MICROBIAL GENOMICS

RESEARCH ARTICLE

Patino *et al.*, *Microbial Genomics* 2021;7:000640

DOI 10.1099/mgen.0.000640

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Revisiting the heterogeneous global genomic population structure of *Leishmania infantum*

Luz H. Patino[1], Adriana Castillo-Castañeda[1], Marina Muñoz[1], Carlos Muskus[2], Matilde Rivero-Rodríguez[3,4], Alveiro Pérez-Doria[3], Eduar E. Bejarano[3] and Juan David Ramírez[1,*]

Abstract

Leishmania infantum is the main causative agent responsible for visceral leishmaniasis (VL), a disease with global distribution. The genomic structure and genetic variation of this species have been widely studied in different parts of the world. However, in some countries, this information is still yet unknown, as is the genomic behaviour of the main antigens used in VL diagnosis (rK39 and rK28), which have demonstrated variable sensitivity and specificity in a manner dependent on the geographic region analysed. The objective of this study was to explore the genomic architecture and diversity of four Colombian *L. infantum* isolates obtained in this study and to compare these results with the genetic analysis of 183 *L. infantum* isolates from across the world (obtained from public databases), as well as to analyse the whole rK39 and rK28 antigen sequences in our dataset. The results showed that, at the global level, *L. infantum* has high genetic homogeneity and extensive aneuploidy. Furthermore, we demonstrated that there are distinct populations of *L. infantum* circulating in various countries throughout the globe and that populations of distant countries have close genomic relationships. Additionally, this study demonstrated the high genetic variability of the rK28 antigen worldwide. In conclusion, our study allowed us to (i) expand our knowledge of the genomic structure of global *L. infantum*; (ii) describe the intra-specific genomic variability of this species; and (iii) understand the genomic characteristics of the main antigens used in the diagnosis of VL. Additionally, this is the first study to report whole-genome sequences of Colombian *L. infantum* isolates.

DATA SUMMARY

The four genomes sequenced for this study can be found in the EBI European Nucleotide Archive (ENA) under project number PRJEB42749. The 183 sequences used in our comparative analyses were download from ENA under accession numbers PRJEB2473, PRJNA494801, PRJEB15063, PRJEB2600, PRJNA589999, PRJNA607007, PRJNA658892 and PRJEB3283. All sequences are summarized along with accession numbers in Table S1 (available in the online version of this article).

INTRODUCTION

Visceral leishmaniasis (VL) is considered the most severe form of leishmaniasis in the world [1, 2]. This disease is endemic on the Indian subcontinent, and in East Africa and the Americas (mainly in Brazil) [1, 3]. VL pathology is caused mainly by parasites of the *Leishmania donovani* complex (*L. infantum/chagasi* and *L. donovani*) [4], and *L. infantum* is the species responsible for VL in the Americas [5–7].

Since 2007, when the whole genome of *L. infantum* was sequenced and published for the first time, various studies using DNA-sequencing technology revealed important findings for this species. Those research efforts facilitated an understanding of its genetic structure [8]; the analysis of the genomic variations associated with the clinical features, hosts and treatment responses [9, 10]; the identification of markers of drug resistance [11]; the determination of the genomic diversity in geographically restricted regions [9, 10, 12, 13] and globally [14]; as well as the low

Received 24 March 2021; Accepted 22 June 2021; Published 07 September 2021

Author affiliations: [1]Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia; [2]Programa de Estudios y Control de Enfermedades Tropicales (PECET), Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia; [3]Grupo de Investigaciones Biomédicas, Universidad de Sucre, Sincelejo, Colombia; [4]Candidata a doctor en Medicina Tropical, Universidad de Cartagena-SUE Caribe, Colombia.

***Correspondence:** Juan David Ramírez, juand.ramirez@urosario.edu.co

Keywords: Aneuploidy; genomic diversity; *Leishmania*; whole genome sequencing.

The four samples sequenced for this study were deposited in the EBI European Nucleotide Archive (ENA) under project number PRJEB42749.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary tables and four supplementary figures are available with the online version of this article.

plasticity and intra-specific genomic variability that characterize this *Leishmania* species [10, 14, 15]. To date, these studies have focused mainly in specific geographical areas, such as Brazil, leaving the genomic diversity of this species in other regions of the world unclear. Although the incidence of VL is low in other regions, there has been an increasing number of cases in recent years [3], such as in Colombia, which has reported 244 cases of VL in the last 10 years [16].

Because of the clinical and epidemiological importance of VL worldwide, different serological and molecular tools currently exist that allow fast and effective diagnosis [17, 18]. One of the serological techniques used is the immunochromatographic rapid diagnostic test, which is based on two antigens: the rK39 recombinant antigen, derived from part of an *L. infantum* kinesin-related gene [19], and the rK28 recombinant antigen, developed by fusing multiple tandem repeat sequences of *L. donovani* haspb1 and k39 kinesin genes [20]. Although these techniques are rapid and easy to perform, their sensitivity is variable, and the diagnostic accuracy has been shown to depend on the geographic region. Although the rK39 antigen test has been reported to have high sensitivity in Indian populations (95–97.0%), in other parts of the world, the reported sensitivities are variable or generally lower, as is the case in populations in East Africa, where a sensitivity of 75–85.3% has been reported [17, 20–22]; France, where the test has a sensitivity of 88.5% [23]; Brazil, where it has a sensitivity of 84.7–93% [22, 24]; and finally Colombia, which has reported a sensitivity of 82.9–91.5% [18]. Regarding the recombinant antigen (rK28), reports have described a high sensitivity for the diagnosis of VL in countries such as Sudan, Ethiopia, India and Brazil, in which sensitivities between 94 and 100% were observed [25–27]. The reason for the different test performances in these areas is unclear, but one reason may be the ability of different *Leishmania* strains to induce the production of specific host antibodies [23, 24]. Another possibility is the molecular diversity of the rK39/rK28 sequences among global strains, which could induce a differential antibody response in the host, antibodies that could have a low affinity by the recombinant K39/K28 antigen commonly used in the immunochromatographic tests, making this test decrease its diagnostic performance. Based on this hypothesis, we believe it is necessary to evaluate the currently unknown rK39/rK28 sequences of strains from different parts of the world.

To date, no studies have analysed the whole genomes of *L. infantum* strains in Colombia or analysed and compared the genomic sequences of the main antigens used in VL diagnosis (rK39/rK28) throughout the world. Therefore, the focus of this study was to analyse the whole genome of four different *L. infantum* isolates collected from humans and dogs in two regions of Colombia, evaluate the genomic variation among the isolates and diversity across the globe (ploidy changes, structural variations and single-point mutations), and analyse the phylogenomic relationships (based on the nuclear/mitochondrial genomes and the recombinants antigens rK39/rK28) within 183 publicly available *L. infantum* genomes from the five continents corresponding to the known broad geographic distribution of the parasite.

Impact Statement

Leishmania infantum is a protozoan parasite responsible for one of the most severe forms of leishmaniasis in the world, the visceral leishmaniasis (VL). In this work we used DNA-sequencing to analyse the genetic structure of this parasite and to describe the genomic behaviour of the main antigens used in VL diagnosis (rK39 and rK28 antigen). The results obtained allow us to expand the knowledge about the intraspecific genomic variability of this parasite to global level and to understand the possible reasons why the tests used in the VL diagnosis present a variable sensitivity and specificity according to the geographical area evaluated.

METHODS

Ethics statement

The genomes were obtained from isolates collected from human patients and dogs (*Canis lupus familiaris*). To collect and use these samples, this study was approved by the Ethics Committee of the Universidad de Antioquia (Number VRI3445/2010) and the Universidad de Sucre in accordance with resolution numbers 36836 and 2–2012, respectively. Written informed consent was obtained from the patients from whom the strains were isolated.

Study population

A total of 187 whole genomes, which were retrieved from ENA databases, were analysed. These included four new Colombian genomes sequenced in this investigation, two of which were isolated from humans and two from dogs with clinical signs of VL. The other 183 were downloaded from public databases: 36 isolates from Asia (Turkey *N*=13, Israel *N*=8, China *N*=6, Cyprus *N*=5, Uzbekistan *N*=2 and Palestine *N*=2) [12, 14], 17 from Europe (France *N*=5, Italy *N*=5, Portugal *N*=2 and Spain *N*=5) [14, 28], 9 from North Africa (Morocco *N*=2, Zaghouan *N*=1, Nabeul *N*=1, Kairouan *N*=5) [13, 14, 29], and 121 from Central and South America (Brazil *N*=118, Honduras *N*=2 and Panama *N*=1) [9, 14, 15, 28, 30]. All metadata on the 187 genomes, including ENA accession numbers for individual samples, are summarized in Table S1. The public dataset was downloaded using the tools described in https://github.com/EnzoAndree/getENA.

Leishmania infantum culture conditions, DNA extraction and species identification

The promastigotes from four Colombian clinical isolates were grown in Schneider medium supplemented with 10% (v/v) foetal bovine serum and cultured at 26 °C with 5% CO_2. Approximately 1×10^6 promastigotes in the late logarithmic growth phase were cultured and harvested by centrifugation for subsequent DNA extraction, which was conducted using the High Pure PCR Template Preparation Kit (Roche Life Science, Mannheim, Germany) in accordance with the

manufacturer's instructions. The DNA concentrations were determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the DNA quality and integrity were determined by 1% agarose gel electrophoresis. Once the quality of the DNA was verified, each sample was divided into two aliquots: one for species identification and the other for whole-genome sequencing. Species identification was performed using direct Sanger sequencing of genes encoding cytochrome b molecules and heat shock protein 70, as described previously [31, 32]. The amplification products were purified with EXOSAP (Affymetrix, Santa Clara, CA, USA) and sequenced using the dideoxy-terminal method on an automated capillary sequencer (AB3730; Applied Biosystems, Foster City, CA, USA). Subsequently, the sequences were subjected to BLASTN [33] for a similarity search with the *Leishmania* sequences deposited in GenBank [31]. The HSP-70 sequences obtained from the four Colombian clinical isolates were aligned with HSP70 sequences from other *Leishmania* species using CLUSTAL W. The analysis was done with sequences of *L. peruviana* (HF586368.1), *L. braziliensis* (GU071173.1), *L. guyanensis* (EU599093.1), *L. panamensis* (FN395055.1), *L. mexicana* (HF586413.1), *L. amazonensis* (HF586354.1) and *L. infantum* (JN676923.1). Later, maximum-likelihood (ML) trees were inferred using IQ-TREE 2 [34]. The robustness of the nodes was evaluated using the bootstrap method (BT, with 1000 replicates). Finally, the obtained tree was visualized and edited using the Interactive Tree Of Life V4 (http://itol.embl.de) [35].

Whole-genome sequencing of clinical isolates and DNA mapping

The extracted whole-genome DNA was sequenced on a HiSeq X-Ten system (Illumina; Novogene Bioinformatics Technology, Beijing, PR China). Briefly, mate-paired libraries were constructed by end repair (350 bp insert size) and subjected to paired-end sequencing (2×150 bp read length). Paired reads were discarded when reads with adapter contamination, >10% uncertain nucleotides, or >50% low-quality nucleotides (base quality <5) were identified [36]. The paired-end Illumina reads of four Colombian clinical isolates and of the 183 global genomes (data downloaded from DDBJ/ENA/GenBank (http://www.ebi.ac.uk/ena) were mapped to the reference *L. infantum* JPCM5 [37] and assembled with the SMALT program (version 0.7.4) (http://www.sanger.ac.uk/science/tools/smalt-0). The JPCM5 genome was used as a reference in this study because so far it is the most comprehensive reference genome available, with a high quality, a robust assembly and without discontinuities and undetermined sequence. The mapping involved the following parameters: exhaustive search option (*–x* and *–y* of 0.8), a reference hash index of 13 bases, and a sliding step of 3. An identity threshold of y=0.8 was used to prevent the mapping of non-*Leishmania* reads to the reference sequences. The reads file merging, sorting, and elimination of PCR duplicates were implemented with SAMtools (version 0.1.18) and Picard (version 1.85: http://broadinstitute.github.io/picard/) [38].

Phylogenomic inferences

The SNP alignments from whole nuclear and mitochondrial genomes for the complete dataset (N=187), as well as the alignment obtained for each antigen gene, rK39 (GenBank accession number: L07879), rK28 (GenBank accession number: HM594686) and K9/K26 (GenBank accession numbers: AF131227/AF131228, respectively), were used to evaluate the phylogenomic relationships among each *L. infantum* isolate. Maximum-likelihood (ML) trees were inferred using IQ-TREE 2 [34]. The robustness of the nodes was evaluated using the bootstrap method (BT, with 1000 replicates). The obtained tree was visualized and edited using the Interactive Tree Of Life V4 (http://itol.embl.de) [35]. To detect recombination signatures in the 187 genomes analysed, phylogenetic networks were built in SplitsTree5 [39] using the neighbor-net method. For all inferences, we included *L. infantum* JPCM5 as the reference genome [37] and, for the outgroup, we used the *L. donovani* (Ld ERR013297) genome assembly from DDBJ/ENA/GenBank database (http://www.ebi.ac.uk/ena) under Run accession ERR013297.

Evaluation of chromosome and gene copy number variations (CNVs)

For the chromosomal somy estimation, the median read depth of each chromosome was initially calculated (di). All positions with a read depth of >1 standard deviations away from this initial median were then removed, and the di was recalculated from high-quality reads. Subsequently, the median depth of the 36 chromosomes (dm) of *L. infantum* was calculated, and the somy (S-value) of each chromosome was obtained with the following formula: S=2×di/dm [40]. The somy values calculated from sequencing data were averaged across the potentially variable somy of these cells. For this reason, somy values may be non-integers, representing the mean value of a mixed population. The ranges of monosomy, disomy, trisomy, tetrasomy and pentasomy were then used to define the full-cell normalized chromosome depth or somy (S) as S<1.5, 1.5≤S<2.5, 2.5≤S<3.5, 3.5≤S<4.5 and 4.5≤S<5.5, respectively, as previously described [41].

To evaluate the CNVs at the gene level, we defined average haploid depth per gene without somy effect as dHG and the full cell depth with somy effect as dFG. Their relationship was defined as dFG=S× dHG. We evaluated the gene and chromosome copy numbers by considering their statistical significance. Significance was set at a z-score cutoff of >2 and an adjusted P-value (Student's t-test) of <0.05. The heatmaps were created using the Heatmap3 package in R [42]. Finally, the genes that presented CNVs were subjected to Gene Ontology enrichment analyses using TriTrypDB tools (http://tritrypdb.org), and Fisher's exact test was used to maintain the FDR below 0.05. The GO terms were submitted to REVIGO [43].

Variant prediction calling and SNP filtering

With the purpose to detect SNPs in the 187 samples analysed, the reads of each genome were aligned to the reference

L. infantum JPCM5 using the SMALT program (version 0.7.4) (http://www.sanger.ac.uk/science/tools/smalt-0). The Picard program (version 1.85) (http://broadinstitute.github.io/picard/) was used for merging and sorting bam files and marking duplicated reads, as previously described [41]. SNP calling was completed in GATK v4.0.2 https://gatk.broadinstitute.org/hc/en-us. More specifically, we used GATK's HaplotypeCaller to produce genotype VCF files for every isolate, combine GVCFs to merge the genotype VCF files of all isolates, genotype GVCFs to perform joint genotyping, and finally, select variants to filter the SNPs. We realigned around indels to remove these and retrieved only the SNPs. Low-quality SNPs were excluded using VariantFiltration when QUAL<500, DP <5, QD <2.0, FS >60.0, MQ <40.0, MQRankSum < −12.5, or ReadPosRankSum < −8.0. All candidate SNPs were visually inspected in the Integrative Genomic Viewer (IGV_2_3_47) [41] and SAMtools to avoid false positives. Considering the total number of SNPs and using the snp-distans programme (https://github.com/tseemann/snp-dists), we generated a pairwise distance matrix, and the results obtained were graphically represented. The SnpEff programme (version v4.1) [41] was used to classify all SNPs based on their functional impact, and from this selection, we identified the unique/shared SNPs between/within the five geographical groups evaluated; the data were included in an Excel matrix, which was used to perform the comparative analysis. Finally, from the allele frequency estimation data, we determined the homozygous and heterozygous variants. Allele shifts of <0.2 or >0.80 were considered homozygous variants, while allele shifts between 0.2 and 0.8 were heterozygous variants [44]. Once the variants were identified, they were counted using an Excel matrix.

Genetic diversity and differentiation analysis

The genetic population structure of *L. infantum* was inferred from whole-genome SNP alignments using FastStructure v 1.0 [45]. The appropriate number of model components that explained the structure in the dataset was determined by the application of the algorithm for multiple choices of K, ranging from $K=1$ to $K=10$, considering nine independent simulations, which were conducted to estimate the cross-validation error values [46] (http://dalexander.github.io/admixture/admixture-manual.pdf). The expected admixture proportions inferred by FastStructure were graphically represented by Distruct plots using the mean of the variational posterior distribution. Later, we grouped the 187 genomes included in this study according to their geographical distribution (Colombia, the Americas, Asia, North Africa and Europe). We evaluated and compared the genetic diversity between/within each of the groups using DnaSP software v.5.0. Finally, to evaluate the degree of genetic differentiation and gene flow among the five populations, we calculated the pairwise F_{ST} value using Arlequin3.5 software [47]. On the basis of the data obtained, we generated a pairwise distance matrix, which was graphically represented.

rK39 and rK28 analysis

The Short Read Sequencing Typing 2 tool [48] was used to extract the rK39 and rK28 gene sequences from the whole genomes of the 187 isolates. Because rK28 was developed by fusing multiple tandem repeat sequences of three proteins, K26, LdK39 and K9, we also extracted and analysed these independently. The consensus sequences of each gene were aligned using the multiple sequence alignment programme MAFFT v7 (Suita, Osaka, Japan) [49], additionally the rK39 gene sequence was translated to amino acids (aa) using UGENE software [50], the DNA and amino acid sequences were grouped according to geographical origin (Colombia, the Americas, Asia, North Africa and Europe). The genomic diversity of each group of sequences was calculated. Diversity indices, such as number of haplotypes (h), haplotype diversity (Hd), and nucleotide diversity (π), were estimated using DnaSP software v.5.0. Neutrality indices, Tajima's D and Fu's Fs, were calculated using Arlequin v. 3.5.2.2 [47]. Additionally, to determine the potential evolutionary selection that occurred in these genes, the dN/dS rates were assessed using the Datamonkey programme (http://www.datamonkey.org) [51]. Finally, the nucleotide sequences were used to perform phylogenetic analysis following the parameters previously described in the Phylogenomic inferences section.

RESULTS

Identification of *Leishmania infantum*

The phylogenetic analysis based on HSP-70 sequences showed that the four Colombian Clinical isolates were closely related with *L. infantum*, while species of subgenera *Viannia* formed an independent cluster (Fig. S1). Once the Colombian samples were typed as *L. infantum*, we submitted the isolates to whole-genome sequencing. The Colombian genomes were sequenced to a mean depth of coverage of 127X (Table S2).

Nuclear and mitochondrial phylogenomic inferences

Two alignments were used to conduct phylogenomic analyses for 187 genome sequences (including four new genomes obtained in this study). The first corresponded to SNPs from nuclear genomes and the second to SNPs from mitochondrial genomes (Fig. 1). For both cases, *L. donovani* (Ld ERR013297) genomes were used as the outgroup. The phylogenetic reconstruction based on nuclear genome SNP alignments (Fig. 1a) showed *L. infantum* to be separated into various groups, some of which coincided with geographic origin. Tree topology revealed the existence of two clusters with well-supported nodes (with bootstrap ≥90.0). Cluster 1 (represented by a grey line outside the circle) was divided into three subgroups (1a to 1 c): subgroups 1a and 1b (highlighted in dark and medium beige, respectively) represented the five sequences from Cyprus included in this study, and subgroup 1c (highlighted in light beige) included 12 of the 13 sequences from Turkey, including the divergent sequence LI_ERR328068. Cluster 2 (represented by a black line outside the circle), like cluster 1, was subdivided in three subgroups (2a to 2c): subgroup 2a

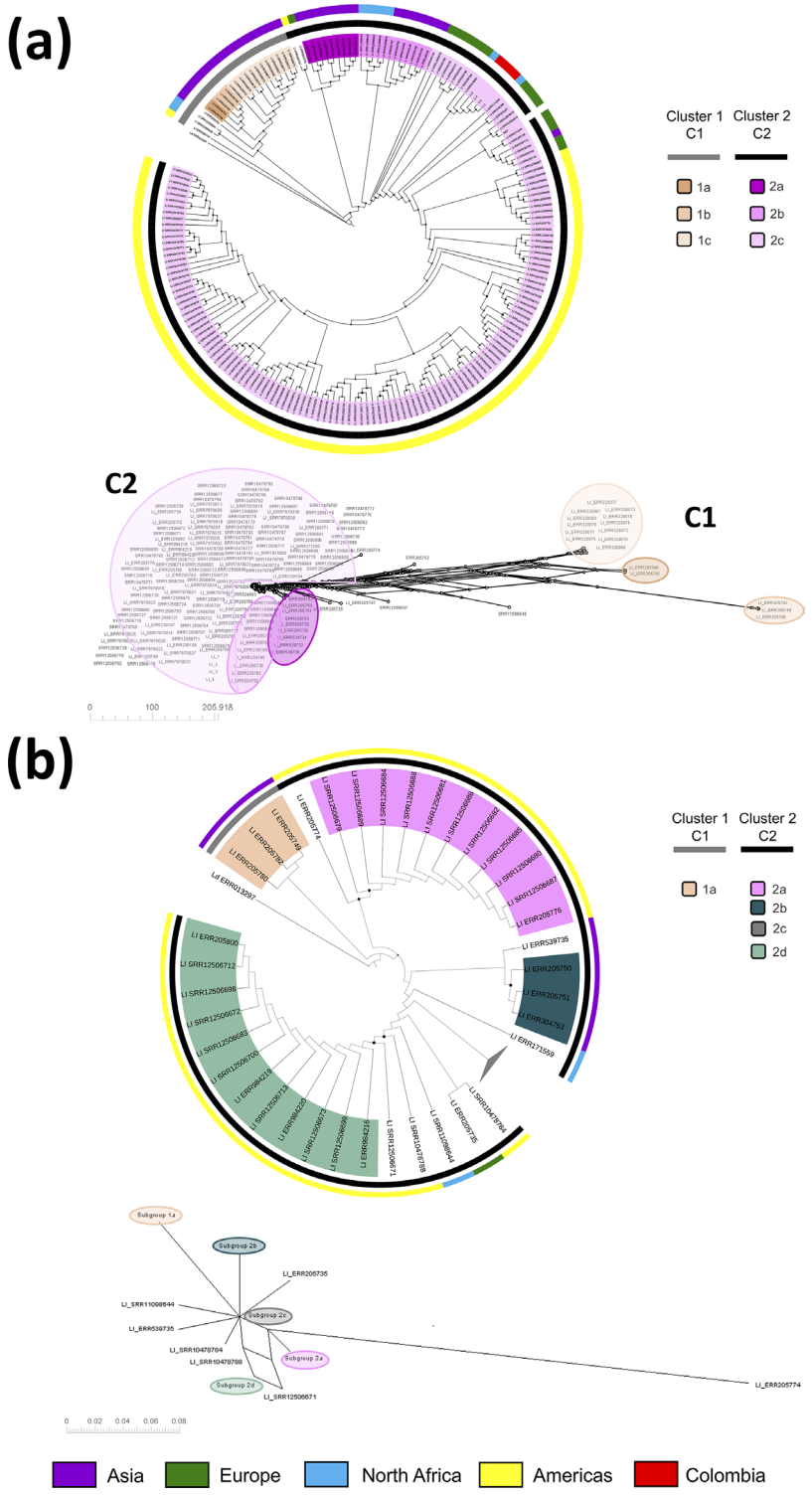


Fig. 1. Nuclear and mitochondrial phylogenies of 187 genomes of *L. infantum*. The trees located at the top of the figure represent the phylogenomic analysis based on SNP alignments for nuclear (a) and mitochondrial (maxicircle) (b) genomes of 187 genomic sequences analysed in this study. We used *L. infantum* JPCM5 as the reference genome (REF) and *L. donovani* (Ld ERR013297) as the outgroup. The first lines (grey and black) located on the outside of the circle represent the identified clusters, and the external lines (different colours) represent the geographical regions. Black dots represent well-supported nodes (bootstrap ≥90). The figures located at the bottom of each tree represent the phylogenetic network (neighbor-net) constructed in SplitsTree 5 based on nuclear and mitochondrial (maxicircle) genome alignments. The grey triangle in the mitochondrial tree represents genomes with identical sequences (total of 151 genomes).

(highlighted in strong magenta) included all sequences from Uzbekistan and China used in this study; subgroup 2b (highlighted in violet) included five sequences from North Africa (Kairouan), sequences from Palestine and Israel, including one divergent sequence from Israel (LI_ERR205797); and the most heterogeneous subgroup, subgroup 2c (highlighted in light violet), included sequences from Europe, South and Central America, North Africa, one sequence from Israel (LI_ERR230194), and the four sequences from Colombia. We observed that Colombian genomes were closely related to sequences from Morocco and Italy. Finally, we highlighted the results obtained from genome LI_ERR205774 (Brazil), which showed high divergence compared with other sequences from the same or closer regions and genomes LI_SRR11098647, LI_SRR11098648 (North Africa), LI_ERR205735 (France), LI_ERR205787 (Panama) and LI_ERR205752 (Turkey), which were identified as outliers (Fig. 1a).

In the tree topology based on mitochondrial genome SNP alignment (Fig. 1b), we observed clustering events that showed little concordance with those observed in the topology of the nuclear genome. In general, we observed two clusters with well-supported nodes (with bootstrap ≥90.0). Cluster 1 (represented by a grey line outside the circle) was represented by the single subgroup 1a (highlighted in dark beige), which included three sequences from Cyprus. Cluster 2 (represented by a black line outside the circle) was divided into four subgroups (2a to 2d): subgroup 2a (highlighted in light purple) included sequences from Brazil, subgroup 2b (highlighted in teal) included sequences from Asia (Uzbekistan and Israel), subgroup 2c (denoted by a grey triangle) was represented by 151 genomes with identical sequences and distributed in different geographical regions, and subgroup 2d (highlighted in mint green) was represented by sequences from Brazil (Fig. 1b). Finally, we identified seven genomes that were considered outliers, including the highly divergent genome LI_ERR205774 (Brazil) (Fig. 1b). All findings were supported by phylogenetic tree topologies obtained in SplitsTree5 [38] (using the neighbor-net method), in which the members of these groups were consistently clustered together (bottom panel Fig. 1a, b).

Chromosome/gene copy number variation within Colombian genomes and between worldwide genomes

Chromosome copy number variation

Initially we analysed the somy values for the four Colombian clinical isolates obtained in this study. The results showed moderate aneuploidy in some genomes. In the LI_2 genome, we observed an unchanged (disomic) karyotype in 35/36 chromosomes, contrasting with what was observed in the three remaining genomes, which showed some degree of aneuploidy (Fig. S2). Analysis of the LI_1 genome provided evidence of an increase in the somy in 8/36 chromosomes (Chrs 1, 5, 13, 15, 20, 22, 26 and 31); the LI_3 genome showed aneuploidy in 5/36 chromosomes (Chrs 5, 8, 20, 23, and 31), and the LI_4 genome showed changes in nine chromosomes: Chrs 6, 7, 9, 22, 23 26 and 29 (trisomic); Chr 20 (tetrasomic); and Chr 31 (pentasomic) (Fig. S2). The allele frequency counts for each predicted heterozygous SNP did not exhibit discordance between read depths and allele frequencies, confirming the accuracy of the previously described somy profiles.

Later, the somy values for the 183 currently available genomes were estimated, and the results were compared to values obtained for the Colombian isolates. In general, the results evidenced moderate aneuploidy in the *L. infantum* genomes, and some chromosomes showed similar behaviour in all geographical regions, namely chromosomes 31 (tetrasomic) and 23 (trisomic) (Fig. 2a). Additionally, we found that some chromosomes presented behaviours particular to some geographical regions; for example, chromosome 5 was trisomic in all geographical regions except Europe; chromosome 20 remained trisomic in all groups except the American genomes; and finally, chromosomes 33 and 35 were trisomic only in the American and European genomes, respectively (Fig. 2b). Interestingly, the American *L. infantum* genomes (mainly from Brazil) showed different behaviour in the ploidy; 25% of genomes evaluated (30/121) evidenced a low number of copies per chromosome compared with the remaining 75%, which presented moderate ploidy (Fig. 2a, b).

Gene copy number variations

After analysing the variations at the chromosomal level, we evaluated the CNVs for each gene (z-score cutoff >2 and adjusted P-value <0.05) and compared the occurrence in the four Colombian genomes. The results obtained did not reveal significant differences among them: the genome LI_4 showed the highest number of genes with CNVs (187 genes), followed by genome LI_3 (183 genes), genome LI_1 (177 genes) and finally genome LI_2 (174 genes). Later, with the purpose of analysing the genes with CNVs shared among the four Colombian genomes, we averaged the values obtained among the genomes, calculated standard deviations σ and the corresponding z-score, the results revealed 188 genes shared among them; 143 (76%) were annotated with known functions and 45 genes (24%) as hypothetical proteins (Fig. 3a). Their functional annotations included mainly proteophosphoglycan, amastin surface glycoprotein, GP63-leishmanolysin, tb-292 membrane associated protein-like protein, elongation factor 1-alpha, beta tubulin, and heat shock protein 83–17 (Table S3). The next step was to evaluate and compare gene CNVs between the 187 genomes analysed, including the Colombian genomes. Initially, we averaged the values of the isolates for each geographical region and made a comparison between them. The initial results evidenced a moderate increase in the number of genes with CNVs in the Colombian genomes (188 genes) compared with genomes from other regions (the Americas: 124 genes, Asia: 131 genes, North Africa: 122 genes, and Europe: 137 genes) (Fig. 3a). Later, we studied in detail the genes with CNV in each geographical region, and the results demonstrated that the gene encoding proteophosphoglycan (LINF_350010200) presented significant CNVs in the 187 genomes. Additionally, we identified other genes with CNVs in some regions; for example, CNVs were observed in genes encoding rRNA, tb-292 membrane-associated protein-like

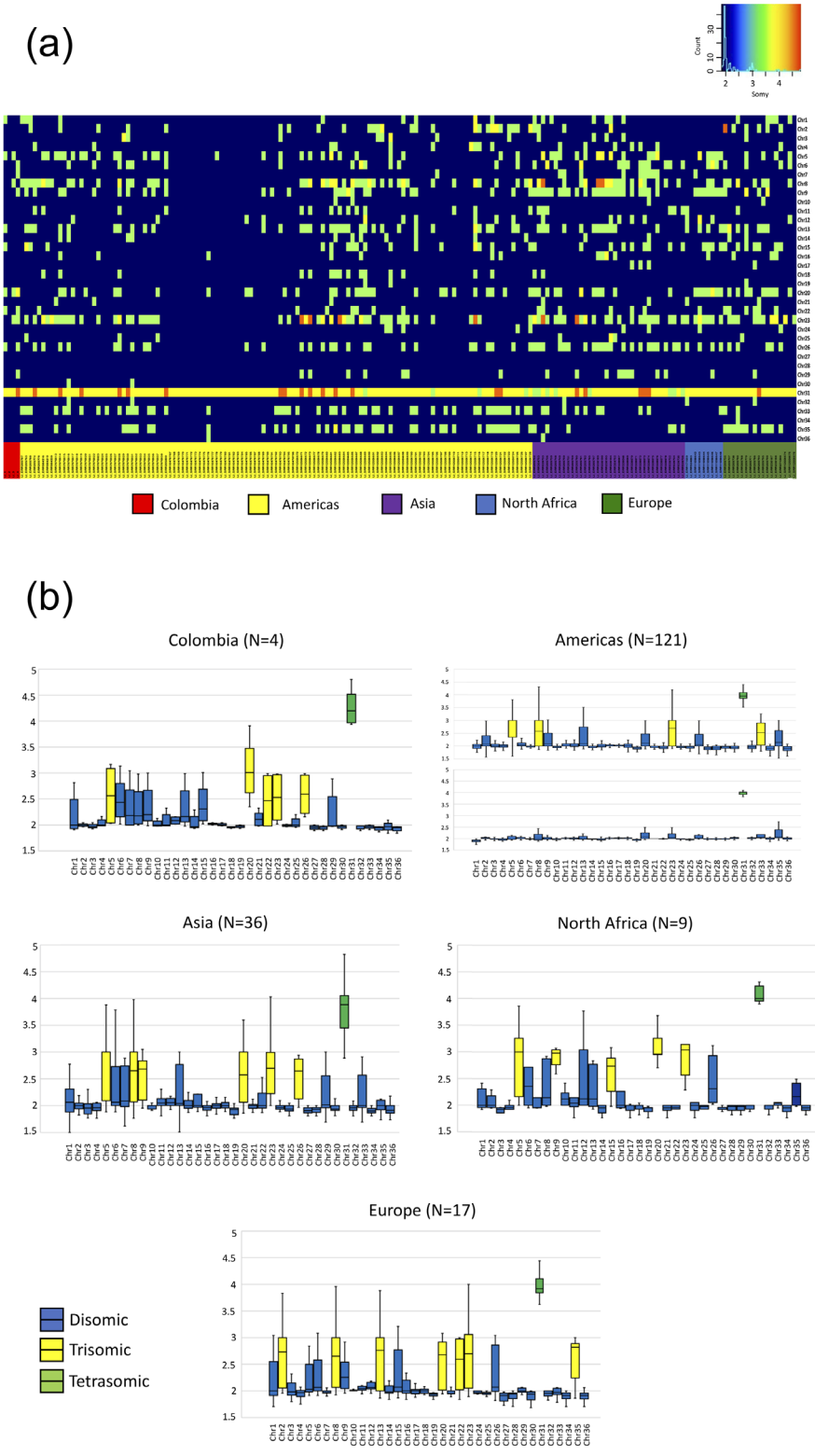
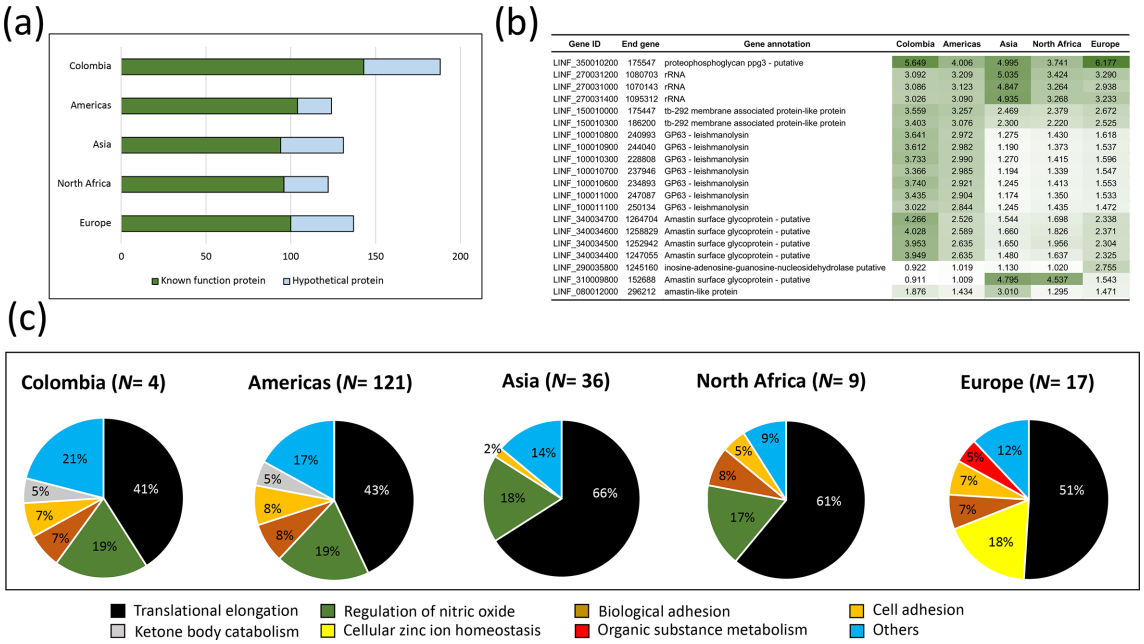


Fig. 2. Evaluation of chromosomal copy number in 187 *L. infantum* genomes. (a) Heatmap shows the estimated copy numbers of the 36 chromosomes (*y*-axis) in the 187 genomes (*x*-axis). Disomic (blue), trisomic (green), tetrasomic (yellow), and pentasomic (orange). (b) Average copy number by chromosome in the five geographical regions evaluated. In the American panel, the top graphic represents the genomes with moderate ploidy (91 genomes), and in the bottom panel, it represents the genomes that presented low ploidy (30 genomes).



Gene ID	End gene	Gene annotation	Colombia	Americas	Asia	North Africa	Europe
LINF_350010200	175547	proteophosphoglycan ppg3 - putative	5.649	4.006	4.995	3.741	6.177
LINF_270031200	1080703	rRNA	3.092	3.209	5.035	3.424	3.290
LINF_270031000	1070143	rRNA	3.086	3.123	4.847	3.264	2.938
LINF_270031400	1095312	rRNA	3.026	3.090	4.935	3.268	3.233
LINF_150010000	175447	tb-292 membrane associated protein-like protein	3.559	3.257	2.469	2.379	2.672
LINF_150010300	186200	tb-292 membrane associated protein-like protein	3.403	3.076	2.300	2.220	2.525
LINF_100010800	240993	GP63 - leishmanolysin	3.641	2.972	1.275	1.430	1.618
LINF_100010900	244040	GP63 - leishmanolysin	3.612	2.982	1.190	1.373	1.537
LINF_100010300	228808	GP63 - leishmanolysin	3.733	2.990	1.270	1.415	1.596
LINF_100010700	237946	GP63 - leishmanolysin	3.366	2.985	1.194	1.339	1.547
LINF_100010600	234893	GP63 - leishmanolysin	3.740	2.921	1.245	1.413	1.553
LINF_100011000	247087	GP63 - leishmanolysin	3.435	2.904	1.174	1.350	1.533
LINF_100011100	250134	GP63 - leishmanolysin	3.022	2.844	1.245	1.435	1.472
LINF_340034700	1264704	Amastin surface glycoprotein - putative	4.266	2.526	1.544	1.698	2.338
LINF_340034600	1258829	Amastin surface glycoprotein - putative	4.028	2.589	1.660	1.826	2.371
LINF_340034500	1252942	Amastin surface glycoprotein - putative	3.953	2.635	1.650	1.956	2.304
LINF_340034400	1247055	Amastin surface glycoprotein - putative	3.949	2.635	1.480	1.637	2.325
LINF_290035800	1245160	inosine-adenosine-guanosine-nucleosidehydrolase putative	0.922	1.019	1.130	1.020	2.755
LINF_310009800	152688	Amastin surface glycoprotein - putative	0.911	1.009	4.795	4.537	1.543
LINF_080012000	296212	amastin-like protein	1.876	1.434	3.010	1.295	1.471

Fig. 3. Evaluation of gene copy number variation (CNV) and gene ontology enrichment analyses in 187 *L. infantum* genomes distributed in five geographical regions. (a) Number of genes that presented CNVs by geographical region. (b) Name and ID of genes that presented the highest CNVs. (c) Gene Ontology enrichment analyses for biological processes found in the five geographical regions.

protein, GP63-leishmanolysin, and putative amastin surface glycoprotein in Colombia and the Americas; genes encoding rRNA, tb-292 membrane associated protein-like protein, and putative amastin surface glycoprotein (LI_310009800) in Asia and North Africa; and those encoding rRNA, tb-292 membrane associated protein-like protein, putative amastin surface glycoprotein, and putative inosine-adenosine-guanosine-nucleoside hydrolase in Europe (Fig. 3b). Finally, Gene Ontology enrichment analysis of the genes that presented CNVs indicated that between 41 and 66% of the enriched terms in isolates from the five geographical regions were associated with translational elongation (with the highest percentages in Asia and North Africa [66 and 61%, respectively]), and 19% of genes were associated with the regulation of nitric oxide (except the isolates from Europe). Curiously, only the European isolates presented genes with CNVs that were involved in cellular zinc ion homeostasis (Fig. 3c).

Variant prediction calling and SNP filtering

Initially, we analysed the four new Colombian genomes and compared them with the reference genome. The results revealed the total number of SNPs in the four isolates ranged from 1147 (LI_3 isolate) to 1182 (LI_4 isolate) without significant differences among them. The four genomes were characterized by low heterozygosity (~80% of SNP loci were homozygous) (Fig. S3a). Of the total number of SNPs identified, 25% had potential effects on gene function (high and moderate impact), and of these, ~98% were synonymous variants. However, in the four Colombian genomes, we identified some variants with high functional impact (stop_gained and stop_lost), the SNPs associated with stop_gained were detected in genes encoding hypothetical proteins and, in a gene, encoding to known function protein as putative ubiquitin-conjugating enzyme. Interestingly, the four Colombian genomes shared the same SNPs associated with stop_lost, which were identified in the gene encoding kinetoplast-associated protein-like protein (LINF_270007500) (Table 1). When we evaluated the number of SNPs with functional impact for each chromosome per isolate; we observed that the distribution of SNPs among the chromosomes was homogeneous, with chromosomes 12 and 36 having the highest number of SNPs (Fig. S3b). Finally, we analysed the shared/unique SNPs among the four Colombian genomes, and 215 SNPs with high and moderate functional impact were found to be shared. Of these, 56% (121 SNPs) were found in genes encoding proteins of known function (Table S4). When we evaluated and compared the unique SNPs in all isolates, we observed that the LI_2 and LI_4 genomes had the greatest numbers of unique SNPs (25 and 28, respectively) not found in the other analysed isolates (Fig. S3c).

Later, we focused our analysis on the 183 genomes from across the world and compared those from the different geographical regions. Initially, we considered the SNP density (including in Colombian genomes) and observed a variable density ranging from 2.7 SNP/Kb to 0.013 SNPs/Kb. The genomes with greatest variability were from Cyprus (2.7 to 2.2 SNPs/Kb), followed by genomes from Turkey (1.8 to 1.4 SNPs/Kb), one genome from Brazil: LI_ERR205774

Table 1. List of genes that showed SNPs with high functional impact, in the four Colombian genomes analysed

(a) SNP position on the gene. (b) Change of nucleotide respect to reference (Ref/Alt).

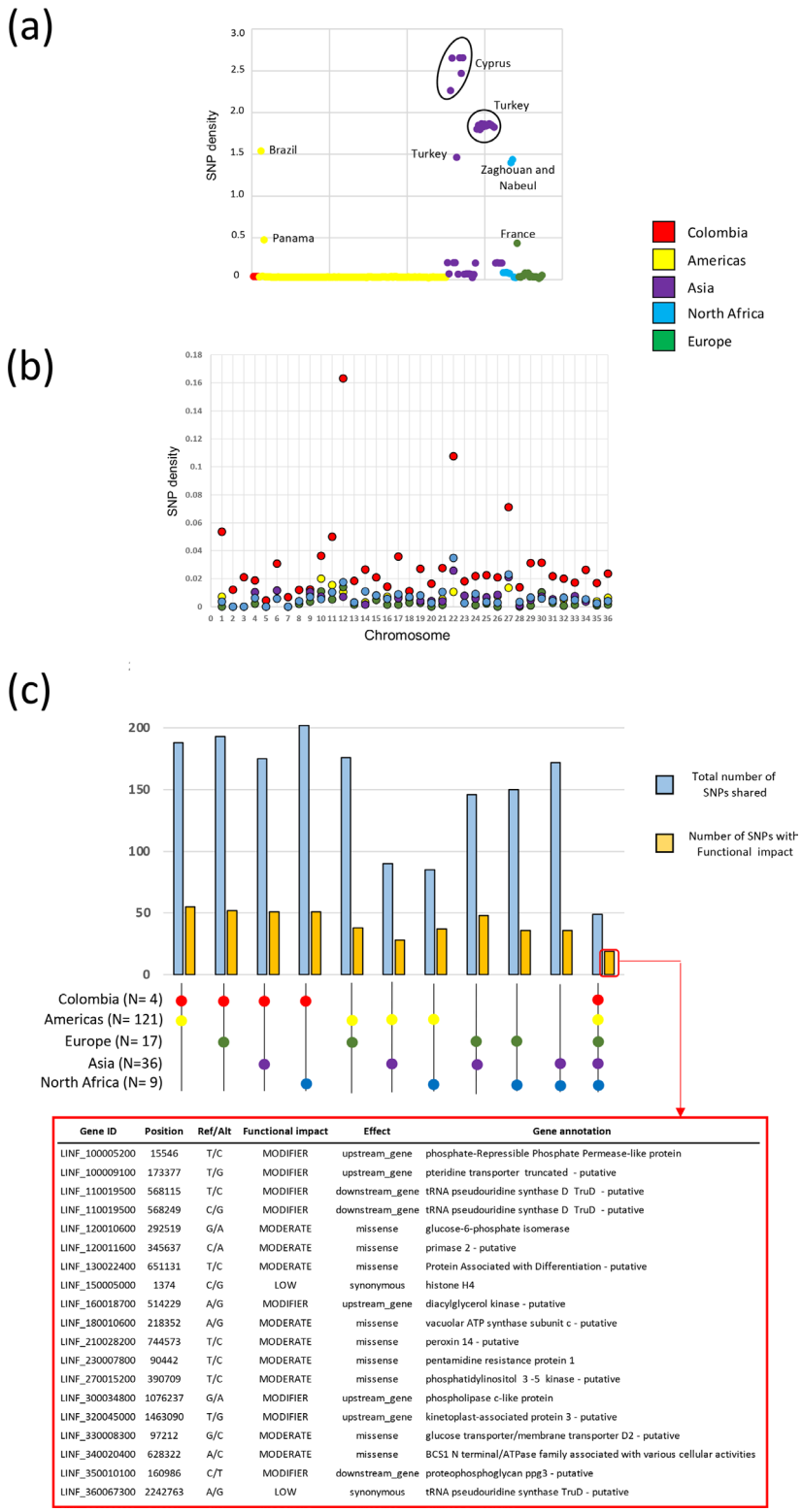
LI_1				
Gene ID		**a**	**b**	**Gene annotation**
LINF_270007500	stop_Lost	62087	A/C	kinetoplast-associated_protein-like_protein
LI_2				
Gene ID		**a**	**b**	**Gene annotation**
LINF_200019900	stop_gained	648 231	G/A	hypothetical_protein_-_conserved
LINF_270007500	stop_gained	61967	C/A	kinetoplast-associated_protein-like_protein
LINF_270007500	stop_lost	62087	A/C	kinetoplast-associated_protein-like_protein
LINF_330039100	stop_gained	1344679	C/G	hypothetical_protein_-_conserved
LI_3				
Gene ID		**a**	**b**	**Gene annotation**
LINF_270007500	stop_lost	62087	A/C	kinetoplast-associated_protein-like_protein
LINF_330039100	stop_gained	1344679	C/G	hypothetical_protein_-_conserved
LI_4				
Gene ID		**a**	**b**	**Gene annotation**
LINF_130021300	stop_gained	598485	C/T	ubiquitin-conjugating_enzyme_-_putative
LINF_270007500	stop_lost	62087	A/C	kinetoplast-associated_protein-like_protein
LINF_360069500	stop_gained	2335305	C/A	glycerophosphoryl_diester_phosphodiesterase_-_putative

(1.5 SNPs/Kb), and the genomes from Zaghouan and Nabeul (1.43 and 1.40 SNPs/Kb, respectively). The genomes with the lowest variability were Brazilian (Figs 4a and S4).

The results showed the largest number of variants in chromosomes 12, 22 and 27 and less variability in chromosomes 2, 5 and 7 (Fig. 4b). Finally, we analysed the number of shared SNPs within and between each group evaluated. The results showed the greatest number of shared SNPs in the Colombian genomes (862 shared SNPs; *N*=4), and there were moderate numbers of shared SNPs in the other geographical regions: the Americas (205 SNPs; *N*=121), Asia (187 SNPs; *N*=36), North Africa (209 SNPs; *N*=9) and Europe (121 SNPs; *N*=17). During this analysis, we discovered that all genomes belonging to certain countries shared SNPs (high functional impact) that were not found in other regions (e.g. the 121 genomes from Brazil shared a mutation in the gene LINF 080015600, histone deacetylase [putative], which was not found in other countries), and the same occurred with genomes from Honduras, China and Cyprus (Table 2). We evaluated the shared SNPs between the different geographical regions and found the regions sharing the most SNPs were Colombia–North Africa and Colombia–Europe (202 and 193 shared SNPs, respectively), and the group with the least shared SNPs was North Africa–America (85 SNPs). Additionally, we identified 49 shared SNPs between the 187 genomes analysed, 38% of which (19 SNPs) were located in genes with known functions. We highlighted those SNPs located in genes involved in resistance to pentamidine (LINF_230007800), in pteridine/glucose transport (LINF_100009100 LINF_330008300), and in glycolysis and gluconeogenesis, that is, glucose 6 phosphate isomerase (LINF_120010600), as well as genes associated with the intracellular signal pathway, such as phosphatidylinositol 3 kinase (LINF_270015200) (Table of Fig. 4c)

Genetic diversity and differentiation analysis

The population organization of the 187 isolates was analysed with FastStructure software, which indicated the existence of four different populations (*K*=4). The largest population—POP1 (highlighted in blue)—comprised 100% of Colombian, Honduran and Brazilian genomes (except LI_ERR205774) and genomes from Europe (France, Italy, Spain and Portugal), North Africa (Morocco and Kairouan) and Asia (Israel and Palestine). POP2 (highlighted in green) represented 100% of the genomes from China and Uzbekistan, one genome from France, and curiously, the unique Panamanian genome. POP3 (highlighted in orange) included all genomes from Turkey (except LI_ERR205752) and some genomes from Cyprus, while POP4 (highlighted in light purple) included the remaining genomes from Cyprus. Additionally, we identified four highly heterogeneous genomes: LI_ SRR11098647 (Nabeul), LI_ SRR11098648 (Zaghouan), LI_ERR205752 (Turkey) and LI_ERR205774 (Brazil) (Fig. 5). Later, the 187 genomes were grouped according to their geographical location



Gene ID	Position	Ref/Alt	Functional impact	Effect	Gene annotation
LINF_100005200	15546	T/C	MODIFIER	upstream_gene	phosphate-Repressible Phosphate Permease-like protein
LINF_100009100	173377	T/G	MODIFIER	upstream_gene	pteridine transporter truncated - putative
LINF_110019500	568115	T/C	MODIFIER	downstream_gene	tRNA pseudouridine synthase D TruD - putative
LINF_110019500	568249	C/G	MODIFIER	downstream_gene	tRNA pseudouridine synthase D TruD - putative
LINF_120010600	292519	G/A	MODERATE	missense	glucose-6-phosphate isomerase
LINF_120011600	345637	C/A	MODERATE	missense	primase 2 - putative
LINF_130022400	651131	T/C	MODERATE	missense	Protein Associated with Differentiation - putative
LINF_150005000	1374	C/G	LOW	synonymous	histone H4
LINF_160018700	514229	A/G	MODIFIER	upstream_gene	diacylglycerol kinase - putative
LINF_180010600	218352	A/G	MODERATE	missense	vacuolar ATP synthase subunit c - putative
LINF_210028200	744573	T/C	MODERATE	missense	peroxin 14 - putative
LINF_230007800	90442	T/C	MODERATE	missense	pentamidine resistance protein 1
LINF_270015200	390709	T/C	MODERATE	missense	phosphatidylinositol 3 -5 kinase - putative
LINF_300034800	1076237	G/A	MODIFIER	upstream_gene	phospholipase c-like protein
LINF_320045000	1463090	T/G	MODIFIER	upstream_gene	kinetoplast-associated protein 3 - putative
LINF_330008300	97212	G/C	MODERATE	missense	glucose transporter/membrane transporter D2 - putative
LINF_340020400	628322	A/C	MODERATE	missense	BCS1 N terminal/ATPase family associated with various cellular activities
LINF_350010100	160986	C/T	MODIFIER	downstream_gene	proteophosphoglycan ppg3 - putative
LINF_360067300	2242763	A/G	LOW	synonymous	tRNA pseudouridine synthase TruD - putative

Fig. 4. Overview of SNPs identified in 187 *L. infantum* genomes. (a) Density of SNPs in each of the 187 genomes. (b) Relationship between SNP density and chromosomes in each of the five geographical regions. (c) Number of shared SNPs (total number represented by blue bars, and number of SNPs with functional impact represented by orange bars) between the five geographical regions. The coloured dots located in the lower part of the graphic represent each region. The table describes the shared genes with known function between all geographical regions. Ref/Alt represents the change in nucleotide with respect to reference.

Table 2. List of genes with unique SNPs representing high functional impact in each region analysed.

(a) SNP position on the gene. (b) Change of nucleotide respect to reference (Ref/Alt).

Brazil				
Gene ID		**a**	**b**	**Gene annotation**
LINF 080015600	stop gained	437460	C/A	histone deacetylase - putative
Honduras				
Gene ID		**a**	**b**	**Gene annotation**
LINF 080011500	stop gained	266177	C/T	hypothetical protein - conserved
LINF 270033100	stop gained	1130864	T/G	tubulin-tyrosine ligase-like protein
LINF 330041100	stop gained	1482887	G/A	Holliday-junction resolvase-like of SPT6/SH2 domain
China				
Gene ID		**a**	**b**	**Gene annotation**
LINF 170016300	stop gained	445751	G/T	ATP-dependent RNA helicase - putative
LINF 360033600	stop gained	1067054	G/A	related to multifunctional cyclin-dependent kinase pho85-like protein
Cyprus				
Gene ID		**a**	**b**	**Gene annotation**
LINF 260014500	stop gained	276029	G/T	hypothetical protein - conserved
LINF 340005800	stop gained	30935	C/T	hypothetical protein - conserved

(five groups in total) and, considering these groups, we evaluated and compared the genetic diversity between each of them. The results demonstrated low diversity in the Colombian genomes (π=0.0015; N=4) and high diversity in Asian genomes (π=0.2535; N=36) compared with genomes from the Americas (π=0.0052; N=121), North Africa (π=0.0952; N=9) and Europe (π=0.0148; N=17). To evaluate the degree of genetic differentiation and gene flow among the five populations, we calculated the F_{ST} values. The results revealed low levels of genetic differentiation between North African/Colombian (F_{ST}=0.0608) and European/Colombian populations (F_{ST}=0.1023) and strong genetic divergence between American/Asian populations (F_{ST}=0.5663) and between American/North African populations (F_{ST}=0.5131); all values obtained were associated with significant P-values (<0.05) (Table S5).

Two antigens commonly used in VL diagnosis—rK39 and rK28—were analysed in this study. The nucleotide sequences of each of them were extracted and compared between the 187 genomes, additionally; the amino acid sequence of rK39 protein was analysed. Phylogenetic analysis based on the rK39 gene revealed the potential existence of two clusters in well-supported nodes (with bootstrap ≥90.0). Cluster 1 (represented by a grey line outside the circle) was divided in two groups (subgroups 1a [highlighted in dark green] and 1b [highlighted in blue]) that included 83/121 (68%) sequences from Brazil, four sequences from Colombia, and some sequences from North Africa and Europe. Cluster 2 (represented by a black line outside the circle) was represented by four groups: subgroup 2a (highlighted in light violet) included sequences from North Africa and Asia; subgroup 2b (highlighted in yellow) included

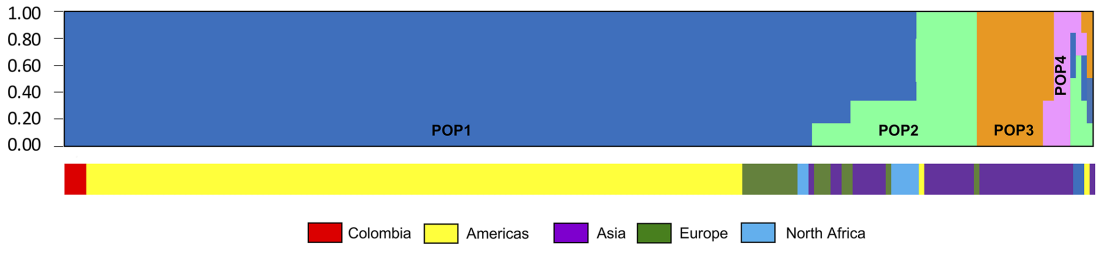


Fig. 5. Estimated population structure of 187 *L. infantum* genomes. Bar graph generated by FastStructure software inferring the genetic structure of 187 genomes of *L. infantum*. Each genome is represented by a single vertical line in the barplots; each colour represents one population (K=4), and the lengths of the colour segments show the genomes' estimated proportion in that population. The bars at the bottom of the graph represent the geographical origin.

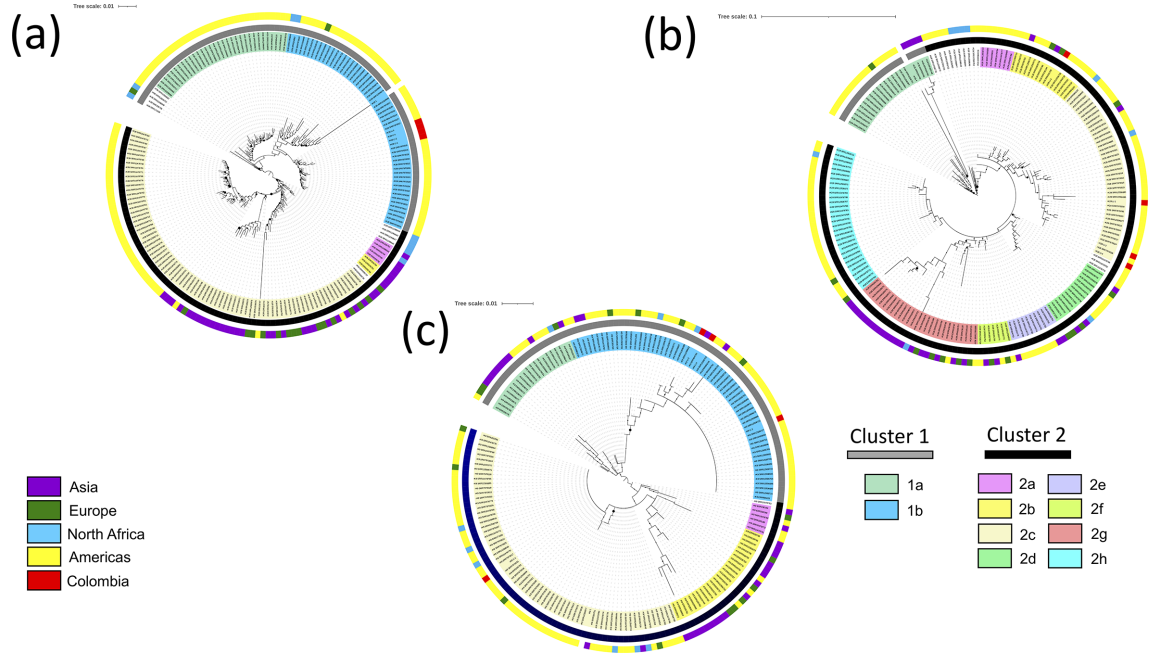


Fig. 6. Phylogenetic analysis of rK39, rK28 and K9 genes obtained from 187 *L. infantum* genomes. The trees represent the phylogenetic analysis based on rK39 (a), rK28fus (b) and K9 (c) genes. We used the sequences of rK39 (L07879), rK28 (HM594686) and K9 (AF131227) genes from GenBank as a reference (Ref). The colours inside the circle represent the subgroups, the first lines (grey and black) on the outside represent the identified clusters, and the external lines (different colours) represent the geographical regions. The black dots represent well-supported nodes (bootstrap ≥90).

three sequences from Asia; subgroup 2c (highlighted in light beige), which was the most heterogeneous, included sequences from three of the five geographical regions analysed (Asia, Europe, Americas [Brazil and Panama]); and subgroup 2d (highlighted in light green) was represented mainly by sequences from Brazil, the two sequences from Honduras, and some from Asia and Europe (Fig. 6a). The analysis performed on the amino acid sequence of rK39 protein, specifically to the region used as target in the diagnostic tests (39 amino acid repeated 6.5 times), revealed a non-synonymous substitution in the amino acid 27, we observed a change of a methionine (M) or serine (S) by a threonine (T). This substitution was identified in four different sites of the protein (residues 720, 759, 836 and 876) (Fig. 7)

Considering that rK28 is a chimeric antigen composed of three amino acid sequences from different proteins, we analysed not only the complete chimaera (rK28fus) but also each component protein separately (K9, K26 and LdK39). Analysis of the rK28fus protein revealed the formation of two well-supported clusters. Cluster 1 (represented by a grey line outside the circle) was represented by 17/121 (14%) sequences from Brazil and four samples from Asia. Cluster 2 (represented by a black line outside the circle), which showed high heterogeneity, was divided into eight groups: subgroup 2a (highlighted in violet) included sequences from Brazil; subgroup 2b (highlighted in yellow) included sequences from Brazil, one sequence from Colombia, and some sequences from Europe and Asia; subgroup 2 c (highlighted in light beige) was represented mainly by sequences from Brazil, three sequences from Colombia, two sequences from North Africa, and one sequence from Asia and Europe; subgroups 2d–2g were the most heterogeneous subgroups and included sequences from various geographical regions; and subgroup 2h (highlighted in turquoise) was represented mainly by sequences from Brazil and one sequence from each of North Africa and Europe (Fig. 6b). As observed for the two previous genes, K9 was grouped into two clusters. Cluster 1 was divided into two groups (subgroups 1a and 1b) and cluster 2 was divided into three groups (subgroups 2a–2c), with each one revealing highly heterogeneous grouping regarding the geographical distribution (Fig. 6c). Finally, the analysis of K26 and LdK39 proteins revealed great differences between each of them and the sequences analysed, and the variation generated gaps that did not allow us to make an adequate phylogenetic analysis.

Additionally, we evaluated and compared the genomic sequence diversity of rK39, rK28fus and K9 while considering geographical origin. Of the three genes, rK28fus presented the highest number of haplotypes in all geographical regions (Table 3). On the basis of the haplotype (Hd) and nucleotide diversity indexes (π) the genes rK39 and rK28fus, the North African sequences presented the greatest genetic (Hd=1) and nucleotide diversities (π=0.014 and π=0.015, respectively), contrary to what was observed for gene K9, which contained the greatest genetic

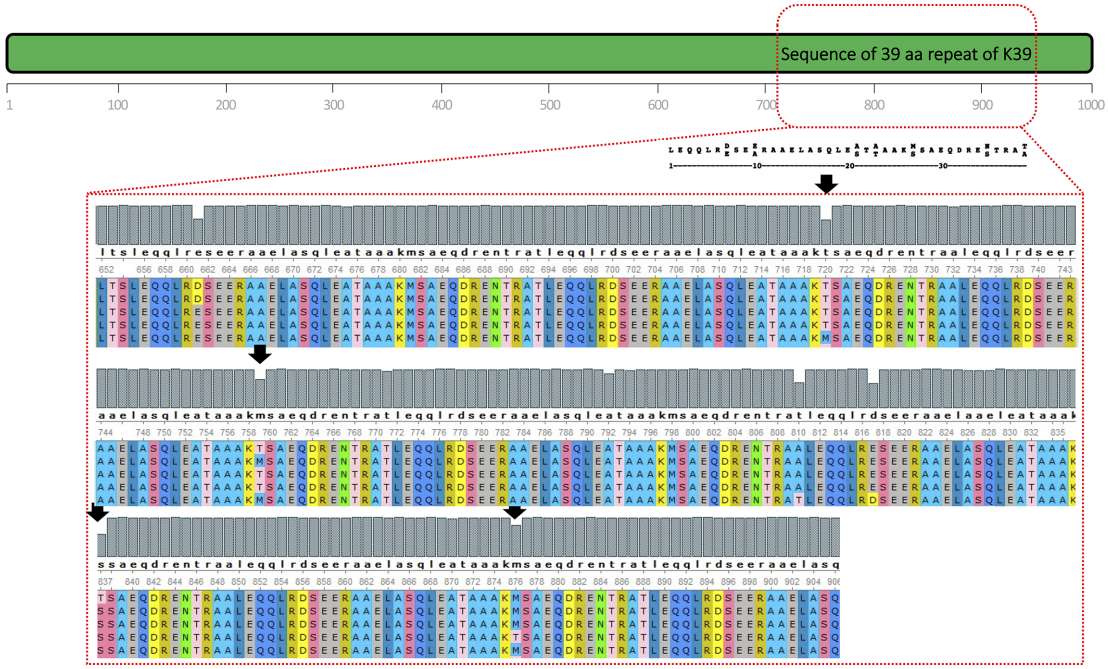


Fig. 7. Multiple alignment of rK39 protein between the 187 *L. infantum* genomes. The top bar represents the sequence of K39 protein (amino acid 1–950). In this sequence, the dotted square shows the region used as target in the immunochromatographic diagnostic test (39 amino acid repeated 6.5 times). The bottom shows the multiple alignment of the target region, among the 187 *L. infantum* genomes. The non-synonymous substitutions found between the genomes are represented with black arrows.

(Hd=1) and nucleotide diversities (π=0.015) in Colombia. Tajima's D index was then used to determine the extent of neutral selection. Although the values were not statistically significant (P>0.05), the Tajima tests for rK39 showed negative values for all geographical regions, except Asia. For rK28, negative Tajima values were observed for the Americas, North Africa and Europe, and positive values were observed for Colombia and Asia. Finally, the results for K9 were negative for genomes from Colombia and Asia and positive for those from Europe, America and North Africa. Regarding Fu's Fs index, the values for genes rK39 and rK28fus were negative and significant (P<0.05) in sequences from the Americas, Asia, North Africa and Europe; and for gene K9, negative and significant values were observed in sequences from the Americas, Asia and Europe (Table 3). Finally, a test for positive selection was carried out by analysing the dN/dS ratio. Interestingly, the results revealed a ratio of >1.0 for the rK28fus gene in all geographical regions (except Colombia) and ratios of <1.0 for rK39 and K9 (Table 3).

DISCUSSION

The objective of this study was to understand the phylogenomic relationships, genetic structure and intraspecific variability of *L. infantum* worldwide. A phylogenetic analysis of the nuclear genomes allowed us to identify the presence of different *L. infantum* phylogroups circulating in continents around the world (Fig. 1a). These findings were consistent with the population structures defined by STRUCTURE and the *F* statistics (F_{ST}) (Fig. 5, Table S5), which revealed the existence of genetically distinct *L. infantum* populations, with the first being represented by genomes from the Americas, Europe, and some from Asia and Africa; the second and third populations represented by Chinese and Turkish genomes; and a fourth population represented by genomes from Cyprus; although the clustering of Turkish genomes was probably related to the organism from which it was isolated (sandfly vector). Overall, however, *L. infantum* demonstrated low genetic variability (Fig. 4a), the results obtained in this study suggest there is genetic diversification at a global level in this species, diversification that could be generated by the immense human migrations between countries and/or by the adaptation of the parasite to the distinct genetic structures of the sandfly vector/reservoirs, which have been demonstrated to play an important role in the genetic structure of *L. infantum* [52, 53]. Additionally, we found evidence of the close phylogenomic relationships between genomes from different geographical regions (e.g. between genomes from China, Panama and France; between genomes from Italy, Portugal and Israel; and between genomes from Colombia, Europe and North Africa) (Figs 1a and 5), which is relevant, considering that the appearance of *L. infantum* strains in different regions could generate fitness with potential implications in the spread of virulent strains, the exploitation of different and new vectors and hosts, resistance to anti-leishmanial agents, and adaptation to new ecological niches.

Table 3. Diversity indexes calculated for the rK39, rK28fus and K9 genes in the different geographical regions evaluated

	rK39 gene					rK28fus gene					rK9 gene				
	Colombia	**Americas**	**North Africa**	**Asia**	**Europe**	**Colombia**	**Americas**	**North Africa**	**Asia**	**Europe**	**Colombia**	**Americas**	**North Africa**	**Asia**	**Europe**
Sample size	4	121	9	36	17	4	121	9	36	17	4	121	9	36	17
Haplotypes	3	66	9	32	16	3	82	9	34	16	4	33	5	25	12
No. of sites with substitutions (S)	4	57	10	29	41	6	36	23	47	19	10	14	12	25	13
Haplotype diversity	0.833	0.974	1.000	0.992	0.993	0.833	0.985	1.000	0.993	0.992	1.000	0,863	0,833	0,971	0,949
Nucleotide diversity (π)	0.002	0.008	0.014	0.011	0.011	0.004	0.011	0.015	0.014	0.006	0.015	0,014	0,013	0,014	0,013
Tajima's D	−0.781	−1.222	−0.437	0.778	−1.147	0.179	−0.554	−0.498	0.031	−0.651	−0.528	2.427	0,269	−0,251	0,645
Fs	0.133	−24.98 *	−6.462 *	−22.722*	−7.431*	0.888	−24.758*	`−3.539*	−17.949*	−11.825*	−0.480	−11.816*	1.043	−15.078*	−4.210*
dN/dS index	0.321	0.409	0.247	0.180	0.453	0.382	1.940	4.01	1.053	1.42	0.838	0,991	0,924	0,876	0,993

*$P<0.05$.

The close relationships between the Colombian genomes and the European/North African genomes (Fig. 1a), the low genetic divergence among genomes from these regions (F_{ST} = 0.10230 and 0.0608, respectively) (Table S5), the high number of shared SNPs (Fig. 4c), and the population structure obtained (POP1) (Fig. 5) support the hypothesis that *L. infantum* was introduced several times and at several points along the American continent from Europe, possibly during Spanish and Portuguese colonizations [53–56]. From the evidence, we propose that North African strains were introduced into Colombia possibly during the arrival of African immigrants/slaves [57], as occurred with *Plasmodium falciparum*, which came to the Americas via the trans-Atlantic slave trade [58]. We theorize that the establishment of North African genotypes in Colombia but not in regions geographically close to Colombia (e.g. Brazil), as well as the distant genomic relationship between the Colombian and Brazilian genomes (Fig. 1a, Table S5), could be associated with the geographical distribution of vectors and reservoirs species found in both countries (*Pintomyia evansi* and *Didelphis marsupialis* in Colombia [59, 60] and *Lutzomyia cruzi* and *D. albiventris* [61, 62] in Brazil). The geographical distributions of some of these vectors are restricted and limited to specific biomes (sylvatic, domestic or peridomestic), which could affect not only the dynamics of *L. infantum* transmission but also the epidemiology of the disease. Together, these findings led us to elucidate the possible origin of *L. infantum* strains in Colombia and to understand the mode of dispersion of this parasite in the American territory. However, we believe that genotyping more strains from South America (mainly from Colombia) would help to test this hypothesis. Additionally, the phylogenetic analysis of mitochondrial (maxicircle) genomes allowed us to identify the high level of intra-specific conservation of *L. infantum* worldwide (Fig. 1b) and once again demonstrated the utility of this region as suitable phylogenetic markers for *Leishmania* typing and for exploring evolutionary relationships within the trypanosomatids [63, 64].

We subsequently analysed the structural changes at the chromosome/gene level. Previously, various studies have described aneuploidy as an important mechanism of adaptation used by Old and New World *Leishmania* species [36, 40, 41, 65, 66]; however, few studies have focused on analysing whether the structural changes observed in species are maintained across different geographical regions. The results obtained confirm not only the widely described [10, 12–15, 41, 65, 67] extensive aneuploidy of *L. infantum* but also the conservation of these structural changes at the global level (Figs 2 and S2), which suggests that, regardless of the genetic background of the host and the eco-epidemiological niche in which the parasite exists, the chromosomal architecture of *L. infantum* needed to undergo extensive changes during possible adaptation mechanisms. On the other hand, and despite observing a low number of genes with CNVs in our dataset (Fig. 3a), the ontology enrichment analysis revealed an increase, at the global level, in genes directly involved with the virulence and infectivity of the parasite (e.g. genes associated with translational elongation). There was also an increase (except in European genomes) in genes involved in defenses against oxidative stress (regulation of nitric oxide) [68] (Fig. 3c). These findings highlighted and confirmed the importance of these genes in *L. infantum* as possible mechanisms of survival and adaptation in both the vector and the mammal host [69]. Finally, we identified an increase in genes involved in the transport and homeostasis of zinc (a component crucial in the cellular functions of this parasite) [70]. Despite these genes having been previously reported in *L. infantum* [70] and other *Leishmania* species [71, 72], it is interesting that the increase was observed only in the European genomes. Considering these findings, we believe that further studies are necessary to determine if this variation depends on the geographical area analysed or if it is the result of the culture conditions.

To evaluate the intraspecific genomic variability of *L. infantum* worldwide, we analysed the presence of nucleotide-level variations (SNPs) in the 187 genomes. The results demonstrated that, in general, *L. infantum* presents low genetic diversity compared with other *Leishmania* species [41, 66, 73]. When we analysed and compared this diversity among the five geographical regions, we found *L. infantum* from South America had the lowest genetic heterogeneity (Figs 4a and S4), a property previously ascribed to Brazilian genomes [9, 10, 14, 74]. We consider that this low diversity could be associated with (i) the relatively recent arrival of *L. infantum* to the Americas [53, 55]; (ii) the number of vectors and/or animal reservoirs involved in the transmission cycle [75, 76] where *L. infantum* could have easily adapted, thus favouring certain genotypes and promoting the stochastic loss of others (as has been described in *Trypanosoma brucei* [77] and suggested for other *Leishmania* species, such as *L. panamensis* [72]); or (iii) the low number of samples included in the study (Colombia *N*=4), which could influence the accuracy of the results. Therefore, additional studies involving representative sampling from Colombia are necessary to confirm these findings. Additionally, the SNP analysis allowed us to identify, at the global level, great variations in the chromosomes (Fig. 4b) containing genes involved in parasite virulence or defence against harmful oxidative effects, such as the A2 family genes found on chromosome 22 required for visceralization [78] and recently assembled for *L. infantum* [37], and heat shock protein 70 and trypanothione synthetase genes found on chromosome 27, which are associated with resistance and metabolism antimony. Another interesting finding was the identification of 29 SNPs shared between the 187 genomes that were located in genes associated with drug resistance, metabolism and signal pathways (Fig. 4c). These results agree with the results of previous studies [14, 79] and suggest that some of these changes could be the result of positive/purifying selection events that bestowed certain advantages upon the parasite in its adaption to ecological niches or promoted the transmission cycle.

Finally, the data for the 187 genomes included in the study were used to analyse the genetic diversity of the antigens used in VL diagnosis (rK28 and rK39). The results revealed there is high molecular divergence at the global level among the genomes (Fig. 6), which agrees with studies conducted on *L. donovani*,

demonstrating the extensive genetic diversity among genotypes and geographical origins [21, 80, 81]. The rK39 nucleotide sequence revealed a moderate genetic heterogeneity of this protein mainly in samples from Europe and Africa (Fig. 6a); genetic variability that could generate a differential antibody response in the host, impacting directly the immunochromatographic tests performance. The results obtained in this study and the hypothesis proposed is in line with the low sensitivity of these diagnostic tests, reported in East Africa, France and Spain, that have showed sensitivity percentages of 75–85.3%, 88.5 and 78%, respectively [17, 20–23, 82]. Contrasting the results observed in Colombia and some regions of Brazil where the sensitivity percentage ranked between 85–93% [18, 24], findings that could agree with the low genomic variability (Fig. 6a) described in this study.

Despite the three markers (rK28/K9 and rK39) showed genetic variability in many of the geographical regions evaluated, we found rK28fus had the highest variability, which was evidenced in the highest number of clustering events (Fig. 6b), the highest haplotypic diversity and the number of substitution sites (*S*), as well as the highest neutrality test (Tajima's D and Fu's Fs) and d*N*/d*S* index scores (Table 3). Despite some studies describe that the rK28-based tests present a high sensitivity (95–100%) in countries such as Sudan, Ethiopia and India suggesting a drastic genetic drift in the sequences evaluated. Herein, the results demonstrate that rK28 is a protein highly diverse worldwide, mainly in countries of North Africa and Asia. This variability could impact indirectly in the sensitivity of diagnostic tests. Likewise, to analyse target sequence of rK39 protein used in the diagnostic tests, an interesting non-synonymous substitution in the amino acid 27 of the protein was observed (change of a methionine [M] or serine [S] by a threonine [T]) (Fig. 7). Considering the biochemical characteristics of the amino acid and the fundamental role they play in both the structure and the functionality of the protein, we consider that additional studies should be performed to evaluate the possible consequences of the change observed.

The high polymorphism found in the rK28 marker at the global level, the low phylogenomic relationships observed among sequences from geographically close countries (e.g. Brazil and Colombia) (Fig. 6b), the limited specificity of this marker in some geographical regions [26, 83], and the discrepancies in the specificity found among kits based on this or other antigens (e.g. rK39) [18] used in different geographical regions [20, 26] highlight the need for diagnostic tests designed that contain as base specific antigens from each geographical region. More specific tests will contribute considerably to the rapid and effective global diagnosis of canine and human VL, as well as decrease the frequency of false-negative reports (~10.5%) [84] and the degree of cross-reactivity with other diseases (e.g. malaria, intestinal schistosomiasis, and tuberculosis) observed [26, 85].

In summary our whole-genome sequence data represents, to date, the biggest global distribution of *L. infantum* analysis. The results obtained for this species demonstrate the low variation at the gene level, the extensive aneuploidy, the low levels of allelic diversity, and the presence of different phylogroups of *L. infantum* circulating in the world. Furthermore, we found a close relationship between the Colombian *L. infantum* genomes and the North African and European genomes. Finally, this was the first study to use whole-genome sequencing that demonstrated the high genetic variability of the main antigens used in the diagnosis of VL at the global level. These findings are relevant, as they facilitate an understanding of the genetic behaviour of these antigens according to geographical area, and the data indicate there is a need to develop for new diagnostic targets, which would contribute to improving the rapid diagnostic techniques for VL.

Funding information
This research was funded by Dirección de Investigación e Innovación from Universidad del Rosario.

Acknowledgements
We would like to thank the Programa de Control y Estudio de Enfermedades Tropicales (PECET) and Universidad Sucre (Colombia) for the assistance with the sampling. We also thank Dr. Hideo Imamura for assistance in the analyses. This work was funded by DIRECCIÓN DE INVESTIGACIÓN E INNOVACIÓN from Universidad del Rosario. Juan David Ramírez González, Ph.D. is a Latin American fellow in the Biomedical Sciences supported by The Pew Charitable Trusts. We thank Suzanne Leech, Ph.D., from Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript.

Author contributions
Conceptualization: L.H.P and J.D.R. Methodology: L.H.P, M.M and A.C. Formal analysis: L.H.P and M.M. Investigation: L.H.P and J.D.R. Supervision: J.D.R. Writing – original draft: L.H.P and J.D.R. Writing – review and editing: L.H.P, J.D.R. A.C, M.M, C.M, M.R. and E.B.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
This study was approved by the Ethics Committee of the Universidad de Antioquia (Number VRI3445/2010) and the Universidad de Sucre in accordance with the resolution numbers 36836 and 2–2012, respectively. Written informed consent was obtained from the patients from which the strains were isolated.

References

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P. Leishmaniasis worldwide and global estimates of its incidence. *PloS one* 2012;7:e35671.
2. Burza S, Croft SL, Boelaert M. Leishmaniasis. *Lancet* 2018;392:951–970.
3. Organization. PAH. Epidemiological report of the Americas. Report N° 7 Dec-2019. 2019.
4. Chappuis F, Sundar S, Hailu A, Ghalib H, Rijal S. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? *Nat Rev Microbiol* 2007;5:873–882.
5. Pennisi MG, Persichetti MF. Feline leishmaniosis: Is the cat a small dog? *Vet Parasitol* 2018;251:131–137.
6. Dantas-Torres F. *Canine leishmaniosis* in South America. *Parasites Vectors* 2009;2:S1.
7. Kaszak I, Planellas M, Dworecka-Kaszak B. *Canine leishmaniosis* - an emerging disease. *Ann Parasitol* 2015;61:69–76.

8. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* 2007;39:839–847.
9. Carnielli JBT, Crouch K, Forrester S, Silva VC, Carvalho SFG. A Leishmania infantum genetic marker associated with miltefosine treatment failure for visceral leishmaniasis. *EBioMedicine* 2018;36:83–91.
10. Teixeira DG, Monteiro GRG, Martins DRA, Fernandes MZ, Macedo-Silva V. Comparative analyses of whole genome sequences of *Leishmania infantum* isolates from humans and dogs in northeastern Brazil. *Int J Parasitol* 2017;47:655–665.
11. Yasur-Landau D, Jaffe CL, David L, Doron-Faigenboim A, Baneth G. Resistance of *Leishmania infantum* to allopurinol is associated with chromosome and gene copy number variations including decrease in the S-adenosylmethionine synthetase (METK) gene copy number. *Int J Parasitol Drugs Drug Resist* 2018;8:403–410.
12. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M. Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated Leishmania population. *PLoS Genet* 2014;10:e1004092.
13. Bussotti G, Benkahla A, Jeddi F, Souiai O, Aoun K. Nuclear and mitochondrial genome sequencing of North-African Leishmania infantum isolates from cured and relapsed visceral leishmaniasis patients reveals variations correlating with geography and phenotype. *Microb Genom* 2020;6.
14. Franssen SU, Durrant C, Stark O, Moser B, Downing T. Global genome diversity of the *Leishmania donovani* complex. *eLife* 2020;9.
15. Carvalho KSS, da Silva Júnior WJ, da Silveira Regueira Neto M, Silva VC, de Sá Leitão Paiva Júnior S, *et al*. Application of Next Generation sequencing (NGS) for descriptive analysis of 30 genomes of Leishmania infantum isolates in middle-north Brazil. *Sci Rep* 2020;10:12321.
16. SIVIGILA. Boletin epidemiologico. *Sistema Nacional de Vigilancia en Salud Publica* 2019.
17. Kuhne V, Rezaei Z, Pitzinger P, Buscher P. Systematic review on antigens for serodiagnosis of visceral leishmaniasis, with a focus on East Africa. *PLoS Negl Trop Dis* 2019;13:e0007658.
18. Herrera G, Castillo A, Ayala MS, Florez C, Cantillo-Barraza O. Evaluation of four rapid diagnostic tests for canine and human visceral *Leishmaniasis* in Colombia. *BMC Infect Dis* 2019;19:747.
19. Burns JM, Shreffler WG, Benson DR, Ghalib HW, Badaro R. Molecular characterization of a kinesin-related antigen of *Leishmania chagasi* that detects specific antibody in African and American visceral leishmaniasis. *Proc Natl Acad Sci USA* 1993;90:775–779.
20. Pattabhi S, Whittle J, Mohamath R, El-Safi S, Moulton GG. Design, development and evaluation of rK28-based point-of-care tests for improving rapid diagnosis of visceral leishmaniasis. *PLoS Negl Trop Dis* 2010;4.
21. Bhattacharyya T, Boelaert M, Miles MA. Comparison of visceral leishmaniasis diagnostic antigens in African and Asian *Leishmania donovani* reveals extensive diversity and region-specific polymorphisms. *PLoS Negl Trop Dis* 2013;7:e2057.
22. Cunningham J, Hasker E, Das P, El Safi S, Goto H. A global comparative evaluation of commercial immunochromatographic rapid diagnostic tests for visceral leishmaniasis. *Clin Infect Dis* 2012;55:1312–1319.
23. Abass E, Kang C, Martinkovic F, Semiao-Santos SJ, Sundar S. Heterogeneity of *Leishmania donovani* parasites complicates diagnosis of visceral leishmaniasis: comparison of different serological tests in three endemic regions. *PloS one* 2015;10:e0116408.
24. da Silva MRB, Brandao NAA, Colovati M, de Sousa MMP, de Lima LC. Performance of two immunochromatographic tests for diagnosis of visceral leishmaniasis in patients coinfected with HIV. *Parasitol Res* 2018;117:419–427.
25. Mukhtar M, Abdoun A, Ahmed AE, Ghalib H, Reed SG. Diagnostic accuracy of rK28-based immunochromatographic rapid diagnostic tests for visceral leishmaniasis: a prospective clinical cohort study in Sudan. *Trans R Soc Trop Med Hyg* 2015;109:594–600.
26. Bezuneh A, Mukhtar M, Abdoun A, Teferi T, Takele Y. Comparison of point-of-care tests for the rapid diagnosis of visceral leishmaniasis in East African patients. *Am J Trop Med Hyg* 2014;91:1109–1115.
27. Venturin GL, Bragato JP, Silva KL, de Lima VM. Recombinant K28 antigen in ELISA in the diagnosis of canine visceral leishmaniosis. *Parasite Immunol* 2015;37:670–673.
28. Schwabl P, Boite MC, Bussotti G, Jacobs A, Andersson B. Colonization and genetic diversification processes of *Leishmania infantum* in the Americas. *Commun Biol* 2021;4:139.
29. Brotherton MC, Bourassa S, Leprohon P, Legare D, Poirier GG. Proteomic and genomic analyses of antimony resistant *Leishmania infantum* mutant. *PloS one* 2013;8:e81899.
30. Valdivia HO, Almeida LV, Roatt BM, Reis-Cunha JL, Pereira AA. Comparative genomics of canine-isolated *Leishmania* (Leishmania) amazonensis from an endemic focus of visceral leishmaniasis in Governador Valadares, southeastern Brazil. *Sci Rep* 2017;7:40804.
31. Ramirez JD, Hernandez C, Leon CM, Ayala MS, Florez C. Taxonomy, diversity, temporal and geographical distribution of Cutaneous Leishmaniasis in Colombia: A retrospective study. *Sci Rep* 2016;6:28266.
32. Patino LH, Mendez C, Rodriguez O, Romero Y, Velandia D. Spatial distribution, *Leishmania* species and clinical traits of Cutaneous Leishmaniasis cases in the Colombian army. *PLoS Negl Trop Dis* 2017;11:e0005876.
33. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* 2007;2007:pdb
34. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:1530–1534.
35. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
36. Patino LH, Imamura H, Cruz-Saavedra L, Pavia P, Muskus C. Major changes in chromosomal somy, gene expression and gene dosage driven by Sb(III) in Leishmania braziliensis and *Leishmania panamensis*. *Sci Rep* 2019;9:9485.
37. Gonzalez-de la Fuente S, Peiro-Pastor R, Rastrojo A, Moreno J, Carrasco-Ramiro F. Resequencing of the *Leishmania infantum* (strain JPCM5) genome and de novo assembly into 36 contigs. *Sci Rep* 2017;7:18050.
38. Ilg T. Proteophosphoglycans of *Leishmania*. *Parasitol Today* 2000;16:489–497.
39. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–267.
40. Cuypers B, Berg M, Imamura H, Dumetz F, De Muylder G. Integrated genomic and metabolomic profiling of ISC1, an emerging *Leishmania donovani* population in the Indian subcontinent. *Infect Genet Evol* 2018;62:170–178.
41. Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J. Modulation of aneuploidy in *Leishmania donovani* during adaptation to different *in vitro* and *in vivo* environments and its impact on gene expression. *mBio* 2017;8.
42. Zhao S, Guo Y, Sheng Q, Shyr Y. Advanced heat map and clustering analysis using heatmap3. *Biomed Res Int* 2014;2014:986048.
43. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* 2011;6:e21800.
44. Tihon E, Imamura H, Van den Broeck F, Vermeiren L, Dujardin JC. Genomic analysis of Isometamidium Chloride resistance in Trypanosoma congolense. *Int J Parasitol Drugs Drug Resist* 2017;7:350–361.
45. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 2014;197:573–589.
46. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 2005;14:2611–2620.
47. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10:564–567.

48. Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC genomics* 2012;13:338.
49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
50. Okonechnikov K, Golosova O, Fursov M, team U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 2012;28:1166–1167.
51. Poon AF, Frost SD, Pond SL. Detecting signatures of selection from DNA sequences using Datamonkey. *Methods Mol Biol* 2009;537:163–183.
52. Ribolla PE, Gushi LT, Pires E, Costa CH, Costa DL. *Leishmania infantum* genetic diversity and lutzomyia longipalpis mitochondrial haplotypes in Brazil. *Biomed Res Int* 2016;2016:9249217.
53. Boite MC, Spath GF, Bussotti G, Porrozzi R, Morgado FN. Transatlantic spill over: deconstructing the ecological adaptation of *Leishmania infantum* in the Americas. *Genes (Basel)* 2019;11.
54. Leblois R, Kuhls K, Francois O, Schonian G, Wirth T. Guns, germs and dogs: On the origin of *Leishmania chagasi*. *Infect Genet Evol* 2011;11:1091–1095.
55. Lukes J, Mauricio IL, Schonian G, Dujardin JC, Soteriadou K. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc Natl Acad Sci USA* 2007;104:9375–9380.
56. Kuhls K, Alam MZ, Cupolillo E, Ferreira GE, Mauricio IL. Comparative microsatellite typing of new world leishmania infantum reveals low heterogeneity among populations and its recent old world origin. *PLoS Negl Trop Dis* 2011;5:e1155.
57. Nozais JP. [Hypotheses on the origin of certain parasites on the Latin American continent]. *Bull Soc Pathol Exot Filiales* 1985;78:401–412.
58. Tagliamonte MS, Yowell CA, Elbadry MA, Boncy J, Raccurt CP. Genetic markers of adaptation of plasmodium falciparum to transmission by american vectors identified in the genomes of parasites from Haiti and South America. *mSphere* 2020;5.
59. Gonzalez C, Cabrera OL, Munstermann LE, Ferro C. [Distribution of *Leishmania infantum* vector species in Colombia]. *Biomedica* 2006;26:64–72.
60. Ardila MM, Carrillo-Bonilla L, Pabon A, Robledo SM. Surveillance of phlebotomine fauna and *Didelphis marsupialis* (Didelphimorphia: Didelphidae) infection in an area highly endemic for visceral leishmaniasis in Colombia. *Biomedica* 2019;39:252–264.
61. Falcao de Oliveira E, Galati EAB, Oliveira AG, Rangel EF, Carvalho BM. Ecological niche modelling and predicted geographic distribution of *Lutzomyia cruzi*, vector of *Leishmania infantum* in South America. *PLoS Negl Trop Dis* 2018;12:e0006684.
62. Silva EM, Alves LC, Guerra NR, Farias MP, Oliveira EL. *Leishmania* spp. in *Didelphis* spp. from Northeastern Brazil. *J Zoo Wildl Med* 2016;47:942–944.
63. Kaufer A, Barratt J, Stark D, Ellis J. The complete coding region of the maxicircle as a superior phylogenetic marker for exploring evolutionary relationships between members of the *Leishmaniinae*. *Infect Genet Evol* 2019;70:90–100.
64. Camacho E, Rastrojo A, Sanchiz A, Gonzalez-de la Fuente S, Aguado B. Leishmania mitochondrial genomes: maxicircle structure and heterogeneity of minicircles. *Genes (Basel)* 2019;10.
65. Patino LH, Muskus C, Munoz M, Ramirez JD. Genomic analyses reveal moderate levels of ploidy, high heterozygosity and structural variations in a Colombian isolate of *Leishmania* (Leishmania) amazonensis. *Acta Trop* 2019;203:105296.
66. Patino Luz HM, Carlos M, David RJ. Genomic diversification, structural plasticity, and hybridization in *Leishmania* (Viannia) braziliensis. *Front Cell Infect Microbiol* 2020;10.
67. Bussotti G, Gouzelou E, Cortes Boite M, Kherachi I, Harrat Z. Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *mBio* 2018;9.
68. Olekhnovitch R, Ryffel B, Muller AJ, Bousso P. Collective nitric oxide production provides tissue-wide immunity during *Leishmania* infection. *J Clin Invest* 2014;124:1711–1722.
69. Timm T, Annoscia G, Klein J, Lochnit G. The Eukaryotic Elongation Factor 1 Alpha (eEF1alpha) from the Parasite *Leishmania infantum* is modified with the immunomodulatory substituent phosphorylcholine (PC). *Molecules* 2017;22.
70. Carvalho S, Barreira da Silva R, Shawki A, Castro H, Lamy M. LiZIP3 is a cellular zinc transporter that mediates the tightly regulated import of zinc in *Leishmania infantum* parasites. *Mol Microbiol* 2015;96:581–595.
71. Saini S, Bharati K, Shaha C, Mukhopadhyay CK. Zinc depletion promotes apoptosis-like death in drug-sensitive and antimony-resistance *Leishmania donovani*. *Sci Rep* 2017;7:10488.
72. Patino LH, Munoz M, Muskus C, Mendez C, Ramirez JD. Intraspecific genomic divergence and minor structural variations in *Leishmania* (Viannia) panamensis. *Genes (Basel)* 2020;11.
73. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM. Gene expression in leishmania is regulated predominantly by gene dosage. *mBio* 2017;8.
74. Segatto M, Ribeiro LS, Costa DL, Costa CH, Oliveira MR. Genetic diversity of Leishmania infantum field populations from Brazil. *Mem Inst Oswaldo Cruz* 2012;107:39–47.
75. Diniz SA, Silva FL, Carvalho Neta AC, Bueno R, Guerra RM. Animal reservoirs for visceral leishmaniasis in densely populated urban areas. *J Infect Dev Ctries* 2008;2:24–33.
76. Gonzalez C, Paz A, Ferro C. Predicted altitudinal shifts and reduced spatial distribution of Leishmania infantum vector species under climate change scenarios in Colombia. *Acta Trop* 2014;129:83–90.
77. Oberle M, Balmer O, Brun R, Roditi I. Bottlenecks and the maintenance of minor genotypes during the life cycle of *Trypanosoma brucei*. *PLoS Pathog* 2010;6:e1001023.
78. McCall LI, Zhang WW, Matlashewski G. Determinants for the development of visceral leishmaniasis disease. *PLoS Pathog* 2013;9:e1003053.
79. Cordeiro AT, Michels PA, Delboni LF, Thiemann OH. The crystal structure of glucose-6-phosphate isomerase from *Leishmania mexicana* reveals novel active site features. *Eur J Biochem* 2004;271:2765–2772.
80. Kuhls K, Keilonat L, Ochsenreither S, Schaar M, Schweynoch C. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. *Microbes Infect* 2007;9:334–343.
81. Alam MZ, Kuhls K, Schweynoch C, Sundar S, Rijal S. Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of *Leishmania donovani* strains in the Indian subcontinent. *Infect Genet Evol* 2009;9:24–31.
82. Bangert M, Flores-Chavez MD, Llanes-Acevedo IP, Arcones C, Chicharro C. Validation of rK39 immunochromatographic test and direct agglutination test for the diagnosis of Mediterranean visceral leishmaniasis in Spain. *PLoS Negl Trop Dis* 2018;12:e0006277.
83. WHO DEST. Visceral leishmaniasis rapid diagnostic test performance. 2010.
84. Oliveira Dos Santos Maciel M, Soares MF, Costa SF, Bragato JP, Rebech GT, *et al*. Plasmonic rK28 ELISA improves the diagnosis of canine Leishmania infection. *Parasite Immunol* 2020;42:e12684.
85. Vaish M, Bhatia A, Reed SG, Chakravarty J, Sundar S. Evaluation of rK28 antigen for serodiagnosis of visceral Leishmaniasis in India. *Clin Microbiol Infect* 2012;18:81–85.