Parasite Genotype Is a Major Predictor of Mortality from Visceral Leishmaniasis

Cooper Alastair Grace,[a] Kátia Silene Sousa Carvalho,[b] Mayara Ingrid Sousa Lima,[a,c] Vladimir Costa Silva,[b] João Luís Reis-Cunha,[a] Matthew J. Brune,[a] Sarah Forrester,[a] Conceição de Maria Pedrozo e Silva de Azevedo,[d] Dorcas Lamounier Costa,[b,e,f] Doug Speed,[g] Jeremy C. Mottram,[a] Daniel C. Jeffares,[a] Carlos H. N. Costa[b,e]

aYork Biomedical Research Institute, Department of Biology, University of York, York, United Kingdom
bDepartment of Community Medicine and Institute of Tropical Diseases Natan Portela, Federal University of Piauí, Teresina, Brazil
cDepartment of Biology, Postgraduate Programs in Health Sciences and Postgraduate Program in Health and Environment, Federal University of Maranhão, São Luís, Maranhão, Brazil
dDepartment of Medicine and Postgraduate Program in Health Sciences, Federal University of Maranhão, São Luís, Maranhão, Brazil
eIntelligence Center for Emerging and Neglected Diseases, Teresina, PI, Brazil
fMother Child Department, Federal University of Piauí, Teresina, PI, Brazil
gCentre for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark

Cooper Alastair Grace, Kátia Silene Sousa Carvalho, Mayara Ingrid Sousa Lima, and Vladimir Costa Silva contributed equally to this work. Author order was determined in order of their collective contribution to data analysis, manuscript writing and laboratory work.

ABSTRACT Visceral leishmaniasis (VL) is a potentially fatal disease caused mainly by *Leishmania infantum* in South America and *Leishmania donovani* in Asia and Africa. Disease outcomes have been associated with patient genotype, nutrition, age, sex, comorbidities, and coinfections. In this study, we examine the effects of parasite genetic variation on VL disease severity in Brazil. We collected and sequenced the genomes of 109 *L. infantum* isolates from patients in northeastern Brazil and retrieved matching patient clinical data from medical records, including mortality, sex, HIV coinfection, and laboratory data (creatinine, hemoglobin, and leukocyte and platelet counts). We identified genetic differences between parasite isolates, including single nucleotide polymorphisms (SNPs), small insertions/deletions (indels), and variations in genic, intergenic, and chromosome copy numbers (copy number variants [CNVs]). To describe associations between the parasite genotypes and clinical outcomes, we applied quantitative genetics methods of heritability and genome-wide association studies (GWAS), treating clinical outcomes as traits that may be influenced by parasite genotype. Multiple aspects of the genetic analysis indicate that parasite genotype affects clinical outcomes. We estimate that parasite genotype explains 83% chance of mortality (narrow-sense heritability [h^2] = 0.83 ± 0.17) and has a significant relationship with patient sex (h^2 = 0.60 ± 0.27). Impacts of parasite genotype on other clinical traits are lower ($h^2 \leq$ 0.34). GWAS analysis identified multiple parasite genetic loci that were significantly associated with clinical outcomes; 17 CNVs were significantly associated with mortality, two with creatinine, and one with bacterial coinfection, jaundice, and HIV coinfection, and two SNPs/indels and six CNVs were associated with age, jaundice, HIV and bacterial coinfections, creatinine, and/or bleeding sites. Parasite genotype is an important factor in VL disease severity in Brazil. Our analysis indicates that specific genetic differences between parasites act as virulence factors, enhancing risks of severe disease and mortality. More detailed understanding of these virulence factors could be exploited for novel therapies.

IMPORTANCE Multiple factors contribute to the risk of mortality from visceral leishmaniasis (VL), including, patient genotype, comorbidities, and nutrition. Many of these factors are influenced by socioeconomic biases. Our work suggests that the virulence of the infecting parasite is an important risk factor for mortality. We pinpoint some specific genomic markers that are associated with mortality, which can

lead to a greater understanding of the molecular mechanisms that cause severe VL disease, to the identification of genetic markers for virulent parasites, and to the development of drug and vaccine therapies.

Leishmaniasis is a neglected tropical disease caused by protozoan parasites of the genus *Leishmania*. Visceral leishmaniasis (VL), the most severe form, is caused by *Leishmania infantum* in South America, the Mediterranean basin, the Middle East, and Central Asia and by *Leishmania donovani* in East Africa and the Indian subcontinent. VL is characterized by fever, weight loss, hepatosplenomegaly, anemia, pancytopenia, hypoalbuminemia, and hypergammaglobulinemia and is frequently fatal if untreated (1, 2). Severe disease presents as hepatic dysfunction with jaundice, edema, and dyspnea (3). Death is usually associated with hemorrhagic phenomena, which could be caused by bacterial coinfection (4, 5) as a consequence of cytokine-driven systemic inflammation (6). *Leishmania*/HIV results in greater risks of relapse and fatality (7, 8). Despite the range of clinical presentations, the parasite factors that determine differential virulence in humans are not well understood.

There are approximately 5,000 VL cases/year in Brazil, 90% of the cases registered in the Americas (9–12). There are large numbers of asymptomatic *L. infantum* infections, with various studies identifying 1 to 6% of the general population in regions of endemicity (7, 13, 14). VL case fatality rates in Brazil average 7%, the highest in the world, and show an increasing trend (12). Hence, understanding the factors that lead to severe disease and mortality is a priority for leishmaniasis research. Disease outcomes are known to be associated with patient age, sex, comorbidities (4), adverse effects of treatment, and coinfections such as HIV (15). Genome-wide association studies (GWAS) have shown to human genotypes influence VL disease severity in East Africa, Brazil, and the Indian subcontinent (16), including a strong influence from the HLA-DR-DQ region within the major histocompatibility complex (17). However, apart from studies of drug resistance (10, 18), the influence of natural parasite genetic variation on disease severity has not been investigated using genome-scale methods.

To investigate this, we collected 109 *L. infantum* isolates from the states of Piauí and Maranhão in Brazil, which have large VL burdens (19, 20). We used genome-scale quantitative genetic analyses to investigate the effects of parasite genotype on multiple clinical indicators of disease severity and mortality. Our analysis showed that parasite genotype influences multiple aspects of disease severity and has a strong relationship with patient mortality. We identified multiple parasite genetic loci that affect mortality. These virulence factors are key points for vaccine and drug discovery (21) that will be complementary to reverse vaccinology methods (22).

RESULTS

Sampling and clinical traits. We collected *L. infantum* isolates for genomic sequencing and clinical data from 109 patients with VL in Brazil. The isolates were collected from patients in the states Piauí and Maranhão, where VL is highly endemic (Fig. 1a). Each parasite isolate was matched to the source patient with information on age; sex; the binary clinical factors mortality (recovered/died), bleeding, presence of edema, HIV status, jaundice, vomiting, dyspnea, and bacterial coinfection; laboratory clinical metrics (leukocyte and platelet counts and hemoglobin and creatinine levels); and estimated probabilities of mortality from clinical and clinical+laboratory models (Kala-Cal* and Kala-Cal** respectively) (5). All these factors varied among patients (see Fig. S13 at https://figshare.com/projects/Brazil-VL-GWAS/126331; see Tables S2 and S3 in the supplemental material), and mortality was well predicted by the Kala-Cal* model (see Fig. S14 at the address above). In this cohort, no trait, except age, was associated with HIV status (see Fig. S15 at the address above). The only other between-trait correlations were that leukocyte and creatinine counts declined with patient age (see Fig. S16 at the address above).
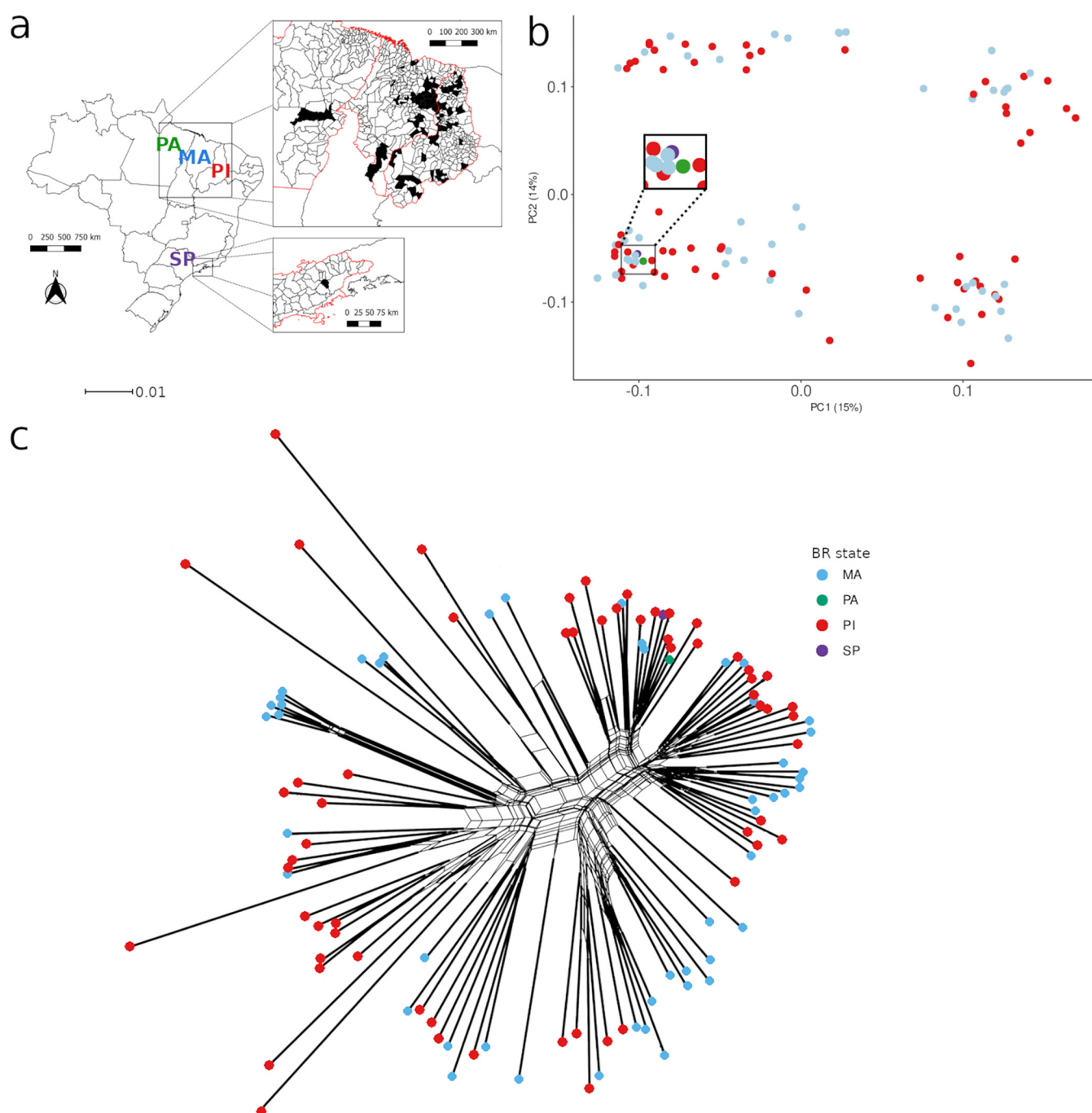
FIG 1 Sample origins and population structure. (a) Map showing residential locations of VL patients who provided the *L. infantum* strains. MA, Maranhão; PI, Piauí; PA, Pará; SP, São Paulo. Pará and São Paulo yielded a single isolate each. (b) Principal-component coordinates, derived from SNP variation, show that genetic relatedness is not clustered by state. Each filled circle indicates the principal-component-derived relatedness between isolates. Circles are colored by their state of origin. (c) Phylogenetic network analysis of all 109 strains, with colored circles indicating state of origin using the color scheme in panel b. Branches with thick lines have placement with 100% bootstrap support. The scale bar indicates the number of substitutions per site. The network file and version with isolate names are available at https://figshare.com/projects/Brazil-VL-GWAS/126331.

Genomic variation. To characterize genomic variation in these 109 *L. infantum* isolates, we identified single nucleotide polymorphism (SNP) and insertion/deletion (indel) variants, gene copy number variations (GCNVs), intergenic copy number variations (IGCNVs), and chromosome copy number variations (CCNVs) (Table 1). In accordance with other analyses (19, 23, 24), *L. infantum* populations contain 10-fold less genetic diversity than *L. donovani* populations in East Africa (nucleotide diversity [π] in Brazil = 4.3×10^{-5}; Africa $\pi = 42 \times 10^{-5}$) and half as much diversity as the Indian

TABLE 1 *L. infantum* genomic variability of 109 isolates

Genomic variation	Count
SNVs[a]	14,203
SNPs	4,468
Indels	9,735
Variant sites with MAF >5%	3,526
Biallelic variants	3,004
Genes	8,748
Genes with CV of >1% and coverage at >20% of genome coverage	8,218
Intergenic regions	8777
Intergenic regions with CV of >1% and coverage at >20% of genome coverage	8,219

[a]SNVs, small nucleotide variants; MAF, minor allele frequency; CV, coefficient of variation.

subcontinent (India $\pi = 8 \times 10^{-5}$). Analysis of between-strain relatedness with principal-component analysis (PCA) and phylogenetics showed very little population structuring, with no clear distinction between isolates based on state of origin (Fig. 1b and c). The majority of chromosomes were disomic (see Fig. S2 and S3 at https://figshare.com/projects/Brazil-VL-GWAS/126331), whereas a high number of chromosomal duplications and GCNVs were observed in several isolates, as expected for *Leishmania* and other eukaryotic genomes (10, 25).

Effects of genetic variation in *Leishmania infantum* on clinical factors. To quantify the effect of parasite genetic variation, we estimated the narrow-sense (additive) heritability, using clinical factors as traits (Table S2) and parasite genetic variants as genotypes (see Materials and Methods). This analysis estimated the proportion of variation in patient clinical factors that is attributed to parasite genetic variation. Possible kinship between parasites was inferred by LDAK using genetic variants (see Materials and Methods). We utilized four categories of genetic variants in this analysis: (i) SNP/indel variants, (ii) CCNV, (iii) GCNV, and (iv) IGCNV. Heritability estimation was performed using each genotype category independently, followed by a composite model that estimates the combined contribution of all variants. As we sought to examine the effects of parasite variation, host genotypes were not considered in this analysis.

Our results show that parasite genotype has significant effects on multiple clinical factors (Fig. 2). Due to the small sample size ($n = 109$), all estimates have relatively large standard deviations (Table S4). Mortality was strongly affected by parasite genetic variation, with a total heritability of 0.83 (standard deviation, 0.17; 95% confidence interval, 0.66 to 1.00), impacted by SNPs/indels and GCNVs. Sex was also strongly associated with SNP/indel variation (0.60 ± 0.27; 95% confidence interval, 0.33 to 0.87), indicating that parasite genotype influences the susceptibility of males and females to VL differentially. Strain-specific differences that affect male and female disease severity may be the cause of the higher occurrence of infections in males (15). The majority of the other traits were affected primarily by SNP/indel variations, rather than CNVs. Approximately a third of clinical traits ($n = 5$) are affected by independent heritability components, while the majority of traits ($n = 11$) are complex, with multiple types of genotypic variation contributing to host phenotype. This analysis shows that even within Brazil, where *L. infantum* genetic diversity is relatively low, differences between parasites influence disease progression.

The effect of parasite genotype on mortality could be due to one lineage of particularly virulent parasites. To assess this possibility, we visualized this clinical outcome on a phylogenetic tree of isolates (see Fig. S4 at https://figshare.com/projects/Brazil-VL-GWAS/126331). Isolates from patients who died are dispersed throughout the phylogeny, indicating that no specific lineage is responsible for mortality. This is expected, if multiple loci affecting mortality were segregating within the population and these loci are unlinked by recombination between strains (see Fig. S1 at https://figshare.com/projects/Brazil-VL-GWAS/126331). In this scenario, no single phylogeny will cluster these unlinked variants.
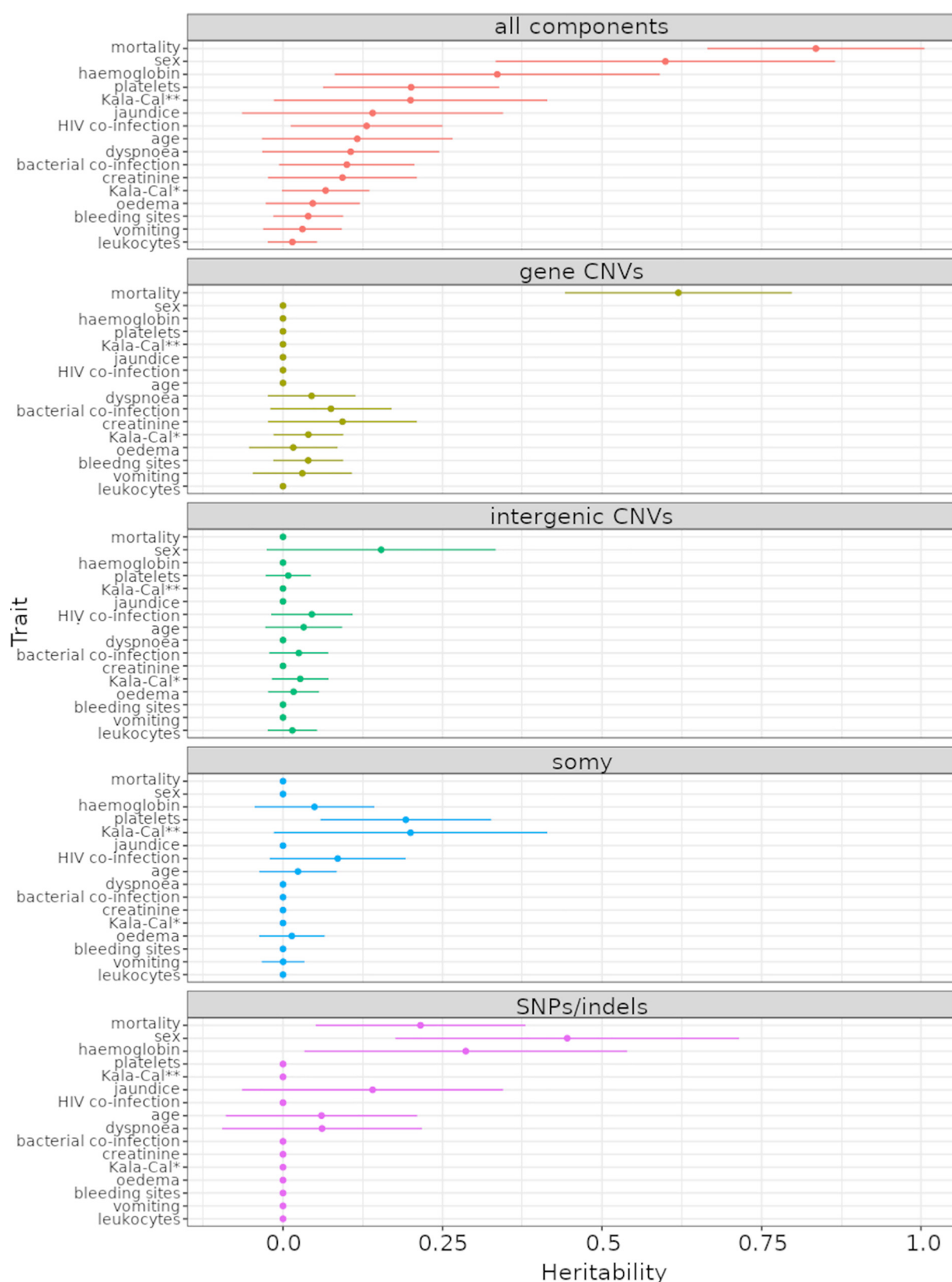
FIG 2 Heritability estimates from the composite model. Heritability estimates used normalized trait values. Horizontal lines represent standard deviations for each heritability estimate. Here, most analyses used 109 samples. (Top) Estimate of total heritability from all types of genetic variants combined. Panels below this show the heritability contributed by the various types of genetic variants. Due to linkage between different categories of variants (SNPs, gene copy variants, etc.), the total heritability explained is not expected to be the sum of all variant types.

TABLE 2 Genomic variations associated with clinical traits

GWAS component	No. of significant associations	Trait(s)
SNPs/indels	2	Age; jaundice
GCNVs	11	Mortality; creatinine; jaundice
IGCNVs	12	Bacterial coinfection; creatinine; mortality; HIV coinfection
CCNVs	0	

To identify the genetic loci that affect clinical outcomes, we conducted GWAS, using all four categories of genetic variants as predictors of clinical traits. In all analyses apart from chromosome copies (CCNVs), association P values were inflated above expected values, indicating that multiple genetic loci affect clinical outcomes (QQ plots, Fig. S5 and S12 at https://figshare.com/projects/Brazil-VL-GWAS/126331). We observed no association between CCNV and any of the evaluated traits, which is in accordance with chromosome copy number variation (somy) being associated only with short-term environmental adaptation (26) that may vary after culturing (27). Given the small sample size available ($n = 109$), we expect to discover only cases where genotypes have strong effects on clinical outcomes. Nevertheless, we found multiple parasite genetic variants that were associated with clinical outcomes above our permutation-derived thresholds (Table 2; Tables S5, S6, and S8).

Two small variants were significantly associated with clinical factors. An indel variant was significantly associated with the presence of jaundice, and a SNP variant was associated with patient age ($P = 5.2 \times 10^{-7}$ and $P = 2.6 \times 10^{-6}$) (Table S5; also, see Fig. S5 at https://figshare.com/projects/Brazil-VL-GWAS/126331). Both of these variants were located within intergenic regions (see supplemental text B at the address above). Gene copy variants were significantly associated with three traits: mortality (Fig. 3), creatinine levels, and jaundice (see Fig. S7 at the address above). Variations in intergenic DNA copies were associated with four clinical factors: bacterial coinfection, creatinine, mortality and HIV coinfection (see Fig. S8, S9, and S21 to S23 in supplemental text C at the address above). The association scores, chromosomal positions, and annotation of the 11 significant genes and 12 intergenic regions are displayed in Tables S6 and S8, respectively. As expected, given the evidence for recombination in these isolates, we found that some strains contain multiple virulence-associated genotypes and others contain only a few (Table S7), which may result in a range of virulence levels.

The progression of disease is likely to be influenced by interactions between patient factors, comorbidities, and parasite genotype. For example, it is possible that particular parasite genotypes are more likely to infect HIV-positive patients that are immunosuppressed. To examine this possibility, we conducted analysis of gene and intergenic copy variants that affect mortality, using all other clinical traits as covariates. This analysis frequently produced lower (more significant) association P values (Tables S9 and S10). The most significant association using covariates was between the gene LINF_340051500 (phosphatidylinositol 3-kinase; *tor2*) copy variant and mortality ($P = 6.24 \times 10^{-9}$), using Kala-Cal** risk prediction as a covariate. Since the Kala-Cal** mortality prediction uses many clinical characteristics, it likely captures many factors of the relationship between clinical factors, parasite, and mortality. The gene LINF_270006900 (hypothetical protein) was still significantly associated with mortality when each other trait was individually used as a covariate and when all the other traits excluding Kala-Cal* and Kala-Cal** were simultaneously used as covariates (Table S9). While these results do not define the precise relationships between parasite, host, and factors, they show that genetic association studies have considerable promise.

DISCUSSION

We used genome data from 109 *L. infantum* isolates from northeastern Brazil coupled with detailed patient clinical data to study the effects of parasite variation on VL disease. We found that the chance of mortality and patient sex were strongly associated with
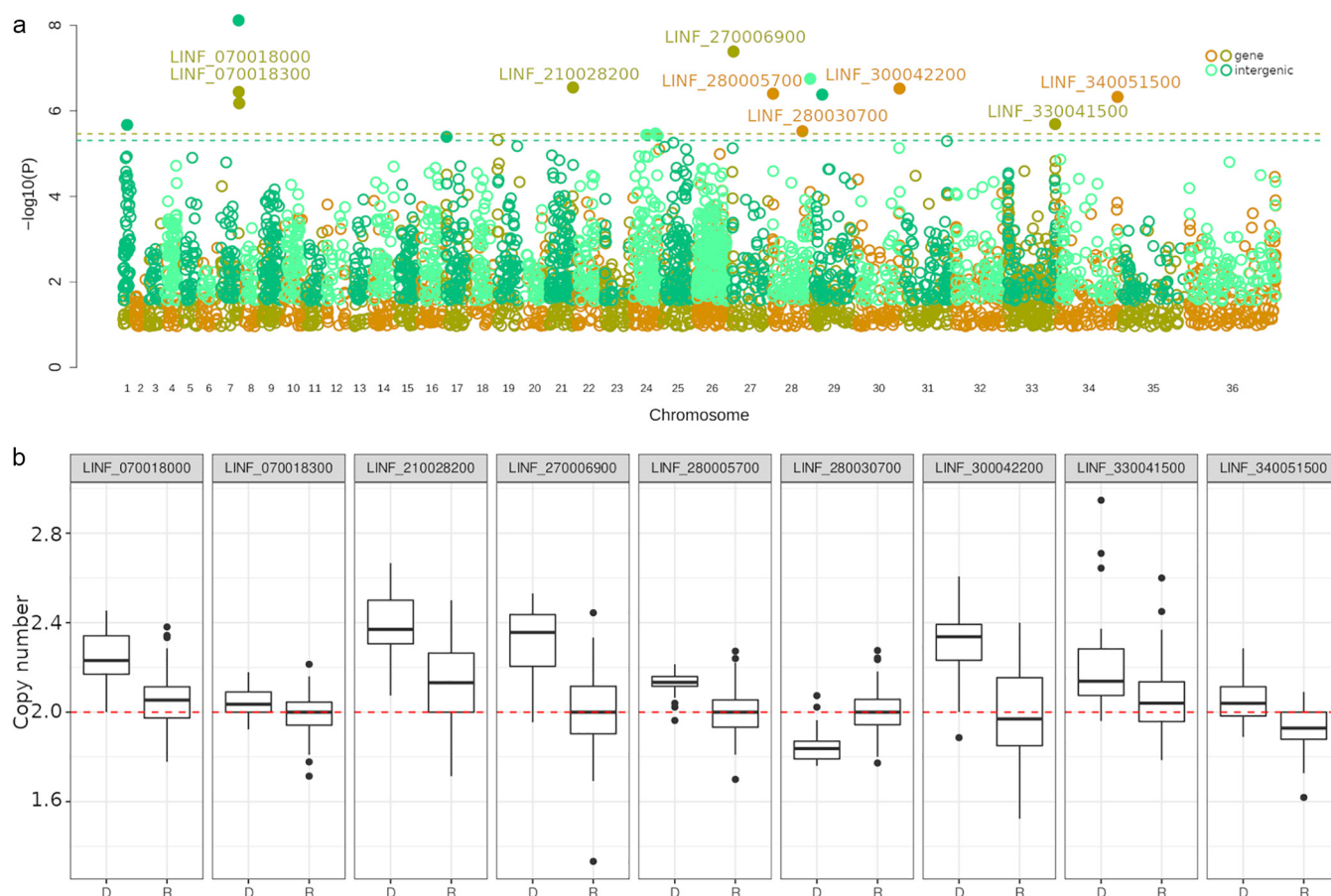
FIG 3 Association between genetic variation and mortality. (a) Manhattan plot displaying significant associations between GCNVs (yellow/orange), IGCNVs (dark/light green), and the mortality phenotype. Horizontal dashed lines indicate significance thresholds, calculated with 1,000 permutations (see Materials and Methods). Significant variants are highlighted by filled circles. Significant GCNVs are also named. $-\text{Log}_{10}$ P values were filtered to remove the 50% least significant values to aid figure clarity. (b) The copy numbers of significant GCNVs are displayed as box plots, indicating copies from isolates where the patient either died (D) or recovered (R). The dashed line indicates expected disomy. Mann-Whitney tests confirmed that copy numbers for these genes are significantly different in parasites causing patient mortality: LINF_070018000, $P = 1.26 \times 10^{-7}$; LINF_070018300, $P = 2.99 \times 10^{-3}$; LINF_210028200, $P = 1.60 \times 10^{-8}$; LINF_270006900, $P = 1.17 \times 10^{-8}$; LINF_280005700, $P = 1.89 \times 10^{-8}$; LINF_280030700, $P = 5.76 \times 10^{-8}$; LINF_300042200, $P = 4.59 \times 10^{-8}$; LINF_330041500, $P = 1.45 \times 10^{-3}$; LINF_340051500, $P = 1.58 \times 10^{-7}$.

parasite genotype (heritability, 0.83 ± 0.17 and 0.60 ± 0.27, respectively). Although our sample size is very small, we found multiple associations that are supported by permutation-derived thresholds for several clinical traits that are not correlated with one another. Both these significant GWAS associations (Fig. 3) and the inflation of P values in the SNP QQ plots indicate that multiple genetic factors contribute to different aspects of clinical outcomes. While this sample size is very small, similarly small sample sizes have identified significant association with microbial GWAS (28, 29).

We expect that disease outcomes and clinical manifestations are a combination of the interaction between the parasite, patient, environmental factors, and treatment. While no study has investigated the effects of human genetic variation on disease severity or mortality, several studies have evaluated the effects of human genetic variation on both visceral and cutaneous leishmaniasis susceptibility using GWAS (17).

The VL analysis identified a single locus with high significance; the human leukocyte antigen (HLA) haplotypes in the HLA-DR-DQ region ($P = 2.76 \times 10^{-17}$) were the only genetic determinant of susceptibility to VL (17) that was consistent between populations in Brazil and India. Much weaker association hits (top hits at a P value of $<5 \times 10^{-5}$) were observed for susceptibility to cutaneous leishmaniasis (CL). Our analysis of parasite genetic variants is entirely consistent with these results, if we consider that host-parasite interactions determine clinical severity. The most striking result of

our analysis is the magnitude of the contribution from the parasite genotype that we detected, without considering the human genotype. We consider this result reasonable from two perspectives. First, the effects of human genotypes on disease severity have not been examined directly (only disease susceptibility), and second, it is the parasite that causes disease, while host genotypes merely react to, and modify, disease severity.

Parasite-derived drug resistance and drug pharmacokinetics are also important factors (30). Genome-scale approaches have identified loci in *L. donovani* that are associated with antimonial resistance, which affect clinical outcomes (31, 32). Similarly, the low efficacy of miltefosine in Brazil, compared to initial efficacies in the Indian subcontinent, has been associated with the deletion of the miltefosine sensitivity locus (MSL), which that has been detected only in *L. infantum* isolates from Brazil (10). In our analysis, we did not observe any effects of MSL gene deletions on mortality or any other trait (GWAS *P* values for mortality for all MSL genes $> 1 \times 10^{-2}$), consistent with MSL being related to miltefosine resistance rather than mortality. The lack of association between the MSL locus and mortality is expected for our data set, as none of the evaluated patients were treated with miltefosine.

We consider it unlikely that the associations of mortality and sex with parasite genotype described here are due to drug resistance. This cohort is not ideal for such an analysis, since patients were treated with a variety of drugs, depending on disease severity (meglumine antimoniate, amphotericin B, liposomal amphotericin B, or pentamidine) (Table S3). Deaths from VL usually happen in the first week of admission. Drug treatments, even when resistance is present, are effective at this point, and resistance is noted only weeks after the beginning of the therapy. At present, no clear relationship between treatment failure and susceptibility to these drugs has been demonstrated in Brazil. Preliminary *in vitro* susceptibility analysis with some of the *L. infantum* isolates examined here confirms the lack of association between patient mortality and parasite drug susceptibility (M. Lima, unpublished data). However, host sex is implicated in VL incidence (33, 34) and increased parasitic load (35) but not in host mortality (4, 5). Therefore, the links of *L. infantum* genome to mortality and sex seems to follow distinct pathogenic pathways.

Our GWAS analysis indicated that a variety of different genetic variants were associated with clinical outcomes (Table 2). We discuss the potential implications of each observed SNP/indel and GCNV in supplemental texts B and C (https://figshare.com/projects/Brazil-VL-GWAS/126331), respectively. While we estimate that the parasite genotype has a strong influence on the chance of mortality ($h^2 = 0.83 \pm 0.17$; 95% confidence interval, 0.66 to 1.00), and some proportion of this effect appears to be derived from SNPs (Fig. 2), we did not observe any SNPs that were significantly associated with mortality (Fig. 3). This is possibly due to mortality being influenced by multiple alleles, each with small effects. While the significance values in our GWAS analysis appear fairly modest compared to GWAS analysis of human genetic variants that affect VL susceptibility (17), several considerations need to be taken into account. First, we have a smaller sample size ($n = 109$), so we cannot expect such strongly significant associations from a GWAS. Our results remain significant after permutation, because the *Leishmania* genome is smaller than human/dog genomes, which reduces the burden of multiple test correction. Nevertheless, some of our associations are very strong, particularly when covariate analysis is carried out. The gene LINF_340051500 was supported by a robust *P* value of 6.4×10^{-9} when Kala-Cal** was used as a covariate. Similarly, the genes LINF_270006900 and LINF_330042400 were supported when each trait besides Kala-Cal* and Kala-Cal** was simultaneously used as a covariate (*P* values of 5.9×10^{-7} and 1.19×10^{-6}, respectively) (Table S9), and LINF_270006900 was also supported in all cases when each other trait was individually used as a covariate (*P* values ranging from 8.08×10^{-7} to 1.25×10^{-8}) (Table S9).

Many of the variants associated with clinical outcomes defy mechanistic explanation. This is unsurprising given the poor state of basic knowledge about *Leishmania*-host interactions. Intergenic copy variants are particularly challenging to explain

mechanistically, as the function(s) of intergenic regions in *Leishmania* genomes is not well understood. However, intergenic regions in the human genome and yeast genomes are frequently associated with phenotypes, consistent with known functional elements within intergenic regions of these species (29, 36).

We found no associations between chromosome number (CCNVs) and phenotypic traits. This is in accordance with previous work suggesting that karyotypic levels change rapidly between insect and mammalian hosts and also when parasites are cultured, while GCNVs are drivers of long-term adaptation in the field (26). As the DNA from the parasites was obtained after only two passages, and significant changes in parasite virulence due to culturing are observed only after 20 to 30 passages (37), we believe that SNPs and GCNV alterations due to culturing would be minimal. Furthermore, even if small (SNP/indel) variations are altered within cultured parasites, such variations were previously associated with a reduction in virulence, which is restored with passages in mice (37, 38). Hence, genetic changes with parasite culture would be expected to reduce the strength of our GWAS/heritability associations between parasite genotype and clinical outcomes, rather than create artifactual ones. The strong relationships between parasite genotype and clinical factors observed in the present work indicate that the genome data obtained from cultured parasites relate meaningfully to outcomes.

We envisage several implications of our results. Most importantly, the strong influence of parasite genetic variation on mortality indicates that *L. infantum* isolates vary in their virulence, so the risk of mortality depends at least in part on the isolate that causes the infection. Variation in parasite virulence within Brazilian *L. infantum* isolates may explain the range of clinical VL symptoms, from symptomatic infections (39), to rare occurrences of highly virulent infections that result in a high fatality rate for VL in Brazil (40). While it may be possible to use parasite genotypes as early predictors of severe infections, this is unlikely to be economically viable or practically feasible at present, given the complexity of the genotypes associated with mortality and the requirement for parasite culture. It is more practical to use clinical data for risk assessments, which are well-powered and feasible in resource-poor settings (5) (see Fig. S14 at https://figshare.com/projects/Brazil-VL-GWAS/126331). However, studies of this type will lead to an enhanced understanding of the molecular mechanisms that lead to severe VL disease. The pathway to this understanding will be a combination of assessing genetic markers, the parasite genes that are implicated, and the host genes enrolled in the immune responses that the different alleles elicit. As many *Leishmania* genes lack detailed gene functional annotations, laboratory analysis will be required to obtain mechanistic insights.

MATERIALS AND METHODS

Study design and sample collection. The objective of this study was to search for relationships between parasite genotypes and patient clinical factors and outcomes, including mortality. To gather parasite genotypes, we collected *L. infantum* parasite samples from 109 diagnostic aspirates of 109 patients in the states of Piauí and Maranhão (Fig. 1), cultured them, and sequenced their genomes. DNA extraction from promastigote parasites was performed after two culture passages, to minimize the impact of culture adaptation. Each parasite genome is matched to clinical patient data. The corresponding patient and clinical data were extracted from medical records and routine patient follow-up. This included patient factors (age and sex), clinical characteristics (mortality, bleeding, edema, HIV coinfection, jaundice, vomiting, dyspnea, and bacterial coinfection), and laboratory clinical metrics (leukocyte and platelet counts and creatinine and hemoglobin levels). We also estimated probabilities of mortality from clinical or clinical + laboratory models (Kala-Cal* and Kala-Cal**) (Tables S1 and S2) (5). Our data analysis employed quantitative genetics methods of narrow-sense (additive) heritability and genome-wide association studies (GWAS) to associate specific parasite genetic variants (SNPs, indels, and gene and chromosomal CNVs) with patient factors and clinical characteristics.

Ethics. Piauí samples and patient data were obtained as a part of a broad study from the Federal University of Piauí and Federal University of Maranhão, respectively. Patient recruitment was performed at the Institute of Tropical Medicine Natan Portella, approved by the Research Ethics Committee of the Federal University of Piauí (approval ID 0116/2005), and in the Reference Hospital for Infectious Diseases of the Maranhão, approved by the Research Ethics Committee of the University Hospital of the Federal University of Maranhao (approval ID 2.793.599) and the Research Ethics Committee of the Federal University of Maranhao (approval ID 3.921.086). Both projects were approved by the Department of

Biology Ethics Committee at the University of York (approval IDs DJ201803 and DJ202009). All methods were performed according to approved guidelines and regulations. All participants or their legal guardians signed a written informed consent form. Clinical, epidemiological, and laboratory data from patients were obtained during hospital admission or medical follow-up or from medical records. All patients were treated with one of the following protocols, based on disease severity: (i) pentavalent antimony (20 mg/kg/day; maximum, 800 mg/day) for 21 days, (ii) amphotericin B deoxycholate (1 mg/kg/day for 14 to 20 days), or (iii) liposomal amphotericin B (AmBisome; 3 mg/kg/day for 7 days for immunocompetent and 14 days for immunocompromised patients). Parasites were isolated from bone marrow aspirates before the start of patient treatment, and culturing was performed at the Laboratory of Leishmaniasis at the Universidade Federal do Piauí or at the Laboratory of Genetics and Molecular Biology at the Universidade Federal do Maranhão.

Sequencing and bioinformatic analyses. Genome sequencing was performed on Illumina HiSeq 2500, generating paired-end 150-nucleotide (nt) reads. All but eight of the libraries had genome coverage of $\geq$30$\times$ (median, 49$\times$). (Table S1). Sequencing of 73 isolates was described previously (23). Sequencing reads were mapped onto the *L. infantum* JPCM5 v.46 (https://tritrypdb.org) reference genome using bwa v.0.7.17 (41). Read duplicates were removed with SAMtools v.1.9 (42). For each strain, SNPs and indels were called using Genome Analysis Toolkit (GATK) program HaplotypeCaller v.4.1.0.0 (43) and the "discovery" genotyping mode of Freebayes v.1.3.2 (https://github.com/ekg/freebayes), accepting only calls discovered by both methods. We retained only biallelic variants, with read depth within 0.3 to 1.7$\times$ the chromosome coverage, excluding any variants called on repetitive regions (see the supplemental methods at https://figshare.com/projects/Brazil-VL-GWAS/126331). All genomes were initially compared using phylogenetic approaches and PCA (see the supplemental methods). The gene, intergenic, and chromosome copy number variation of the 109 isolates was also assessed. This was performed by comparing variations in read depth coverage (RDC) between isolates (see the supplemental methods).

Genetic analysis; estimating heritability and GWAS. Heritability analysis is traditionally performed by analyzing individuals for whom we have pedigree information. The basic principle is that if a trait is heritable, then pairs of individuals who are more closely related (i.e., more genetically similar) will tend to be more phenotypically similar than pairs who are distantly related (i.e., less genetically similar). This enables us to estimate heritability based on the phenotypic correlations between twins, full siblings, parent-offspring pairs, etc. The advent of SNP genotyping means that it is no longer necessary to know the pedigree of the individuals, as family relationships can instead be inferred from SNP values. Further, it becomes possible to include information from unrelated pairs of individuals (in which case, the analysis leverages the fact that even across unrelated individuals, some pairs will by chance be more genetically similar than others).

SNPs/indels and gene, intergenic region, and chromosome CNVs were used as genotypes to estimate the total effect of parasite genetic variation on the patient clinical traits (narrow-sense heritability), using LDAK v.5.1 (44). In this analysis, clinical data were coded as either binary or quantitative traits (Table S2). PLINK v.1.9 (45) was used to create the binary files. Gene, intergenic, and chromosome CNVs were generated from RDC values normalized by genome coverage for each gene, intergenic region, or chromosome and coded as quantitative genotypes. Gene CNVs varied from 0 to 3.8 copies per haploid genome copy (mean, 1.017). The contribution of genotype class to each clinical trait was estimated using LDAK restricted maximum likelihood (REML) (see the supplemental methods at https://figshare.com/projects/Brazil-VL-GWAS/126331).

We used GWAS to identify specific genomic variants that were statistically associated with each clinical trait. The LDAK function –linear was used to conduct mixed-model GWAS, using the kinship matrix derived from SNP/indel variants to control for unequal strain relatedness. Significance thresholds were determined via trait permutation. For each trait, values were permuted 1,000 times, and parasite genotype-human trait associations were estimated. The lowest P value was determined from each replicate permutation analysis, and the 5th percentile was used as the significance threshold for associations (i.e., 5% family-wise false discovery rate) for each trait. P-value thresholds varied from 5.2 to 7.3 $-\log_{10}$. GWAS for SNPs/indels used only biallelic sites with call quality of $\geq$2,000 and population-wide minor allele frequency (MAF) of >5% across all 109 isolates (i.e., an allele must be present in at least 11 of the 218 haploid chromosomes to be used; n = 3,526). GWAS for CNVs used only genes with a mean genome coverage of >20% and a coverage coefficient of variation (CV) (SD/mean) of >1% (n = 8,219 genes; see the supplemental methods at https://figshare.com/projects/Brazil-VL-GWAS/126331).

Data availability. Genome sequencing data produced here are available from NCBI's sequencing read archive under the BioProject ID PRJNA781413. All other sequencing data used are listed in Table S1. The supplemental material, including scripts used to run associations and permutations as well as the script to generate the Manhattan, QQ, correlation, and box plots and the alignments and resulting phylogenetic tree files, is available at https://figshare.com/projects/Brazil-VL-GWAS/126331.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
TABLE S1, XLSX file, 0.02 MB.
TABLE S2, XLSX file, 0.05 MB.
TABLE S3, XLSX file, 0.01 MB.
TABLE S4, XLSX file, 0.01 MB.

TABLE S5, XLSX file, 0.1 MB.
TABLE S6, XLSX file, 0.01 MB.
TABLE S7, XLSX file, 0.1 MB.
TABLE S8, XLSX file, 0.01 MB.
TABLE S9, XLSX file, 0.01 MB.
TABLE S10, XLSX file, 0.01 MB.

REFERENCES

1. Ribeiro CJN, Dos Santos AD, Lima SVMA, da Silva ER, Ribeiro BVS, Duque AM, Peixoto MVS, Dos Santos PL, de Oliveira IM, Lipscomb MW, de Araújo KCGM, de Moura TR. 2021. Space-time risk cluster of visceral leishmaniasis in Brazilian endemic region with high social vulnerability: an ecological time series study. PLoS Negl Trop Dis 15:e0009006. https://doi.org/10.1371/journal.pntd.0009006.

2. Burza S, Croft SL, Boelaert M. 2018. Leishmaniasis. Lancet 392:951–970. https://doi.org/10.1016/S0140-6736(18)31204-2.

3. WHO. 2010. Control of the leishmaniases WHO technical report series 949. World Health Organization, Geneva, Switzerland.

4. Costa CHN, Werneck GL, Costa DL, Holanda TA, Aguiar GB, Carvalho AS, Cavalcanti JC, Santos LS. 2010. Is severe visceral leishmaniasis a systemic inflammatory response syndrome? A case control study. Rev Soc Bras Med Trop 43:386–392. https://doi.org/10.1590/s0037-86822010000400010.

5. Costa DL, Rocha RL, Chaves EdBF, Batista VGdV, Costa HL, Costa CHN. 2016. Predicting death from kala-azar: construction, development, and validation of a score set and accompanying software. Rev Soc Bras Med Trop 49:728–740. https://doi.org/10.1590/0037-8682-0258-2016.

6. Costa DL, Rocha RL, Carvalho RMA, Lima-Neto AS, Harhay MO, Costa CHN, Barral-Neto M, Barral AP. 2013. Serum cytokines associated with severity and complications of kala-azar. Pathog Glob Health 107:78–87. https://doi.org/10.1179/2047773213Y.0000000078.

7. Porcino GN, Carvalho KSS, Braz DC, Costa Silva V, Costa CHN, de Miranda Santos IKF. 2019. Evaluation of methods for detection of asymptomatic individuals infected with Leishmania infantum in the state of Piauí, Brazil. PLoS Negl Trop Dis 13:e0007493. https://doi.org/10.1371/journal.pntd.0007493.

8. Silva de Lima UR, Vanolli L, Moraes EC, Ithamar JS, Pedrozo e Silva de Azevedo CDM 2019. Visceral leishmaniasis in Northeast Brazil: what is the impact of HIV on this protozoan infection? PLoS One 14:e0225875. https://doi.org/10.1371/journal.pone.0225875.

9. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M, WHO Leishmaniasis Control Team. 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7:e35671. https://doi.org/10.1371/journal.pone.0035671.

10. Carnielli JB, Crouch K, Forrester S, Silva VC, Carvalho SF, Damasceno JD, Brown E, Dickens NJ, Costa DL, Costa CH, Dietze R, Jeffares DC, Mottram JC. 2018. A Leishmania infantum genetic marker associated with miltefosine treatment failure for visceral leishmaniasis. EBioMedicine 36:83–91. https://doi.org/10.1016/j.ebiom.2018.09.029.

11. Pan American Health Organization. 2019. Leishmaniasis. Epidemiological report of the Americas, March 2019. Leishmaniasis report 7.

12. Cota G, Erber AC, Schernhammer E, Simões TC. 2021. Inequalities of visceral leishmaniasis case-fatality in Brazil: a multilevel modeling considering space, time, individual and contextual factors. PLoS Negl Trop Dis 15: e0009567. https://doi.org/10.1371/journal.pntd.0009567.

13. Lima ID, Queiroz JW, Lacerda HG, Queiroz PVS, Pontes NN, Barbosa JDA, Martins DR, Weirather JL, Pearson RD, Wilson ME, Jeronimo SMB. 2012. Leishmania infantum chagasi in northeastern Brazil: asymptomatic infection at the urban perimeter. Am J Trop Med Hyg 86:99–107. https://doi.org/10.4269/ajtmh.2012.10-0492.

14. Silva LP, Montenegro S, Werkauser R, Sales KGdS, Soares FCS, Costa VMA, Bezerra AC, Pinto MBdA, Ferreira SM, Neitzke-Abreu HC, Dantas-Torres F, Lima Junior MSdC 2020. Asymptomatic Leishmania infection in blood donors from a major blood bank in Northeastern Brazil: a cross-sectional study. Rev Inst Med Trop Sao Paulo 62:e92. https://doi.org/10.1590/s1678-9946202062092.

15. Maia-Elkhoury ANS, Sierra Romero GA, O B Valadas SY, L Sousa-Gomes M, Lauletta Lindoso JA, Cupolillo E, Ruiz-Postigo JA, Argaw D, Sanchez-Vazquez MJ. 2019. Premature deaths by visceral leishmaniasis in Brazil investigated through a cohort study: a challenging opportunity? PLoS Negl Trop Dis 13: e0007841. https://doi.org/10.1371/journal.pntd.0007841.

16. Mohamed HS, Ibrahim ME, Blackwell JM. 2019. Genetic susceptibility to visceral leishmaniasis, p 71–85. In Ibrahim ME, Rotimi CN (ed), The genetics of African populations in health and disease. Cambridge University Press, Cambridge, United Kingdom.

17. Blackwell JM, Fakiola M, Castellucci LC. 2020. Human genetics of Leishmania infections. Hum Genet 139:813–819. https://doi.org/10.1007/s00439-020-02130-w.

18. Ponte-Sucre A, Gamarro F, Dujardin J-C, Barrett MP, López-Vélez R, García-Hernández R, Pountain AW, Mwenechanya R, Papadopoulou B. 2017. Drug resistance and treatment failure in leishmaniasis: a 21st century challenge. PLoS Negl Trop Dis 11:e0006052. https://doi.org/10.1371/journal.pntd.0006052.

19. Carvalho KSS, da Silva Júnior WJ, da Silveira Regueira Neto M, Silva VC, de Sá Leitão Paiva Júnior S, Balbino VQ, Costa DL, Costa CHN. 2020. Application of next generation sequencing (NGS) for descriptive analysis of 30 genomes of Leishmania infantum isolates in Middle-North Brazil. Sci Rep 10:12321. https://doi.org/10.1038/s41598-020-68953-9.

20. de Azevedo TS, Lorenz C, Chiaravalloti-Neto F. 2019. Risk mapping of visceral leishmaniasis in Brazil. Rev Soc Bras Med Trop 52:e20190240. https://doi.org/10.1590/0037-8682-0240-2019.

21. Casadevall A, Pirofski L-A. 2009. Virulence factors and their mechanisms of action: the view from a damage-response framework. J Water Health 7 (Suppl 1):S2–S18. https://doi.org/10.2166/wh.2009.036.

22. Moxon R, Reche PA, Rappuoli R. 2019. Editorial: Reverse vaccinology. Front Immunol 10:2776. https://doi.org/10.3389/fimmu.2019.02776.

23. Grace CA, Forrester S, Silva VC, Sousa Carvalho KS, Kilford G, Chew YP, James S, Costa DL, Mottram JC, Costa CCHN, Jeffares DC. 2021. Candidates for balancing selection in Leishmania donovani complex parasites. Genome Biol Evol 13:evab265. https://doi.org/10.1093/gbe/evab265.

24. Franssen SU, Durrant C, Stark O, Moser B, Downing T, Imamura H, Dujardin J-C, Sanders MJ, Mauricio I, Miles MA, Schnur LF, Jaffe CL, Nasereddin A, Schallig H, Yeo M, Bhattacharyya T, Alam MZ, Berriman M, Wirth T, Schönian G, Cotton JA. 2020. Global genome diversity of the Leishmania donovani complex. Elife 9:e51243. https://doi.org/10.7554/eLife.51243.

25. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin J-C, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. Genome Res 21:2129–2142. https://doi.org/10.1101/gr.122945.111.

26. Bussotti G, Gouzelou E, Côrtes Boité M, Kherachi I, Harrat Z, Eddaikra N, Mottram JC, Antoniou M, Christodoulou V, Bali A, Guerfali FZ, Laouini D, Mukhtar M, Dumetz F, Dujardin J-C, Smirlis D, Lechat P, Pescher P, El Hamouchi A, Lemrani M, Chicharro C, Llanes-Acevedo IP, Botana L, Cruz I, Moreno J, Jeddi F, Aoun K, Bouratbine A, Cupolillo E, Späth GF. 2018. Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. mBio 9:e01399-18. https://doi.org/10.1128/mBio.01399-18.

27. Domagalska MA, Imamura H, Sanders M, Van den Broeck F, Bhattarai NR, Vanaerschot M, Maes I, D'Haenens E, Rai K, Rijal S, Berriman M, Cotton JA, Dujardin J-C. 2019. Genomes of Leishmania parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. PLoS Negl Trop Dis 13:e0007900. https://doi.org/10.1371/journal.pntd.0007900.

28. Power RA, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet 18:41–50. https://doi.org/10.1038/nrg.2016.132.

29. Jeffares DC, Rallis C, Rieux A, Speed D, Převorovský M, Mourier T, Marsellach FX, Iqbal Z, Lau W, Cheng TMK, Pracana R, Mülleder M, Lawson JLD, Chessel A, Bala S, Hellenthal G, O'Fallon B, Keane T, Simpson JT, Bischof L, Tomiczek B, Bitton DA, Sideri T, Codlin S, Hellberg JEEU, van Trigt L, Jeffery L, Li J-J, Atkinson S, Thodberg M, Febrer M, McLay K, Drou N, Brown W, Hayles J, Carazo Salas RE, Ralser M, Maniatis N, Balding DJ, Balloux F, Durbin R, Bähler J. 2015. The genomic and phenotypic diversity of Schizosaccharomyces pombe. Nat Genet 47:235–241. https://doi.org/10.1038/ng.3215.

30. Alves F, Bilbe G, Blesson S, Goyal V, Monnerat S, Mowbray C, Ouattara GM, Pecoul B, Rijal S, Rode J, Solomos A, Strub-Wourgaft N, Wasunna M, Wells S, Zijlstra EE, Arana B, Alvar J. 2018. Recent development of visceral leishmaniasis treatments: successes, pitfalls, and perspectives. Clin Microbiol Rev 31:e00048-18. https://doi.org/10.1128/CMR.00048-18.

31. Imamura H, Downing T, Van den Broeck F, Sanders MJ, Rijal S, Sundar S, Mannaert A, Vanaerschot M, Berg M, De Muylder G, Dumetz F, Cuypers B, Maes I, Domagalska M, Decuypere S, Rai K, Uranw S, Bhattarai NR, Khanal B, Prajapati VK, Sharma S, Stark O, Schönian G, De Koning HP, Settimo L, Vanhollebeke B, Roy S, Ostyn B, Boelaert M, Maes L, Berriman M, Dujardin J-C, Cotton JA. 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5:e12613. https://doi.org/10.7554/eLife.12613.

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

33. Sharma MC, Gupta AK, Saran R, Sinha SP. 1990. The effect of age and sex on incidence of kala-azar. J Commun Dis 22:277–278.

34. Gouvêa MV, Werneck GL, Costa CHN, de Amorim Carvalho FA. 2007. Factors associated to Montenegro skin test positivity in Teresina, Brazil. Acta Trop 104:99–107. https://doi.org/10.1016/j.actatropica.2007.07.010.

35. Zacarias DA, Rolão N, de Pinho FA, Sene I, Silva JC, Pereira TC, Costa DL, Costa CHN. 2017. Causes and consequences of higher Leishmania infantum burden in patients with kala-azar: a study of 625 patients. Trop Med Int Health 22:679–687. https://doi.org/10.1111/tmi.12877.

36. Giral H, Landmesser U, Kratzer A. 2018. Into the wild: GWAS exploration of non-coding RNAs. Front Cardiovasc Med 5:181. https://doi.org/10.3389/fcvm.2018.00181.

37. Moreira D, Santarém N, Loureiro I, Tavares J, Silva AM, Amorim AM, Ouaissi A, Cordeiro-da-Silva A, Silvestre R. 2012. Impact of continuous axenic cultivation in Leishmania infantum virulence. PLoS Negl Trop Dis 6:e1469. https://doi.org/10.1371/journal.pntd.0001469.

38. Ali KS, Rees RC, Terrell-Nield C, Ali SA. 2013. Virulence loss and amastigote transformation failure determine host cell responses to Leishmania mexicana. Parasite Immunol 35:441–456. https://doi.org/10.1111/pim.12056.

39. Costa CHN, Werneck GL, Rodrigues L, Santos MV, Araújo IB, Moura LS, Moreira S, Gomes RBB, Lima SS. 2005. Household structure and urban services: neglected targets in the control of visceral leishmaniasis. Ann Trop Med Parasitol 99:229–236. https://doi.org/10.1179/136485905X28018.

40. Ruiz-Postigo JA, Jain S, Mikhailov A, Maia-Elkhoury AN, Valadas S, Warusavithana S, Osman M. 2021. Global leishmaniasis surveillance: 2019–2020, a baseline for the 2030 roadmap/Surveillance mondiale de la leishmaniose: 2019–2020, une periode de reference pour la feuille de route a l'horizon 2030. Wkly Epidemiol Rec 96:401–420.

41. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Bur-rows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.

42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

43. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A frame-work for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. https://doi.org/10.1038/ng.806.

44. Speed D, Holmes J, Balding DJ. 2020. Evaluating and improving heritabil-ity models using summary statistics. Nat Genet 52:458–462. https://doi.org/10.1038/s41588-020-0600-y.

45. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Sec-ond-generation PLINK: rising to the challenge of larger and richer data-sets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8.