

This is a repository copy of *A Leishmania infantum genetic marker associated with miltefosine treatment failure for visceral leishmaniasis.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/135849/>

Version: Accepted Version

Article:

Mottram, Jeremy Charles orcid.org/0000-0001-5574-3766, Jeffares, Daniel orcid.org/0000-0001-7320-0706, Brambilla Carnielli Trindade, Juliana et al. (10 more authors) (2018) A *Leishmania infantum* genetic marker associated with miltefosine treatment failure for visceral leishmaniasis. *EBioMedicine*. ISSN 2352-3964

<https://doi.org/10.1016/j.ebiom.2018.09.029>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A *Leishmania infantum* genetic marker associated with miltefosine treatment failure for visceral leishmaniasis.

Juliana B. T. Carnielli D.Sc.,^{1,2,3} Kathryn Crouch PhD,³ Sarah Forrester PhD,² Vladimir Costa Silva D.Sc.,⁴ Sílvia F. G. Carvalho D.Sc.,⁵ Jeziel D. Damasceno D.Sc.,⁶ Elaine Brown BSc,² Nicholas J. Dickens PhD,³ Dorcas L. Costa D.Sc.,⁴ Carlos H. N. Costa D.Sc.,⁴ Reynaldo Dietze M.D.,^{1,7} Daniel C. Jeffares PhD,² Jeremy C. Mottram PhD,^{2,3}

¹ Laboratório de Leishmanioses, Núcleo de Doenças Infecciosas, Universidade Federal do Espírito Santo, Vitória-ES, Brazil.

² Centre for Immunology and Infection, Department of Biology, University of York, United Kingdom.

³ Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, United Kingdom.

⁴ Laboratório de Pesquisas em Leishmanioses, Instituto de Doenças Tropicais Natan Portella, Universidade Federal do Piauí, Teresina-PI, Brazil.

⁵ Hospital Universitário Clemente de Faria, Universidade Estadual de Montes Claros, Montes Claros-MG, Brazil.

⁶ Laboratório de Biologia Molecular de *Leishmania*, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto-SP, Brazil.

⁷ Instituto de Higiene e Medicina Tropical, Universidade NOVA de Lisboa –Lisbon, Portugal.

Brief title: Miltefosine prognostic marker in VL

Corresponding Author

Jeremy C. Mottram

Centre for Immunology and Infection. Department of Biology. University of York. Wentworth Way
Heslington, York, United Kingdom. YO10 5DD.

Phone: 01904 328869

jeremy.mottram@york.ac.uk

Alternate Corresponding Author

Juliana B. T. Carnielli

Laboratório de Leishmanioses, Núcleo de Doenças Infecciosas. Centro de Ciências da Saúde.
Universidade Federal do Espírito Santo. Avenida Marechal Campos, 1468, Maruípe, Vitória – ES,
Brazil. 29043-900.

Phone: +55 27 33357210.

julianacarnielli@gmail.com

Abstract

Background: Miltefosine has been used successfully to treat visceral leishmaniasis (VL) in India, but it was unsuccessful for VL in a clinical trial in Brazil.

Methods: To identify molecular markers that predict VL treatment failure whole genome sequencing of 26 *L. infantum* isolates, from cured and relapsed patients allowed a GWAS analysis of SNPs, gene and chromosome copy number variations.

Findings: A strong association was identified ($p=0.0005$) between the presence of a genetically stable *L. infantum* Miltefosine Sensitivity Locus (MSL), and a positive response to miltefosine treatment. The risk of treatment failure increased 9.4-fold (95% CI 2.11–53.54) when an isolate did not have the MSL. The complete absence of the MSL predicted miltefosine failure with 0.92 (95% CI 0.65–0.996) sensitivity and 0.78 (95% CI 0.52–0.92) specificity. A genotyping survey of *L. infantum* ($n=157$) showed that the frequency of MSL varies in a cline from 95% in North East Brazil to less than 5% in the South East. The MSL was found in the genomes of all *L. infantum* and *L. donovani* sequenced isolates from the Old World ($n=671$), where miltefosine can have a cure rate higher than 93%.

Interpretation: Knowledge on the presence or absence of the MSL in *L. infantum* will allow stratification of patients prior to treatment, helping to establish better therapeutic strategies for VL treatment.

Funding: CNPq, FAPES, GCRF MRC and Wellcome Trust.

Keywords: Visceral leishmaniasis; Miltefosine treatment failure; Whole-genome sequencing; Miltefosine Susceptibility Locus; Prognostic marker.

Research in Context

Evidence before this study

Miltefosine has been used with success to treat visceral leishmaniasis caused by *L. donovani* on the Indian subcontinent. However, the clinical trial conducted in Brazil resulted in about 40% of patients with *L. infantum* infection relapsing after the miltefosine treatment. Previous work has demonstrated that miltefosine resistance can be selected easily in vitro. These *Leishmania* laboratory lines have been used to understand the molecular basis of miltefosine resistance; however, studies using field isolates with different responses to miltefosine are still needed to understand the underlying mechanism in natural populations of *Leishmania*.

Added value of this study

This study generated genome sequences of 26 *L. infantum* isolates, obtained from patients with different miltefosine treatment outcomes. This provides extensive information on genomic variation (SNPs, InDels, and gene and chromosome copy number variation) unique to the parasites of Brazil including a new molecular marker of miltefosine treatment failure (MSL). The MSL contains four genes (3'-nucleotidase/nucleases *LinJ.31.2370* and *LinJ.31.2380*; helicase-like protein *LinJ.31.2380*; and 3,2-trans-enoyl-CoA isomerase *LinJ.31.2400*). The MSL has not been identified in any studies where miltefosine resistance has been generated in *Leishmania* in the laboratory.

Implications of all the available evidence

Our genomic analysis showed MSL as a potential molecular marker to predict miltefosine treatment outcome in visceral leishmaniasis. Our work highlights the importance of genotyping this locus in *Leishmania* field isolates to contribute to the rational use of miltefosine and design new therapeutic strategies for the treatment of VL.

Introduction

Visceral leishmaniasis (VL) is a neglected of tropical diseases that is endemic in more than 65 countries, with major foci in the Indian subcontinent, East Africa and Latin America. VL has a case-fatality rate of ~10% from an estimated 200,000-400,000 cases per year. VL is the most severe form of a complex of leishmaniasis diseases and is caused by *Leishmania donovani* and *L. infantum* (synonymy with *L. chagasi*) [1]. The number of cases of VL in the Indian subcontinent is decreasing and elimination has been seen as achievable with effective deployment of existing control measures [2], although challenges remain [3]. VL in Brazil is a zoonotic disease caused by *L. infantum*, with the dog as the primary reservoir. Brazil reports around 3,000 new case of VL a year; the disease was originally restricted to remote rural areas in Brazil, but has now become prevalent in urban centers, greatly enhancing the population at risk of infection especially in immunodeficient individuals [1, 4].

VL treatment relies on a few available drugs and in Brazil first line treatment is a 20-day course of meglumine antimoniate (20 mg Sb⁵⁺/kg/day), with liposomal amphotericin B (3 mg/kg/day for 7 days) or amphotericin B deoxycholate (1 mg/kg/day for 14 days) being second line treatments [5]. There are no new drugs in clinical trials and in light of increasing treatment failure [6] efforts are currently focussed on improving current treatments with the available drugs, including in combinations [7]. The identification of an effective and safe oral drug, miltefosine (hexadecylphosphocholine), was an important advance in leishmaniasis therapy [8]. Miltefosine was the first oral drug approved for VL treatment in India [9, 10]. While miltefosine was able to produce a clinical cure in about 94% of VL caused by *L. donovani* when first introduced in India [9, 10], a phase 2 dose-ranging trial in Brazil showed that the cure rate was much lower (~60%). The reason for the low cure rate is unknown, but may be associated with a natural resistance to miltefosine within the circulating population of *L. infantum* in Brazil. Whilst resistance to miltefosine is easily generated in the laboratory, and some resistance mechanisms have been determined [11-14], the molecular basis involved in miltefosine treatment failure are required, so that diagnostic markers can be developed to inform clinical practice. The purpose of this genome-wide association study (GWAS) was to investigate the *L. infantum* genomic variation associated with miltefosine treatment failure in Brazil.

We pinpoint new molecular components that correlate with miltefosine treatment failure and that can be used in clinical settings to establish the best therapeutic strategies for VL treatment.

Materials and Methods

Study design, patients and parasites

The genome-wide association (GWA) study, designed to identify genetic markers of miltefosine treatment failure, was performed with 26 pre-treatment *L. infantum* isolates (14 from cured and 12 from relapsed patients) recovered out of the 42 VL patients enrolled in the clinical trial designed to evaluate the efficacy and toxicity of miltefosine in treatment of VL in Brazil (Montes Claros, Minas Gerais and Teresina, Piauí) in 2005-2007 (Figure 1A). The patients were treated with 2.5 mg/kg/day of miltefosine for 28 days (14 patients) or 42 days (28 patients), and were followed for at least six months, but at most one year after treatment. Patients were considered cured if no signs and symptoms of the disease were present at the time of examination. Relapse was defined as a patient who was considered cured, but upon follow-up, showed reappearance of clinical signs and positive parasitology. The protocol was approved by the Comissão Nacional de Ética em Pesquisa (CONEP D-18506-Z019) and are registered with ClinicalTrials.gov, number NCT00378495. Ethical clearance, which was a waiver for informed consent from patients, for utilization in research of the *L. infantum* isolates obtained from patients enrolled in miltefosine Brazilian trial was obtained from the institutional review board of the Centro de Ciências da Saúde, Universidade Federal do Espírito Santo (CEP-066/2007), Brazil.

Genomic Analysis

Genome sequencing was performed with paired-end Illumina reads. Details of DNA preparation, sequencing, genotyping, and bioinformatics analysis are described in the appendix and Figure 1B. Briefly, reads were aligned to the resequenced *L. infantum* JPCM5 reference genome, downloaded from <http://leish-esp.cbm.uam.es> version 1. Copy number variations (CNV) at chromosome and gene

level were calculated according to Rogers et al [15], using OrthoMCL to assign genes to 7,822 multi-copy gene clusters (orthologous groups, OGs). Single nucleotide polymorphisms (SNPs) and small insertion-deletion polymorphism (InDel) calls were generated by GATK HaplotypeCaller [16] and Freebayes [17], accepting only variants called by both callers.

CNVs may be unstable within hosts and/or within culture [18]. To examine whether they were sufficiently stable to be useful as prognostic markers we estimated the heritability of these 7,822 OGs, using a restricted maximum likelihood method (REML) implemented in LDAK [19], using a kinship matrix derived from SNPs and InDels. OGs that were inherited in a stable manner with SNPs (n=757) were defined as those whose heritability was two standard deviations above zero. GWAS to test for associations between OGs and treatment outcome were performed with the 757 stable OGs using custom R scripts. Associations were detected using Mann-Whitney tests, corrected for multiple tests by permuting the cure/relapse trait.

Genome-wide association analyses (GWAS) were performed using SNPs and InDels to test for statistical associations with miltefosine treatment outcome using LDAK [19], with the 1752 variants that had a minor allele count greater than 1. The SNP and InDel kinship matrix as above was used to control for unequal relatedness of strains. We estimated the P-value threshold by permuting the cure/relapse trait for individuals 1,000 times, recording the lowest P-value, and using the 5% quantile (50th lowest value) of these values as the threshold. Variants passing this threshold (4.7×10^{-6}) therefore had a 5% error rate.

Genetic analysis of the MSL locus

For technical validation of NGS data, PCR amplification of the MSL on chromosome 31 was accomplished according to PCR strategy shown in figure 2A and supplementary table 1. The geographic distribution of genetic marker (MSL, Miltefosine Sensitivity Locus) highlighted by GWAS was investigated in the 26 isolates from the miltefosine trial and in another 131 *L. infantum* isolates from different regions of Brazil: by PCR in 111 isolates collected as part of VL diagnostic process in Brazil; and by analysis of 20 whole-genome parasite sequences available on Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>). The MSL frequency was also determined in

L. infantum or *L. donovani* from the Old World, using 671 whole-genome parasite sequences on the SRA.

Correlation between complete absence of MSL and miltefosine treatment failure was assessed by contingency table analysis (Fisher's exact test). Relative risk and Sensitivity/Specificity were estimated using the Koopman asymptotic score and Wilson-Brown tests, respectively.

Homogeneity of *L. infantum* clinical isolates

A number of isolates from the miltefosine trial showed heterogeneous MSL genotypes by PCR. To resolve these all heterogeneous samples (n=7) and homogeneous controls (n=3) were cloned, and re-screened for the MSL presence. The cloning, genomic DNA extraction, and MSL screening processes are described in the appendix. The natural mechanism of MSL deletion was investigated by cloning and sequencing the MSL flanking sequence in 21 *L. infantum* MSL⁻ or MSL^{+/-} isolates.

Results

A phase 2 dose-ranging trial to assess efficacy and safety of orally administered miltefosine in patients with VL in Brazil followed 42 patients: 14 from Montes Claros–MG and 28 from Teresina–PI sites (Figure 1A and Table 1). In Montes Claros, a standard dose and treatment length of 2·5 mg/Kg/day, with the maximum dose of 100 mg/day, for 28 days was used. All patients from Montes Claros presented initial cure, although eight relapsed after the treatment, showing a final cure rate of 43%. Because of the high relapse rate observed in Montes Claros an extended treatment of 42 days with the same dose was carried out in Teresina. In this site two patients did not respond to the miltefosine treatment and from the others who presented initial cure seven relapsed after the treatment, resulting in a final cure rate of 68%. These data together revealed a cure rate of ~60% in these two Brazilian regions, lower than that found for Indian VL when miltefosine was first used in India (>90%) [9, 10, 20]. The cure rate between the pediatric and the adult patients did not exhibit a significant difference (treatment failure rate of 52·2% [12/23] and 26·3% [5/19] in pediatric and adult patients, respectively, with Chi-square p=0·09).

To investigate the molecular basis of miltefosine treatment failure genome sequences were obtained from 26 pre-treatment *L. infantum* isolates (Table 1). Comparison to the Spanish reference strain (*L. infantum* JPCM5) identified 16,268 genetic variants, including 11,010 SNPs. Excluding variants that were fixed in all Brazilian isolates (that represent differences between the Brazilian population and the Spanish reference), left 1,969 variants that were polymorphic within Brazilian isolates (413 InDels; 1,535 SNPs; and 21 others) (Table 2 and Supplementary Table 2).

This relatively small number of variants made GWAS analysis feasible, because it reduced the statistical burden of multiple test correction. Genome-wide association analysis was performed using these 1,969 variants and clinical data to associate variants with relapse/cure treatment outcomes. The minimum association P-value was 3.1×10^{-3} . As this did not pass the empirical P-value threshold of 4.7×10^{-6} we conclude that none of these variants were significantly associated with cure/relapse. Whole-chromosome copy number analysis revealed different degrees of ploidy due to the high plasticity of *L. infantum* (Supplementary Figure 1). As with the small variants (SNPs and InDels), the diversity in whole-chromosome copy number was not able to distinguish the isolates from cured and relapsed VL patients and did not segregate with any other property of these samples such as geographic location.

Association analysis with heritable Orthologous Groups (OG) identified 59 that exhibited a significant difference in gene dose between the pre-treatment isolates from the cured and relapsed groups ($p < 0.05$, Mann-Whitney tests, permutation-corrected) (Supplementary Tables 3). The most significantly associated OG and two other highly significantly OGs were located in a 12.7 kb region on supernumerary chromosome 31 (the miltefosine sensitivity locus, MSL), which was present in the majority of isolates from cured patients, but absent in almost all isolates from relapsed patients (Figure 2 and Supplementary Figure 2).

The MSL contains four genes, two of which are tandem duplicates (*LinJ.31.2370* and *LinJ.31.2380*, encoding for 3'-nucleotidase/nuclease, putative). *LinJ.31.2390* (encoding for helicase-like protein) is a single-copy gene with no paralogues outside of the cluster, whilst *LinJ.31.2400* (encoding for 3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative) has a paralog upstream of the MSL

(*LinJ.31.2320*). The gene dosage for the MSL locus was between 1.5 – 2.3, which is lower than expected for genes on a tetraploid chromosome and is possibly due to the presence of mixed populations in the isolates (see below). The dosage of all orthologous groups within the MSL were highly heritable (heritability >90%), indicating that gene⁺ dosages segregate consistently with the SNPs, rather than fluctuating rapidly within the population, or within culture (Supplementary Table 3, Supplementary Figure 3). The MSL is a genetically stable, variable gene copy number marker that is strongly associated with treatment outcome (Supplementary Table 3, supplementary Figure 3). The presence of the MSL locus in the 26-sequenced *L. infantum* isolates was technically validated by PCR, which confirmed the genome sequencing data (Figure 2). Fisher's Exact test revealed that there is a very strong association (p=0.0005) between the presence of the MSL in the *L. infantum* genome and cure with miltefosine treatment, with a reciprocal relative risk of 9.43 (95% CI 2.11 to 53.54) to treatment failure when an isolate has a complete deletion of the MSL.

The PCR analyses revealed that seven *L. infantum* isolates (MA02A, MA05A, PI04A, PI05A, PI08A, PI09A and PI10A) from cured patients exhibited a mixed genomic profile, producing both MSL⁺ and MSL⁻ products, consistent with a reduced gene dose observed in the sequencing data for these isolates. To determine whether the heterogeneity resulted from loss of MSL on some chromosomes in individual cells or whether the population was a mix of genotypes, these isolates were cloned and analysed for the presence of the MSL. In addition, three isolates in which heterogeneity was not observed in the initial screen (MA01A, MG11A and MG14A) were cloned and analysed. This revealed that all the cloned cells were either MSL⁺ or MSL⁻ (Supplementary Figure 4), showing that the observed heterogeneity is due to a mix of genotypes in the original isolate. However, in some of the clones PCR analysis revealed a low intensity DNA fragment at ~1.2 kb in addition to the more prominent MSL DNA fragment at 14 kb. This DNA fragment was extracted and sequenced, revealing the novel junction formed after MSL deletion. It can be inferred that loss of an allele of the MSL can occur at low level within individual cells in a culture population.

The mechanism of MSL loss from *L. infantum* was investigated by sequencing the novel junctions formed after deletion of this locus. This revealed that the deletion occurred in exactly the same

position for all isolates (Supplementary Figure 5). The novel junction formed when the MSL is excised corresponds to the repetitive elements that make up the repeat alignment group RAG337 described by Ubeda et al [21]. We can speculate that deletion of the MSL occurred by homologous recombination using repetitive sequence flanking the locus (Figure 3), and that it rose to the relatively high frequency we observe in Brazil either by genetic drift or by selection for some trait other than miltefosine pressure.

To estimate the percentage of *L. infantum* parasites that might respond to treatment with miltefosine in different regions of Brazil, 131 more *L. infantum* isolates were screened for the presence of the MSL (Figure 4). Overall, 43% of these *L. infantum* isolates contained the MSL. The prevalence of the MSL was highest in states of the Northeast of Brazil, including Piauí, Maranhão and Rio Grande do Norte (74%), whereas isolates from the states of Espírito Santo, Minas Gerais and Bahia had a low MSL frequency (0%, 5% and 25%, respectively). In contrast the MSL was found in publically available whole genome sequence data for 671 Old World *L. donovani*/*L. infantum* isolates. In the vast majority of these Old World isolates, the gene dose for *LdBPK_312390* (the orthologue of *LinJ.31.2390*) is close to 4, which is as expected for this tetrasomic chromosome. The same pattern was observed for *LdBPK_312380*. Complete loss of these two genes together as described in *L. infantum* was not observed in the 671 genomes analysed, suggesting that the loss of the MSL locus occurred within Brazil.

Discussion

The genome-wide study with 26 New World Brazilian *L. infantum* isolates revealed 16,268 small genetic variants. Of these, 11,010 SNPs are present in all these Brazilian isolates, but are different from the Spanish Old World reference strain JPCM5. The low genetic diversity within Brazil suggests that these isolates are a recent population, in agreement with a relatively recent Old world origin of New World *L. infantum* [22]. The resistance to miltefosine observed *in vitro* has been proposed to occur via changes in membrane permeability [14] or a decrease in intracellular drug accumulation [23]. In contrast, our study did not associate any individual SNP or InDel with miltefosine treatment

failure, including in genes that code for the miltefosine transporter (*LinJ.13.1590*) and its β -subunit Ros3 (*LinJ.32.1040*), which have been associated with resistance to miltefosine in *Leishmania* spp. induced *in vitro* [11], and reported once in a *L. infantum* clinical isolate [24]. The absence of genetic variation in the miltefosine transporter in *L. infantum* isolates characterized here indicate that the genetic basis of miltefosine resistance induced *in vitro* are not responsible for miltefosine treatment failure in Brazil. It would be informative to carry out a prospective study using PCR detection of the MSL to predict miltefosine treatment outcome, but the drug is not currently licensed for use in Brazil.

The genome analysis revealed the large extent of chromosome and gene copy number variation among the *L. infantum* isolates. These data corroborate previous findings, which have established variable degrees of aneuploidy in strains and species of *Leishmania* grown in culture and in animal models of infection [15, 25, 26]. Although aneuploidy in *Leishmania* ssp. parasites has been observed previously in drug resistance selected in promastigotes *in vitro* [27, 28], our findings support the lack of correlation between aneuploidy and miltefosine treatment failure of clinical isolates [29]. It is likely that some aneuploidy variation arises through culture of promastigote parasites derived from the clinical isolates, as well as change in parasite's environment [18], complicating the analysis. Together with aneuploidy, gene amplification by expansion and contraction of genes in tandem arrays and by generation of extrachromosomal elements also contributes to gene-dosage fluctuation. In *Leishmania* parasites, which lack regulated transcription, these gene amplifications function as a mechanism to increase gene expression [18]. Variation in gene dose has also been associated with drug resistance in *Leishmania* parasites induced *in vitro* [21, 28], and naturally found in field [25], although for the first time, we document the gene dosage fluctuation associated with miltefosine treatment failure in VL caused by *L. infantum*.

In our study, 59 gene arrays, including genes contained within MSL, exhibited a significant difference in gene dosage ($p < 0.05$) between isolates from cured and relapsed patients. These loci all demonstrated high heritability (> 0.95), indicating stable inheritance with SNPs. Genes from the MSL locus were the most significantly associated with treatment outcome. Patients infected with an MSL⁺ or MSL^{+/-} *L. infantum* isolate will likely respond to treatment with miltefosine, whilst those infected

with MSL⁻ have 9.4 fold greater risk of relapsing after miltefosine treatment. The MSL locus was not completely predictive of outcome, indicating that there are other parasite genetic, environmental or host genetic factors involved. The sample size in this study is too small to detect subtle parasite genetic factors. Furthermore, the presence of a mixed population (MSL⁺ / MSL⁻) in some clinical isolates from cured patients hints at a possible fitness cost for MSL⁻ parasites when in the presence of MSL⁺ parasites following miltefosine treatment.

The MSL deletion process most likely occurs by homologous recombination using the direct repetitive sequence SIDER2 (Short Interspersed DEgenerate Retroposon) [21] that flanks the MSL. This repeat sequence is widely distributed in the *Leishmania* spp. genome, continuously undergoes rearrangement and is known to play a role in post-transcriptional control of gene expression [21, 30]. The heritability analysis, however, suggests that the MSL locus is relatively stable within these regions of Brazil. Thus, isolates that contain the MSL locus maintain it within the population, whereas once the MSL is lost, the locus remains absent, and is not readily reintroduced into the population. Our genome analysis of 671 Old World *L. donovani* and *L. infantum* isolates revealed the presence of MSL in all strains, indicating that the MSL locus is also relatively stable in these regions. The frequency of the MSL locus varies considerably within Brazil. In the south east (Espírito Santo and Minas Gerais), the locus is rare (4.4% are MSL⁺), whereas in northeast region (Rio Grande do Norte, Maranhão and Piauí) 74.1% of isolates are MSL⁺. We would expect that miltefosine efficacy would differ geographically, and miltefosine will be more an effective treatment of VL patients in Northeast Brazil.

In summary, a simple PCR test for the MSL allows the prediction of miltefosine treatment outcome in VL patients infected by *L. infantum*, allowing the establishment of more appropriate and personalized drug treatment for visceral leishmaniasis in Brazil.

Acknowledgements/Funding

This work was supported by: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, Brazil [grant number 478080/2009]; Fundação de Pesquisa do Estado do Espírito Santo-FAPES, Brazil [grant number 70984379/2015]; and the Global Challenges Research Fund and MRC [grant number MR/P024483]. JBTC was supported by CAPES fellowship [Proc. nº 99999.014030/2013-06]. JDD was supported by FAPESP [Proc. 14/00751-9]. The Wellcome Centre for Molecular Parasitology is supported by core funding from the Wellcome Trust [104111]. The funders had no role in study design, data collection, data analysis, interpretation, and writing of the manuscript.

Author contributions

Conceived and designed the experiments: JBTC, KC, JDD, JCM. Performed experiments: JBTC, JDD, VCS, EB. Analyzed data: JBTC, KC, SF, DCJ, NJD, JCM. Gathered the clinical sample and data: DLC, CHNC, SFGC, RD. Wrote the manuscript: JBTC, KC, SF, DCJ, JCM. All authors read and approved the final manuscript.

References

1. Alvar J, Velez ID, Bern C, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **2012**; 7(5): e35671.
2. Singh OP, Hasker E, Boelaert M, Sundar S. Elimination of visceral leishmaniasis on the Indian subcontinent. *Lancet Infect Dis* **2016**; 16(12): e304-e9.
3. Dorlo TPC, Ostyn BA, Uranw S, Dujardin JC, Boelaert M. Treatment of visceral leishmaniasis: pitfalls and stewardship. *Lancet Infect Dis* **2016**; 16(7): 777-8.
4. Martins-Melo FR, Lima Mda S, Ramos AN, Jr., Alencar CH, Heukelbach J. Mortality and case fatality due to visceral leishmaniasis in Brazil: a nationwide analysis of epidemiology, trends and spatial patterns. *PLoS One* **2014**; 9(4): e93770.
5. Romero GA, Boelaert M. Control of visceral leishmaniasis in latin america-a systematic review. *PLoS Negl Trop Dis* **2010**; 4(1): e584.
6. Ponte-Sucre A, Gamarro F, Dujardin JC, et al. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. *PLoS Negl Trop Dis* **2017**; 11(12): e0006052.
7. Romero GAS, Costa DL, Costa CHN, et al. Efficacy and safety of available treatments for visceral leishmaniasis in Brazil: A multicenter, randomized, open label trial. *PLoS Negl Trop Dis* **2017**; 11(6): e0005706.
8. Sundar S, Rosenkaimer F, Makharia MK, et al. Trial of oral miltefosine for visceral leishmaniasis. *Lancet* **1998**; 352(9143): 1821-3.
9. Jha TK, Sundar S, Thakur CP, et al. Miltefosine, an oral agent, for the treatment of Indian visceral leishmaniasis. *N Engl J Med* **1999**; 341(24): 1795-800.
10. Sundar S, Jha TK, Thakur CP, et al. Oral miltefosine for Indian visceral leishmaniasis. *N Engl J Med* **2002**; 347(22): 1739-46.
11. Perez-Victoria FJ, Sanchez-Canete MP, Castanys S, Gamarro F. Phospholipid translocation and miltefosine potency require both *L. donovani* miltefosine transporter and the new protein LdRos3 in *Leishmania* parasites. *J Biol Chem* **2006**; 281(33): 23766-75.
12. Deep DK, Singh R, Bhandari V, et al. Increased miltefosine tolerance in clinical isolates of *Leishmania donovani* is associated with reduced drug accumulation, increased infectivity and resistance to oxidative stress. *PLoS Negl Trop Dis* **2017**; 11(6): e0005641.
13. Perez-Victoria JM, Perez-Victoria FJ, Parodi-Talice A, et al. Alkyl-lysophospholipid resistance in multidrug-resistant *Leishmania tropica* and chemosensitization by a novel P-glycoprotein-like transporter modulator. *Antimicrob Agents Chemother* **2001**; 45(9): 2468-74.

14. Rakotomanga M, Saint-Pierre-Chazalet M, Loiseau PM. Alteration of fatty acid and sterol metabolism in miltefosine-resistant *Leishmania donovani* promastigotes and consequences for drug-membrane interactions. *Antimicrob Agents Chemother* **2005**; 49(7): 2677-86.
15. Rogers MB, Hilley JD, Dickens NJ, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* **2011**; 21(12): 2129-42.
16. McCormick RF, Truong SK, Mullett JE. RIG: Recalibration and interrelation of genomic sequence data with the GATK. *G3 (Bethesda)* **2015**; 5(4): 655-65.
17. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**; 1207: 3907.
18. Dumetz F, Imamura H, Sanders M, et al. 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(3).
19. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **2012**; 91(6): 1011-21.
20. Sundar S, Singh A, Rai M, et al. Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use. *Clin Infect Dis* **2012**; 55(4): 543-50.
21. Ubeda JM, Raymond F, Mukherjee A, et al. Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite *Leishmania*. *PLoS Biol* **2014**; 12(5): e1001868.
22. Lukes J, Mauricio IL, Schonian G, et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc Natl Acad Sci U S A* **2007**; 104(22): 9375-80.
23. Perez-Victoria FJ, Gamarro F, Ouellette M, Castanys S. Functional cloning of the miltefosine transporter. A novel P-type phospholipid translocase from *Leishmania* involved in drug resistance. *J Biol Chem* **2003**; 278(50): 49965-71.
24. Cojean S, Houze S, Haouchine D, et al. *Leishmania* resistance to miltefosine associated with genetic marker. *Emerg Infect Dis* **2012**; 18(4): 704-6.
25. Downing T, Imamura H, Decuypere S, et al. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res* **2011**; 21(12): 2143-56.
26. Prieto Barja P, Pescher P, Bussotti G, et al. Haplotype selection as an adaptive mechanism in the protozoan pathogen *Leishmania donovani*. *Nat Ecol Evol* **2017**; 1(12): 1961-9.
27. Ubeda JM, Legare D, Raymond F, et al. Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biol* **2008**; 9(7): R115.

28. Leprohon P, Legare D, Raymond F, et al. Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Res* **2009**; 37(5): 1387-99.
29. Imamura H, Downing T, Van den Broeck F, et al. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *Elife* **2016**; 5.
30. Bringaud F, Muller M, Cerqueira GC, et al. Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathog* **2007**; 3(9): 1291-307.

FIGURE LEGENDS

Figure 1: Flow charts of phase-two Brazilian miltefosine trial and *L. infantum* parasites recovered from enrolled patients (A), and the key steps used in the bioinformatics analysis (B). SNPs, Single Nucleotide Polymorphisms; InDels, Insertions and Deletions; CNV, Copy Number Variation; GWAS, Genome-Wide Association Study.

Figure 2: Technical validation of bioinformatics data of structural variation. (A) PCR strategy to verify the presence of miltefosine sensitivity locus (MSL) and their genes individually: *LinJ.31.2370*, *LinJ.31.2380*, *LinJ.31.2390* and *LinJ.31.2400*. (B) Results of PCR reaction presented in “A”. The * highlights the *L. infantum* isolates from cured patients that showed mixed genetic profile for the presence of MSL, and were subsequently cloned. The ^ highlights the *L. infantum* isolates used as controls for the cloning process. The presence of the complete MSL in the *L. infantum* genome is demonstrated by a PCR product of ~14kb, whilst loss is demonstrated by a ~1.2kb DNA fragment.

Figure 3: Identification of repeated sequences flanking MSL in the *Leishmania* genome and possible mechanisms for MSL loss. (A) Position of repeats sequences on chromosome 31 of *L. infantum* JPCM5, identified by sequence of novel junction formed after MSL deletion in *L. infantum* isolates, (upper map). Repetitive elements with >90% identity to the *L. infantum* RAG337 can be found in equivalent positions on chromosome 31 of *L. donovani* BPK282A1 and flank a region that is syntenic with the MSL (lower map). (B) Model for loss of the MSL: Homologous recombination between two directed-repeat sequences flanking the MSL (RAG337-1' and RAG337-2'), leading to circular extrachromosomal element.

Figure 4: Geographical distribution of MSL in *L. infantum* circulating in different regions of Brazil. ES, Espírito Santo; MG, Minas Gerais; BA, Bahia; MA, Maranhão; PI, Piauí; RN, Rio Grande do Norte.

TABLES

Table 1 - Clinical profiles of the VL patients treated with miltefosine and genome sequencing summary

Isolate ID	Location ^a	Treatment length ^b (days)	Clinical outcome	Presence of MSL ^c	Coverage fold	Mapping ^d (%)
MHOM/BR/06/MA01A	Paraibano-MA	42	Cure	+	66.2	98.89
MHOM/BR/05/MA02A	Codó-MA	42	Cure	+/-	66.8	99.06
MHOM/BR/06/MA03A	Timon-MA	42	Cure	+	91.2	98.72
MHOM/BR/06/MA04A	Codó-MA	42	Cure	-	46.9	98.85
MHOM/BR/05/MA05A	Timon-MA	42	Cure	+/-	41.3	98.98
MHOM/BR/06/MA07A	Caxias-MA	42	Relapse	-	27.1	98.94
MHOM/BR/05/MG11A	Montes Claros-MG	28	Relapse	-	75.1	98.67
MHOM/BR/05/MG12A	Montes Claros-MG	28	Cure	-	43.4	98.87
MHOM/BR/05/MG13A	São Francisco-MG	28	Relapse	-	47.5	99.15
MHOM/BR/05/MG14A	Montes Claros-MG	28	Cure	+	57.7	98.99
MHOM/BR/05/MG15A	Porteirinha-MG	28	Relapse	-	64.5	98.74
MHOM/BR/05/MG16A	São Francisco-MG	28	Relapse	-	83.5	98.99
MHOM/BR/05/MG17A	Montes Claros-MG	28	Relapse	-	65.1	98.83
MHOM/BR/05/MG18A	Montes Claros-MG	28	Relapse	-	48.6	98.54
MHOM/BR/05/MG19A	Catuni-MG	28	Cure	-	56.3	98.44
MHOM/BR/06/PI01A	José de Freitas-PI	42	Relapse	-	92.7	98.95
MHOM/BR/06/PI02A	Valença do Piauí-PI	42	Relapse	-	60.0	98.63
MHOM/BR/06/PI03A	Cabeceiras-PI	42	Relapse	-	65.3	98.57
MHOM/BR/05/PI04A	Valença do Piauí-PI	42	Cure	+/-	46.7	98.82
MHOM/BR/06/PI05A	Valença do Piauí-PI	42	Cure	+/-	41.3	99.02
MHOM/BR/06/PI07A	Piripiri-PI	42	Cure	+	47.1	99.03
MHOM/BR/05/PI08A	Altos-PI	42	Cure	+/-	46.1	99.07
MHOM/BR/05/PI09A	Nova Santa Rita-PI	42	Cure	+/-	57.8	99.01
MHOM/BR/06/PI10A	Teresina-PI	42	Cure	+/-	76.9	98.63
MHOM/BR/06/PI11A	José de Freitas-PI	42	Relapse	-	42.2	98.94
MHOM/BR/05/PI12A	Lima Campos-PI	42	Relapse	+	58.6	98.68

^a City-States in Brazil where the *L. infantum* isolates were collected: MA, Maranhão; MG, Minas Gerais; PI, Piauí. ^b Miltefosine therapy schedule. Patients received about 2.5mg/kg/day. ^c Genotyping of *L. infantum* isolates for presence of MSL⁺, homogeneous population for presence of MSL⁻, homogeneous population for absence of MSL^{+/-}, heterogeneous population for presence of MSL. ^d Mapping of sequences reads from *L. infantum* isolates to *L. infantum* JPCM5 reference genome.

Table 2 – Summary of genetic variants identified in *L. infantum* isolates from cure and relapse patients.

Variants in 26 <i>L. infantum</i> sequenced genomes		Variable CNVs between cure and relapse (12 OGs with lowest p-values)									
		Ortholog Group	Gene ID	Chr ^a	Product Description	Ref Hap ^b	Mean – Gene Dosage		Heritability	Mann Whitney P-value ^c	Perm. p-value ^d
						Cure Group	Relapse Group				
SNP sites	16,268	OG5_183927	<i>LinJ.31.2390</i>	LinJ.31	helicase-like protein	1	2.97	0.33	1.00	0.00	0.0005
Monomorphic SNPs	11010	OG5_183871	<i>LinJ.31.0050</i>	LinJ.31	MFS/sugar transport protein, putative	1	3.46	4.04	1.00	0.00	0.0013
Monomorphic InDels	5755	OG5_128720	<i>LinJ.31.2370, LinJ.31.2380</i>	LinJ.31	3'-nucleotidase/nuclease, putative 3'-nucleotidase/nuclease precursor, putative	2	5.38	0.72	1.00	0.00	0.0015
SNPs	5258	OG5_148411	<i>LinJ.14.1300</i>	LinJ.14	hypothetical protein, conserved	1	2.33	1.92	1.00	0.00	0.0033
MNPs	20	OG5_133169	<i>LinJ.34.3390</i>	LinJ.34	Complex 1 protein (LYR family), putative	1	2.39	1.05	1.00	0.01	0.0046
Others	30	OG5_145899	<i>LinJ.13.0890</i>	LinJ.13	hypothetical protein, conserved	1	2.62	1.99	1.00	0.01	0.007
Segregating variants	1,969	OG5_140412	<i>LinJ.31.3090</i>	LinJ.31	hypothetical protein, conserved	1	3.76	4.29	1.00	0.01	0.0074
Variant sites	1,969	OG5_148059	<i>LinJ.19.0630</i>	LinJ.19	histone H3 variant V	1	0.56	3.20	1.00	0.01	0.0077
InDels	413	OG5_171427	<i>LinJ.01.0840</i>	LinJ.01	potassium channel subunit-like protein	1	1.96	1.75	1.00	0.01	0.0078
SNPs	1535	OG5_148814	<i>LinJ.28.0780</i>	LinJ.28	hypothetical protein, conserved	1	0.85	2.19	1.00	0.01	0.0087
MNPs	18	OG5_148000	<i>LinJ.01.0070</i>	LinJ.01	BSD domain containing protein, putative	1	2.18	1.79	1.00	0.01	0.0088
Others	3	OG5_184157	<i>LinJ.36.4130</i>	LinJ.36	hypothetical protein, unknown function	1	2.16	1.83	1.00	0.01	0.0093

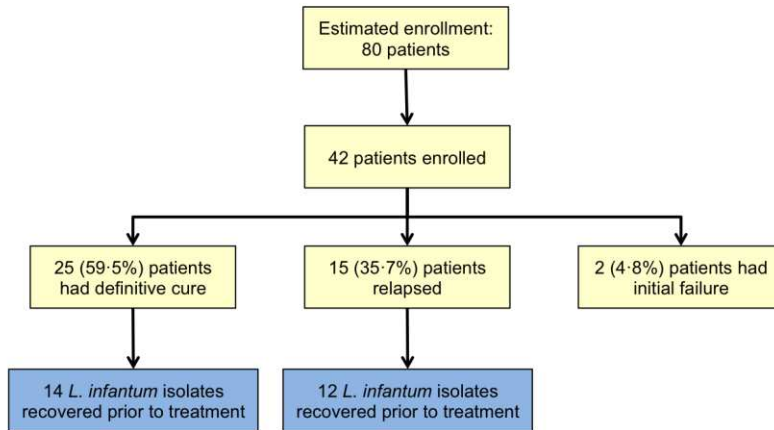
^a Chr, chromosome. ^b Ref Hap., haploid copy number in reference *L. infantum* JPCM5. ^c Mann-Whitney p, p-value of Mann-Whitney analysis. ^d Perm. p, p-value after permutation analysis.

MNPs, Multiple Nucleotide Polymorphisms. SNPs, Single Nucleotide Polymorphisms. InDels, Insertions and Deletions. Others, variants not covered in the table. Monomorphic SNPs or InDels correspond to common variants to all Brazilian *L. infantum* isolate analysed here.

Figures

Figure 1

a



b

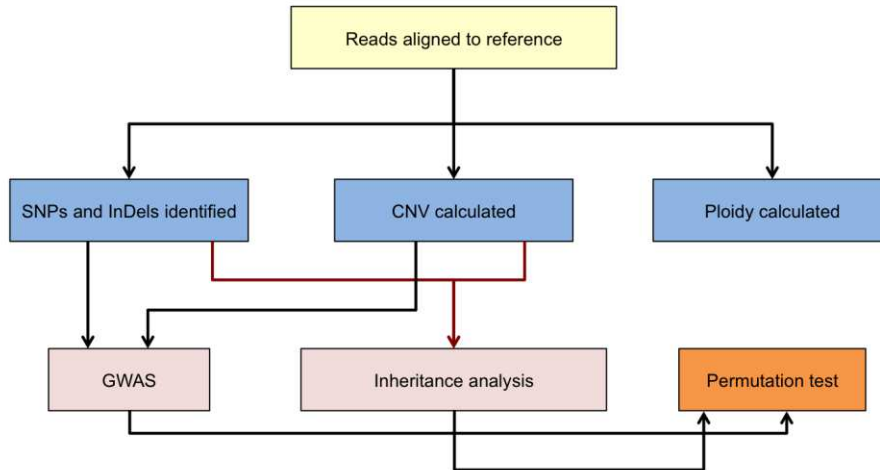


Figure 3

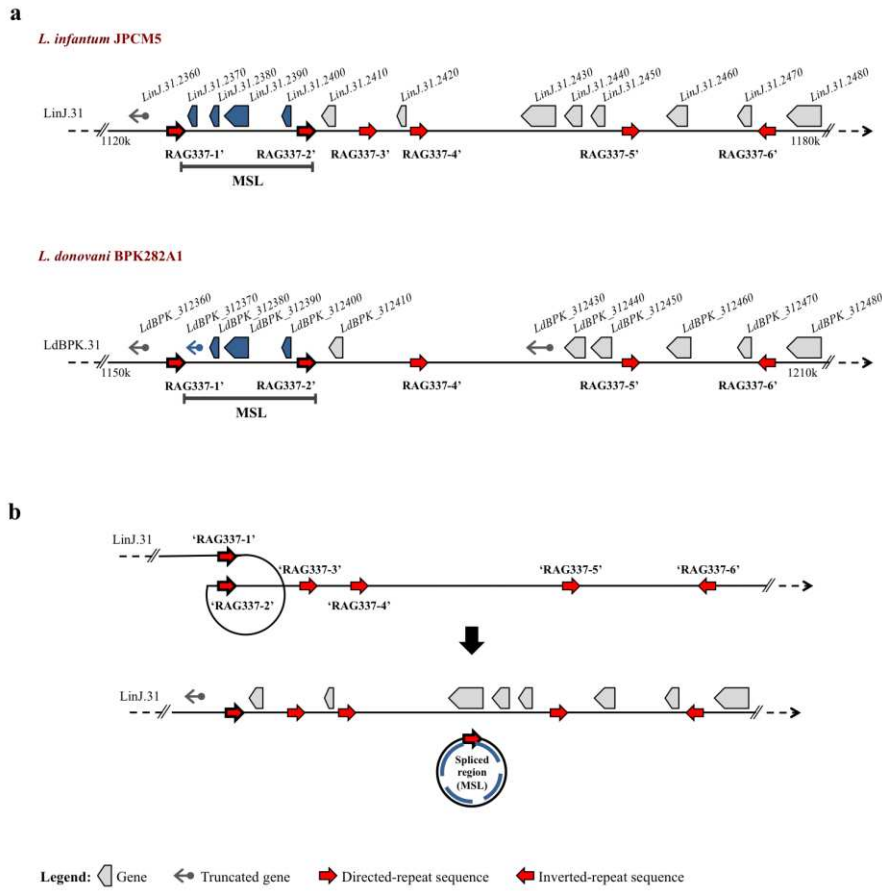
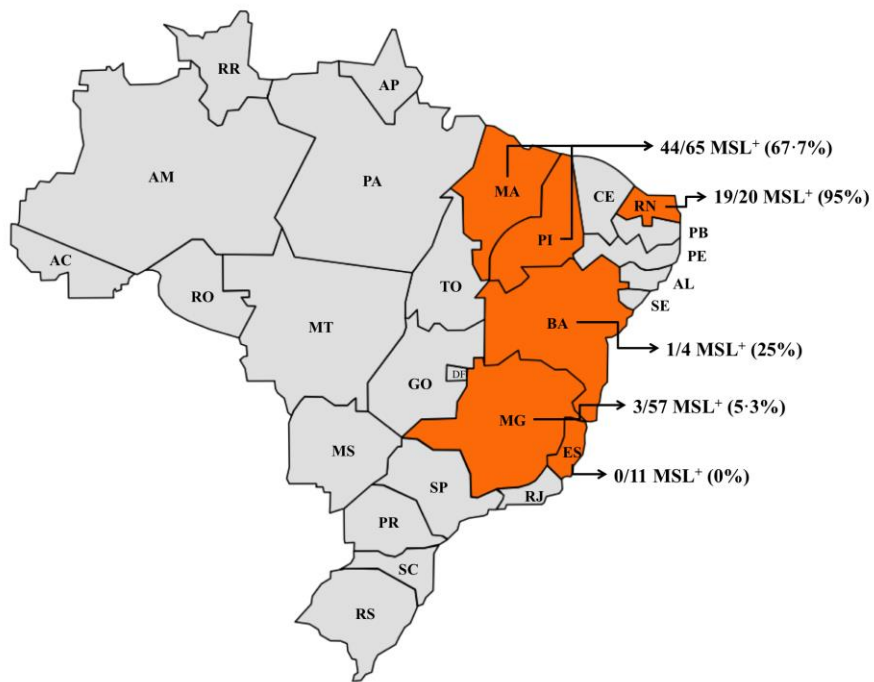


Figure 4



APPENDIX

A *Leishmania infantum* genetic marker associated with miltefosine treatment failure for visceral leishmaniasis in Brazil

Juliana B. T. Carnielli D.Sc., Kathryn Crouch PhD, Sarah Forrester PhD; Vladimir Costa Silva D.Sc., Silvio F. G. Carvalho D.Sc., Jeziel D. Damasceno D.Sc., Elaine Brown BSc, Nicholas J. Dickens PhD, Dorcas L. Costa D.Sc., Carlos H. N. Costa D.Sc., Reynaldo Dietze M.D., Daniel Jeffares PhD, Jeremy C. Mottram PhD.

CONTENTS

Supplementary Table 1: Sequence of oligonucleotides and PCR conditions.....	3
Supplementary Table 2: Summary of variants identified in each <i>L. infantum</i> isolate from patients enrolled in Brazilian miltefosine trial. SNPs were filtered for minor allele count >1 to monomorphic variants. This table forms the basis for the matrix that was used to perform the heritability analysis.	4
Supplementary Table 3: The 59 gene clusters that were highly heritable, had a coefficient of variance >0 and a permuted p-value <0.05. These clusters defined were variable between cured and relapsed patients, and had a heritability score that was 2 standard deviations above, and resulted in 757 orthologue groups. The following 59 had a multiple test corrected p-value of <0.05 following 10,000 permutations in plink. The three OGs within MSL are shown in bold.	5
Supplementary Figure 1: Aneuploidy in natural populations of <i>L. infantum</i> . The heatmap shows the copy-number status of the 36 chromosomes for the 26 pre-treatment isolates as disomic (yellow), trisomic (orange), tetrasomic (red), and pentasomic (dark red). Branches on the left represent the phylogenetic analysis carried out, using R, to cluster isolates according to aneuploidy similarity. The isolates MG12A, MG13A, MG15A, MG16A, MG18A, and PI11A exhibited the same pattern of ploidy across all chromosomes, being disomic in 35 chromosomes and tetrasomic in chromosome 31. The other 20 isolates displayed large polysomic diversity with unique patterns of aneuploidy.....	7
Supplementary Figure 2: Visual inspection of the aligned reads in a genome browser of the sequencing coverage of chromosome 31 region that contain the genes from MSL locus (<i>LinJ.31.2370</i> , <i>LinJ.31.2380</i> , <i>LinJ.31.2390</i> and <i>LinJ.31.2400</i>) of the <i>L. infantum</i> isolates obtained before the treatment. This image was generated in the Integrative Genomics Viewer software (IGV, v.2.3.40). Generally, a lack of coverage in a re-sequenced isolate must be interpreted carefully, as it can often represent a region that does not replicate or map well. Visual inspection of the aligned reads in a genome browser, however, supported the calculated predictions with the presence of reads that are split over the predicted deletion indicating that this is a genuine deletion and not a poorly sequenced region.	8
Supplementary Figure 3: The top two panels show the $-\log_{10}p$ value for the permuted p values (empirical) on the left side, and the uncorrected p values on the right side. Pink points represent variable OGs that do not contain genes within the MSL locus, blue points are the OGs that are within the MSL locus. This figure shows that the OGs within the MSL locus are both highly heritable, and highly significant. The bottom figure emphasises this and shows that while most of the variable OG groups that are significant after correction, the MSL genes are rank higher than the other OGs represented.	9
Supplementary Figure 4: Genotyping of MSL in the clones obtained from <i>L. infantum</i> isolates that exhibited homogeneous (^) and heterogeneous (*) genomic profile for MSL. The □ indicate clones that were re-cloned. The clones are identified by <i>L. infantum</i> isolate ID followed by clone ID (C1 – C16). PCR	

products were obtained by reaction that simultaneously detects MSL and its deletion, using OL4621/OL4622 primers and Long PCR Enzyme Mix (Figure 2 and Supplementary Table 1). 10

Supplementary Figure 5: Alignment of sequences obtained from the novel junction formed after MSL deletion of 21 *L. infantum* clinical isolates, the PP75 *L. infantum chagasi* reference strain, *L. infantum* JPCM5 reference and a consensus sequence from all 26 *L. infantum* isolates from Brazilian miltefosine trial. The alignment was carried out in CLC Genomics Workbench v.7. 12

SUPPLEMENTARY METHODS 13

Supplementary Table 1: Sequence of oligonucleotides and PCR conditions.

Oligonucleotides		PCR Conditions		
ID	Sequence	AT (°C)	NC	PCR reaction mixture
OL4613	5'-ATCTAGATTATAAAATCCAGTGCGATCG-3'	61	35	30 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 0.2 U of Phusion® and 1x HF PCR buffer.
OL4614	5'-TATAAGCTTCTGTGCATCACTCTTGTTAATGCG-3'			
OL4615	5'-ATCTAGACTAGAGGGCGACGTGCTCAT-3'	60	35	30 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 400 µM betain, 0.2 U of Phusion® and 1x HF PCR buffer.
OL4616	5'-TATAAGCTTATGGCTCGAGCTCGTTTCC-3'			
OL4617	5'-ATCTAGACTGCTACGCGCTCCTGTG-3'	60	35	30 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 400 µM betain, 0.2 U of Phusion® and 1x HF PCR buffer.
OL4618	5'-TATAAGCTTATGACCCTGCAGTGCGAT-3'			
OL4619	5'-ATCTAGACAGATTGCAGAATTCACGC-3'	63	35	30 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 0.2 U of Phusion® and 1x HF PCR buffer.
OL4620	5'-TATAAGCTTGCGTGGTTATATACGTGAGCG-3'			
OL4621	5'-AGTTGAGTCTGCTCCGGTG-3'	63	35	30 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 0.2 U of Phusion® and 1x HF PCR buffer. Or
OL4622	5'-TTCACGTCACGGCCAAAG-3'			10 ng of genomic DNA, 0.5 µM of Fw and Rv primers, 0.2 mM dNTPs, 2.5 U of Long PCR Enzyme Mix, 4% of DMSO and 1x PCR buffer.

AT, Annealing temperature; NC, Number of cycles; Fw, Forward; Rv, Reverse.

The conditions for initial denaturation, denaturation, annealing and extension was followed as recommended by the manufacturer of Phusion® High-Fidelity DNA Polymerase (New England BioLabs®inc.) and Long PCR Enzyme Mix (Thermo Scientific).

The genomic regions amplified by each set of oligonucleotides are shown in the figure 2.

Supplementary Table 2: Summary of variants identified in each *L. infantum* isolate from patients enrolled in Brazilian miltefosine trial. SNPs were filtered for minor allele count >1 to monomorphic variants. This table forms the basis for the matrix that was used to perform the heritability analysis.

<i>L. infantum</i> isolate ID	Number of SNPs	Number of SNPs unique to isolate	Homozygous alternative allele count	Number of InDels
MA01A	737	21	745	189
MA02A	686	0	703	188
MA03A	726	8	735	190
MA04A	737	34	745	185
MA05A	736	7	710	192
MA07A	777	47	786	184
MG11A	719	32	722	187
MG12A	720	11	688	200
MG13A	770	60	782	188
MG14A	596	78	554	215
MG15A	663	10	641	195
MG16A	702	18	718	185
MG17A	708	21	710	188
MG18A	769	27	786	191
MG19A	742	36	753	189
PI01A	701	17	692	182
PI02A	696	17	663	188
PI03A	740	22	725	178
PI04A	739	3	762	190
PI05A	718	3	676	203
PI07A	767	11	780	188
PI08A	724	2	690	202
PI09A	698	1	703	188
PI10A	699	18	718	189
PI11A	743	32	764	181
PI12A	737	20	776	189

SNPs, Single Nucleotide Polymorphisms
InDels, Insertions and Deletions

Supplementary Table 3: The 59 gene clusters that were highly heritable, had a coefficient of variance >0 and a permuted p-value <0.05. These clusters defined were variable between cured and relapsed patients, and had a heritability score that was 2 standard deviations above, and resulted in 757 orthologue groups. The following 59 had a multiple test corrected p-value of <0.05 following 10,000 permutations in plink. The three OGs within MSL are shown in bold.

Ortholog Group	Gene ID	Chr ^a	Product Description	Ref Hap ^b	Mean – Gene Dosage		Heritability	Mann Whitney P-value ^c	Permuted p-value ^d
					Cure Group	Relapse Group			
OG5_183927	<i>LinJ.31.2390</i>	LinJ.31	helicase-like protein	1	2.97	0.33	1.00	0.00	0.0005
OG5_183871	<i>LinJ.31.0050</i>	LinJ.31	MFS/sugar transport protein, putative	1	3.46	4.04	1.00	0.00	0.0013
OG5_128720	<i>LinJ.31.2370</i>, <i>LinJ.31.2380</i>	LinJ.31	3'-nucleotidase/nuclease, putative 3'-nucleotidase/nuclease precursor, putative	2	5.38	0.72	1.00	0.00	0.0015
OG5_148411	<i>LinJ.14.1300</i>	LinJ.14	hypothetical protein, conserved	1	2.33	1.92	1.00	0.00	0.0033
OG5_133169	<i>LinJ.34.3390</i>	LinJ.34	Complex 1 protein (LYR family), putative	1	2.39	1.05	1.00	0.01	0.0046
OG5_145899	<i>LinJ.13.0890</i>	LinJ.13	hypothetical protein, conserved	1	2.62	1.99	1.00	0.01	0.007
OG5_140412	<i>LinJ.31.3090</i>	LinJ.31	hypothetical protein, conserved	1	3.76	4.29	1.00	0.01	0.0074
OG5_148059	<i>LinJ.19.0630</i>	LinJ.19	histone H3 variant V	1	0.56	3.20	1.00	0.01	0.0077
OG5_171427	<i>LinJ.01.0840</i>	LinJ.01	potassium channel subunit-like protein	1	1.96	1.75	1.00	0.01	0.0078
OG5_148814	<i>LinJ.28.0780</i>	LinJ.28	hypothetical protein, conserved	1	0.85	2.19	1.00	0.01	0.0087
OG5_148000	<i>LinJ.01.0070</i>	LinJ.01	BSD domain containing protein, putative	1	2.18	1.79	1.00	0.01	0.0088
OG5_184157	<i>LinJ.36.4130</i>	LinJ.36	hypothetical protein, unknown function	1	2.16	1.83	1.00	0.01	0.0093
OG5_142238	<i>LinJ.29.2020</i>	LinJ.29	hypothetical protein, conserved	1	1.70	2.21	1.00	0.01	0.0106
---	<i>LinJ.29.0650</i>	LinJ.29	BET1-like protein, putative	1	3.54	0.48	1.00	0.01	0.0108
OG5_166727	<i>LinJ.24.2230</i>	LinJ.24	ubiquitin-conjugating enzyme, putative	1	1.62	2.32	1.00	0.01	0.0109
OG5_139387	<i>LinJ.30.0990</i>	LinJ.30	hypothetical protein, conserved	1	2.45	2.08	1.00	0.01	0.0111
OG5_183108	<i>LinJ.01.0690</i>	LinJ.01	hypothetical protein, conserved	1	1.55	2.21	1.00	0.01	0.0115
OG5_154553	<i>LinJ.34.0110</i>	LinJ.34	PrimPol-like protein 2, putative	1	1.81	2.14	1.00	0.01	0.0126
OG5_148246	<i>LinJ.04.0990</i>	LinJ.04	hypothetical protein, conserved	1	2.18	1.90	1.00	0.01	0.0129
OG5_128581	<i>LinJ.34.2100</i>	LinJ.34	clathrin coat assembly protein AP17, putative	1	0.00	1.70	1.00	0.01	0.0133
OG5_150010	<i>LinJ.15.1230</i> , <i>LinJ.15.1240</i> , <i>LinJ.15.1250</i> , <i>LinJ.15.1260</i>	LinJ.15	nucleoside transporter 1, putative	4	4.09	5.94	1.00	0.01	0.0136
OG5_173492	<i>LinJ.25.2420</i>	LinJ.25	hypothetical protein, conserved	1	1.92	2.25	1.00	0.01	0.0137
OG5_128561	<i>LinJ.35.3500</i>	LinJ.35	DNA-repair protein, putative	1	2.66	2.04	1.00	0.01	0.0145
OG5_148988	<i>LinJ.35.1750</i>	LinJ.35	hypothetical protein, conserved	1	4.06	0.96	1.00	0.01	0.0146
OG5_128477	<i>LinJ.27.0770</i>	LinJ.27	Pep3/Vps18/deep orange family/Region in Clathrin and VPS, putative	1	2.04	1.82	1.00	0.01	0.016
OG5_130243	<i>LinJ.31.2400</i>, <i>LinJ.31.2320</i>	LinJ.31	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative	2	6.51	4.27	0.97	0.02	0.0172
OG5_127434	<i>LinJ.29.2650</i>	LinJ.29	NOL1/NOP2/sun family, putative	1	2.01	1.76	1.00	0.02	0.0181
OG5_183124	<i>LinJ.03.0170</i>	LinJ.03	zinc-finger of acetyl-transferase ESCO, putative	1	2.31	1.90	1.00	0.02	0.0195
OG5_152719	<i>LinJ.32.2260</i>	LinJ.32	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain), putative	1	2.23	1.74	1.00	0.02	0.0198

Ortholog Group	Gene ID	Chr ^a	Product Description	Ref Hap ^b	Mean – Gene Dosage		Heritability	Mann Whitney P-value ^c	Permutated p-value ^d
					Cure Group	Relapse Group			
OG5_146076	<i>LinJ.35.0910</i>	LinJ.35	hypothetical protein, conserved	1	2.54	0.75	1.00	0.02	0.0208
OG5_129632	<i>LinJ.36.5180</i>	LinJ.36	hypothetical protein, conserved	1	2.02	1.80	1.00	0.02	0.0217
OG5_146086	<i>LinJ.35.3160</i>	LinJ.35	Present in the outer mitochondrial membrane proteome 7	1	2.36	2.09	1.00	0.02	0.0221
OG5_151516	<i>LinJ.17.0900</i>	LinJ.17	RNA-binding protein, putative	1	2.19	1.83	1.00	0.03	0.0225
OG5_183138	<i>LinJ.04.0150</i>	LinJ.04	hypothetical protein, conserved	1	2.34	1.98	1.00	0.02	0.0232
OG5_148094	<i>LinJ.21.1780</i>	LinJ.21	Sec8 exocyst complex component specific domain containing protein, putative	1	2.13	1.86	1.00	0.03	0.024
OG5_141729	<i>LinJ.31.1230</i>	LinJ.31	Protein of unknown function (DUF3638)/Protein of unknown function (DUF3645)/Zn-finger in Ran binding protein and others, putative	1	4.06	4.34	1.00	0.03	0.0245
OG5_143911	<i>LinJ.11.0860</i>	LinJ.11	hypothetical protein, conserved	1	2.01	2.24	1.00	0.03	0.0252
OG5_146653	<i>LinJ.23.0910</i>	LinJ.23	50S ribosome-binding GTPase, putative	1	2.68	2.26	1.00	0.03	0.0255
OG5_127854	<i>LinJ.07.1300</i>	LinJ.07	proteasome regulatory non-ATP-ase subunit, putative	1	1.80	2.19	1.00	0.03	0.0256
OG5_130133	<i>LinJ.24.2330</i>	LinJ.24	ATP:cob(D)alamin adenosyltransferase, putative	1	1.54	1.94	1.00	0.03	0.0256
OG5_172620	<i>LinJ.12.0430</i>	LinJ.12	hypothetical protein, unknown function	1	1.94	2.34	1.00	0.03	0.0257
OG5_129179	<i>LinJ.34.3300</i>	LinJ.34	peroxisome biosynthesis protein-like protein	1	1.99	1.81	1.00	0.03	0.0257
OG5_127231	<i>LinJ.21.0960</i>	LinJ.21	metallo-peptidase, Clan MG, Family M24	1	2.11	1.78	1.00	0.03	0.0265
OG5_146716	<i>LinJ.29.2550</i>	LinJ.29	3'5'-cyclic nucleotide phosphodiesterase, putative	1	2.13	1.89	1.00	0.03	0.0273
OG5_126846	<i>LinJ.36.5320</i>	LinJ.36	Casein kinase II subunit beta, putative	1	2.16	1.60	1.00	0.03	0.0289
OG5_183338	<i>LinJ.13.1230</i>	LinJ.13	hypothetical protein, conserved	1	2.42	1.96	1.00	0.03	0.0293
OG5_150308	<i>LinJ.11.1300</i>	LinJ.11	hypothetical protein, conserved	1	1.70	2.28	1.00	0.03	0.0299
OG5_154559	<i>LinJ.34.0430</i>	LinJ.34	hypothetical protein, conserved	1	2.24	1.84	1.00	0.03	0.0305
OG5_166755	<i>LinJ.31.0420</i>	LinJ.31	cysteine peptidase, Clan CA, family C2, putative	1	4.08	4.36	1.00	0.04	0.0331
OG5_127896	<i>LinJ.33.2930</i>	LinJ.33	GTP-binding protein, putative	1	2.50	2.13	1.00	0.04	0.038
OG5_151510	<i>LinJ.16.1580</i>	LinJ.16	hypothetical protein, conserved	1	1.88	2.19	1.00	0.04	0.0405
OG5_148684	<i>LinJ.24.1900</i>	LinJ.24	hypothetical protein, conserved	1	1.55	2.31	1.00	0.04	0.0411
OG5_152166	<i>LinJ.03.0500</i>	LinJ.03	hypothetical protein, conserved	1	2.00	2.19	1.00	0.05	0.0427
OG5_148941	<i>LinJ.33.1340</i>	LinJ.33	hypothetical protein, conserved	1	2.57	2.02	1.00	0.05	0.043
OG5_126828	<i>LinJ.03.0970</i>	LinJ.00	choline/ethanolamine phosphotransferase, putative	1	1.47	1.80	1.00	0.05	0.0438
OG5_127099	<i>LinJ.25.1210</i>	LinJ.25	ATP synthase subunit beta, mitochondrial, putative	1	3.03	3.75	1.00	0.05	0.0461
OG5_126880	<i>LinJ.18.0090</i>	LinJ.18	alpha glucosidase II subunit, putative	1	1.89	2.05	1.00	0.05	0.0463
OG5_145891	<i>LinJ.12.0210</i>	LinJ.12	hypothetical protein, conserved	1	2.00	2.27	1.00	0.05	0.0475
OG5_129435	<i>LinJ.35.2980</i>	LinJ.35	dolichol kinase, putative	1	2.47	2.05	1.00	0.05	0.0478

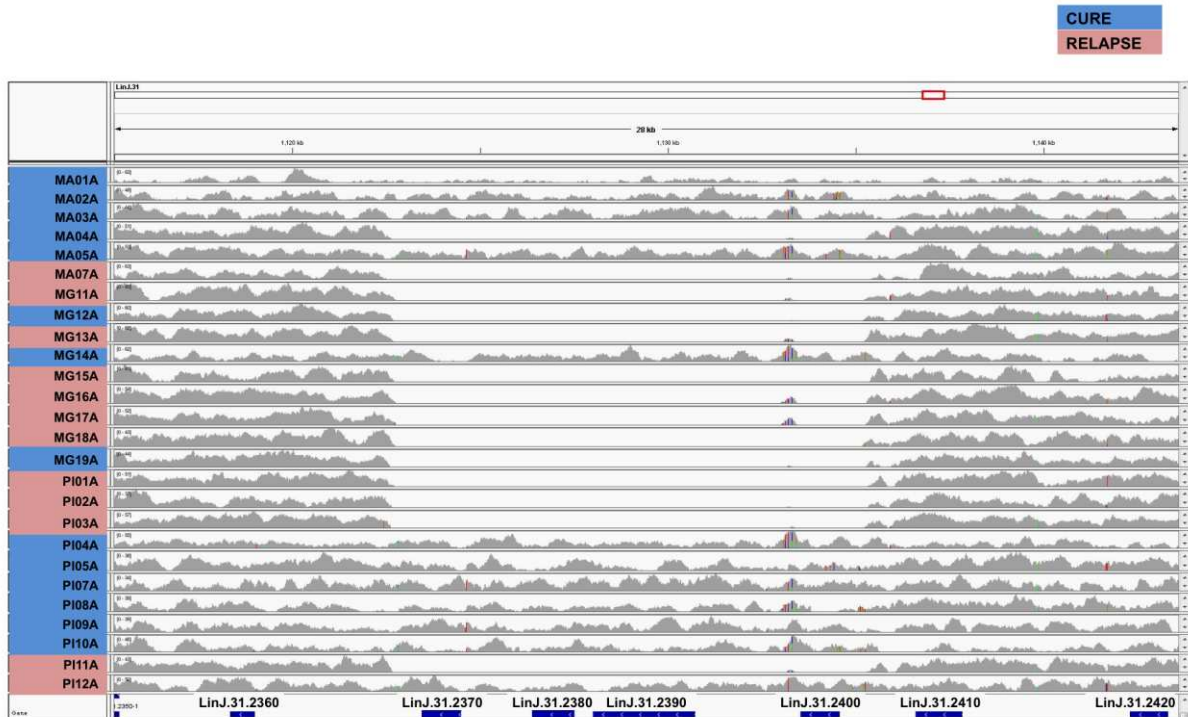
^a Chr, chromosome.

^b Ref Hap., haploid copy number in reference *L. infantum* JPCM5.

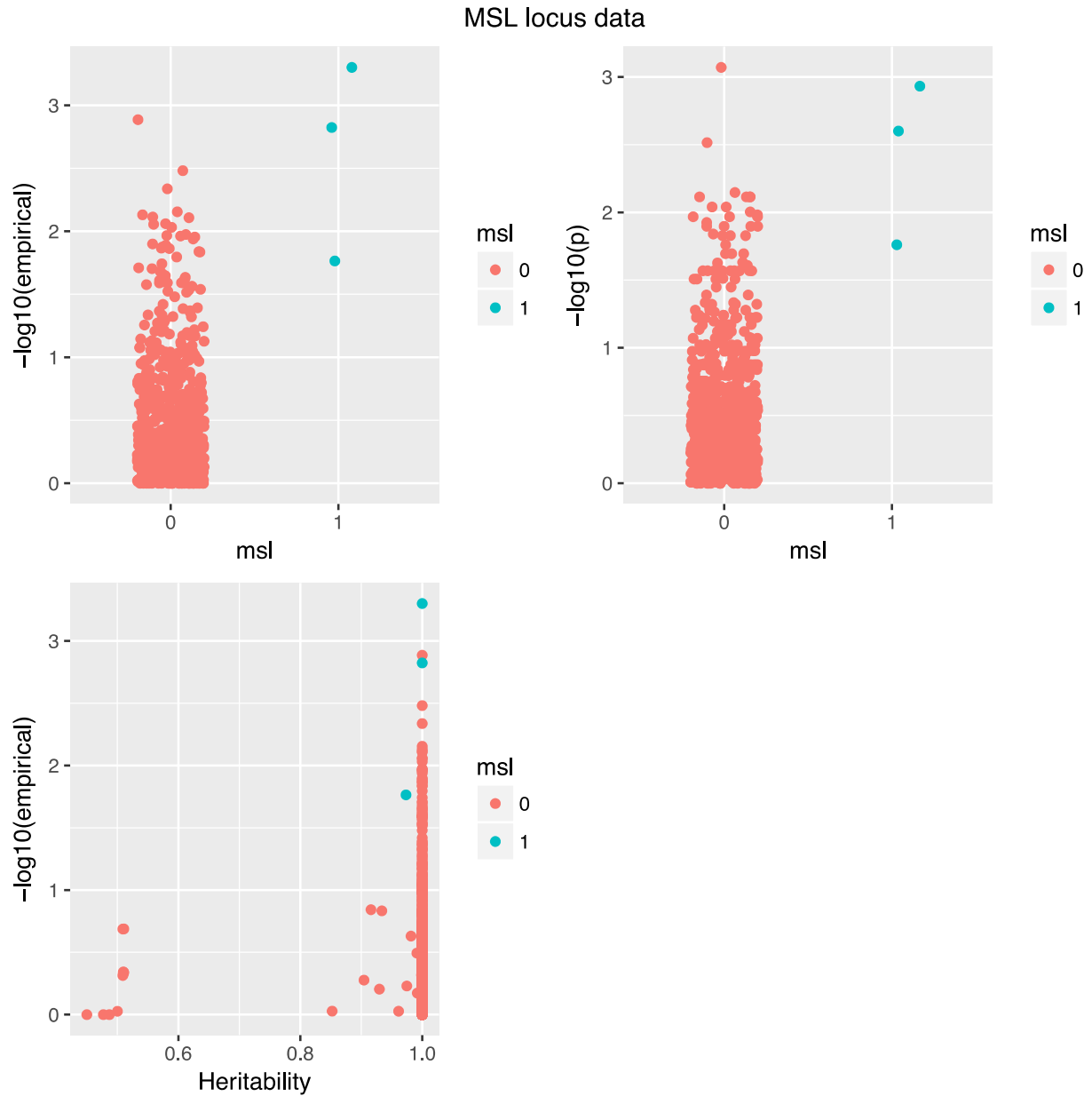
^c Mann-Whitney p, p-value of Mann-Whitney analysis

^d Perm. p, p-value after permutation analysis.

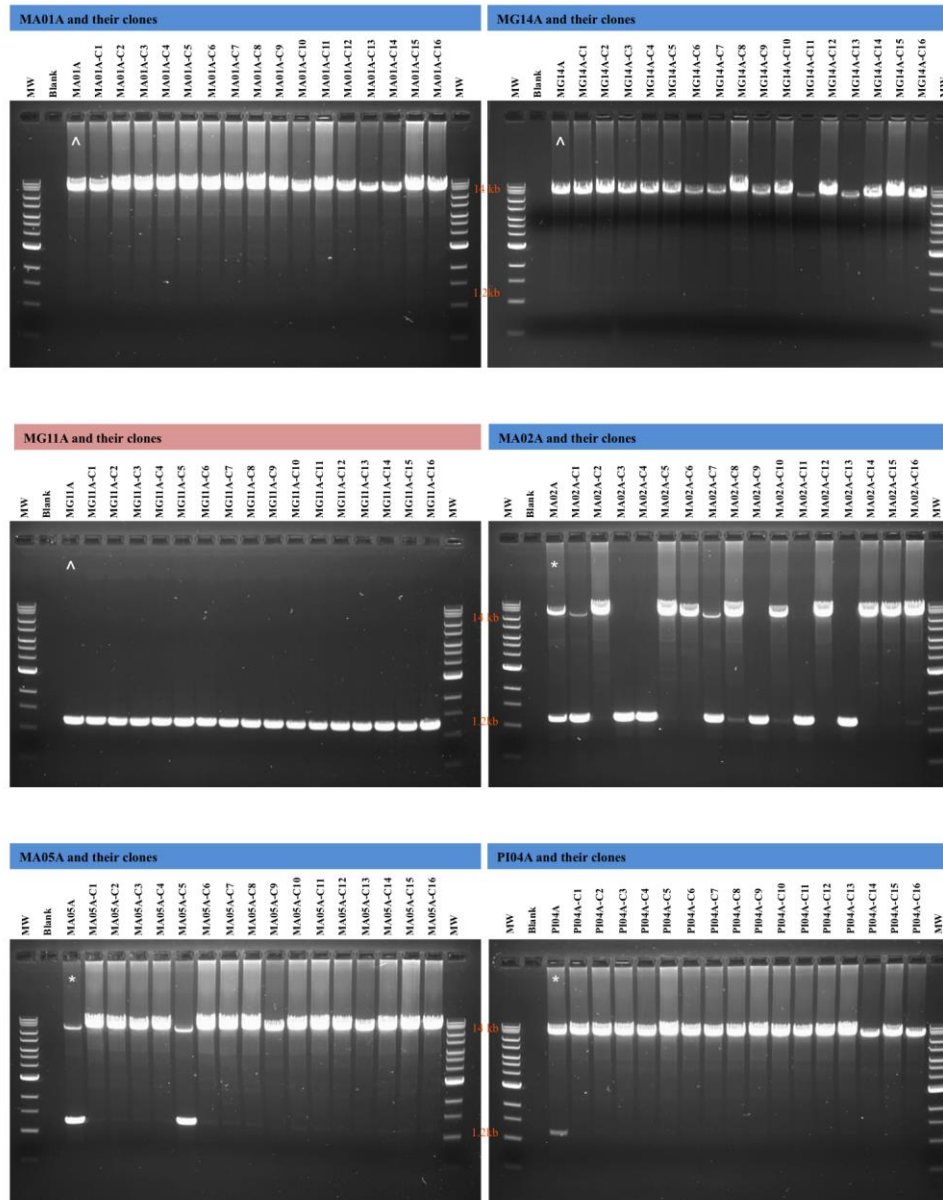
Supplementary Figure 2: Visual inspection of the aligned reads in a genome browser of the sequencing coverage of chromosome 31 region that contain the genes from MSL locus (*LinJ.31.2370*, *LinJ.31.2380*, *LinJ.31.2390* and *LinJ.31.2400*) of the *L. infantum* isolates obtained before the treatment. This image was generated in the Integrative Genomics Viewer software (IGV, v.2.3.40). Generally, a lack of coverage in a re-sequenced isolate must be interpreted carefully, as it can often represent a region that does not replicate or map well. Visual inspection of the aligned reads in a genome browser, however, supported the calculated predictions with the presence of reads that are split over the predicted deletion indicating that this is a genuine deletion and not a poorly sequenced region.



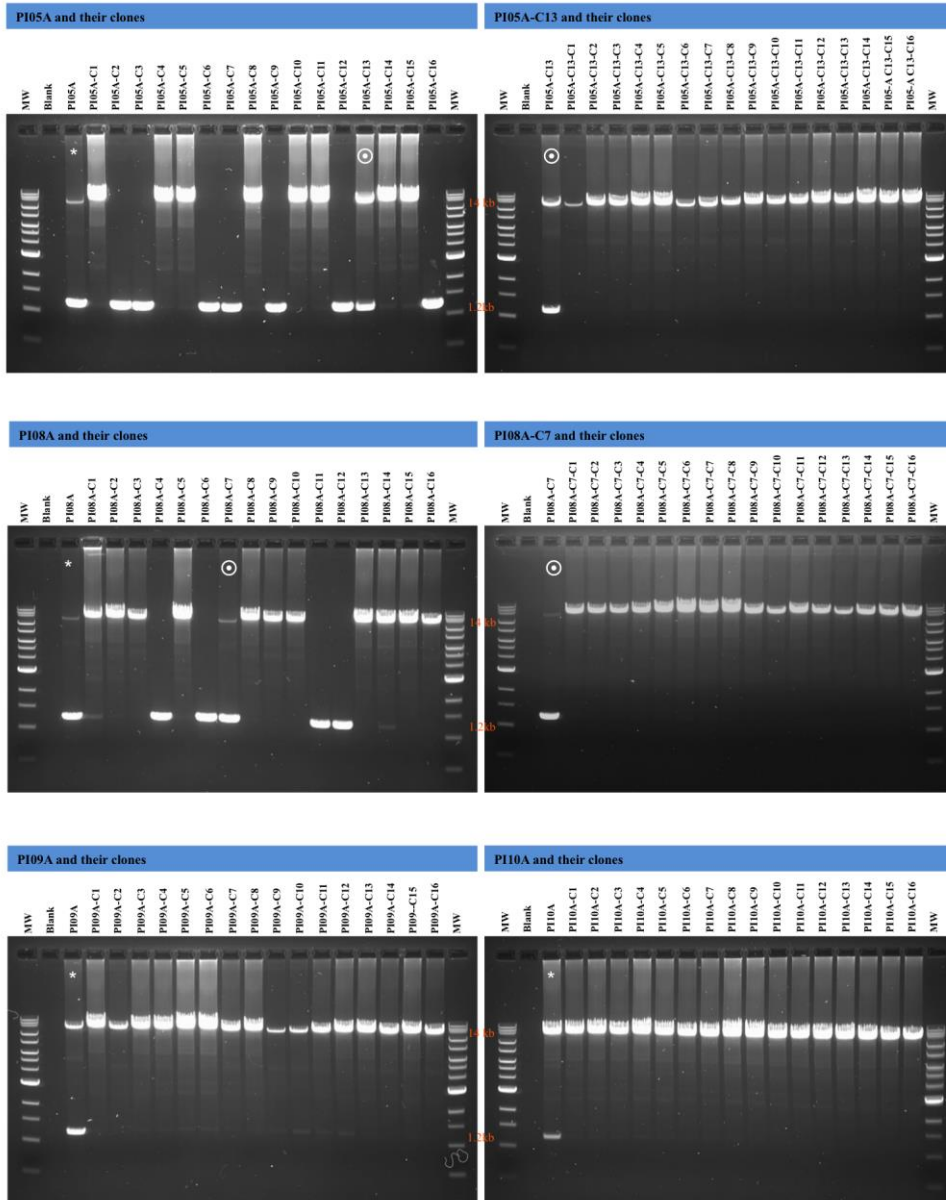
Supplementary Figure 3: The top two panels show the $-\log_{10}p$ value for the permuted p values (empirical) on the left side, and the uncorrected p values on the right side. Pink points represent variable OGs that do not contain genes within the MSL locus, blue points are the OGs that are within the MSL locus. This figure shows that the OGs within the MSL locus are both highly heritable, and highly significant. The bottom figure emphasises this and shows that while most of the variable OG groups that are significant after correction, the MSL genes are rank higher than the other OGs represented.



Supplementary Figure 4: Genotyping of MSL in the clones obtained from *L. infantum* isolates that exhibited homogeneous (^) and heterogeneous (*) genomic profile for MSL. The □ indicate clones that were recloned. The clones are identified by *L. infantum* isolate ID followed by clone ID (C1 – C16). PCR products were obtained by reaction that simultaneously detects MSL and its deletion, using OL4621/OL4622 primers and Long PCR Enzyme Mix (Figure 2 and Supplementary Table 1).



Supplementary Figure 4: continuation



SUPPLEMENTARY METHODS

Study design

The GWAS study, designed to identify genetic markers of miltefosine treatment failure, was performed with 26 pre-treatment *L. infantum* isolates (14 from cured and 12 from relapsed patients) recovered out of the 42 VL patients enrolled in the clinical trial designed to evaluate the efficacy and toxicity of miltefosine in treatment of VL in Brazil (Montes Claros, MG and Teresina, PI) in 2005-2007 (Figure 1). Geographical distribution of genetic marker highlighted by GWAS (MSL) was investigated in the 26 isolates from miltefosine trial and in more 131 *L. infantum* isolates from different regions of Brazil by PCR (isolates collected as part of VL diagnostic process in Brazil) or analysis of whole-genome parasite sequences available on Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>). The MSL frequency was also determined in *L. infantum* or *L. donovani* from old world, using 671 whole-genome parasite sequences on SRA. Finally, we investigate the mechanism by which MSL is lost from *L. infantum* genome.

Correlation between complete absence of MSL and miltefosine treatment failure was assessed by contingency table analysis (Fisher's exact test). Relative risk and Sensitivity/Specificity were estimated using the Koopman asymptotic score and Wilson-Brown tests, respectively.

Patients and parasites

The *Leishmania* isolates were obtained by bone marrow aspirates from 26 out of the 42 patients with mild disease enrolled in a clinical trial designed to evaluate the efficacy and toxicity of miltefosine in treatment of VL in Brazil (Montes Claros, MG and Teresina, PI). The patients were treated with 2.5 mg/kg/day of miltefosine for 28 (14 patients) or 42 days (28 patients), and were followed for a minimum of six months after treatment. Patients were considered cured if no signs and symptoms of the disease were present at the time of examination. Relapse was defined as a patient who was considered cured, but upon follow-up, showed reappearance of clinical signs and positive parasitology. The protocol were approved by Comissão Nacional de Ética em Pesquisa (CONEP D-18506-Z019) and are registered with ClinicalTrials.gov, number NCT00378495. Ethical clearance for utilization in research of the *L. infantum* clinical isolates obtained from patients enrolled in miltefosine Brazilian trial was obtained from the institutional review board of the Centro de Ciências da Saúde, Universidade Federal do Espírito Santo (CEP-066/2007), Brazil. The *Leishmania* isolates were collected before the treatment (Figure 1 and Table 1). These clinical isolates were identified as *L. infantum* based on a PCR-RFLP assay [1].

Other 111 *L. infantum* isolates from Núcleo de Doenças Infecciosas (Universidade Federal do Espírito Santo, Brazil) and Laboratório de Pesquisas em Leishmanioses (Universidade Federal do Piauí, Brazil) parasite banks (isolates collected as part of VL diagnostic process in Brazil), and more 20 whole-genome *L. infantum* sequences from Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>) were used to investigate the geographical distribution of MSL in *L. infantum* circulating in different regions of Brazil.

The MSL frequency was also determined in *L. infantum* or *L. donovani* from old world, using 671 whole-genome parasite sequences available on SRA or in European Nucleotide Archive (<http://www.ebi.ac.uk>).

Parasite culture

Promastigotes were grown in liver infusion tryptose (LIT) medium supplemented with 10% heat-inactivated foetal calf serum (HiFCS) pH 7.5, 25°C. The cultures were initiated by inoculating parasites into culture medium to a final concentration of 10⁶ parasites mL⁻¹. Cell number was determined microscopically using a Neubauer chamber.

Genomic DNA extraction and sequencing

The total DNA was isolated from late-log-phase promastigotes, using DNeasy® Blood & Tissue kit (Qiagen) as recommended by the manufacturer. Libraries were prepared from each DNA sample of the *L. infantum* isolates

using the Nextera DNA Library Preparation kit (Illumina) by the standard protocol. Sequencing was performed on the HiSeq system (Illumina) using paired-end reads of 125 nucleotides.

Bioinformatics analysis

Reads from *L. infantum* isolates were aligned against the resequenced *L. infantum* JPCM5 reference genome, downloaded from <http://leish-esp.cbm.uam.es> version 1. BWA version 0.7.5a-r405 was used to align the reads from clinical isolates to the reference genome, using the is indexing algorithm and mem alignment algorithm [2]. Analysis of copy number variations (CNV) at chromosome and gene level was carried out according to Rogers et al [3]. The estimated ploidy for each chromosome was calculated using median read depth for chr/(median read depth for genome/base ploidy) where base ploidy is the ploidy expected for most chromosomes (2 in *Leishmania*). Normalised coverage (fpkm) was calculated for each gene using Cufflinks version 2.2.1 [4]. The estimation of copy number for each gene was calculated using fpkm/(median read depth for chr/calculated ploidy). Lastly, it was assumed that genes on the same chromosome and with the same ortholog ID are arrays of duplicated genes. Genes were therefore clustered by chromosome and ortholog ID, and data was pooled per cluster. The script reported haploid number and the gene dose (the total number of genes in the array taking the estimated ploidy into account). The nonparametric Mann-Whitney U test was used to test for differences in gene dose in arrays of genes between relapsed and cure groups. To correct for multiple testing, empirical p-values were generated for each variant ortholog group by permutation using plink [5]. Correction for multiple tests was carried out by randomly permuting the data and re-calculating the Mann-Whitney test 10,000 times for each array, using a custom R script. An empirical p-value was calculated as the number of times the test returned a p-value the same or lower than the original test divided by the number of permutations carried out.

Alignments were realigned using the GATK local realignment tool. SNP and InDel predictions were then generated by GATK HaplotypeCaller [6] and Freebayes [7] using these realigned files and only variants identified by both were used for subsequent analysis. GATK's haplotypeCaller was used in discovery genotyping mode with `-emitRefConfidence` and `-maxReadsInRegionPerSample 40`. Bcftools was used to filter these variant calls (`QUAL>30 && MQ>30`). SNPs and InDels were also identified using freebayes with (`--min-alternate-count 5 -dont-left-align-inDels`) and filtered using a minimum quality filter of 30. The initial SNP and InDel calls were done per isolate, these variant call files were then merged and used in conjunction with the alignment files to correct the genotypes using freebayes (`--use-best-n-alleles 2 -standard-filters -genotyping-max-iterations 100 -variant-input and -haplotype-basis-alleles`). These variants were then further filtered using `'QUAL>30 && MQM>30 & MQMR>30`. SnpEff was used to annotate and predict the effects of genetic variants, and SNPSIFT was used to extract variants that resulted in coding changes [8].

Due to the significant gene dose differences between relapse and cured patients, we explored whether these copy number variants were heritable (that is, segregated consistently with SNPs, or changed rapidly independently of SNPs). For this analysis, we used SNPs called from the 26 pre-treatment *L. infantum* isolates, removing SNPs that were present in all Brazilian isolates. To estimate heritability, a kinship matrix was constructed from these SNPs using Linkage Disequilibrium Adjusted Kinships (LDAK) version 5.0 [9]. OGs were classified as variable if the ratio of the standard deviation of the gene dosage to mean was greater than 0. Heritability scores were then calculated by scoring each *L. infantum* isolate against the 7,822 variable OG dosages, which were treated as phenotypes, and the genotypes were derived from the SNP kinship matrix. The level of heritability of each of the 7,822 gene clusters and the cure/relapse phenotype was then estimated using a restricted maximum likelihood method (REML) implemented in LDAK. From these 7,822 variable OGs, only 59 had a multiple corrected p value of <0.05 , when performed as mentioned above. The heritability scores for these OGs are shown in supplementary table 6. GWAS was then performed using this kinship matrix to control for unequal relatedness of strains. Traits were permuted 1,000 times to determine a genome-wide significance threshold.

Technical validation of the miltefosine treatment failure marker (MSL) from NGS data

For validation of NGS data, PCR amplification of the MSL in chromosome 31 was accomplished according to PCR strategy showed in Figure 2A and Supplementary Table 1 for all 26 *L. infantum* isolates from Brazilian miltefosine trial. Mainly: (i) with two sets of primers using the Phusion[®] High-Fidelity DNA Polymerase (New England

BioLabs®inc.). OL4613/OL4614, or OL4615/OL4616, or OL4617/OL4618, or OL4618/OL4619 amplified the gene *LinJ.31.2370*, *LinJ.31.2380*, *LinJ.31.2390*, and *LinJ.31.2400* of MSL, respectively, whereas OL4621/OL4622 amplified the novel junction formed after the deletion of MSL; (ii) or with the set of primers OL4621/OL4622 using the Long PCR Enzyme Mix (Thermo Scientific), that simultaneously amplified the MSL and/or new junction formed after MSL deletion. The total reaction mixture was made up to 25 µL by addition of the genomic DNA, extracted as described above.

Homogeneity of *L. infantum* clinical isolates

All *L. infantum* isolates from Brazilian miltefosine trial that showed both the presence of the MSL and the novel junction after MSL deletion (n=7: MA02A, MA05A, PI04A, PI05A, PI08A, PI09A, and PI10A) were cloned to evaluate their homogeneity. Three other *L. infantum* isolates (MA01A, MG11A, and MG14A) were used as control of cloning process. Parasites from early passages (maximum passage 3) were plated on SDM-79 agar supplemented with 5 µg.mL⁻¹ of hemin, 10 µM of 6-biopterin, and 10% of HiFCS for 6-12 days. Sixteen single colonies from each isolate were picked from the plates and independently subcultured. Genomic DNA of each clone was extracted and screened for the detection of MSL and the novel junction originating from MSL deletion, by PCR amplification as described above.

Investigation of the mechanism of MSL deletion

The natural mechanism of MSL deletion was investigated using all 21 *L. infantum* (MA02A, MA04A, MA05A, MA07A, MG11A, MG12A, MG13A, MG15A, MG16A, MG17A, MG18A, MG19A, PI01A, PI02A, PI03A, PI04A, PI05A, PI08A, PI09A, PI10A, and PI11A) isolates from Brazilian miltefosine trial that presented deletion of MSL.

To investigate the mechanism of MSL deletion, the PCR-amplified products, corresponding to novel junction formed after the MSL deletion (from the set of primers OL4621 and OL4622), were subcloned into the pGEM-T easy vector (Promega) for sequencing. All sequences obtained plus correspondent sequences of the *L. infantum* JPCM5 and of the consensus sequence (from all *L. infantum* isolates) were aligned by CLC Genomics Workbench (version 7.5.1). Repeat sequences flanking the MSL in chromosome 31 were located in the genome sequence of *L. infantum* JPCM5 (v5) using Blastn on TriTrypDB, based: in the sequence formed by the novel junction after MSL deletion; and in repeated sequences reported by Ubeda et al [10]. Blast hits were filtered for identities and lengths higher than 85% and 200 nucleotides, respectively.

References

1. Segatto M, Ribeiro LS, Costa DL, et al. Genetic diversity of *Leishmania infantum* field populations from Brazil. *Mem Inst Oswaldo Cruz* **2012**; 107(1): 39-47.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**; 25(14): 1754-60.
3. Rogers MB, Hilley JD, Dickens NJ, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* **2011**; 21(12): 2129-42.
4. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **2010**; 28(5): 511-5.

5. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **2007**; 81(3): 559-75.
6. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **2010**; 20(9): 1297-303.
7. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* **2012**; 1207: 3907.
8. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **2012**; 6(2): 80-92.
9. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **2012**; 91(6): 1011-21.
10. Ubeda JM, Raymond F, Mukherjee A, et al. Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite *Leishmania*. *PLoS Biol* **2014**; 12(5): e1001868.