup. The alignment between the PH8 genome and the L. *mexicana* reference genome also allowed the identification of clusters of shared and species-specific genes. A core set of 7612 orthologous gene clusters shared between the genomes of PH8 and L. *mexicana* was identified, while 96 protein-coding gene clusters was found to be specific to L. *amazonensis* and 141 to L. *mexicana*. Applied to L. *amazonensis*, *L. mexicana*, *L. donovani*, *L. infantum*, L. *major* and *L. braziliensis* protein data sets, this analysis resulted in 8025 clusters, 7120 of them shared by L. *amazonensis* and other *Leishmania* species (Fig. S5). The five largest clusters identified in *Leishmania* spp. were also identified in L. *amazonensis* (Table S2): amastin-like proteins (Cluster 1), dyneins (Cluster 2) glycoprotein GP63 (Cluster 3), ATP-binding cassette protein (Cluster 4), and histone H4 (Cluster 5).

Similar to other Trypanosomatid genomes [12,56,57], we identified repeated sequences of different sizes widely distributed throughout the L. *amazonensis* genome. These low-complexity sequences (Table S3) occupy 9.57% of the PH8 strain genome and are classified as retroelements (1.58%), DNA transposons (1.06%), Rolling circles (0.21%), Small RNAs (0.41%), Satellites (0.16%), Simple (1.87%), low complexity (0.27%) or unclassified repeats (4.01%), agreeing with the proportion of repetitive content estimated for references such as L. *braziliensis* and L. *infantum* [12]. Like other Trypanosomatid genomes, several multigene families organized in tandem repeats were found scattered in different chromosomes of the PH8 genome (Fig. 1).

Table 2
Genome annotation statistics of the PH8 strain.

General		
	Annotation transfer	Ab initio prediction
Total genes	8225	8317
coding percentage (%)	47.19	48.92
Protein coding genes		
Genes	7437	7999
Pseudogenes*	788	318
CDS average size (bp)	1828	1846
G + C content (%)	60.15	62.05
RNAs		
ncRNA	270	138
tRNA	82	92
rRNA	13	15

* Pseudogenes were inferred based on the presence of premature stop codons and/or incomplete reading frames.

A distinctive feature shared by Kinetoplastids is the network structure of the mitochondrial DNA composed of dozens of 20-40 kb maxicircles and thousands of 0.5-2.0 kb minicircles, known as kinetoplast DNA (or kDNA) [58]. Maxicircles are functionally similar to the eukaryotic mitochondrial genome, whereas minicircles encode small RNAs (gRNA) involved in RNA editing of several maxicircle transcripts [59]. Assembled as a 18,137 bp molecule, the L. *amazonensis* maxicircle was found to encode 12S and 9S rRNAs as well as several components of the electron transport chain, including cytochrome *b* (CYb), cytochrome oxidase subunits I, II and III (COI, COII and COIII), and NADH dehydrogenase subunits 1, 4, 5 7 (ND1, ND4, ND5 and ND7). Four open reading frames (ORFs) of unknown function (MURF1, MURF2 (ND2), MURF4 (ATPase6), MURF5 (uS3m)) and five putative sequences encoding gDNAs were also identified (Fig. 2). To generate translatable mRNAs, several mitochondrial protein-coding transcripts need to be edited through the addition or removal of uridine residues that correct frameshifts and introduce start codons. For some mRNAs, >50% of the final length of the molecule can be edited. i.e., have uridines inserted or deleted [60,61]. Whereas most gRNAs were identified in different locations in the maxicircle, distant from the genes they are related to, the gRNA acting on the COII gene was found next to it. The presence of guanine-rich regions described as being involved in RNA editing [62] was also observed.

The coding region of the maxicircle DNA of L. *amazonensis* is highly similar to the coding regions of the maxicircles of L. *mexicana*, *L. infantum*, *L. donovani*, L. *braziliensis* and *L. major*, presenting and 91.88%, 87.75%, 89.05%, 84.41%, 87.30% identity, respectively. Synteny is also conserved between different species, however, in the maxicircle genome of PH8, ORFs encoding the ND3, ND8 and ND9 proteins were not detected. Instead, various patterns of G-rich regions were present at these positions, interrupting the transcription of mRNAs, similar to was observed in L. *tarentolae* and L. *donovani* [63,64]. Furthermore, unlike L. *infantum*, *L. donovani*, *L. braziliensis* and *L. major*,

Table 1
Comparison of L. *amazonensis* genome assemblies available on public databases.

L. amazonensis strain	Plataform and coverage	Total size (Mb)	No. of contigs or scaffolds	GC content (%)	N50 (bp)	Largest contig or scaffold length	References
PH8	Pacbio (60×)/Illumina (48×)	31,970,850	77	59.62	1,069,653	3,400,190	This work
M2269*	454, Illumina (96×)	29,029,348	2.627	59.26	19,306	113,027	20
RZOD01	Pacbio, Illumina (75×)	33,504,997	92	59.71	850,106	3,425,950	22
UA301	Illumina (99×)	32,156,470	34	59.50	1,135,553	3,336,136	17

* This is the assembly that is currently available in TritrypDB and considered as the L. *amazonensis* reference genome.

the maxicircle of L. *amazonensis* has an orthologous gene of ATPase subunit 6, called MURF4, which was also described in *T. brucei* and L. *tarentolae* [65].

3.3. Chromosomal and gene copy number amplification

Several studies have shown that aneuploidy is highly frequent among *Leishmania* species. Nine polyploid chromosomes have been described in L. *infantum*, four in L. *mexicana*, two in L. *braziliensis*, and one in L. *major* [13]. A recent study with two L. *amazonensis* isolates obtained from dogs with clinical manifestations of the visceral disease indicated that most chromosomes have a depth coverage of reads compatible with a haploid number, except chromosome 30. Using the median read density of each chromosome normalized by the median read depth of the whole genome, we showed that most PH8 chromosomes have a haploid copy number of one, whereas chromosomes 1, 2, 5, 14, 16, 23, and 32 have a ploidy close to 1.5 and chromosome 30, a ploidy of 2 (Fig. 3A). Chromosomal copy number in the PH8 strain is similar, but not identical, to the results described for the UA301 strain, in which 9 trisomal chromosomes and 1 tetrasomal (chromosome 30) were identified [16].

Different degrees of mosaic aneuploidy had already been reported among *Leishmania* species [13,15,17] and within the same species [20,66]. Allelic frequency distribution analysis for all PH8 chromosomes confirms our ploidy estimates (Fig. 3B). Haploid chromosomes, which present the ratio between the median read density of the chromosome and the median read depth of the genome equal to 1, showed peaks for

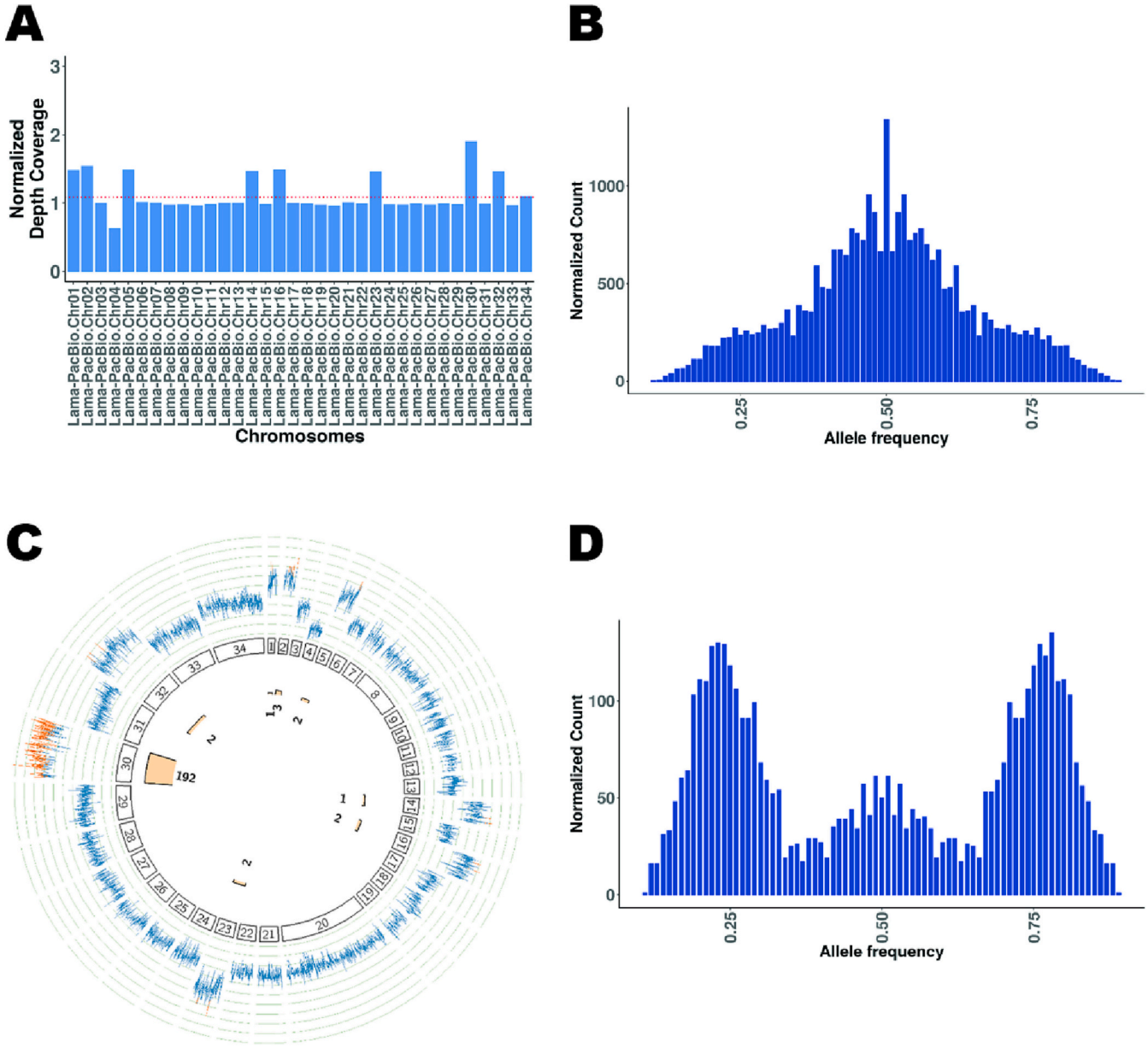

Fig. 3. Chromosome and gene copy number variation in the L. *amazonensis* PH8 strain. (A) Columns represent the estimated somy for each chromosome. The mean genome ploidy is indicated by a dotted red line. (B) Allele frequency of heterozygous SNPs for all chromosomes. The X axis contains allelic frequency values and the Y axis corresponds to the number of occurrences with that frequency. (C) Read depth coverage for each chromosome (internal boxes) along with the number of expanded genes. The mean read depth is shown as a line plot for each chromosome in blue for disomic regions and red for expanded genomic regions. The internal histogram displays the total number of gene expansions identified in each chromosome. (D) Allele frequency of heterozygous SNPs for chromosome 30. The X axis contains allelic frequency values and the Y axis corresponds to the number of occurrences with that frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

heterozygous SNP sites only in 0.5, while polyploids chromosomes this ratio close to 1.5 exhibit peaks at 0.3 or 0.6 (Fig. S6). Despite having extra copies, the read depth was homogeneously distributed throughout the entire chromosome 30, confirming the hypothesis of a complete amplification of this chromosome [20] (Fig. 3C). Consistent with a tetrasomic chromosomes, the divergence in somy for chromosome 30 (median read depth >1.5) showed allelic frequency peaks at 0.25 and 0.75, (Fig. 3D).

Gene duplication has been proposed as a mechanism to increase gene expression since transcriptional regulation control is an absence in trypanosomatid [67]. Using normalized read depth analysis to assess gene copy number variation in the PH8 strain, we identified 205 genes presenting a somy >1.8, 104 of them encoding hypothetical proteins with unknown functions (Table S4). Multigene families encoding amino acid transporters, serine-threonine phosphatase, ferredoxins, kinases, calpains, and GP63 metalloproteases are among the supernumerary genes with known function, several of them, being characterized as virulence factors and are discussed in the next section.

3.4. Characterization of multigene families encoding virulence factors

Genes encoding virulence factors in *Leishmania* are targets for several studies that have largely contributed to a better understanding of *Leishmania* pathogenesis. These molecules may act during cell invasion, escape from the host immune response and dissemination the disease to different host tissues [68]. As indicated above, several of these virulence factors are encoded by multigene families composed of a large number of copies, widely varying among species [13]. We have selected four multigene families encoding known virulence factors to analyze their complete sequence repertoire in the L. *amazonensis* PH8 genome and to compare with other *Leishmania* species deposited in the TritrypDB database (Table 3). For all multigene families analysed, A2 proteins, amastins and GP63 metalloproteases, we identified a greater number of members in the PH8 genome, compared to the genome of the L. *amazonensis* M2269 strain, which is the only L. *amazonensis* genome annotated so far (Table 3). This result indicates that several genes were not previously annotated possibly because this study was based on a highly fragmented genome [18] and a more complete gene repertoire was obtained with the current assembly of the PH8 genome.

Our analysis allowed a complete characterization of the complete gene set encoding A2 proteins in L. *amazonensis*. The A2 gene was initially described as an amastigote-specific sequence present in the L. *donovani* genome and encoding a 22-kDa protein containing repetitive amino acid sequences that are strongly recognized by sera from patients with VL [5]. Studies with L. *major* and L. *infantum* soon revealed that A2 is an amastigote-specific marker associated with species that cause VL [69]. A2 function is likely related to the survival of the parasite in macrophages of visceral organs since L. *donovani* amastigotes deficient in A2 protein are attenuated with respect to survival in visceral organs of infected mice [70–72]. In contrast to L. *major*, in which only one A2 pseudogene is present, as an ORFs with only 159 nucleotides [73], A2 genes are completely absent in other species of the *L. tropica* complex (L. *tropica* and *L. aethiopica*) and *L. braziliensis*[74]. Two copies of A2 genes are present in the genome of L. *mexicana*, 4 copies in L. *donovani* and 1 copy in L. *infantum*. Among the 9 sequences with homology to A2 present in the PH8 genome, 5 have complete ORFs whereas 4 are pseudogenes. Two of the five A2 genes were found on chromosome 22, the other three were found in small scaffolds that were not incorporated in the final assembly (scaffold278, Scaffold282 and Scaffold294, Fig. S7). The A2 proteins encoded by these genes possess between 600 and 1089 amino acids, depending on the number of amino acid repeats, which, similar to the L. *infantum* A2, constitute the largest portion of the protein. The L. *amazonensis* A2 protein structure is similar to A2 from other *Leishmania* species, with a N-terminal sequence followed by the repeated module of 10 amino acids VGP[Q/L]SVGPQS that occur 40 to 90 times [74,75]. Distinct from all other A2 proteins, a second repeat, SLLAR, is found in the L. *amazonensis* PH8 A2 protein (Fig. 4).

Alignment of amino acid sequences of A2 proteins of L. *amazonensis*, *L. mexicana*, L. *infantum* and *L. donovani* showed differences in the number of repeats that varies from 47 to 59 *in* L. *donovani* [76], but it is limited to 41 repeats in 4 out of 5 genes existing in L. *amazonensis* PH8 strain (Fig. 4 and Fig. S8). One A2 protein encoded in Scaffold 278 (Lama IV) has the G to D change in amino acid at the 7th position of the repeat that is present in the A2 proteins of L. *donovani* and is described as being important for the correct folding and function of A2 proteins. The presence of the A2 gene in the PH8 strain is consistent with previous reports showing that L. *amazonensis* expresses A2 proteins [77] and that isolates of L. *amazonensis* were identified in infected dogs with clinical manifestations of VL and TL [5]. Faced with these findings and together with the gene-editing protocol described here, the role of these genes in virulence and visceralization capacity in infections caused by L. *amazonensis* can be better investigated.

Amastins are another family of amastigote-specific proteins encoded by multigene families present in the genome of different *Leishmania* species and which have been characterized as virulence factors. Also present in the genome of other members of the Trypanosomatid family, amastin genes have undergone a major diversification in the genus *Leishmania* where four subfamilies, α, β, γ and δ-amastins, have been characterized [46]. Present at the surface of amastigotes, previous studies in which the expression of L. *braziliensis* δ-amastins has been knocked down by RNA intereference showed that they are essential for parasite growth in vitro and in vivo. Based on their structural homology with the thigh junction protein claudin and the knock down studies with L. *braziliensis* δ-amastins, it has been suggested that they mediate interactions between the surface of amastigotes and the membrane of the parasitophorous vacuole within the infected macrophage [78].

Compared to other *Leishmania* species, amastins have a smaller number of copies in L. *amazonensis*, with a total of 29 amastin genes and 2 pseudogenes distributed in 9 of the 34 chromosomes as well as in 2 small scaffolds that were not incorporated in the final assembly (Fig. 1 and Fig. S7). Amastin genes are organized in small tandem arrays varying from 2 to 4 genes and the size of the encoded proteins varies between 174 aa and 546 aa. Similar to the amastin gene family organization found in the genomes of *T. cruzi*, L. *infantum* and *L. braziliensis* [78–80], only δ-amastins were found associated with the tuzin gene orthologs. The 11 amino acid conserved amastin signature proposed by Rochette et al. in 2005 [79] was found in all PH8 amastin sequences (Fig. S9). Grouped into all four subfamilies, similar to other Leishmania species, δ-amastins constitute the largest sub-family (Fig. S10). Again, compared to previous analyses of the L. *amazonensis* amastin gene repertoire [18], several additional amastin gene copies have been identified in the PH8 genome (Table 3).

Besides amastins and A2, the metalloprotease GP63 has also been identified as a virulence factor in different *Leishmania* species. GP63 or leishmanolysin belongs to the family of zinc-dependent surface metalloproteases involved with events that facilitate the interaction of the parasite with macrophage receptors, allowing its subsequent

Table 3
Members belonging to multigene families present in the genome of L. *amazonensis* and other *Leishmania* species.

Family Species	A2 proteins	Amastins	GP63
L. amazonensis PH8 strain	9 (4)	31 (2)	11(2)
L. amazonensis 2269 strain	0	20 (2)	2
L. mexicana	2	48	7
L. infantum	1	68	13
L. donovani	4	33	3
L. major	63	0	6
L. braziliensis	55	0	33

*Numbers in parentheses indicate pseudogenes.

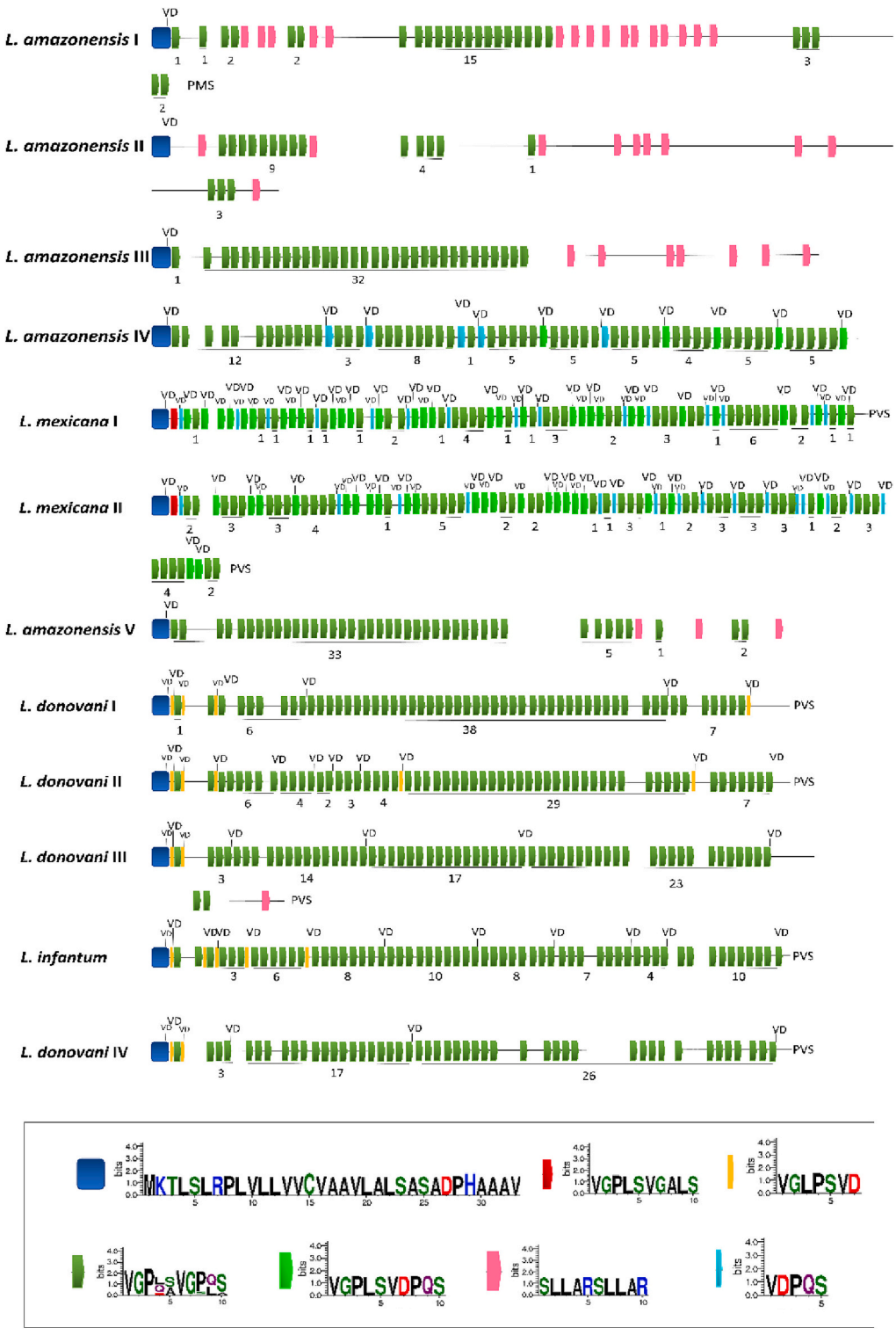

Fig. 4. Schematic alignment of A2 proteins annotated in different *Leishmania* species. Blue squares and coloured arrows represent, respectively, conserved regions at the N-terminus and repeat units present along each protein. Repeats sequences represented by green arrows are VGP(L/Q)SVG(P/S)QS or VGPQ(A/S) VGPLS or VGPEAVGPLS or VGPLSVGPQ (A/S) (dark green), VGPLSVDPQS (light green); yellow arrows represent VGLPSVD repeats, pink arrow represent SLLAR repeats and light blue arrows represent VDPQS repeats. The black line represents gaps in alignment that include other repeats not represented in L. *amazonensis* sequences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phagocytosis [81]. Intracellularly, GP63 has been found to be involved in the cleavage and degradation of various kinases and transcription factors. In the extracellular environment, GP63 acts by degrading extracellular components favouring parasite migration as well by inactivating the complement system cascade, increasing the parasite's resistance to complement-mediated lysis [82]. Secreted through exosomes, GP63 reduces cell activation and microbicidal activity against amastigotes and promastigotes [81,83]. Eleven genes encoding GP63 proteins were identified in the PH8 genome, 2 of them being pseudogenes (Table 3). Like the gene organization found in the L. *mexicana*, most of these genes are located on chromosome 10 as part of a single cluster of 7 genes, with other genes encoding GP63 located on chromosomes 28 and 30. Alignment of all 9 GP63 genes identified the HEXXH conserved domain, characteristic of the metalloprotease family of enzymes, responsible for the binding to a zinc atom, and the SRYD adhesive motif which is important in the binding of macrophage surface receptors (Fig. S11).

3.5. Characterization of genes encoding proteins related to heme and iron metabolism

Iron and heme are essential in many conserved metabolic pathways in *Leishmania*. Despite that, and even though, during their life cycle, these parasites are exposed to large shifts in nutrient availability, *Leishmania* lacks the capacity for heme synthesis and are also unable to store iron or heme [84–87]. In the past few years, several proteins involved in *Leishmania* iron and heme metabolism were identified as virulence factors given their importance for disease establishment and progression [85]. Among the L. *amazonensis* proteins that have been characterized as essential components for iron homeostasis are the ferrous iron transporter LIT1 and the ferric reductase 1 (LFR1), which reduces the translocated iron and the iron exporter LIR1, that controls iron levels in the cytosol [88–91]. Iron may also be acquired by heme uptake via LHR1 and LFLVCRb, or directly by hemoglobin via endocytosis of the hemoglobin receptor (HbR) [92–94]. The L. *amazonensis* mitoferrin 1 (LMIT1) was also identified as a transporter of iron to the parasite mitochondria [95], to be used as cofactors of Leishmania superoxide dismutase A (SODA) and ascorbate peroxidase (APX) [96,97]. Identified as an unusual mitochondrial ABC transporter, LABCB3 was shown to be required for the maturation of cytosolic iron/sulfur clusters in L. *major* [98]. Iron is also a cofactor of the SODB glycosomal isoform, shown to be an essential gene in *Leishmania* [99,100]. Since all these iron and heme-metabolism-related proteins are associated with parasite virulence, we screened the PH8 genome and identified 14 iron and heme-metabolism-related genes. Table S5 shows a comparison of the L. *amazonensis* gene repertoire related to iron metabolism with their orthologs present in the genomes from other *Leishmania* species, *T. brucei* and *T. cruzi*.

The genes encoding proteins of heme and iron metabolism of L. *amazonensis* are highly conserved among L. *mexicana*, *L. major*, L. *donovani* and *L. infantum*, but have diverged to a greater degree from the L. *braziliensis* genes. The first putative ferrous iron transporter identified in Trypanosomatids parasites [91], two copies in tandem (LIT1-1 and LIT1-2) are located on the tetraploid PH8 chromosome 30, suggesting that the extra copies favour L. *amazonensis* fitness. This genome location is similar to L. *mexicana* but distinct from the L. *major* genes, which are located on chromosome 31 (LmjF31.3060 and LmjF31.3070). It is noteworthy that, amastigote intracellular replication was completely abolished in LIT1 null mutants [91]. It is also worth mentioning that either *T. brucei*, which does not have an intracellular stage or *T. cruzi*, which grows as intracellular amastigotes in the cell cytosol does not possess orthologues of the LIT1 gene. Moreover, neither *T. cruzi* nor the *T. brucei* genome encodes homologues of the *Leishmania* heme transporters LFLVCRb or HbR. Also present in the genome of all Trypanosomatids parasites, iron-dependent superoxide dismutases (SODs) have been characterized as virulence factors in several *Leishmania* species as well as in *Trypanosoma* spp. Reverse genetics has been used to address the roles of L. *amazonensis* SODA, the mitochondrial SOD isoform and the mitochondrial iron importer LMIT1. Attempts to generate SODA null mutants showed it is an essential gene and parasites lacking one SODA allele have impaired capacity of differentiation into metacyclic promastigotes, they failed to replicate in macrophages and were severely attenuated in their ability to generate cutaneous lesions in mice [97]. Similarly, LMIT1 null mutants are not viable, consistent with the importance of iron for the assembly of Fe—S cluster proteins [95]. The characterization of the complete set of genes involved with iron and heme metabolism in L. *amazonensis* provided here is an important step towards a complete functional study that will fill one of the many gaps in our understanding of the *Leishmania* infection and may ultimately reveal new strategies for disease control.

3.6. Editing the Miltefosine transporter gene in L. amazonensis

Because it is more efficient and more accurate than any other gene manipulation method, functional genomics in protozoan parasites has largely benefited in the past few years from genome editing based on the clustered regularly interspaced short palindromic repeats (CRISPR) technology. A central component of CRISPR, the Cas9 nuclease edits genes by precisely cutting the DNA guided to a specific location within the target genome by base pairing with a small RNA (sgRNA). Once generated, the double-strand break (DSB) is repaired through the Non-Homologous End-Joining (NHEJ) or the Homologous Recombination Repair (HR) pathways [53]. Since genome analyses revealed that essential components of the NHEJ pathway are absent in Trypanosomatids, it has been proposed that, in the absence of a donor sequence to promote HR, DSB damage is repaired by the error-prone microhomology-mediated end joining (MMEJ). To establish an efficient protocol for CRISPR genome editing in L. *amazonensis*, we tested two strategies to generate knockout mutants of the miltefosine transporter (MT). The MT gene is a single copy gene, present in the L. *amazonensis* chromosome 13 that encodes a transmembrane protein of 1104 amino acids. The MT gene was selected as a target locus since disruption of both alleles results in knockout parasites that can be easily selected by their miltefosine-resistant phenotype. For the first strategy, we transfected L. *amazonensis* promastigotes with the pLDCN plasmid containing the *S. pyogenes* Cas9 nuclease coding region (SpCas9) with a nuclear localization signal and selected G418-resistant parasites. After confirming the expression of Cas9 by western blot (Fig. 5A), we transfected the Cas9 expressing cell line with in vitro transcribed sgRNA containing MT sequences and an oligonucleotide that serves as a donor sequence for HR repair. The sgRNA has a 20-nucleotide complementary to a sequence in the central part of the coding region of the MT gene. The oligonucleotide donor DNA contains 30 bp of homology arms flanking the MT gene, 3 stop codons *in tandem* and a restriction site for the *Hin*dIII endonuclease (Fig. S12A). Forty-eight hours after transfection, total DNA was purified from parasites that were cultivated in the presence of miltefosine as well as from wild type (WT) promastigotes. After PCR amplification with a pair of primers annealing in the MT gene and digestion of the amplicons with HindIII, the DNA fragments were submitted to agarose gel electrophoresis. As shown in Fig. 5B, amplified fragments corresponding to the MT sequence that was cleaved with Hind III confirmed the disruption of the MT gene in the population of transfected, miltefosine resistant parasites by the insertion of the stop codons in the MT gene as a result of HR with the donor DNA sequence.

As a second strategy to generate L. *amazonensis* cell lines with an edited genome, we tested a transfection protocol to deliver, through nucleofection, a ribonucleoprotein (RNP) complex composed of purified recombinant Cas9 complexed with in vitro transcribed sgRNA. Besides allowing genome editing without the need of creating a parasite expressing the Cas9 nuclease, transfection with the recombinant enzyme, which remains active only for a limited period, minimizes the possibility of off-target effects. *E. coli* expressing the *S. aureus* Cas9 (SaCas9) was used to obtain the purified nuclease as previously described [53]. An in vitro transcribed sgRNA with homology to a specific region of the gene of the L. *amazonensis* MT gene but containing scaffold sequences that are specific for SaCas9, was used to transfected L. *amazonensis* promastigotes with the RNP complex together with a donor DNA oligonucleotide. The nuclease activity of the SaCas9 protein was previously confirmed by incubating the enzymes complexed with the sgRNA with a PCR fragment containing the sequence of the *T. cruzi* GP72 gene (Fig. 5C). Since the DNA donor fragment contains MT sequences with stop codons and an *Xho*I restriction site (Fig. S12B), we used the same PCR/digestion strategy to confirm the insertion of stop codons in the MT gene. Because the L. *amazonensis* MT gene has one XhoI restriction site, digestion of PCR products amplified from DNA extracted from untransfected parasites generate 2 fragments whereas PCR products from transfected, miltefosine resistant parasites generated 3 fragments (Fig. 5D).

One of the main advantages of delivering Cas9 in the form of an RNP complexed with sgRNA is that it allows any cell to be manipulated

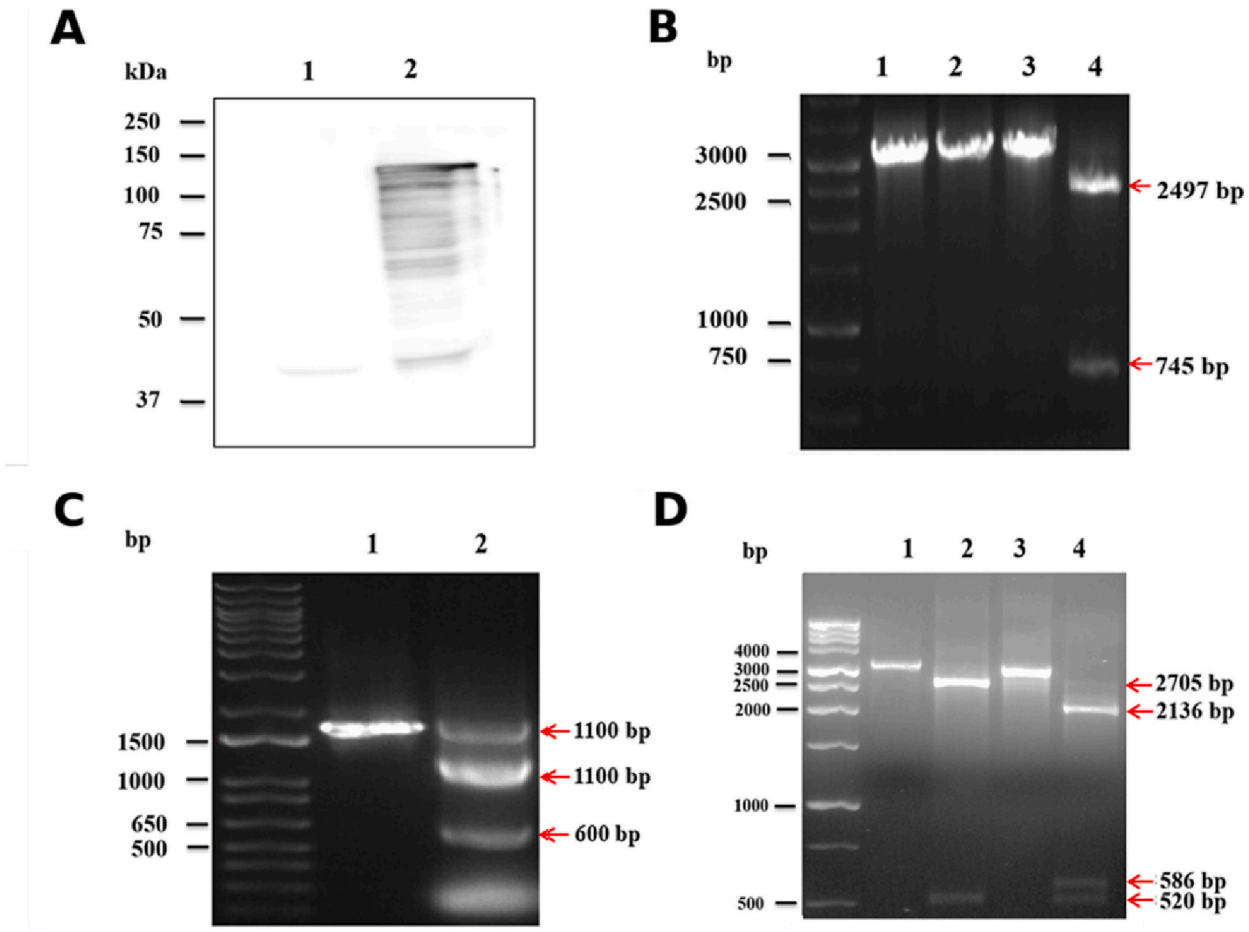

Fig. 5. Disruption of the L. *amazonensis* Miltefosine Transporter (MT) gene using two CRISPR editing strategies. (A) Western blot analysis of total protein extracts from promastigotes of wild type (1) or *L. amazonensis* transfected with the pLDCN plasmid containing the SpCas9 coding sequence (2) using an anti-Cas9 antibody. (B) DNA isolated from parasites expressing Cas9 and transfected with sgRNA targeting the MT gene together with a donor oligonucleotide containing MT sequences and a *Hin*dIII restriction site was used for PCR amplification with primers for the MT gene. After digestion of the 3.2Kbp PCR product with HindIII, the fragments were separated in agarose gel electrophoresis. Undigested (1) and digested (2) PCR products from parasites expressing Cas9 not transfected were used as negative controls. Undigested (3) and digested (4) PCR products from parasites expressing Cas9 and transfected with sgRNA targeting the MT gene and the donor oligonucleotide showed two fragments (2497 and 745 bp) after HindIII digestion. (C) In vitro activity assay to test the activity of purified recombinant SaCas9 protein. A 1,7 kb amplicon containing sequences of the *T. cruzi* GP72 gene was analysed by agarose gel electrophoresis before (1) and after (2) incubation with the SaCas9 protein and in vitro transcribed sgRNA targeting theGP72 gene. Digestion of the GP72 amplicon generates two fragments (1100 bp and 600 bp), as well as the undigested fragment. The small fragments at the bottom of the gel correspond to the sgRNA. (D) DNA isolated from parasites transfected with recombinant SaCas together with in vitro transcribed sgRNA targeting the MT gene and a donor oligonucleotide containing a *Xho*I site was used to amplify a 3.2 Kbp containing part of the MT gene. After digestion with XhoI, the PCR products were separated by agarose gel electrophoresis. Undigested (1) and digested (2) PCR products from non-transfected parasite DNA showed the existence of a restriction site for the XhoI endonuclease in the MT gene and the generation of two fragments (2705 and 520 bp) after XhoI digestion. Undigested (3) and digested (4) PCR products from transfected parasite DNA showed the insertion of a second XhoI restriction site and the generation of three fragments (2136, 586 and 520 bp) after XhoI digestion.

immediately without the need to create a transfected population expressing the nuclease. Because it is smaller than other Cas9 proteins, transfection with Cas9 from *S. aureus* results is highly efficient. Furthermore, as the transfected recombinant enzyme is only active for a limited period, off-target effects are likely minimized when RNP transfection is performed. Together, these results demonstrated that L. *amazonensis* mutants with a single copy gene disrupted by the insertion of stop codons can be generated using fast and highly efficient protocols based on two distinct CRISPR editing strategies. Together with efficient genome editing protocols, the availability of a fully sequenced and annotated genome opens endless possibilities for in depth studies on L. *amazonensis* genes, particularly when multigene families are the genes of interest. One relevant question that can be now addressed is related to the role of A2 genes that were characterized in L. *amazonensis* as a factor involved with visceral organ tropism. Using the CRISPR protocol, we have begun to create knock out cell lines for the A2 gene containing the repetitive motifs similar to L. *donovani*. Besides gene deletion methods to study of gene function, other genetic manipulation strategies based on CRISPR technology such as the generation of L. *amazonensis* overexpressing a specific protein or parasites with reporter or tagged genes, will also contribute to a better understanding of the molecular mechanisms involved in host-parasite interaction and the establishment of the infection. Considering also that *Leishmania* species vary considerably in their sensitivity to different drugs, the genome manipulation methodology proposed here will also allow us to advance in understanding the impact of resistance mechanisms on the therapeutic approach and discovery of new drugs, among other aspects [101]. Being an uncommon and still little-known parasite, which can cause almost all types of manifestations of leishmaniasis, the work presented here opens new doors for much needed studies, which may lead to improved methods of control and elimination of leishmaniasis.

4. Conclusions

In this work, we present the assembly of the nuclear and mitochondrial genomes of *Leishmania amazonensis* PH8 strain obtained from a hybrid approach that combines long and short sequencing reads as well as synteny with the *Leishmania mexicana* genome. This complete assembly of a genome from one of the most studied L. *amazonensis* strain allowed us to improve its annotation as shown by the identification of additional genes compared to the previously reported genome of the L. *amazonensis* strain 2269, which is the only strain with an annotated genome available in the TritrypDB database. Besides the analysis of gene content, analyses of chromosomal copy number variation also allowed us to confirm the diploid nature of L. *amazonensis* and the existence of extra copies for some of its chromosomes. Together with the improved annotation and gene organization analyses of multicopy gene families encoding virulence factors, the description of efficient CRISPR protocols to perform gene knockouts opens the possibility of directly addressing the role of these genes and to develop a deeper understanding on the genetic basis of parasite virulence.

Funding

This work received funds from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) and The Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV).

CRediT authorship contribution statement

Wanessa Moreira Goes: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Carlos Rodolpho Ferreira Brasil:** Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **João Luis Reis-Cunha:** Methodology, Formal analysis, Writing – review & editing. **Anderson Coqueiro-dos-Santos:** Methodology, Software, Formal analysis, Writing – original draft. **Viviane Grazielle-Silva:** Methodology, Investigation. **Júlia de Souza Reis:** Investigation. **Tatiane Cristina Souto:** Investigation. **Maria Fernanda Laranjeira-Silva:** Conceptualization, Investigation, Writing – original draft, Visualization. **Daniella Castanheira Bartholomeu:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Ana Paula Fernandes:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Santuza Maria Ribeiro Teixeira:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Data availability

all data is available

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110661.

References

[1] World Health Organization, Leishmaniasis. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis, 2021 (accessed December 17, 2021).

[2] G. Natarajan, S. Oghumu, S. Varikuti, A. Thomas, A. Satoskar, Mechanisms of immunopathology of leishmaniasis, Pathogen. Leishman.: New Develop. Res. (2013) 1–13, https://doi.org/10.1007/978-1-4614-9108-8_1.

[3] P.A.H.O. (PAHO), LEISHMANIOSES: Informe epidemiológico das Américas, 2021.

[4] M. Akhoundi, K. Kuhls, A. Cannet, J. Votýpka, P. Marty, P. Delaunay, D. Sereno, A historical overview of the classification, evolution, and dispersion of Leishmania parasites and sandflies, PLoS Negl. Trop. Dis. 10 (2016), https://doi.org/10.1371/journal.pntd.0004349.

[5] H.O. Valdivia, L.V. Almeida, B.M. Roatt, J.L. Reis-Cunha, A.A. Sampaio, C. Pereira, R.T. Gontijo, A.B. Fujiwara, M.J. Reis, J.A. Cotton Sanders, Daniella C. Bartholomeu, Comparative genomics of canine-isolated Leishmania (Leishmania) amazonensis from an endemic focus of visceral leishmaniasis in Governador Valadares, southeastern Brazil, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/srep40804.

[6] J.E. Tolezano, S.R.B. Uliana, H.H. Taniguchi, M.F.L. Araújo, J.A.R. Barbosa, J.E. R. Barbosa, L.M. Floeter-Winter, J.J. Shaw, The first records of Leishmania (Leishmania) amazonensis in dogs (Canis familiaris) diagnosed clinically as having canine visceral leishmaniasis from Araçatuba County, São Paulo State, Brazil, Vet. Parasitol. 149 (2007) 280–284, https://doi.org/10.1016/J.VETPAR.2007.07.008.

[7] L. Soong, C.A. Henard, P.C. Melby, Immunopathogenesis of non-healing American cutaneous leishmaniasis and progressive visceral leishmaniasis, Semin. Immunopathol. 34 (6) (2012) 735–751, https://doi.org/10.1007/S00281-012-0350-8.

[8] C.I. De Oliveira, C.I. Brodskyn, The immunobiology of Leishmania braziliensis infection, Front. Immunol. 3 (2012) 145, https://doi.org/10.3389/fimmu.2012.00145.

[9] B.A.S. Pereira, C.R. Alves, Immunological characteristics of experimental murine infection with Leishmania (Leishmania) amazonensis, Vet. Parasitol. 158 (2008) 239–255, https://doi.org/10.1016/J.VETPAR.2008.09.015.

[10] A.C. Ivens, J.M. Blackwell, Unravelling the Leishmania genome, Curr. Opin. Genet. Dev. 6 (1996) 704–710, https://doi.org/10.1016/S0959-437X(96)80024-4.

[11] A. Mannaert, T. Downing, H. Imamura, J.C. Dujardin, Adaptive mechanisms in pathogens: universal aneuploidy in Leishmania, Trends Parasitol. 28 (2012) 370–376, https://doi.org/10.1016/J.PT.2012.06.003.

[12] C.S. Peacock, K. Seeger, D. Harris, L. Murphy, J.C. Ruiz, M.A. Quail, N. Peters, E. Adlem, A. Tivey, M. Aslett, A. Kerhornou, A. Ivens, A. Fraser, M.-A. Rajandream, T. Carver, H. Norbertczak, T. Chillingworth, Z. Hance, K. Jagels, S. Moule, D. Ormond, S. Rutter, R. Squares, S. Whitehead, E. Rabbinowitsch, C. Arrowsmith, B. White, S. Thurston, F. Bringaud, S.L. Baldauf, A. Faulconbridge, D. Jeffares, D.P. Depledge, S.O. Oyola, J.D. Hilley, L.O. Brito, L.R.O. Tosi, B. Barrell, A.K. Cruz, J.C. Mottram, D.F. Smith, M. Berriman, Comparative genomic analysis of three Leishmania species that cause diverse human disease, Nat. Genet. 39 (2007) 839–847, https://doi.org/10.1038/ng2053.

[13] M.B. Rogers, J.D. Hilley, N.J. Dickens, J. Wilkes, P.A. Bates, D.P. Depledge, D. Harris, Y. Her, P. Herzyk, H. Imamura, T.D. Otto, M. Sanders, K. Seeger, J. C. Dujardin, M. Berriman, D.F. Smith, C. Hertz-Fowler, J.C. Mottram, Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania, Genome Res. 21 (2011) 2129–2142, https://doi.org/10.1101/GR.122945.111.

[14] P. Bastien, C. Blaineau, M. Pages, Leishmania: sex, lies and karyotype, Parasitol. Today 8 (1992) 174–177, https://doi.org/10.1016/0169-4758(92)90016-U.

[15] T. Downing, H. Imamura, S. Decuypere, T.G. Clark, G.H. Coombs, J.A. Cotton, J. D. Hilley, S. De Doncker, I. Maes, J.C. Mottram, M.A. Quail, S. Rijal, M. Sanders, G. Schönian, O. Stark, S. Sundar, M. Vanaerschot, C. Hertz-Fowler, J.-C. Dujardin, M. Berriman, Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance, Genome Res. 21 (2011) 2143–2156, https://doi.org/10.1101/gr.123430.111.

[16] L.H. Patino, C. Muskus, M. Muñoz, J.D. Ramírez, Genomic analyses reveal moderate levels of ploidy, high heterozygosity and structural variations in a Colombian isolate of Leishmania (Leishmania) amazonensis, Acta Trop. 203 (2020), 105296, https://doi.org/10.1016/j.actatropica.2019.105296.

[17] Y. Sterkers, L. Lachaud, N. Bourgeois, L. Crobu, P. Bastien, M. Pagès, Novel insights into genome plasticity in eukaryotes: mosaic aneuploidy in *Leishmania*, Mol. Microbiol. 86 (2012) 15–23, https://doi.org/10.1111/j.1365-2958.2012.08185.x.

[18] F. Real, R.O. Vidal, M.F. Carazzolle, J.M.C. Mondego, G.G.L. Costa, R.H. Herai, M. Würtele, L.M. De Carvalho, R.C.E. Ferreira, R.A. Mortara, C.L. Barbiéri, P. Mieczkowski, J.F. Da Silveira, M.R.D.S. Briones, G.A.G. Pereira, D. Bahia, The genome sequence of *Leishmania (Leishmania) amazonensis*: functional annotation and extended analysis of gene models, DNA Res. 20 (2013) 567–581, https://doi.org/10.1093/dnares/dst031.

[19] D.A. Tschoeke, G.L. Nunes, R. Jardim, J. Lima, A.S.R. Dumaresq, M.R. Gomes, L. de M. Pereira, D.R. Loureiro, P.H. Stoco, H.L. de Matos Guedes, A.B. de Miranda, J. Ruiz, A. Pitaluga, F.P. Silva, C.M. Probst, N.J. Dickens, J.C. Mottram, E. C. Grisard, A.M.R. Dávila, The comparative genomics and phylogenomics of Leishmania amazonensis parasite, Evol. Bioinforma. 10 (2014) 131–153, https://doi.org/10.4137/EBO.S13759.

[20] D. Batra, W. Lin, V. Narayanan, L.A. Rowe, M. Sheth, Y. Zheng, V. Loparev, M. de Almeida, Draft genome sequences of Leishmania (Leishmania) amazonensis, Leishmania (Leishmania) mexicana, and Leishmania (Leishmania) aethiopica, potential etiological agents of diffuse cutaneous Leishmaniasis, Microbiol. Resour. Announc. 8 (2019), https://doi.org/10.1128/mra.00269-19.

[21] H.O. Valdivia, L.V. Almeida, B.M. Roatt, J.L. Reis-Cunha, A.A.S. Pereira, C. Gontijo, R.T. Fujiwara, A.B. Reis, M.J. Sanders, J.A. Cotton, D.C. Bartholomeu, Comparative genomics of canine-isolated Leishmania (Leishmania) amazonensis from an endemic focus of visceral leishmaniasis in Governador Valadares, southeastern Brazil, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/srep40804.

[22] S. Kramer, N.C. Kimblin, M. Carrington, Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*, BMC Genomics 11 (2010) 283, https://doi.org/10.1186/1471-2164-11-283.

[23] N.M. El-Sayed, P.J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E.A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D. C. Bartholomeu, B.J. Haas, A.N. Tran, J.R. Wortman, U.C.M. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J.M. Carlton, G.C. Cerqueira, T. Creasy, A.L. Delcher, A. Djikeng, T.M. Embley, C. Hauser, A.C. Ivens, S. K. Kummerfeld, J.B. Pereira-Leal, D. Nilsson, J. Peterson, S.L. Salzberg, J. Shallom, J.C. Silva, J. Sundaram, S. Westenberger, O. White, S.E. Melville, J. E. Donelson, B. Andersson, K.D. Stuart, N. Hall, Comparative genomics of trypanosomatid parasitic protozoa, Science 309 (2005) (1979) 42, https://doi.org/10.1126/science.1112181.

[24] A. Butenko, A.Y. Kostygov, J. Sádlová, Y. Kleschenko, T. Bečvář, L. Podešvová, D. H. MacEdo, D. Žihala, J. Lukeš, P.A. Bates, P. Volf, F.R. Opperdoes, V. Yurchenko, Comparative genomics of Leishmania (Mundinia), BMC Genomics 20 (2019) 1–12, https://doi.org/10.1186/s12864-019-6126-y.

[25] S. Andrews, Babraham Bioinformatics - FastQC a Quality Control Tool for High Throughput Sequence Data, (n.d.). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed December 17, 2021).

[26] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

[27] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation, Genome Res. 27 (2017) 722–736, https://doi.org/10.1101/GR.215087.116.

[28] T.D. Otto, IPA: Script to Improve Long Read (Pacbio) Assemblies. https://github.com/ThomasDOtto/IPA, 2007 (accessed June 22, 2020).

[29] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinform. Appl. Note. 27 (2011) 578–579, https://doi.org/10.1093/bioinformatics/btq683.

[30] F. Nadalin, F. Vezzi, A. Policriti, GapFiller: a de novo assembly approach to fill the gap within paired reads, BMC Bioinform. 13 (2012), https://doi.org/10.1186/1471-2105-13-S14-S8.

[31] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (2014), e112963, https://doi.org/10.1371/journal.pone.0112963.

[32] S. Assefa, T.M. Keane, T.D. Otto, C. Newbold, M. Berriman, ABACAS: algorithm-based automatic contiguation of assembled sequences, Bioinformatics. 25 (2009) 1968–1969, https://doi.org/10.1093/bioinformatics/btp347.

[33] M. Manni, M.R. Berkeley, M. Seppey, F.A. Simão, E.M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, Mol. Biol. Evol. 38 (2021) 4647–4654, https://doi.org/10.1093/molbev/msab199.

[34] S. Nachtweide, M. Stanke, Multi-genome annotation with AUGUSTUS, Methods Mol. Biol. 2019 (1962) 139–160, https://doi.org/10.1007/978-1-4939-9173-0_8.

[35] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics. 26 (2010) 841–842, https://doi.org/10.1093/BIOINFORMATICS/BTQ033.

[36] T.D. Otto, G.P. Dillon, W.S. Degrave, M. Berriman, RATT: rapid annotation transfer tool, Nucleic Acids Res. 39 (2011), https://doi.org/10.1093/nar/gkq1268.

[37] S. Steinbiss, F. Silva-Franco, B. Brunk, B. Foth, C. Hertz-Fowler, M. Berriman, T. D. Otto, Companion: a web server for annotation and analysis of parasite genomes, Nucleic Acids Res. 44 (2016) W29, https://doi.org/10.1093/NAR/GKW292.

[38] E. Camacho, A. Rastrojo, Á. Sanchiz, S. González-de la Fuente, B. Aguado, J. M. Requena, Leishmania mitochondrial genomes: Maxicircle structure and heterogeneity of Minicircles, Genes (Basel) 10 (2019) 758, https://doi.org/10.3390/genes10100758.

[39] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, A.F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families, PNAS. 117 (2020) 9451–9457, https://doi.org/10.1186/s13059-018-1577-z.

[40] M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics 4 (2009), https://doi.org/10.1002/0471250953.BI0410S25.

[41] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics. 25 (2009) 1754, https://doi.org/10.1093/BIOINFORMATICS/BTP324.

[42] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics. 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

[43] L. Viana de Almeida, J. Luís Reis-Cunha, A. Coqueiro-dos-Santos, G. Flávia Rodrigues-Luís, R. de Paula Baptista, S. de Oliveira Silva, M. Norma de Melo, D. Castanheira Bartholomeu, Comparative genomics of Leishmania isolates from Brazil confirms the presence of Leishmania major in the Americas, Int. J. Parasitol. 51 (2021) 1047–1057, https://doi.org/10.1016/J.IJPARA.2021.05.009.

[44] W. Wang, D. Peng, R.P. Baptista, Y. Li, J.C. Kissinger, R.L. Tarleton, Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease, PLoS Pathog. 17 (2021), e1009254, https://doi.org/10.1371/JOURNAL.PPAT.1009254.

[45] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier, Nucleic Acids Res. 33 (2005), https://doi.org/10.1093/NAR/GKI442.

[46] A.P. Jackson, The Evolution of Amastin Surface Glycoproteins in Trypanosomatid Parasites, 2009, https://doi.org/10.1093/molbev/msp214.

[47] K. Katoh, K. Misawa, K.I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066, https://doi.org/10.1093/NAR/GKF436.

[48] S. Guindon, F. Delsuc, J.F. Dufayard, O. Gascuel, Estimating maximum likelihood phylogenies with PhyML, Methods Mol. Biol. 537 (2009) 113–137, https://doi.org/10.1007/978-1-59745-251-9_6.

[49] F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fit models of protein evolution, Bioinformatics. 21 (2005) 2104–2105, https://doi.org/10.1093/BIOINFORMATICS/BTI263.

[50] I.M. Wallace, O. O'Sullivan, D.G. Higgins, C. Notredame, M-coffee: combining multiple sequence alignment methods with T-coffee, Nucleic Acids Res. 34 (2006) 1692, https://doi.org/10.1093/NAR/GKL091.

[51] M. Arenas, Trends in substitution models of molecular evolution, Front. Genet. 6 (2015) 319, https://doi.org/10.3389/fgene.2015.00319.

[52] W.-W. Zhang, P. Lypaczewski, G. Matlashewski, Optimized CRISPR-Cas9 genome editing for Leishmania and its use to target a multigene family, induce chromosomal translocation, and study DNA break repair mechanisms, MSphere. 2 (2017), https://doi.org/10.1128/mSphere.00340-16.

[53] G.A. Burle-Caldas, M. Soares-Simões, L. Lemos-Pechnicki, W.D. DaRocha, S.M. R. Teixeira, Assessment of two CRISPR-Cas9 genome editing protocols for rapid generation of Trypanosoma cruzi gene knockout mutants, Int. J. Parasitol. 48 (2018) 591–596, https://doi.org/10.1016/J.IJPARA.2018.02.002.

[54] D.C. Bartholomeu, S.M.R. Teixeira, A.K. Cruz, Preto, genomics and functional genomics in *Leishmania* and *Trypanosoma cruzi*: statuses, challenges and perspectives, Mem. Inst. Oswaldo Cruz 116 (2021) 1–21, https://doi.org/10.1590/0074-02760200634.

[55] A.M. Giani, G.R. Gallo, L. Gianfranceschi, G. Formenti, Long walk to genomics: history and current approaches to genome sequencing and assembly, Comput. Struct. Biotechnol. J. 18 (2020) 9–19, https://doi.org/10.1016/J.CSBJ.2019.11.002.

[56] J.M. Ubeda, F. Raymond, A. Mukherjee, M. Plourde, H. Gingras, G. Roy, A. Lapointe, P. Leprohon, B. Papadopoulou, J. Corbeil, M. Ouellette, Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite Leishmania, PLoS Biol. 12 (2014), e1001868, https://doi.org/10.1371/JOURNAL.PBIO.1001868.

[57] S. González-De La Fuente, E. Camacho, R. Peiró-Pastor, A. Rastrojo, F. Carrasco-Ramiro, B. Aguado, J.M. Requena, Complete and de novo assembly of the Leishmania braziliensis (M2904) genome, Mem. Inst. Oswaldo Cruz 114 (2019) 1–6, https://doi.org/10.1590/0074-02760180438.

[58] L. Simpson, The mitochondrial genome of Kinetoplastid Protozoa: genomic organization, transcription, replication, and evolution, Annu. Rev. Microbiol. 41 (1987) 363–380, https://doi.org/10.1146/annurev.mi.41.100187.002051.

[59] S. Hajduk, T. Ochsenreiter, RNA editing in kinetoplastids, RNA Biol. 7 (2010) 229–236, https://doi.org/10.4161/rna.7.2.11393.

[60] J.M. Shaw, J.E. Feagin, K. Stuart, L. Simpson, Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons, Cell. 53 (1988) 401–411, https://doi.org/10.1016/0092-8674(88)90160-2.

[61] L. Simpson, S. Sbicego, R. Aphasizhev, Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business, RNA. 9 (2003) 265–276, https://doi.org/10.1261/rna.2178403.

[62] D.A. Maslov, N.R. Sturm, B.M. Niner, E.S. Gruszynski, M. Peris, L. Simpson, An intergenic G-rich region in Leishmania tarentolae kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12, Mol. Cell. Biol. 12 (1992) 56–67, https://doi.org/10.1128/mcb.12.1.56.

[63] D.A. Maslov, O. Thiemann, L. Simpson, Editing and misediting of transcripts of the kinetoplast maxicircle G5 (ND3) cryptogene in an old laboratory strain of Leishmania tarentolae, Mol. Biochem. Parasitol. 68 (1994) 155–159, https://doi.org/10.1016/0166-6851(94)00160-X.

[64] O.H. Thiemann, D.A. Maslov, L. Simpson, Disruption of RNA editing in Leishmania tarentolae by the loss of minicircle-encoded guide RNA genes, EMBO J. 13 (1994) 5689, https://doi.org/10.1002/J.1460-2075.1994.TB06907.X.

[65] G.J. Bhat, D.J. Koslowsky, J.E. Feagin, B.L. Smiley, K. Stuart, An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6, Cell. 61 (1990) 885–894, https://doi.org/10.1016/0092-8674(90)90199-O.

[66] F. Dumetz, H. Imamura, M. Sanders, V. Seblova, J. Myskova, P. Pescher, M. Vanaerschot, C.J. Meehan, B. Cuypers, G. De Muylder, G.F. Späth, G. Bussotti, J.R. Vermeesch, M. Berriman, J.A. Cotton, P. Volf, J.C. Dujardin, M. A. Domagalska, Modulation of aneuploidy in leishmania donovani during

adaptation to different in vitro and in vivo environments and its impact on gene expression, MBio. 8 (2017), https://doi.org/10.1128/mBio.00599-17.
[67] S.A. Iantorno, C. Durrant, A. Khan, M.J. Sanders, S.M. Beverley, W.C. Warren, M. Berriman, D.L. Sacks, J.A. Cotton, M.E. Grigg, Gene expression in Leishmania is regulated predominantly by gene dosage, MBio. 8 (2017), https://doi.org/10.1128/MBIO.01393-17.
[68] E. Bifeld, J. Clos, The genetics of Leishmania virulence, Med. Microbiol. Immunol. 204 (2015) 619–634, https://doi.org/10.1007/S00430-015-0422-1.
[69] E. Ghedin, F. Bringaud, J. Peterson, P. Myler, M. Berriman, A. Ivens, B. Andersson, E. Bontempi, J. Eisen, S. Angiuoli, D. Wanless, A. Von Arx, L. Murphy, N. Lennard, S. Salzberg, M.D. Adams, O. White, N. Hall, K. Stuart, C. M. Fraser, N.M.A. El-Sayed, Gene synteny and evolution of genome architecture in trypanosomatids, Mol. Biochem. Parasitol. 134 (2004) 183–191, https://doi.org/10.1016/j.molbiopara.2003.11.012.
[70] W.W. Zhang, G. Matlashewski, Loss of virulence in Leishmania donovani deficient in an amastigote-specific protein, A2, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 8807, https://doi.org/10.1073/PNAS.94.16.8807.
[71] W.W. Zhang, G. Matlashewski, Characterization of the A2–A2rel gene cluster in Leishmania donovani: involvement of A2 in visceralization during infection, Mol. Microbiol. 39 (2001) 935–948, https://doi.org/10.1046/J.1365-2958.2001.02286.X.
[72] L.I. McCall, G. Matlashewski, Localization and induction of the A2 virulence factor in Leishmania: evidence that A2 is a stress response protein, Mol. Microbiol. 77 (2010) 518–530, https://doi.org/10.1111/J.1365-2958.2010.07229.X.
[73] W.W. Zhang, S. Mendez, A. Ghosh, P. Myler, A. Ivens, J. Clos, D.L. Sacks, G. Matlashewski, Comparison of the A2 gene locus in Leishmania donovani and Leishmania major and its control over cutaneous infection *, J. Biol. Chem. 278 (2003) 35508–35515, https://doi.org/10.1074/JBC.M305030200.
[74] W.W. Zhang, H. Charest, E. Ghedin, G. Matlashewski, Identification and overexpression of the A2 amastigote-specific protein in Leishmania donovani, Mol. Biochem. Parasitol. 78 (1996) 79–90, https://doi.org/10.1016/S0166-6851(96)02612-6.
[75] C. H, M. G, Developmental gene expression in Leishmania donovani: differential cloning and analysis of an amastigote-stage-specific gene, Mol. Cell. Biol. 14 (1994) 2975–2984, https://doi.org/10.1128/MCB.14.5.2975-2984.1994.
[76] P. Lypaczewski, J. Hoshizaki, W.W. Zhang, L.I. McCall, J. Torcivia-Rodriguez, V. Simonyan, A. Kaur, K. Dewar, G. Matlashewski, A complete Leishmania donovani reference genome identifies novel genetic variations associated with virulence, Sci. Rep. 8 (2018) 1–14, https://doi.org/10.1038/s41598-018-34812-x.
[77] F.A.A. Carvalho, H. Charest, C.A.P. Tavares, G. Matlashewski, E.P. Valente, A. Rabello, R.T. Gazzinelli, A.P. Fernandes, Diagnosis of American visceral leishmaniasis in humans and dogs using the recombinant Leishmania donovani A2 antigen, Diagn. Microbiol. Infect. Dis. 43 (2002) 289–295, https://doi.org/10.1016/S0732-8893(02)00410-8.
[78] R.M.C. de Paiva, V. Grazielle-Silva, M.S. Cardoso, B.N. Nakagaki, R.P. Mendonça-Neto, A.M.C. Canavaci, N. Souza Melo, P.M. Martinelli, A.P. Fernandes, W.D. da Rocha, S.M.R. Teixeira, Amastin knockdown in Leishmania braziliensis affects parasite-macrophage interaction and results in impaired viability of intracellular amastigotes, PLoS Pathog. 11 (2015) e1005296, https://doi.org/10.1371/journal.ppat.1005296.
[79] A. Rochette, F. McNicoll, J. Girard, M. Breton, É. Leblanc, M.G. Bergeron, B. Papadopoulou, Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in Leishmania spp, Mol. Biochem. Parasitol. 140 (2005) 205–220, https://doi.org/10.1016/j.molbiopara.2005.01.006.
[80] P.R. Araújo, S.M. Teixeira, Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in trypanosoma cruzi - a review, Mem. Inst. Oswaldo Cruz 106 (2011) 257–266, https://doi.org/10.1590/S0074-02762011000300002.
[81] M. Olivier, V.D. Atayde, A. Isnard, K. Hassani, M.T. Shio, Leishmania virulence factors: focus on the metalloprotease GP63, Microbes Infect. 14 (2012) 1377–1389, https://doi.org/10.1016/j.micinf.2012.05.014.
[82] A. Isnard, M.T. Shio, M. Olivier, Impact of Leishmania metalloprotease GP63 on macrophage signaling, Front. Cell. Infect. Microbiol. 2 (2012) 72, https://doi.org/10.3389/fcimb.2012.00072.
[83] B. McGwire, K.P. Chang, Genetic rescue of surface metalloproteinase (gp63)-deficiency in Leishmania amazonensis variants increases their infection of macrophages at the early phase, Mol. Biochem. Parasitol. 66 (1994) 345–347, https://doi.org/10.1016/0166-6851(94)90160-0.
[84] L. Kořený, M. Oborník, J. Lukeš, Make it, take it, or leave it: Heme metabolism of parasites, PLoS Pathog. 9 (2013), e1003088, https://doi.org/10.1371/JOURNAL.PPAT.1003088.
[85] Chin Shen Chang, Kwang-Poo Chang, Heme requirement and acquisition by extracellular and intracellular stages of Leishmania mexicana amazonensis, Mol. Biochem. Parasitol. 16 (1985) 267–276, https://doi.org/10.1016/0166-6851(85)90069-6.
[86] A.R. Flannery, C. Huynh, B. Mittra, R.A. Mortara, N.W. Andrews, LFR1 ferric iron reductase of Leishmania amazonensis is essential for the generation of infective parasite forms, J. Biol. Chem. 286 (2011) 23266–23279, https://doi.org/10.1074/jbc.M111.229674.
[87] M.F. Laranjeira-Silva, I. Hamza, J.M. Pérez-Victoria, Iron and Heme Metabolism at the Leishmania–Host Interface, Trends Parasitol. 36 (2020) 279–289, https://doi.org/10.1016/J.PT.2019.12.010.
[88] M.F. Laranjeira-Silva, W. Wang, T.K. Samuel, F.Y. Maeda, V. Michailowsky, I. Hamza, Z. Liu, N.W. Andrews, A MFS-like plasma membrane transporter required for Leishmania virulence protects the parasites from iron toxicity, PLoS Pathog. 14 (2018), e1007140, https://doi.org/10.1371/JOURNAL.PPAT.1007140.
[89] A.R. Flannery, R.L. Renberg, N.W. Andrews, Pathways of iron acquisition and utilization in Leishmania, Curr. Opin. Microbiol. 16 (2013) 716–721, https://doi.org/10.1016/j.mib.2013.07.018.
[90] A. Sarkar, N.W. Andrews, M.F. Laranjeira-Silva, Intracellular iron availability modulates the requirement for Leishmania Iron regulator 1 (LIR1) during macrophage infections, Int. J. Parasitol. 49 (2019) 423–427, https://doi.org/10.1016/J.IJPARA.2019.02.002.
[91] C. Huynh, D.L. Sacks, N.W. Andrews, A Leishmania amazonensis ZIP family iron transporter is essential for parasite replication within macrophage phagolysosomes, J. Exp. Med. 203 (2006) 2363–2375, https://doi.org/10.1084/JEM.20060559.
[92] D.C. Miguel, A.R. Flannery, B. Mittra, N.W. Andrews, Heme uptake mediated by lhr1 is essential for leishmania amazonensis virulence, Infect. Immun. 81 (2013) 3620–3626, https://doi.org/10.1128/IAI.00687-13.
[93] S. Agarwal, R. Rastogi, D. Gupta, N. Patel, M. Raje, A. Mukhopadhyay, Clathrin-mediated hemoglobin endocytosis is essential for survival of Leishmania, Biochim. Biophys. Acta (BBA) – Mol. Cell Res. 1833 (2013) 1065–1077, https://doi.org/10.1016/J.BBAMCR.2013.01.006.
[94] M. Cabello-Donayre, L.M. Orrego, E. Herráez, P. Vargas, M. Martínez-García, J. Campos-Salinas, I. Pérez-Victoria, B. Vicente, J.J.G. Marín, J.M. Pérez-Victoria, Leishmania heme uptake involves LmFLVCRb, a novel porphyrin transporter essential for the parasite, Cell. Mol. Life Sci. 77 (2019) 1827–1845, https://doi.org/10.1007/S00018-019-03258-3.
[95] B. Mittra, M.F. Laranjeira-Silva, J. Perrone Bezerra de Menezes, J. Jensen, V. Michailowsky, N.W. Andrews, A Trypanosomatid Iron transporter that regulates mitochondrial function is required for Leishmania amazonensis virulence, PLoS Pathog. 12 (2016), e1005340, https://doi.org/10.1371/JOURNAL.PPAT.1005340.
[96] L. Xiang, M.F. Laranjeira-Silva, F.Y. Maeda, J. Hauzel, N.W. Andrews, B. Mittra, Ascorbate-dependent peroxidase (APX) from leishmania amazonensis is a reactive oxygen species-induced essential enzyme that regulates virulence, Infect. Immun. 87 (2019), https://doi.org/10.1128/IAI.00193-19/SUPPL_FILE/IAI.00193-19-S0001.PDF.
[97] B. Mittra, M.F. Laranjeira-Silva, D.C. Miguel, J. Perrone Bezerra De Menezes, N. W. Andrews, The iron-dependent mitochondrial superoxide dismutase SODA promotes Leishmania virulence, J. Biol. Chem. 292 (2017) 12324–12338, https://doi.org/10.1074/jbc.M116.772624.
[98] M. Martínez-García, J. Campos-Salinas, M. Cabello-Donayre, E. Pineda-Molina, F. J. Gálvez, L.M. Orrego, M.P. Sánchez-Cañete, S. Malagarie-Cazenave, D. M. Koeller, J.M. Pérez-Victoria, LmABCB3, an atypical mitochondrial ABC transporter essential for Leishmania major virulence, acts in heme and cytosolic iron/sulfur clusters biogenesis, Parasit. Vectors 9 (2016) 1–17, https://doi.org/10.1186/s13071-015-1284-5.
[99] K.A. Plewes, S.D. Barr, L. Gedamu, Iron superoxide dismutases targeted to the glycosomes of Leishmania chagasi are important for survival, Infect. Immun. 71 (2003) 5910–5920, https://doi.org/10.1128/IAI.71.10.5910-5920.2003.
[100] B.J. Davenport, C.G. Martin, S.M. Beverley, D.J. Orlicky, A. Vazquez-Torres, T. E. Morrison, SODB1 is essential for Leishmania major infection of macrophages and pathogenesis in mice, PLoS Negl. Trop. Dis. 12 (2018), e0006921, https://doi.org/10.1371/JOURNAL.PNTD.0006921.
[101] S.L. Croft, S. Sundar, A.H. Fairlamb, Drug resistance in Leishmaniasis, Clin. Microbiol. Rev. (2006), https://doi.org/10.1128/CMR.19.1.111-126.2006.