EVOLUTIONARY GENOMICS OF A ZOONOTIC PARASITE ACROSS THE NEOTROPICAL REALM

Senne Heeren[1,2,3,*], Mandy Sanders[4,†], Jeffrey Jon Shaw[5], Sinval Pinto Brandão-Filho[6], Mariana Côrtes Boité[7], Lilian Motta Cantanhêde[7,8], Khaled Chourabi[7], Ilse Maes[1], Alejandro Llanos-Cuentas[9], Jorge Arevalo[9], Jorge D. Marco[10], Philippe Lemey[2], James A. Cotton[4,11], Jean-Claude Dujardin[1,3], Elisa Cupolillo[7,8,*], Frederik Van den Broeck[1,2,*]

AFFILIATIONS

[1] Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[2] Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium
[3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
[4] Welcome Sanger Institute, Hinxton, United Kingdom
[5] Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo (USP), São Paulo, SP, Brazil
[6] Department of Immunology, Aggeu Magalhães Institute, Fiocruz, Pernambuco, Brazil
[7] Leishmaniasis Research Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro 21040360, Brazil
[8] Instituto Nacional de Ciência e Tecnologia de Epidemiologia da Amazônia Ocidental, INCT EpiAmO, Porto Velho 76812100, Brazil
[9] Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru
[10] Instituto de Patología Experimental, Universidad Nacional de Salta — Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Salta, Argentina
[11] School of Biodiversity, One Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
[†]In Memoriam of Mandy Sanders

Co-senior authors: Frederik Van den Broeck and Elisa Cupolillo

***Correspondence:**
Senne Heeren (sheeren@itg.be)
Frederik Van den Broeck (fvandenbroeck@gmail.com)
Elisa Cupolillo (elisa.cupolillo@ioc.fiocruz.br)

ORCID IDs

SH: 0000-0002-5786-258X
FVDB: 0000-0003-2542-5585
PL: 0000-0003-2826-5353
JCD: 0000-0002-9217-5240
JAC: 0000-0001-5475-3583
JDM: 0000-0002-4408-0941
SPBF: 0000-0003-3768-2810
EC: 0000-0002-0620-3250

ABSTRACT

The Neotropical realm, one of the most biodiverse regions on Earth, houses a broad range of zoonoses that pose serious public health threats. Protozoan parasites of the *Leishmania* (*Viannia*) *braziliensis* species complex cause zoonotic leishmaniasis in Latin America with clinical symptoms ranging from simple cutaneous to destructive, disfiguring mucosal lesions. We present the first comprehensive genome-wide continental study including 257 cultivated isolates representing most of the geographical distribution of this major human pathogen. The *L. braziliensis* species complex is genetically highly heterogeneous, consisting of divergent parasite groups that are associated with different environments and vary greatly in diversity. Apart from several small ecologically isolated groups with little diversity, our sampling identifies two major parasite groups, one associated with the Amazon and the other with the Atlantic Forest biomes. These groups show different recombination histories, as suggested by high levels of heterozygosity and effective population sizes in the Amazonian group in contrast to high levels of linkage and clonality in the Atlantic group. We argue that these differences are linked to strong eco-epidemiological differences between the two regions. In contrast to geographically focused studies, our study provides a broad understanding of the molecular epidemiology of zoonotic parasites circulating in tropical America.

INTRODUCTION

Leishmaniasis is a vector-borne disease that is caused by the protozoan *Leishmania* parasite (Trypanosomatidae) and transmitted by phlebotomine sand flies in tropical regions. It is a spectral disease with many clinical manifestations, including visceral and various forms of cutaneous leishmaniasis (CL) (1). While visceral leishmaniasis is potentially fatal if the patient is not treated, CL is the most common form of the disease and causes a large burden due to social stigma and humiliation (2). It is estimated that globally about 700,000 to 1.2 million CL cases occur each year (3). In South America the annual CL incidence is estimated to lie between 190,000 and 308,000 cases.

One of the most important causative agents of CL and the more severe mucocutaneous disease in South America is *Leishmania* (*Viannia*) *braziliensis*, a zoonotic parasite circulating principally in wild rodents (4, 5). Human infections appear to be a spillover from the sylvatic transmission cycle. In addition, skin lesions due to *L. braziliensis* have also been found in domestic animals such as equines, dogs and cats (6, 7), suggesting a peridomestic transmission cycle in some areas. In ecological terms, *L. braziliensis* has typical generalist characteristics that allow it to occupy a broad range of ecological niches. This is highlighted by i) its high genetic diversity (8–14), ii) its continent-wide distribution, occurring in at least 15 Central and South American countries (13–16), iii) its vast range of different vector (17, 18) and reservoir (4) host species, and (iv) the existence of a handful of genetically divergent ecotypes (8, 9, 11–14, 19). The *L. braziliensis* parasite is thus an ideal model species for understanding the population structure of zoonotic pathogens circulating across the Neotropical realm.

Studies investigating the natural genetic diversity of *L. braziliensis* based on amplified (8) or restriction fragment length polymorphisms (AFLP; RFLP) (12), multilocus microsatellites (9, 20) and whole genome sequence data (11, 13, 14, 21, 22) revealed a high genetic heterogeneity partitioned by the environment. At the continental level, there is a clear distinction between *L. braziliensis* populations circulating in the Amazonian and Atlantic rainforests (23). Parasite molecular heterogeneity appeared to be substantially higher in the Amazon, presumably due to its more diverse vector and reservoir host communities (23). In Peru and Bolivia, studies have shown that the Amazonian *L. braziliensis* is further subdivided into distinct subpopulations that are associated with specific ecoregions (20, 22). In addition, several genetically divergent ecotypes have been reported across South-America (8, 9, 11–13, 19), such as *Leishmania (Viannia) peruviana* that emerged in the Peruvian Andes during forestation changes over the past 150,000 years (13). These observations highlight the extensive diversity of *L. braziliensis* variants infectious to humans.

Most studies on the natural variation of *L. braziliensis* were restricted in terms of the geographic scope (11–13, 20, 22, 23), limiting our knowledge on the evolution of the parasite across its range. Here, the goal of our study was to map the continental genome variation and population structure of *L. braziliensis* within a broad ecological context. This was achieved by using whole genome sequencing data of 257 cryopreserved parasite isolates sampled in Argentina, Bolivia, Brazil and Peru, covering a wide range of ecological regions including Andean, Amazonian and Atlantic forests. Capitalizing on an unprecedented genome dataset

for this major human pathogen, we gain essential knowledge on the molecular epidemiology of cutaneous leishmaniasis in South America.

RESULTS

L. braziliensis consists of genetically divergent ecotypes

Paired-end whole-genome sequence data were generated from promastigote cultures of 188 *Leishmania* isolates and combined with previously generated sequencing data of 69 *Leishmania* isolates, including isolates from different *L. (Viannia)* species for comparative purposes (Supp. Table 1). The median read coverage was 55x (mean = 56x, SD = 21x). A total of 13 genomes were excluded for downstream analyses, either because of i) aberrant allele frequency distributions potentially indicative of mixed infections or contamination (Supp. Fig 1) or ii) a low or fragmented coverage of the accessible genome (i.e. genomic regions with a minimum read depth of 5, mapping quality of 25 and base quality of 25) (Supp. Fig 2A). The resulting dataset consisted of a total of 244 high-quality genomes belonging to the *L. braziliensis* species complex (N=226), *Leishmania* (*Viannia*) *guyanensis* species complex (N = 5), *Leishmania* (*Viannia*) *lainsoni* (N=2), *Leishmania* (*Viannia*) *naiffi* (N=4) and *L.eishmania* (*Viannia*) *shawi* (N=1). Six genomes showed more complex ancestries and were characterized as interspecific hybrid parasites (Supplementary Results).

Inspection of homozygous and heterozygous SNP counts in our panel of 226 *L. braziliensis* genomes revealed four groups of parasites (Supp. Fig 2B; Table 1), including a large group of *L. braziliensis* parasites found within the Amazonian and Atlantic rainforests (here-after referred to as L1) (N=182, including the M2904 reference strain), one group found sporadically in Brazil, Peru and Bolivia that has previously been associated with both human and canine leishmaniasis (here-after referred to as L2) (N=4) (8, 9, 12), one group that has been described solely in the Paudalho municipality (Pernambuco state) in Northeastern Brazil (here-after referred to as L3) (N=9) (11), and the well-described *L. peruviana* ecotype that is found within the Peruvian highlands (N=31) (13). L2 (60,095 SNPs) showed a significantly larger number of SNPs compared to L1 (30,158 SNPs), L3 (25,620 SNPs) and *L. peruviana* (26,024 SNPs) (pairwise Dunn's tests: Supp. Table 2). L2 appeared genomically to be the most divergent *L. braziliensis* group, as indicated by its distant position in the phylogenetic network (Supp. Fig 2C) and the high number of homozygous SNPs (55,773 SNPs) called against the *L. braziliensis* reference (Supp. Fig 2B).

In terms of heterozygous SNPs per isolate, L2 (median 4,406 SNPs) and in particular L3 (median 113 SNPs) and *L. peruviana* (median 98 SNPs) exhibited a significantly lower number compared to L1 (median 13,766 SNPs) (pairwise Dunn's tests: Supp. Table 3; Supp. Fig 2B). Additionally, the population allele frequency spectrum of L1 was dominated by low-frequency variants (i.e. 75% of the alleles having a frequency below 0.1), whereas the majority of SNP loci were entirely fixed in L2 (81.42%), L3 (99.1%) and *L. peruviana* (66.66%) (Supp. Fig 3). Pairwise Dunn's tests on the pairwise genetic distances (Bray-Curtis dissimilarity) confirmed that L2 (L2-L1: $Z = 5.53$; $p = 6.33e-08$), L3 (L3-L1: $Z = 14.85$; $p < 2.2e-16$) and *L. peruviana* (*Lp*-L1: $Z = 48.96$; $p < 2.2e-16$) hold a significantly lower genetic variability relative to L1.

A phylogenetic network based on 695,229 genome-wide SNPs highlighted the extensive diversity in L1 where individual genomes were separated by relatively long branches, in contrast to the L3 and *L. peruviana* genomes that appear terminally as single divergent offshoots (Fig 1A). This was corroborated by PCA: PC1 (29.9%) mainly explained the large diversity in L1, while PC2 (16.6%) and PC3 (10.4%) separated *L. peruviana* and L3, respectively (Fig 1B,C). Ancestry estimation revealed more insight into the divergence of *L. braziliensis* in South America (Fig 1D-F) and its association with the environment (Fig 1G). At the deepest evolutionary level (i.e. K=2) (Fig 1D), there was a clear separation between L1 parasites from the Atlantic (i.e. Eastern Highlands) and Amazonian Forests (i.e. Amazonian–Orinocan Lowlands). *L. peruviana* and L3 appeared as separate parasite groups at K=3 (Fig 1E) and K=4 (Fig 1F), respectively. Both *L. peruviana* from the Peruvian highlands and L3 from the Pernambuco state in Brazil clustered largely with L1 from the Amazonian rainforests at K=2, although the ancestry of L3 seems somewhat more complex (Fig 1D). In addition to the clear distinction between Amazonian and Atlantic L1, we also encountered isolates showing patterns of mixed ancestry between these two distinct populations. These isolates originated geographically from the center of the Amazon, more or less in between the foci of their putative parental lineages.

Pairwise F_{st} calculations confirmed the divergent nature of each parasite group, with estimates ranging from 0.11 to 0.77 (Supp. Table 4). Notably, F_{st} was similar when estimated between the Amazonian L1 group on the one hand and Atlantic L1 (F_{st} = 0.12), L3 (F_{st} = 0.14) or *L. peruviana* (F_{st} =0.11) on the other hand, which may indicate that the Amazonian L1 group represents the ancestral parasite population from which all other parasite groups emerged. Estimates of F_{st} were much higher when compared between Atlantic L1 on the one hand and L3 (F_{st} = 0.39) or *L. peruviana* (F_{st} = 0.40) on the other hand, and between L3 and *L. peruviana* (F_{st} = 0.77) (Supp. Table 4).

Continental population diversity and stucture of *L. braziliensis* L1

The population structure of the L1 group was examined in more detail based on 194,791 SNPs (178,400 bi-allellic) that were called across 182 *L. braziliensis* isolates sampled in Argentina, Bolivia, Brazil and Peru (Supp. Table 1). Our analyses predicted that 1,727 variants (216 SNPs; 1,511 INDELs) have a deleterious impact on the underlying protein sequences. However, the large majority (96.99%) of these deleterious mutations occurred in low frequencies (< 5%) (Supp. Table 5). For population structure analyses, we retained one genome per clonal group (here-after N_{unique} refers to the number of genomes after removing multiple clones) (see methods and below for more details) and removed SNPs showing high LD, resulting in a dataset of 106,188 bi-allelic SNPs called across 119 genomes.

We identified three major parasite groups showing strong spatio-environmental structuring (Fig 2A-C), whereby each isolate was assigned with at least 85% ancestry to their respective group. The Atlantic (ATL) group (N_{all}=66, N_{unique}=20) represents parasites isolated in (North-) Eastern Brazil and North Argentina between 1995 and 2016. The West Amazon (WAM) group (N_{all}=81, N_{unique}=67) contained isolates from Bolivia, Western Brazil and Peru that were sampled between 1990 and 2003. The Central Amazon (CAM) group (N_{all}=22, N_{unique}=20) contains isolates sampled in Bolivia and West/Central Brazil between 1984 and 2015 (Fig 2C). The ATL group (22,414 SNPs) exhibited a significantly lower number of SNPs in comparison to WAM (30,725 SNPs) and CAM (30,254 SNPs) (Kruskal-Wallis test: χ^2 =121.81; df = 2; p < 2.2e-16; pairwise Dunn's tests: Supp. Table 6). Isolates showing less than

85% ancestry for any of the inferred groups were grouped together into a conglomerate (CON) group of parasites showing patterns of mixed ancestry (N_{all}=13, N_{unique}=12) (Fig 2A). Parasites of this polyphyletic group were sampled between 1975 and 2015, originating from Argentina, Brazil and Peru (Fig 2C).

Next to the geographical east-west stratification of *L. braziliensis* L1, there were also indications of ecological differentiation (Fig 2B) as we found a significant association between the three major parasite groups and the biomes where they occur (Chi-squared test of independence: χ^2 = 300.83; df = 18; p = 3.24e-53). More specifically, ATL was predominantly linked with the Atlantic Forest biome in Brazil and the Western Dry Chaco in Argentina, CAM was mainly associated with the Amazonian and Coastal Lowlands while WAM was more associated with the Amazonian Irregular Plains and Piedmont, the Yungas as well as the Central High Andes (Fig 2B). Pairwise mean F_{st} values revealed a clear differentiation between the Amazonian and Atlantic populations ($F_{st(WAM-ATL)}$ = 0.16 ± 0.07; $F_{st(CAM-ATL)}$ = 0.15 ± 0.06; Supp. Table 7), which was higher compared to the differentiation within the two Amazonian populations ($F_{st(WAM-CAM)}$ = 0.06 ± 0.02) (Supp. Table 7). Assuming K=5 populations (as per lowest cross validation error) revealed the sub-structuring of the WAM population which corresponded with the recently described population structure of Amazonian *L. braziliensis* in Peru and Bolivia (22), and which will not be further discussed here (Fig 2 A,D).

We next investigated the distribution of chromosome and gene copy number variants across the different *L. braziliensis* populations. Consistent with previous reports (24–26), we described considerable variation in chromosome copy numbers, including chromosome 31 that was polysomic in all individuals (Supp. Fig 4). A PCA based on gene copy number variations (CNVs) revealed a similar population structure as observed based on SNPs, suggesting that each of the three populations WAM, CAM and ATL carries a specific CNV pattern (Fig 3A). We found significant differences in the number of CNVs between WAM-CAM and WAM-ATL, though not between CAM-ATL (Fig 3B,C; Supp. Table 8). Similarly, the CNV burden (i.e. the proportion of the genome covered by CNVs) ranged between 0.001% to 0.83% of the genome, and was significantly different between amplifications of WAM and ATL, and deletions of WAM and CAM or ATL (Fig 3D,E; log rank tests for survival curve differences: Supp. Table 9). The CNV frequency distributions in each population were skewed towards rare variants, with a median CNV frequency of 5% for WAM, 23% for CAM and 7% for ATL (Supp. Fig 5; Supp. Table 10-12). Nine amplifications were present in more than 90% of the individuals in each of the three populations (ANOVA: F = 268.4, df = 2, p < 2.2e-16; adjusted R-squared = 0.27) (Tukey's HSD test: Supp. Table 13), seven of which coding for beta-tubulins on chromosome 33 (ORTHOMCL4), one coding for GP63 on chromosome 10 (ORTHOMCL1) and one conserved hypothetical protein on chromosome 31 (ORTHOMCL2303).

Contrasting recombination histories in Amazonian and Atlantic *L. braziliensis* L1

We identified 18 clusters of near-identical genomes that constituted 44.5% (N=81) of the isolates (Supp. Table 1). Genomes within each of the 18 clusters exhibited relatively few heterozygous SNP differences (median=256, min=3, max=2720) and virtually no fixed homozygous SNP differences (homSNPs) (median=0, min=0, max=17) (Supp. Table 14). These observations suggest that there is a lack of recombination and chromosomal re-assortment between parasites of the same cluster (hereafter referred to as clonal groups). Exceptions were isolates M2903 and EMM (133 homSNPs), LSC358_2 and LSC582 (127 homSNPs), LSC358_2 and LSC684 (127 homSNPs) and LSC358_2 and LSC392 (126

homSNPs). Close inspection revealed that these homSNPs are localized on chromosomes 20 (first 360kb) and 35 (300kb-410kb) for M2903 and EMM or on chromosome 29 (1Mb-1.2Mb) for LSC358_2, LSC582, LSC684 and LSC392, and are thus likely the result of gene conversion.

We found a strong difference in the number of near-identical genomes between the Amazonian (WAM, CAM) and Atlantic populations (ATL) (Chi-squared test: $\chi^2 = 49.55$; df = 2; p = 1.742e-11; Fig 4A). In particular, ATL (53/66, 80.3%) appeared to have a significantly higher clonal prevalence compared to WAM (22/81, 27.2%) and CAM (4/22, 18.2%) (pairwise Fisher's exact tests: Supp. Table 15). No significant differences were found between the populations WAM, CAM and ATL in terms of the number of near-identical genomes per clonal group (Kruskal-Wallis test: $\chi^2 = 7.89$; df=6; p = 0.25). However, three of the top four largest clonal groups (group 2: N=32, group 3: N=5 and group 16: N=8) belonged to ATL, while group 9 (N=7) belonged to WAM. While all but one clonal group (group 8 found in Peru and Bolivia) were unique to a single country, ten groups were additionally restricted to a single department/state (Fig 4B,C). The remainder of the clonal groups were identified in either two (groups 4, 6, 7, 8 and 17) or three (groups 2 and 16) departments/states.

When only accounting for the unique genomes (i.e. retaining one genome per clonal group) we found that the Amazonian populations (WAM, CAM) were characterized by i) a strong LD decay ($r^2 < 0.2$ within 10bp; Fig 4D; Supp. Table 16) and ii) distributions of inbreeding coefficients (F_{is}) centered around zero (0.042 ±0.21 for WAM and 0.009 ± 0.21 for CAM) (Supp. Fig 6; Supp. Table 17). In contrast, ATL showed a much slower LD decay ($r^2 < 0.2$ from 37.1kb or 101 kb; Fig 4D; Supp. Table 16) and distributions of F_{is} deviating negatively from zero (F_{is} = -0.17 ± 0.32) (Supp. Fig 6; Supp. Table 17). We also observed significant differences between the three populations in the number ($\chi^2 = 33.05$, df = 2, p = 6.67e-08) and proportion ($\chi^2 = 37.33$, df = 2, p = 7.84e-09) of 'loss of heterozygosity' (LOH) regions across their genomes (Fig 4E,F; Supp. Table 18). Overall, ATL showed a much denser LOH pattern throughout the genome (Supp. Fig 7) with an average of 48 LOH blocks, covering on average 18% of the genome (Supp. Table 19). In contrast, WAM and CAM each harbored on average 26 and 14 LOH blocks encompassing about 5.7% and 4.2% of the genome, respectively. These results suggest that a considerable degree of genetic diversity has been lost in ATL.

Finally, we inferred the effective population size (N_e) of each major *L. braziliensis* L1 population with G-PhoCS (Fig 5) and simulated the change of N_e over past generations with MSMC2 (Fig 6). Estimations of N_e were made for different scenarios of historical migration between the Amazon and Atlantic populations: no migration (i), unidirectional migration from the Amazon to the Atlantic (ii) or the Atlantic to the Amazon (iii), and bi-directional migration (iv) (Fig 5, right panel). All estimations were done in triplicate (i.e. using three different sample subsets) per population and per migration scenario (Supp. Table 20). This revealed strong significant differences in N_e between all pairwise combinations (main effects multi-way ANOVA: F = 96.55 on 8 and 423 df; p < 2.2e-16; adjusted $R^2 = 0.64$) (Tukey's HSD test: Supp. Table 21; Supp. Fig 8). Here, ATL consistently exhibited a significantly lower N_e compared to WAM (factor 1.7), CAM (factor 2.6) and AM (i.e. the ancestral Amazonian population prior to the WAM-CAM divergence) (factor 3.1). This pattern remained consistent across the different migration models and replicate runs (Supp. Table 14; Supp. Fig 8).

Simulations of N_e over time (Fig 6) revealed similar patterns whereby ATL showed lower Ne compared to WAM and CAM for the past 2.74 million generations. Nevertheless, all parasite populations showed a slight decline in Ne for the past 3 million generations until approximately 400,000 generations ago (Fig 6). From then on the N_e seemed to rise again for the three populations until 300,000 to 250,000 generations ago where the N_e of WAM and CAM continued to increase whereas ATL exhibited a second and stronger decline. Calculation of the relative cross-coalescence rates (rCCR) between populations revealed mid-point values (i.e. divergence time estimates; see methods) at around 500,000-300,000 generations ago for the split between the two Amazonian populations (Supp. Fig 9), while this was estimated around 5.2 million to 3.4 million generations ago for the split between the Amazonian and Atlantic populations (Supp. Fig 10).

DISCUSSION

We confirmed that the *L. braziliensis* species complex is genetically highly heterogeneous (8, 9, 11–13, 19, 23), consisting of divergent parasite groups that are associated with the environment and vary greatly in diversity. We described two major, widespread and genetically diverse groups, one associated with the Amazon and the other with Atlantic Forest biomes, and several smaller groups with little diversity showing a restricted geographic and environmental distribution. Parasites of the smaller groups showed stable long-term genetic diversification and their origin was accompanied by a strong population bottleneck, as indicated by a genome-wide loss of heterozygosity and fixation of SNP polymorphisms. Ancestry and F_{ST} estimates suggest that the major admixed Amazonian group may represent the ancestral population from which the other groups emerged, as indicated previously for *L. peruviana* (13). This is consistent with historical, biological and epidemiological data suggesting that *L. braziliensis* and its variants preexisted in Amazonia before spreading to other regions through clonal expansion (27). Our data thus add to a growing body of evidence suggesting the existence of distinct evolutionary and ecological groups of zoonotic *L. braziliensis* parasites in South America (15, 19, 23).

Our main goal was to examine the population diversity and structure of the two major genetically diverse parasite groups that are associated with the Amazon and Atlantic Forest biomes. While both forests were connected as a single forest around 30 thousand years ago (kya) (28), they were separated 20 kya after the last major glaciation (29) by more open savannah-like ecosystems (e.g. Cerrados, Gran Chaco, Caatinga) (30) These may thus represent important barriers to natural gene flow of *L. braziliensis* (22), as has been suggested for lianas, didelphids and anuran trypanosomes (31–33). Our demographic models suggest that the two major *L. braziliensis* populations separated 5.2 to 3.4 million generations ago, which would equate to approximately 742 to 340 kya when assuming 7-10 generations per year (34). The two Amazonian populations diverged much earlier, namely between 300 and 500 thousand generations ago (30 to 71 kya). Our results revealed a decline in N_e since about 2.5 million generations ago (357 kya - 250 kya) and a strong increase, in particular for the Amazonian populations, about 250 thousand generations ago (35 kya - 25 kya). The latter estimate coincides largely with the end of the last major glaciation, which may suggest that subsequent habitat expansions may have promoted a resurgence of this major zoonotic parasite in the Amazon. While these calculations should be considered as rough estimates, they indicate that the history of diversification of *L. braziliensis* is limited to the Pleistocene, an

319 epoch that is characterized by a succession of glacial and interglacial climatic cycles that
320 resulted in changes of suitable habitat for *Leishmania* (13).

321 The two major *L. braziliensis* groups in South-America show vastly different recombination
322 histories. The Amazonian group was characterized by high levels of heterozygosity, low
323 linkage disequilibrium and median inbreeding coefficients approximating zero, as would be
324 predicted for a population experiencing predominantly meiotic recombination. In contrast, the
325 Atlantic group was characterized by a high prevalence of near-identical genomes, a slow
326 decay in linkage disequilibrium, negative median inbreeding coefficients and extensive loss of
327 heterozygosity that likely arose from gene conversion events, as would be predicted for a
328 population experiencing predominant clonal propagation. In addition, the effective population
329 size was at least twice as large in the Amazonian groups compared to the Atlantic group. Our
330 results thus clearly show that these protozoan parasites show a broad spectrum of population
331 structures (13, 22, 35–37). Within this context, we examined the impact of *L. braziliensis*
332 population structure on the frequency and burden of CNVs, which are characteristic to and
333 highly heterogeneous in *Leishmania* (24, 25). Our data revealed that CNV distributions were
334 strongly skewed towards low frequency variants in all populations, suggesting that CNVs are
335 deleterious and subject to strong purifying selection in *L. braziliensis*. We hypothesized that
336 CNVs would be more efficiently purged from the large and stable Amazonian parasite
337 populations than from the smaller and endogamous Atlantic populations, as described in the
338 malaria parasite *Plasmodium falciparum* (38). However, our analysis did not demonstrate that
339 differences in N_e or clonality are sufficient to explain differences in CNV burden and frequency
340 in *L. braziliensis*. This might be because i) differences in population structures are not strong
341 enough to result in differences in purifying selection, ii) *Leishmania* is a predominantly diploid
342 organism (in contrast to *P. falciparum* that has a haploid stage) and/or iii) *Leishmania* can
343 easily change chromosome copy numbers to mitigate the impact of deleterious CNVs (39–
344 41).

345 We argue that the observed demographic differences may be linked to strong eco-
346 epidemiological differences between the two Forest biomes, in particular differences in the
347 type of transmission cycles (23, 42) and forest fragmentation (43). *L. braziliensis* from the
348 Amazon is predominantly circulating in wild animals where human infections appear as
349 spillovers from the sylvatic life cycle, while *L. braziliensis* from the Atlantic is predominantly
350 circulating in synanthropic animals and humans (27). Our observation of high parasite diversity
351 in the Amazon compared to the Atlantic Forest is consistent with other studies where
352 sylvatically transmitted parasite populations were more diverse compared to populations
353 predominated by (peri-)domestic transmission (23, 42, 44, 45). In addition, the Amazon Forest
354 is known as a pristine biome and is the largest contiguous forest in the world. While
355 deforestation in the Amazon poses an extensive threat on the Earth's climate and biodiversity
356 (43, 46), the vast majority of the forest remains contiguous (46, 47). In contrast, the Atlantic
357 Forest is known as a degraded biome as it experienced intense deforestation over the past
358 five centuries (48) and is left highly fragmented along the Atlantic coast (43, 46, 47, 49). Hence,
359 the extensive biodiversity, forest integrity and predominant zoonotic transmission in the
360 Amazon may explain the high diversity of different parasite genotypes sampled in this region,
361 while the genetic uniformity of *L. braziliensis* in the Atlantic may be due to extensive forest
362 fragmentation and predominant synanthropic transmission.

In conclusion, our continent-wide sampling revealed that *L. braziliensis* consists of divergent populations that are associated with the environment and vary greatly in diversity and recombination histories. We argue that these differences are linked to anthropogenic environmental disturbances in the Atlantic Forest that altered the originally sylvatic transmission of *L. braziliensis* towards a predominantly (peri-)domestic or synanthropic transmission, resulting in higher proportions of clonal genotypes. These protozoan parasites thus provide an excellent organism to study a broad spectrum of population structures within a single species, and understand the impact of anthropogenic environmental disturbances on the eco-epidemiology of vector-borne diseases (50).

METHODS

Parasite culturing and DNA sequencing

This study included 257 isolates from different *Leishmania* (*Viannia*) species, mainly *L.* (*V.*) *braziliensis*, sampled between 1975 and 2016, originating from seven South American countries: Argentina (N=11), Bolivia (N=27), Brazil (N=115), Colombia (N=3), Panama (N=2), Peru (N=95), Venezuela (N=2) and two of unknown origin. Parasite isolates were grown in Schneider culture medium until the end of the log phase at the Oswaldo Cruz Institute (Rio de Janeiro, Brazil. DNA was extracted from 10^7 – 10^8 parasites/ml using the QIAmp DNA Mini kit (QIAGEN) following the manufacturer's protocol. Similar to previous work (13), DNA was sheared into 400-600 bp fragments through ultrasonication (Covaris inc.) and amplification-free Illumina libraries were prepared. One hundred 50 base pair paired end reads were generated on the HiSeq ×10 according to the manufacturer's standard sequencing protocol. Sequence data are available at Sequence Read Archive (SRA) BioProject PRJEB4442.

Variant detection

Paired-end sequencing reads were mapped against the M2904 reference genome, a long-read assembly (available at: https://tritrypdb.org/) comprising the 35 autosomal chromosomes (32.73Mb) and the complete sequence of the mitochondrial maxicircle (27.69kb). The mapping of the reads was done using SMALT v0.7.4 (avaliable at: https://www.sanger.ac.uk/science/tools/smalt-0). Here we generated the hash index with words of 13 bp long (k=13) that were sampled at every other position in the genome (s=2). Short variants (SNPs and INDELs) were called using GATK's (v.4.0.2) HaplotypeCaller resulting in genotype VCF (gVCF) files for each parasite isolate. Subsequently, all gVCF files were merged using CombineGVCFs afterwhich joint genotyping of all isolates was performed using GenotypeGVCFs. SelectVariants was used to separate SNPs and INDELS which were separately exposed to hard-filtering thresholds using VariantFiltration to exclude low-quality and false-positive variants. SNPs were excluded when: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, or ReadPosRankSum < −8.0 (51), DP < 5 or when SNPs occurred within SNP clusters (clusterSize = 3 and clusterWindowSize = 10). INDELs were excluded when: QD < 2.0, FS > 200.0, or ReadPosRankSum < −20.0 (51). In addition, we determined which intervals in the genome were accessible for genotyping in each isolate using GATK's CallableLoci (parameters: –minDepth 5 –minBaseQuality 25 – minMappingQuality 25). Finally, we only retained variants that were present in the accessible genome by using the -intersect function of BEDOPS (52).

Ancestry of *Leishmania* (*Viannia*) species and their hybrids

A phylogenetic network (NeighborNet), based on uncorrected p-distances (i.e. the proportion of loci where two sequences differ between each other) of genome-wide, concatenated SNPs, was generated, using the NeighborNet (53) and EqualAngle (54) algorithms implemented in SplitsTree v.4.17.0 (55), to infer phylogenetic relationships within the *Leishmania Viannia* subgenus and to identify putative interspecific hybrids (e.g. long terminal branches, reticulated patterns). Hybrid ancestry was subsequently inferred by phylogenetic analysis of (near-) homozygous stretches, as identified by chromosome specific ARDF (alternate allele read depth frequencies at heterozygous sites) along with PCA-based hybrid-ancestry estimation using PCAdmix v.1.0 (56) using *L. braziliensis* L1, *L. guyanensis/L. panamensis* and *L. shawi* as putative ancestral groups.

Ancestry of the *L. braziliensis* species complex

A phylogenetic network was constructed in a similar way as described above by calculating pairwise uncorrected p-distances based on genome-wide, concatenated, bi-allelic SNPs (683,649 SNPs) using SplitsTree (55). The ecotype structure was further investigated through i) a principal component analysis (PCA) using the 'glPCA' function of the Adegenet (v.2.1.7) R package (57); and ii) a simple model-based ancestry estimation, using ADMIXTURE v.1.3.0 (58), without prior LD pruning. Differences among the number of SNPs (homozygous or heterozygous) among the different sub-lineages were tested by means of pairwise Dunn's tests (FSA v.0.9.4 R-package) (59). A similar comparison was done for comparing the inter-individual pairwise genetic distances, calculated as the Bray-Curtis dissimilarity using the vegan (v.2.6-2) R-package (60), among the different ecotypes. Pairwise F_{st} values between the inferred *L. braziliensis* ecotypes were calculated on a per-site basis over all variable sites using vcftools v.0.1.13 (--weir-fst-pop) (61). Individual genomes with >70% ancestry for a specific ecotype in the K=4 ADMIXTURE model were included in the F_{st} calculations.

Identification of near-identical genomes in *L. braziliensis* L1

Similar to a preceding study (22) groups of potential (near-) identical genomes were identified through branch sharing patterns in the phylogenetic network and low pairwise genetic dissimilarity (<0.02; Bray-Curtis dissimilarity (60)). For each group, fixed SNPs were removed after which counted the number of heterozygous and non-reference homozygous SNPs within each group in a pairwise manner. Near-identical genomes are defined by (near) absence of homozygous SNPs and a relatively low number of heterozygous SNPs (Supp. Table 14).

Population genomic analyses of *L. braziliensis* L1

We constructed a phylogenetic network (NeighborNet) for *L. braziliensis* L1 based on pairwise uncorrected *p*-distances, calculated from genome-wide concatenated SNPs using SplitsTree (55). The population genomic diversity and structure of *L. braziliensis* L1 was examined in greater detail by two model-based clustering methods: i) ADMIXTURE v.1.3.0 (58) ii) fineSTRUCTURE v.4.1.1 (62). ADMIXTURE was run on a LD-pruned SNP panel for K = 1-15 populations with a five-fold cross validation procedure. SNP-pruning was done using plink v.1.9 (63) (--indep-pairwise) by retaining SNPs with an r^2 lower than 0.3 within 50 bp

windows sliding over 10 bp. fineSTRUCTURE was used to infer the genomic ancestries among the individual genomes based on haplotype similarity, generating a co-ancestry matrix. Haplotypes were obtained through computational phasing of the genome-wide SNP genotypes, as was done using BEAGLE v.5.2 (64) (default settings). Inferences with fineSTRUCTURE were done after running the algorithm up to 8e06 MCMC iterations (burn-in: 500,000 iterations) and 2e06 maximization steps (for identifying the best tree building state). The ecological association with the population structure was tested using a chi-squared test for independence using the CrossTable function from the gmodels v.2.18.1.1 R package (65). Pairwise F_{st} estimates between the major parasite groups, as inferred by the K=3 ADMIXTURE model, were calculated in a similar way as between the different *L. braziliensis* ecotypes. In addition, we also investigated the Hardy-Weinberg equilibrium (HWE) by calculating inbreeding coefficients (Eq.1); and LD decay was examined using PopLDdecay (66). To this end, we calculated both F_{is} and LD decay accounting for spatio-temporal Wahlund effects by subsetting individual genomes into groups of individuals close in time (year of isolation < 3 years apart) and space (sample locality in the same department). In addition, the LD decay was corrected for the population sizes (Eq.2) (67). The clonal prevalences of the inferred populations were compared by means of a Chi-squared test using the stats R-package (68).

$$F_{IS} = 1 - \frac{H_O}{H_E} \quad \text{(Eq.1)}$$

$$(H_O: observed\ heterozygosity;\ H_E: expected\ heterozygosity)$$

$$r^2_{corrected} = r^2 - \frac{1}{2N} \quad \text{(Eq.2)}$$

$$(N: population\ size)$$

Finally, we identified Loss-of-Heterozygosity (LOH) regions across the genome as regions in non-overlapping 10kb windows (69), for which the following parameters were used (70): minimum number of SNPs = 1; number of heterozygous SNPs = 0; minimum number of contiguous homozygous 10kb windows = 4; maximum number of 10kb gaps allowed within a LOH region = 1/3 of the windows; and maximum number of heterozygous SNPs allowed in a gap region = 2. Differences in the number and proportions of LOH regions among the inferred populations were tested by means of a Kruskal-Wallis test (stats R-package) (68) along with pairwise Dunn's tests with BH corrected p-values (59).

Estimating effective population size (G-PhoCS)

Effective population sizes (Ne) were estimated using G-PhoCS v.1.3.2 (Generalized Phylogenetic Coalescent Sampler) (71). We estimated N_e per chromosome for four different migration models: i) no migration; ii) unidirectional migration from the Amazon to the Atlantic; iii) unidirectional migration from the Atlantic to the Amazon; and iv) bidirectional migration between the Amazon and Atlantic (Fig 5). Sequence input files were generated based on VCF and BED files per chromosome after excluding SNPs with a MAF < 0.05 and all SNPs present in CDS regions. The chromosomal VCF and BED files were then converted into the G-PhoCS input format using the 'vcf_to_gphocs.py' script from the Popgen Pipeline Platform (available at: https://github.com/jaredgk/PPP/blob/master/pgpipe/). As G-PhoCS only allows for a small number of individuals per population, we selected three subsets of five isolates per population (WAM, CAM, ATL) to include in the analyses (Supp. Table 22). Each G-PhoCS analysis was run over 2,500,000 MCMC iterations (excl. burn-in) sampling every 1000 steps and with an

initial burn-in of 500,000 iterations. Additional information on the priors of the G-PhoCS analyses is available in Supp. Table 22. The following priors were used: tau-theta-alpha = 1; tau-theta-beta = 20000; mig-rate-alpha = 0.02; mig-rate-beta = 0.0001; locus-mut-rate = CONST; find-finetunes = TRUE; find-finetunes-num-steps = 100; find-finetunes-samples-per-step = 100; tau-initial$_{WAM-CAM}$ = 0.0005; tau-initial$_{AM-ATL}$=0.001. Convergence of all theta estimates was assessed by examining their effective sample sizes (ESS) using the Tracerer v.2.2.3 R package (72). We only included G-PhoCS runs where all theta values reached convergence (i.e. ESS > 200). The posterior distributions of the population size estimates were converted into effective population sizes, using $\theta = 4N_e\mu$, assuming the genome-wide mutation rate (μ) of *Leishmania* spp. is 1.99e-09 per bp per generation (36). Following Campagna et al. (2015) (73), we limited the interpretations of the N_e estimates to relative differences to rule out potential biases of the assumed mutation rate on the absolute values. Estimates of N_e were compared between parasite populations, as inferred by ADMIXTURE (K=3) and fineSTRUCTURE, by means of a main effects multi-way ANOVA, accounting for the different sample subsets and migration models using the stats R-package (68). Post hoc pairwise comparisons between populations, migration models and sample subsets were performed using the Tukey's HSD (Honest Significant Difference) method (stats R-package (68)).

Estimating effective population size through time (MSMC2)

Inference of N_e through time was performed using MSMC2 (Multiple Sequentially Markovian Coalescent) (74) and auxiliary scripts from the msmc-tools repository ((available at: https://github.com/stschiff/msmc-tools). A mappability mask from the M2904 reference genome was generated using the code from SNPable (available at: http://lh3lh3.users.sourceforge.net/snpable.shtml) and the makeMappabilityMask.py script (msmc-tools). All SNPs were phased using BEAGLE v.5.2 (64) and separated per chromosome per individual (only for a subset of individuals; Supp. Table 22). Chromosome and individual specific mask files were generated using the vcfAllSiteParser.py script (msmc-tools) afterwhich MSMC2 input files were generated using the generate_multihetsep.py script (msmc-tools). Effective population sizes for each *L. braziliensis* population were inferred, in triplicate (Supp Table 23), by running MSMC2 with 500 iterations (-i) and 1*2+21*1+1*2 as time segmentation pattern (-p). The coalescence rate estimates from MSMC2 were scaled to effective population size values (Eq.3). The inferred time segments from the MSMC2 output were rescaled to numbers of generations (Eq.4). Finally, to get an idea of when populations have diverged from each other we calculated the relative cross-coalescence rate (rCCR) between WAM and CAM, and between WAM+CAM and ATL. This was achieved by running additional MSMC2 runs for cross population coalescence rate (λ) estimates and subsequently running combineCrossCoal.py (msmc-tools) with the msmc2 outputs of the cross-population analysis as well as the two separate populations as input files. The rCCR was then calculated based on the two within population coalescence rates and the across population coalescence rate (Eq.5). The rCCR ranges between 0 and 1 where a value of 1 points towards the point when both populations probably coalesced into one population while a value of 0 indicates the point when both populations are assumed to be fully separated. The midpoint (rCCR0.5) can be seen as an estimate for when both populations have sufficiently diverged from each other to consider them as separate populations (74).

$$N_e = \frac{\frac{1}{\lambda}}{2\mu} \qquad \text{(Eq.3)}$$

$$(N_e: effective\ population\ size; \lambda: coalescence\ rate;\ \mu: mutation\ rate)$$

$$g = \frac{t}{\mu} \qquad \text{(Eq.4)}$$

$$(g: number\ of\ generations\ ago; t: time\ segments;\ \mu: mutation\ rate)$$

$$rCCR = \frac{2\lambda_{POP1-2}}{\lambda_{POP1} + \lambda_{POP2}} \qquad \text{(Eq.5)}$$

$$(rCCR: relative\ cross-coalescence\ rate;\ \lambda: coalescence\ rate)$$

Variant annotation and estimation of chromosome and gene copy number variation in *L. braziliensis* L1

Chromosome and gene copy number variations (CNV) were estimated based on the per-site read depths as obtained with SAMtools depth (-a option). Chromosomal somy variation was estimated assuming diploidy by multiplying the haploid copy number (HCN) by two. Here the HCN was calculated as the division of the median chromosomal read depth over the genome-wide read depth. Differences in chromosomal copy numbers were tested using the Wilks' lambda test (MANOVA) using the Vegan (v.2.6-2) R-package (60). In parallel, gene HCNs were calculated by dividing the median read depth per gene, as per coding DNA sequence (CDS), over the genome-wide median read depth. We defined gene CNVs as an increase (z-score > 3; amplification) or decrease (z-score < -3; deletion) in HCN by subtracting the sample specific HCN over the genome-wide median HCN. Subsequently, the difference in the number of CNVs and the CNV burden across the genome was assessed between the three L1 populations population by means of a Kruskal-Wallis test (stats R-package (68)) along with pairwise Dunn's tests with BH corrected p-values (59) and through survival analyses using the Survival (v.3.3-1) (75) and Survminer (v.0.4.9) (76) R-packages, respectively. The potential difference in HCN of the CNVs that were common in all three populations (i.e. occurring in more than 90% of each population) was tested by means of a one-way ANOVA and subsequent Tukey's HSD post-hoc comparison with p-value correction using the BH method. Prior to the ANOVA, several CDS regions were identified as outliers by the Bonferroni outlier test from the car package in R (v.3.1-1) (77). These CDS regions consistently belonged to isolates of clonal group 3 (Supp. Table 1), showing extreme high amplifications. For the purpose of comparing the mean HCN per population in these highly frequent amplifications, we discarded these individuals. In addition, a PCA was performed on the HCN of all CNVs using the 'prcomp' function of the stats R-package (68) Variants were annotated, based on the *L. braziliensis* M2904 annotation file as reference database, using SNPEFF v.5.2 (78) with default parameters.

FUNDING

This work received financial support from the Directie-Generaal Ontwikkelingssamenwerking en Humanitaire Hulp (DGD) (Belgian cooperation and was funded in part by the Wellcome Trust Grant [206194]. F.V.D.B. and S.H. acknowledge support from the Research Foundation Flanders (Grants 1226120N and G092921N). E.C. acknowledges support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Finance Code 001; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Research Fellow, 302622/2017-9; Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, CNE, E26-202.569/2019 and E_ 08/2020 -COLBIO-210.285/2021; GGBN Award: GGI-GGBN-2021-278. L.M.C. acknowledges support from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Pós-Doutorado Nota 10 E-26/205.730/2022 and 205.731/2022.

REFERENCES

1. R. Reithinger, *et al.*, Cutaneous leishmaniasis. *Lancet Infect. Dis.* **7**, 581–596 (2007).

2. H. Nuwangi, *et al.*, Stigma associated with cutaneous and mucocutaneous leishmaniasis: A systematic review. *PLoS Negl. Trop. Dis.* **17**, e0011818 (2023).

3. J. Alvar, *et al.*, Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **7**, e35671 (2012).

4. A. L. R. Roque, A. M. Jansen, Wild and synanthropic reservoirs of Leishmania species in the Americas. *Int. J. Parasitol. Parasites Wildl.* **3**, 251–262 (2014).

5. J. F. Marinho-Júnior, *et al.*, High levels of infectiousness of asymptomatic Leishmania (Viannia) braziliensis infections in wild rodents highlights their importance in the epidemiology of American Tegumentary Leishmaniasis in Brazil. *PLoS Negl. Trop. Dis.* **17**, e0010996 (2023).

6. T. M. Schubach, *et al.*, American cutaneous leishmaniasis in two cats from Rio de Janeiro, Brazil: first report of natural infection with Leishmania (Viannia) braziliensis. *Trans. R. Soc. Trop. Med. Hyg.* **98**, 165–167 (2004).

7. E. L. Yoshida, *et al.*, Human, canine and equine (Equus caballus) leishmaniasis due to Leishmania braziliensis (= L. braziliensis braziliensis) in the south-west region of São Paulo State, Brazil. *Mem. Inst. Oswaldo Cruz* **85**, 133–134 (1990).

8. S. Odiwuor, *et al.*, Evolution of the Leishmania braziliensis species complex from amplified fragment length polymorphisms, and clinical implications. *Infect. Genet. Evol.* **12**, 1994–2002 (2012).

9. G. Van der Auwera, *et al.*, Evaluation of four single-locus markers for Leishmania species discrimination by sequencing. *J. Clin. Microbiol.* **52**, 1098–1104 (2014).

10. S. Li, R. Jovelin, T. Yoshiga, R. Tanaka, A. D. Cutter, Specialist versus generalist life histories and nucleotide diversity in Caenorhabditis nematodes. *Proc. Biol. Sci.* **281**, 20132858 (2014).

11. B. S L Figueiredo de Sá, A. M. Rezende, O. P. de Melo Neto, M. E. F. de Brito, S. P. Brandão Filho, Identification of divergent Leishmania (Viannia) braziliensis ecotypes derived from a geographically restricted area through whole genome analysis. *PLoS Negl. Trop. Dis.* **13**, e0007382 (2019).

12. A. F. Brilhante, *et al.*, Leishmania (Viannia) braziliensis type 2 as probable etiological agent of canine cutaneous leishmaniasis in Brazilian Amazon. *PLoS One* **14**, e0216291 (2019).

13. F. Van den Broeck, *et al.*, Ecological divergence and hybridization of Neotropical Leishmania parasites. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25159–25168 (2020).

14. L. H. Patino, M. Muñoz, L. Cruz-Saavedra, C. Muskus, J. D. Ramírez, Genomic Diversification, Structural Plasticity, and Hybridization in Leishmania (Viannia) braziliensis. *Front. Cell. Infect. Microbiol.* **10**, 582192 (2020).

15. K. Kuhls, *et al.*, Population structure and evidence for both clonality and recombination among Brazilian strains of the subgenus Leishmania (Viannia). *PLoS Negl. Trop. Dis.* **7**, e2490 (2013).

16. G. Herrera, *et al.*, An interactive database of Leishmania species distribution in the Americas. *Sci Data* **7**, 110 (2020).

17. M. Maroli, M. D. Feliciangeli, L. Bichaud, R. N. Charrel, L. Gradoni, Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Med. Vet. Entomol.* **27**, 123–147 (2013).

18. M. Akhoundi, *et al.*, A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLoS Negl. Trop. Dis.* **10**, e0004349 (2016).

19. F. Van den Broeck, *et al.*, Genome Analysis of Triploid Hybrid Leishmania Parasite from the Neotropics. *Emerg. Infect. Dis.* **29**, 1076–1078 (2023).

20. M. B. De los Santos, I. M. Ramírez, J. E. Rodríguez, P. Beerli, H. O. Valdivia, Genetic diversity and population structure of Leishmania (Viannia) braziliensis in the Peruvian jungle. *PLoS Negl. Trop. Dis.* **16**, e0010374 (2022).

21. H. O. Valdivia, *et al.*, Comparative genomic analysis of Leishmania (Viannia) peruviana and Leishmania (Viannia) braziliensis. *BMC Genomics* **16**, 715 (2015).

22. S. Heeren, *et al.*, Diversity and dissemination of viruses in pathogenic protozoa. *Nat. Commun.* **14**, 8343 (2023).

23. E. Cupolillo, *et al.*, Genetic polymorphism and molecular epidemiology of Leishmania (Viannia) braziliensis from different hosts and geographic areas in Brazil. *J. Clin. Microbiol.* **41**, 3126–3132 (2003).

639 24. M. B. Rogers, *et al.*, Chromosome and gene copy number variation allow major
640 structural change between species and strains of Leishmania. *Genome Res.* **21**, 2129–2142
641 (2011).

642 25. H. Imamura, *et al.*, Evolutionary genomics of epidemic visceral leishmaniasis in the
643 Indian subcontinent. *Elife* **5** (2016).

644 26. S. U. Franssen, *et al.*, Global genome diversity of the Leishmania donovani complex.
645 *Elife* **9** (2020).

646 27. M. C. de A. Marzochi, *et al.*, Anthropogenic Dispersal of Leishmania (Viannia)
647 braziliensis in the Americas: A Plausible Hypothesis. *Frontiers in Tropical Diseases* **2**, 21
648 (2021).

649 28. R. M. D. Ledo, G. R. Colli, The historical connections between the Amazon and the
650 Atlantic Forest revisited. *J. Biogeogr.* **44**, 2551–2563 (2017).

651 29. J. Flenley, M. Bush, *Tropical Rainforest Responses to Climatic Change* (Springer
652 Science & Business Media, 2007).

653 30. E. H. Bucher, Chaco and Caatinga — South American Arid Savannas, Woodlands and
654 Thickets in *Ecology of Tropical Savannas*, (Springer Berlin Heidelberg, 1982), pp. 48–79.

655 31. V. A. Thode, I. Sanmartín, L. G. Lohmann, Contrasting patterns of diversification
656 between Amazonian and Atlantic forest clades of Neotropical lianas (Amphilophium,
657 Bignonieae) inferred from plastid genomic data. *Mol. Phylogenet. Evol.* **133**, 92–106 (2019).

658 32. A. Fabrício Machado, *et al.*, Recent past connections between Amazonian and Atlantic
659 forests by comparative phylogeography and paleodistribution models for didelphid mammals.
660 *Evol. Ecol.* (2024). https://doi.org/10.1007/s10682-024-10292-6.

661 33. R. C. Ferreira, *et al.*, Morphological and molecular diversity and phylogenetic
662 relationships among anuran trypanosomes from the Amazonia, Atlantic Forest and Pantanal
663 biomes in Brazil. *Parasitology* **134**, 1623–1638 (2007).

664 34. M. Koffi, *et al.*, Population genetics of Trypanosoma brucei gambiense, the agent of
665 sleeping sickness in Western Africa. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 209–214 (2009).

666 35. J. D. Ramírez, M. S. Llewellyn, Reproductive clonality in protozoan pathogens-truth or
667 artefact? *Mol. Ecol.* **23**, 4195–4202 (2014).

668 36. M. B. Rogers, *et al.*, Genomic confirmation of hybridisation and recent inbreeding in a
669 vector-isolated Leishmania population. *PLoS Genet.* **10**, e1004092 (2014).

670 37. E. Inbar, *et al.*, Whole genome sequencing of experimental hybrids supports meiosis-
671 like sexual recombination in Leishmania. *PLoS Genet.* **15**, 1–28 (2019).

672 38. I. H. Cheeseman, *et al.*, Population Structure Shapes Copy Number Variation in
673 Malaria Parasites. *Mol. Biol. Evol.* **33**, 603–620 (2016).

674 39. Y. Sterkers, *et al.*, Novel insights into genome plasticity in Eukaryotes: mosaic
675 aneuploidy in Leishmania. *Mol. Microbiol.* **86**, 15–23 (2012).

676 40. F. Dumetz, *et al.*, Modulation of Aneuploidy in Leishmania donovani during Adaptation
677 to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *MBio* **8**
678 (2017).

679 41. G. H. Negreira, *et al.*, High throughput single-cell genome sequencing gives insights
680 into the generation and evolution of mosaic aneuploidy in Leishmania donovani. *Nucleic Acids*
681 *Res.* **50**, 293–305 (2022).

682 42. E. Cupolillo, H. Momen, G. Grimaldi Jr, Genetic diversity in natural populations of New
683 World Leishmania. *Mem. Inst. Oswaldo Cruz* **93**, 663–668 (1998).

684 43. M. C. Hansen, *et al.*, High-resolution global maps of 21st-century forest cover change.
685 *Science* **342**, 850–853 (2013).

686 44. M. E. F. Brito, *et al.*, Species diversity of Leishmania (Viannia) parasites circulating in
687 an endemic area for cutaneous leishmaniasis located in the Atlantic rainforest region of
688 northeastern Brazil. *Trop. Med. Int. Health* **14**, 1278–1286 (2009).

689 45. S. P. Brandão-Filho, *et al.*, Wild and synanthropic hosts of Leishmania (Viannia)
690 braziliensis in the endemic cutaneous leishmaniasis locality of Amaraji, Pernambuco State,
691 Brazil. *Trans. R. Soc. Trop. Med. Hyg.* **97**, 291–296 (2003).

692 46. N. M. Haddad, *et al.*, Habitat fragmentation and its lasting impact on Earth's
693 ecosystems. *Sci Adv* **1**, e1500052 (2015).

694 47. H. S. Grantham, *et al.*, Anthropogenic modification of forests means only 40% of
695 remaining forests have high ecosystem integrity. *Nat. Commun.* **11**, 5978 (2020).

696 48. A. M. G. Teixeira, B. S. Soares-Filho, S. R. Freitas, J. P. Metzger, Modeling landscape
697 dynamics in an Atlantic Rainforest region: Implications for conservation. *For. Ecol. Manage.*
698 **257**, 1219–1230 (2009).

699 49. M. C. Ribeiro, J. P. Metzger, A. C. Martensen, F. J. Ponzoni, M. M. Hirota, The Brazilian
700 Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for
701 conservation. *Biol. Conserv.* **142**, 1141–1153 (2009).

702 50. W. M. de Souza, S. C. Weaver, Effects of climate change and human activities on
703 vector-borne diseases. *Nat. Rev. Microbiol.* (2024). https://doi.org/10.1038/s41579-024-
704 01026-0.

705 51. G. A. Van der Auwera, *et al.*, From FastQ data to high confidence variant calls: the
706 Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–
707 11.10.33 (2013).

708 52. S. Neph, A. P. Reynolds, M. S. Kuehn, J. A. Stamatoyannopoulos, Operating on
709 Genomic Ranges Using BEDOPS. *Methods Mol. Biol.* **1418**, 267–281 (2016).

53. D. Bryant, V. Moulton, Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265 (2004).

54. A. W. M. Dress, D. H. Huson, Constructing splits graphs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1**, 109–115 (2004).

55. D. H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

56. A. Brisbin, *et al.*, PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* **84**, 343–364 (2012).

57. T. Jombart, adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

58. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

59. D. H. Ogle, J. C. Doll, A. P. Wheeler, A. Dinno, FSA: Simple fisheries stock assessment methods. R package version 0.9. 4. [Preprint] (2023).

60. J. Oksanen, *et al.*, *Vegan Community Ecology Package Version 2.5-7* (2020).

61. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

62. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).

63. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

64. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).

65. G. R. Warnes, *et al.*, gmodels: Various R Programming Tools for Model Fitting. [Preprint] (2022). Available at: https://CRAN.R-project.org/package=gmodels.

66. C. Zhang, S.-S. Dong, J.-Y. Xu, W.-M. He, T.-L. Yang, PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).

67. D. Shin, S.-H. Kim, J. Park, H.-K. Lee, K.-D. Song, Extent of linkage disequilibrium and effective population size of the Landrace population in Korea. *Asian-australas. J. Anim. Sci.* **31**, 1078–1087 (2018).

68. R Core Team, R: A language and environment for statistical computing. [Preprint] (2021). Available at: https://www.R-project.org/.

744 69. W. Weir, *et al.*, Population genomics reveals the origin and asexual evolution of human
745 infective trypanosomes. *Elife* **5**, e11473 (2016).

746 70. A. Hadermann, *et al.*, Genome diversity of Leishmania aethiopica. *Front. Cell. Infect.*
747 *Microbiol.* **13** (2023).

748 71. I. Gronau, M. J. Hubisz, B. Gulko, C. G. Danko, A. Siepel, Bayesian inference of
749 ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034
750 (2011).

751 72. R. J. C. Bilderbeek, R. S. Etienne, babette: BEAUti 2, BEAST2 and Tracer for R.
752 *bioRxiv* 271866 (2018).

753 73. L. Campagna, I. Gronau, L. F. Silveira, A. Siepel, I. J. Lovette, Distinguishing noise
754 from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Mol.*
755 *Ecol.* **24**, 4238–4251 (2015).

756 74. S. Schiffels, K. Wang, MSMC and MSMC2: The Multiple Sequentially Markovian
757 Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).

758 75. T. Therneau, A Package for Survival Analysis in R_. R package version 3.3-1. 2022.
759 [Preprint] (2022). Available at: https://CRAN.R-project.org/package=survival.

760 76. A. Kassambara, M. Kosinski, P. Biecek, S. Fabian, survminer: Drawing Survival
761 Curves using'ggplot2'. R package version 0.4. 9. 2021. [Preprint] (2021).

762 77. J. Fox, S. Weisberg, An R Companion to Applied Regression. [Preprint] (2019).
763 Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

764 78. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single
765 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain
766 w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

767 79. G. E. Griffith, J. M. Omernik, S. H. Azevedo, Ecological classification of the Western
768 Hemisphere. *Unpublished Report* 49 (1998).

769

TABLES

Table 1: Main characteristics of the identified *L. braziliensis* groups. *Ecoregion classification and data was extracted from Griffith et al. (1998) (89). The level 1 ecoregion 'Eastern Highland' encompasses the Atlantic Forest region (a level 2 ecoregion).

Group	No. isolates	Sampled countries	Sampled ecoregions (level 1)*	Median No. SNPs [min - max]	Median No. heterozygous SNPs	Median No. homozygous SNPs	Refs
L1	182	Argentina, Bolivia, Brazil, Peru	Eastern Highland**, Amazonian-Orinocan Lowland, Northern/Central Andes, Gran Chaco	30,158 [7,385 - 31,892]	13,766	15,654	(13, 14, 22, 23) (13, 14, 22, 23)
L2	4	Bolivia, Brazil, Peru	Amazonian-Orinocan Lowland	60,095 [59,879 - 60,181]	4,406	55,773	(8, 9, 12)
L3	9	Brazil	Eastern highlands	25,620 [25,596 - 25,631]	113	25,509	(11)
L. peruviana	31	Peru	Northern/Central Andes	26,024 [25,586 - 26,482]	98	25,664	(13)

774 **FIGURES**

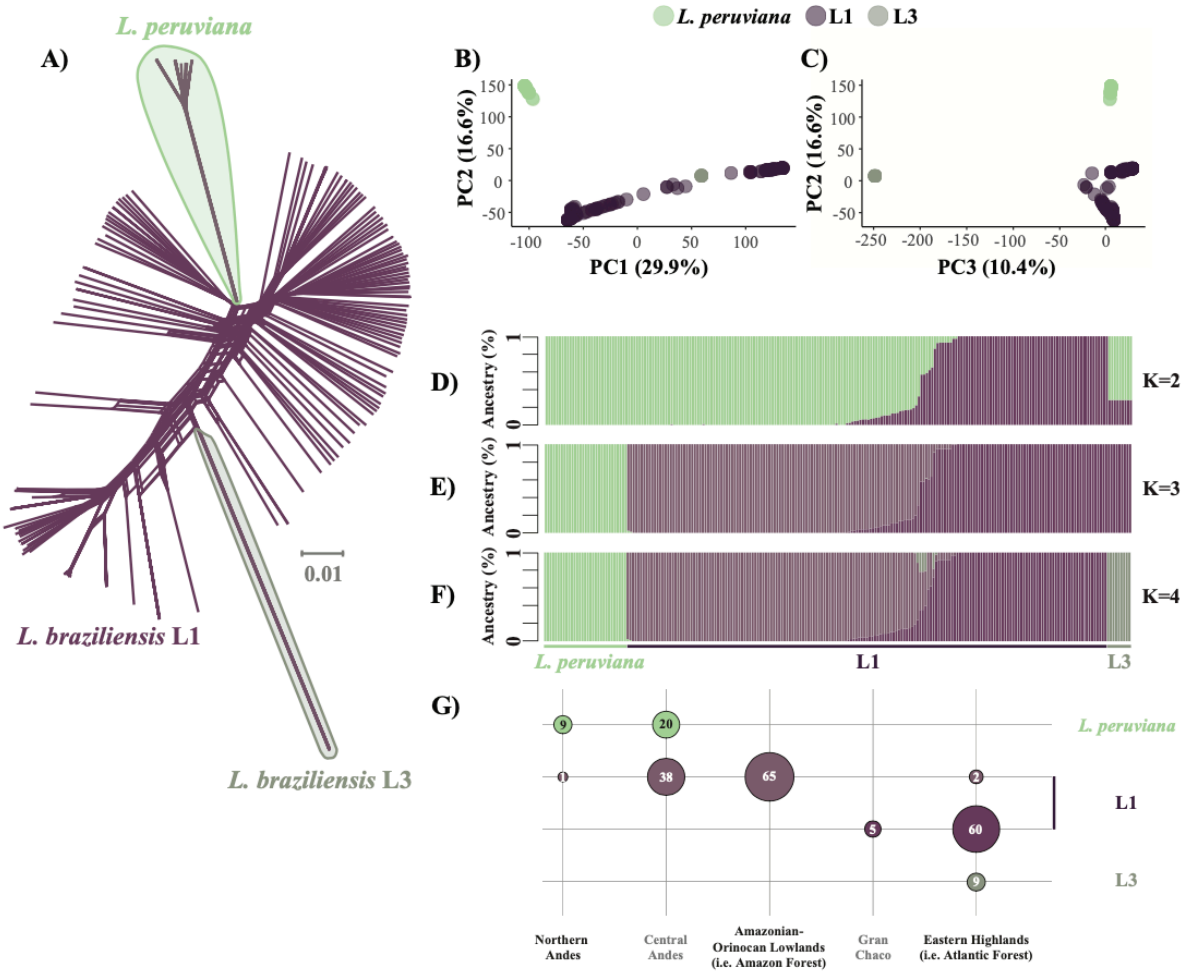
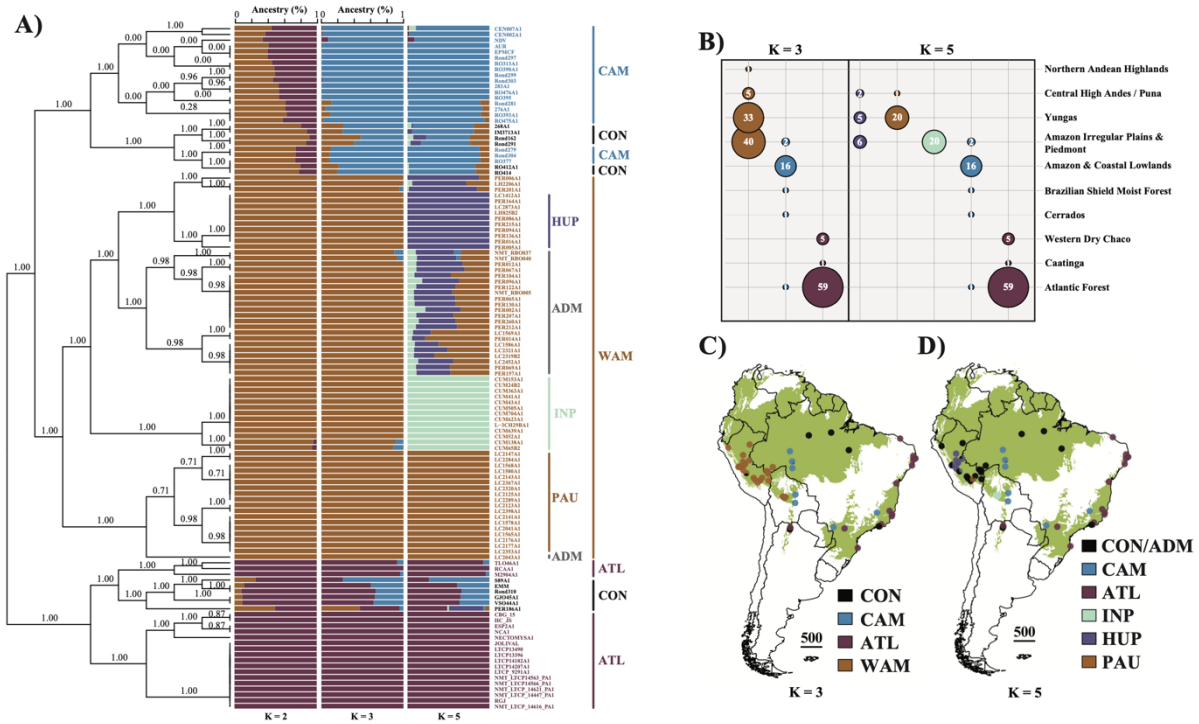
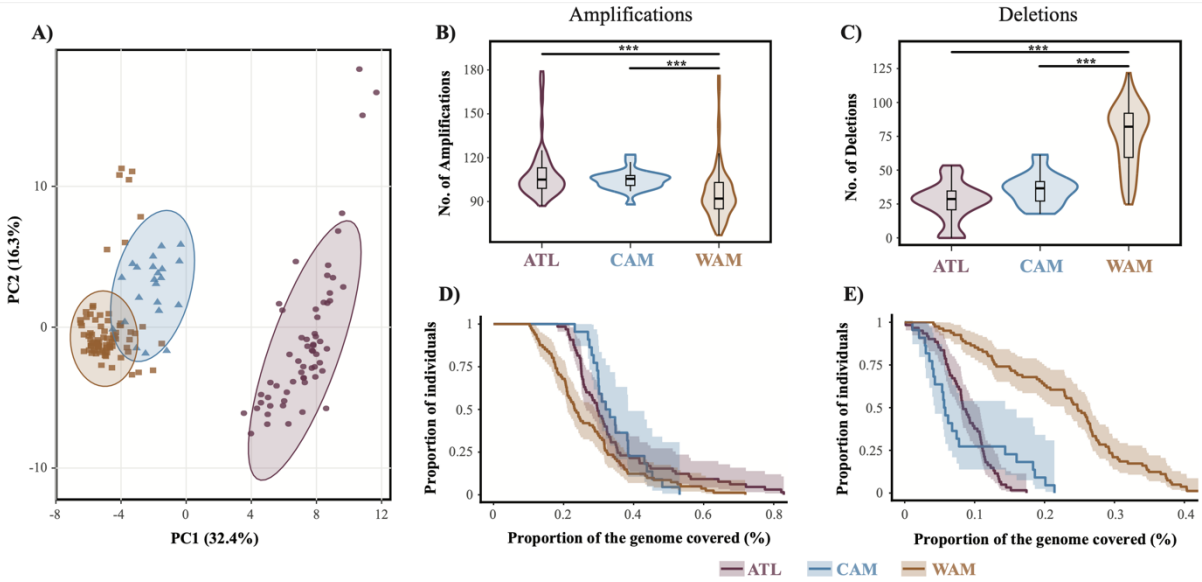


Fig 1. Divergence within the *L. braziliensis* species complex. A) A phylogenetic network, based on 695,229 genome-wide SNPs, showing uncorrected p-distances between 222 isolates of the *L. braziliensis* species complex (incl. L1, L3 and *L. peruviana*). **B-C)** Principal component analysis for the 222 isolates showing the first three PC axes. **D-F)** ADMIXTURE barplots showing the estimated ancestry per isolate assuming K=2 (**D**), K=3 (**E**) and K=4 (**F**) ancestral components. **G)** Sample size distribution of *Leishmania* isolates from each group and per ecoregion. The four colors match the four ancestral components as inferred with ADMIXTURE K=4 (**F**). Only isolates with at least 70% ancestry for a specific ancestral component were included.



775 **Fig 2. Population genomic structure of *L. braziliensis* L1. A)** ADMIXTURE barplots depicting the ancestry per isolate (N_{unique} = 116) assuming K=2, K=3 and K=5 ancestral components. Isolates are labeled according to K=3 ancestral components. Black is used for isolates with uncertain/hybrid ancestry (CON). Outer vertical lines show the major parasite groups (WAM, CAM, ATL, CON) delineated by ADMIXTURE for K=3. Inner vertical lines represent the parasite groups within WAM as inferred by ADMIXTURE for K=5 in this study, which is in accordance with Heeren et al. (2023) (22) (PAU, HUP, INP, ADM). The left bound tree represents the population tree of L1 as inferred by fineSTRUCTURE. Branch support values represent the posterior probability for each inferred clade. **B)** Sample size distribution per ancestral component per ecoregion (level 2; (89)) for all isolates with at least 85% ancestry to a specific group/population. **C-D)** Map of the South American continent showing the L1 population genomic structure, assuming K=3 (C) and K=5 (D) populations. The base map depicts the occurrence of (Sub-) Tropical Moist Broadleaf Forests. Abbreviations: CAM = Central Amazon; WAM = West Amazon; ATL = Atlantic; CON = Conglomerate; PAU = Southern Peru; HUP = Central/Northern Peru; INP = Central Bolivia; ADM = Admixed.



776 **Fig 3. Copy number variations across the major *L. braziliensis* populations. A)** Scatterplot showing the first two principal components as calculated based on haploid copy numbers of all CNVs, after removing isolates of clonal group 3 (see methods). Ellipses represent the 95% confidence boundaries of the major parasite populations in the PCA-space. **B-C)** Violin plots summarizing the number of CNVs per parasite genome. **D-E)** Survival curves depicting the CNV burden per *L. braziliensis* population.

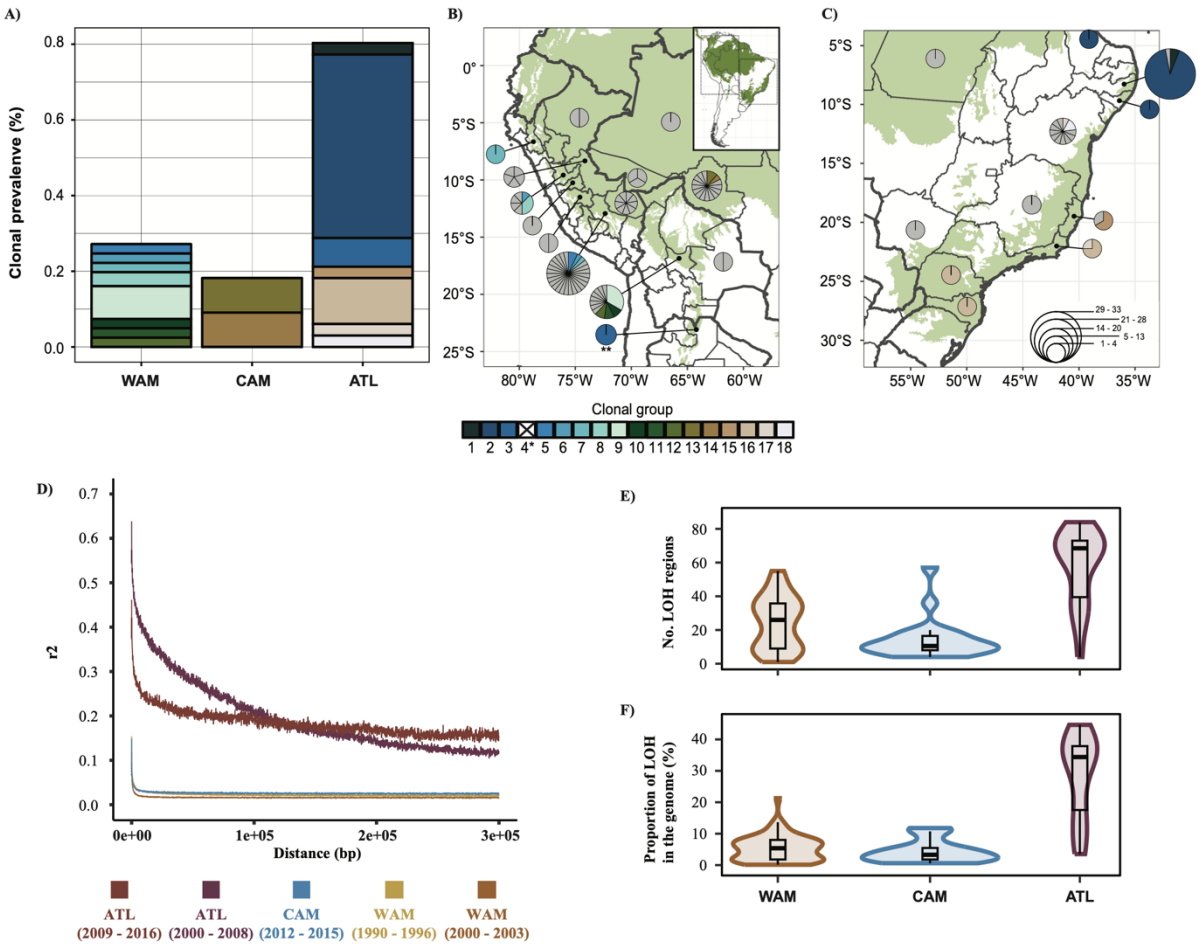


Fig 4. Contrasting clonality and population structure in *L. braziliensis* L1. A) Clonal prevalence per population. **B-C)** Distribution of genotypes in the Amazon and along the Atlantic coast, summarized per department/state of the respective country. The size of each pie indicates the number of genotypes found in each locality with each segment representing a unique genotype. Colored segments indicate the different clonal groups that were identified. Note: * clonal group 4 is not included as it consists of two isolates of the CON group; ** clonal group 3, located in Salta, Argentina belongs to ATL. **D)** Linkage disequilibrium decay of the different *L. braziliensis* populations, accounting for spatio-temporal Wahlund effects and population size. **E)** The number of loss-of-heterozygosity (LOH) regions per major population. **F)** Proportion of LOH regions across the genome per major population.

777

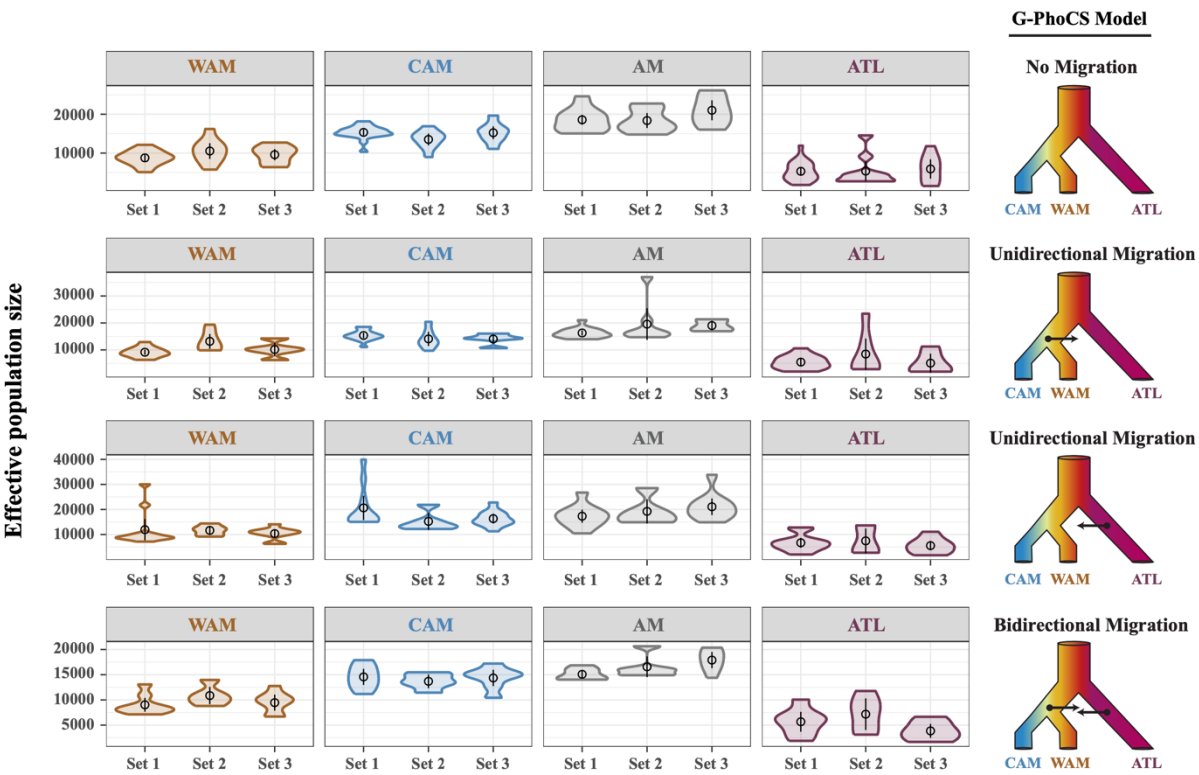


Fig 5. Estimated effective population sizes (Ne) per population for four possible migration scenarios. Each row depicts the Ne estimates per population for a given model of historical migration. WAM, CAM and ATL represent the three major populations as inferred by ADMIXTURE and fineSTRUCTURE (Fig 2A). AM represents the ancestral population prior to the split of WAM and CAM. Four models of historical migration were tested: i) no migration, ii) unidirectional migration from AM to ATL, iii) unidirectional migration from ATL to AM and iv) bidirectional migration between AM and ATL.

778

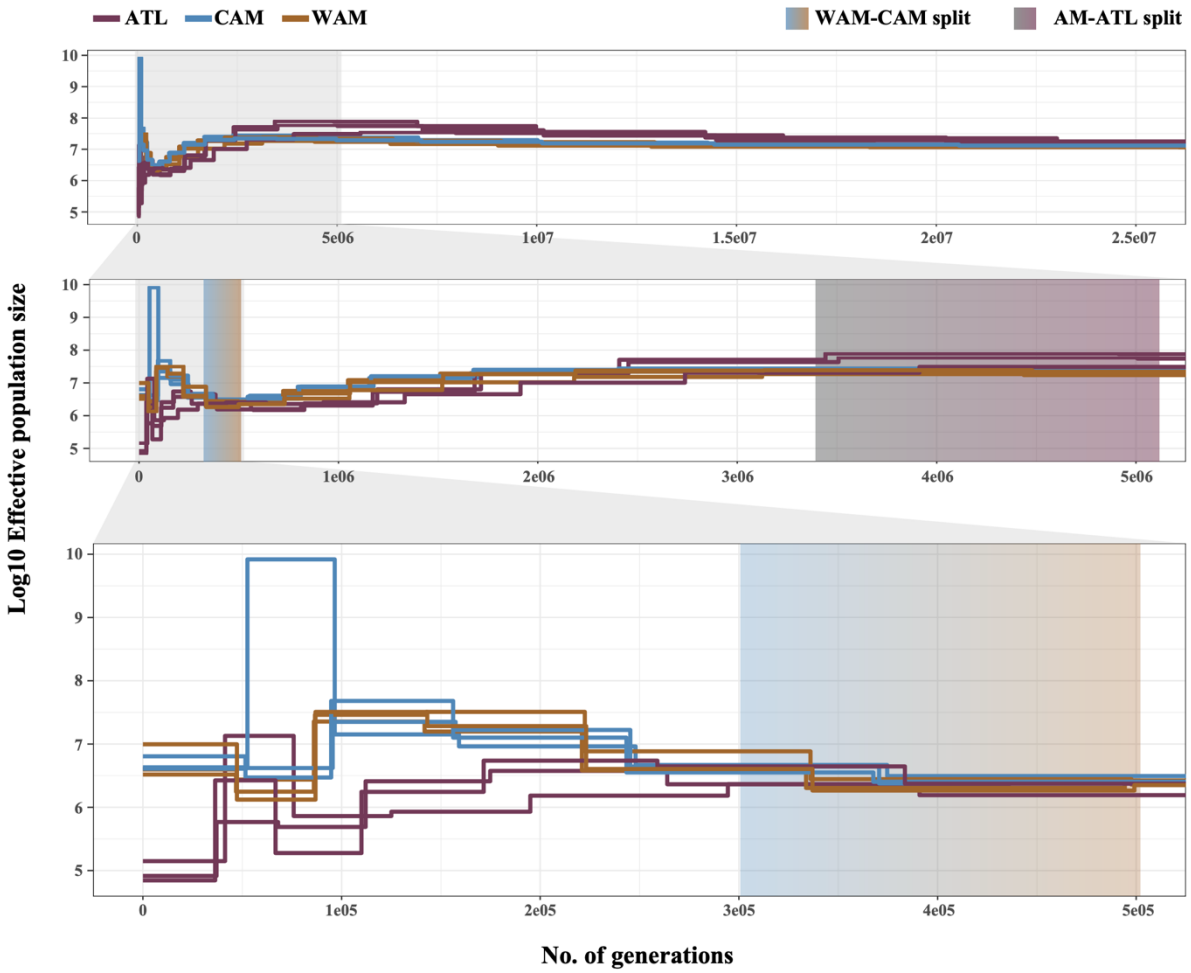


Fig 6. Simulated changes in Ne per population through time (in units of generations ago). Simulations were performed in triplicate; on the same sample subsets per population as figure 5. Gradient boxes depict the estimated time of the first population split (rCCR ≈ 0.5) within the past 25 million generations, between WAM-CAM and AM-ATL based on the relative cross-coalescence rate (Supp. Fig 9,10).

779

Supplementary information for:

Evolutionary genomics of a zoonotic parasite circulating across the Neotropical realm

Senne Heeren*, Mandy Sanders, Jeffrey Jon Shaw, Sinval Pinto Brandão-Filho, Mariana Côrtes Boité, Lilian Motta Cantanhêde, Khaled Chourabi, Ilse Maes, Alejandro Llanos-Cuentas, Jorge Arevalo, Jorge D. Marco, Philippe Lemey, James A. Cotton, Jean-Claude Dujardin, Elisa Cupolillo*, Frederik Van den Broeck*

***Correspondence:**
Senne Heeren (sheeren@itg.be)
Frederik Van den Broeck (fvandenbroeck@gmail.com)
Elisa Cupolillo (elisa.cupolillo@ioc.fiocruz.br

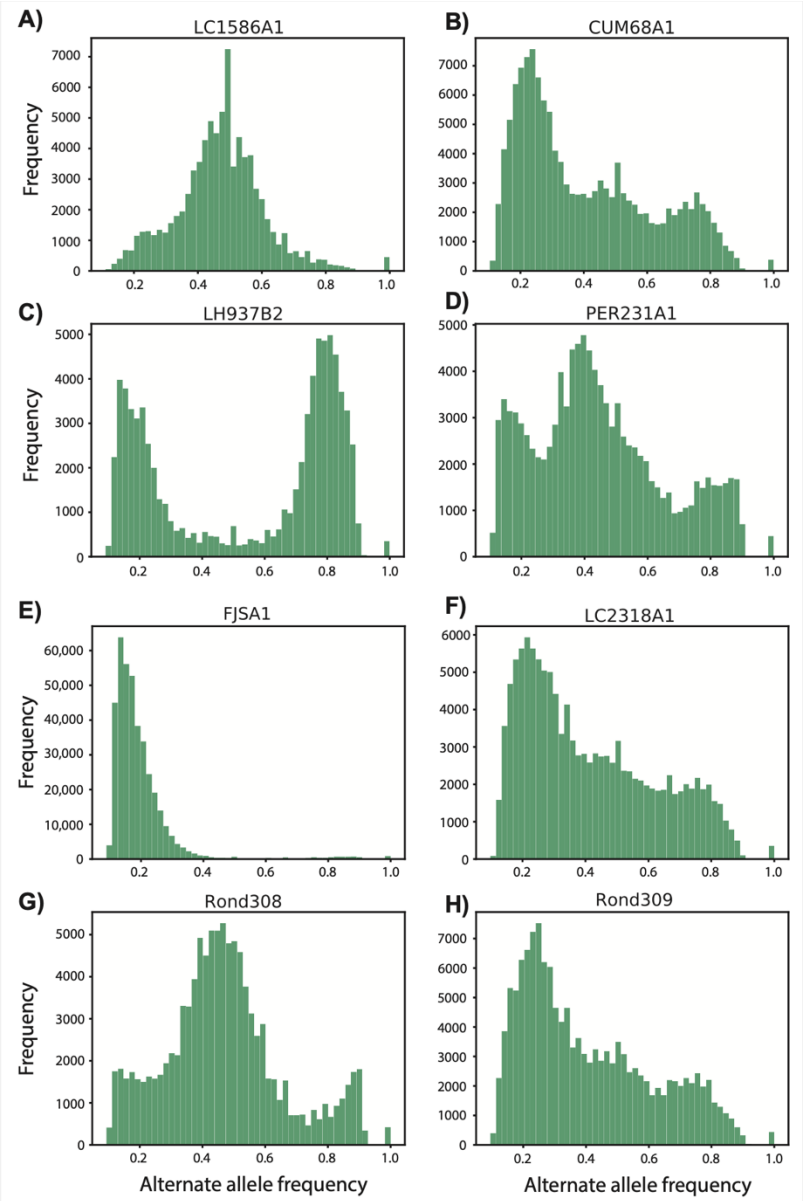
SUPPLEMENTARY RESULTS

Genomic evidence of interspecific hybridization in *Leishmania* (*Viannia*) parasites

Genotyping across the combined accessible genome (25.5 Mb, or 77.7% of the genome) of the 244 genomes disclosed a total of 834,178 bi-allelic Single Nucleotide Polymorphisms (SNPs) called against the reference. Phylogenetic network analyses revealed a similar topological relationship among the major *L. Viannia* species as disclosed earlier with reduced marker sets (1–3). *Leishmania lainsoni* was phylogenetically the most distant species to *L. braziliensis* L1 (Supp. Fig 2B), with an average 290,660 homozygous SNPs called against the *L. braziliensis* M2904 reference, followed by *L. naiffi* (average 205,814 homozygous SNPs), the *L. guyanensis* species complex (average 100,520 homozygous SNPs) and *L. braziliensis* L2 (average 55,709 homozygous SNPs) (Supp. Fig 2B). Two divergent subgroups of the *L. braziliensis* species complex, *L. peruviana* (average 298 heterozygous sites) and *L. braziliensis* L3 (average 106 heterozygous sites), were each devoid of heterozygous sites compared to the remainder of the *L. braziliensis* genomes (average 13,601 heterozygous SNPs) (Supp. Fig 2B,C).

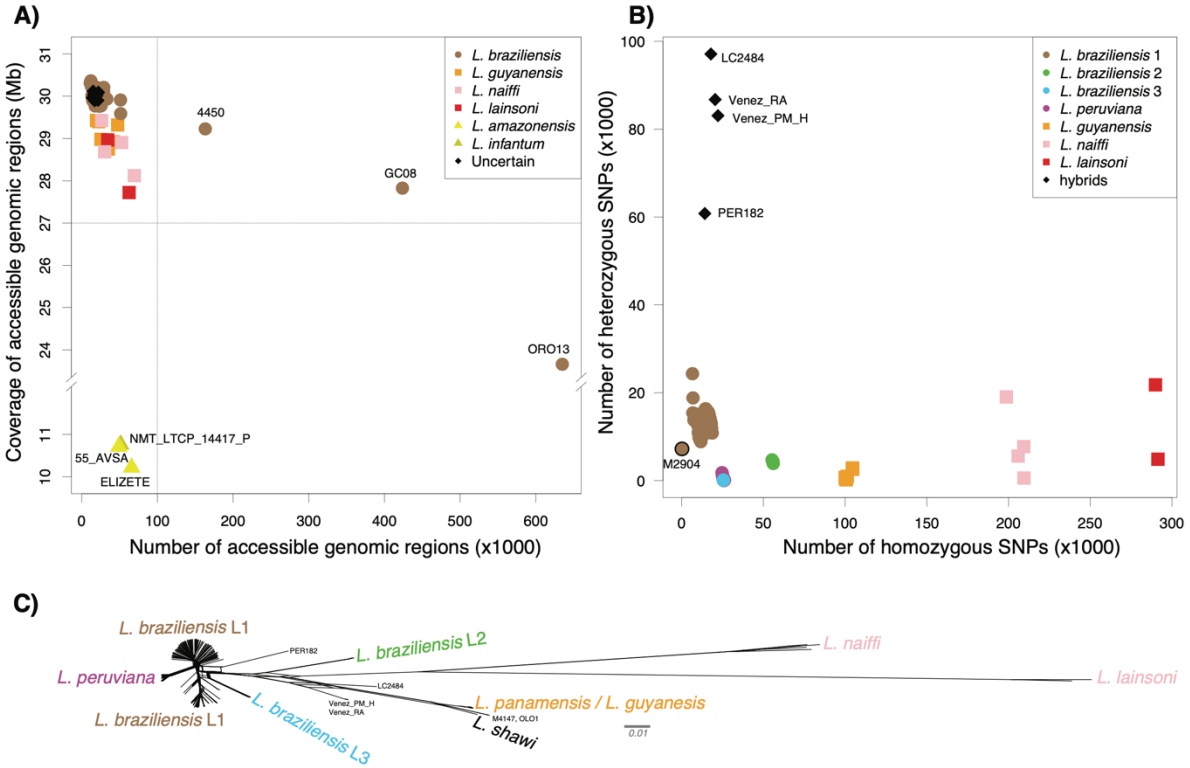
The phylogenetic network further revealed an uncertain reticulated ancestry for six *Leishmania* genomes (LC2484, PER182, Venez_RA, Venez_PM, OLO and M4147) (Supp. Fig 2C). The distribution of alternate allele read depth frequencies (ARDF) was centered around 0.5 for all six genomes (Supp. Fig 10), as expected for diploid organisms (both alleles are represented equally). Inspecting the chromosomal ARDF, all putative hybrid isolates revealed to have several genomic regions (or entire chromosomes) that were either largely homozygous or SNP poor amidst highly heterozygous regions (Suppl. Fig 2B). Generating phylogenetic networks based on the SNPs present in these non-heterozygous regions revealed patterns of i) interspecific hybridization between the *L. braziliensis* and *L. guyanensis* species complexes (LC2484, Venez_RA, Venez-PM_H32) (Supp. Fig 12); and ii) hybridization within the *L. guyanensis* species complex between *L. shawi* and *L. guyanensis*/*L. panemensis* (OLO, M4147) (Supp. Fig 13). No clear indications explaining the high heterozygosity/mixed ancestry of PER182 were found. These patterns of inter- and intraspecific hybridization were further substantiated by PCA-based ancestry estimation (PCAdmix) (Supp. Fig 14) showing more central positions of the hybrid in contrast to their putative parental species.

821 **SUPPLEMENTARY FIGURES**



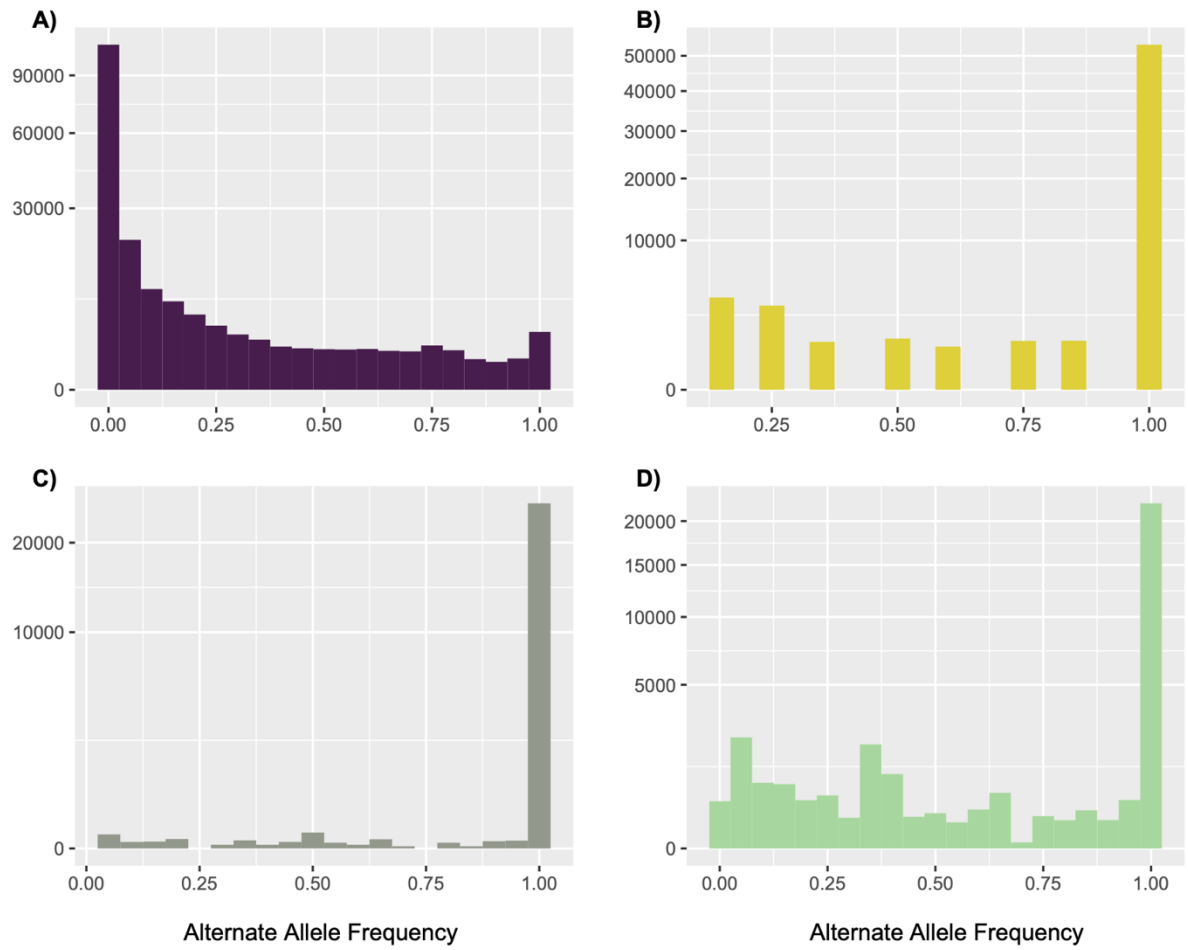
Supp. Fig 1 Removal of isolates based on alternate allele frequency distributions. Assuming diploidy is the ancestral state within the *L.* (*Viannia*) subgenus (2), the genome-wide distribution of alternate allele frequencies at heterozygous sites should approximate 0.5. Visual inspection of frequency plots confirmed a unimodal distribution centered around 0.5 for 244 isolates. This is exemplified by isolate LC1586A1 **(A)**. Seven isolates were removed for downstream analysis because they showed vastly different distributions **(B-H)** which might be the result of mixed infections.

822



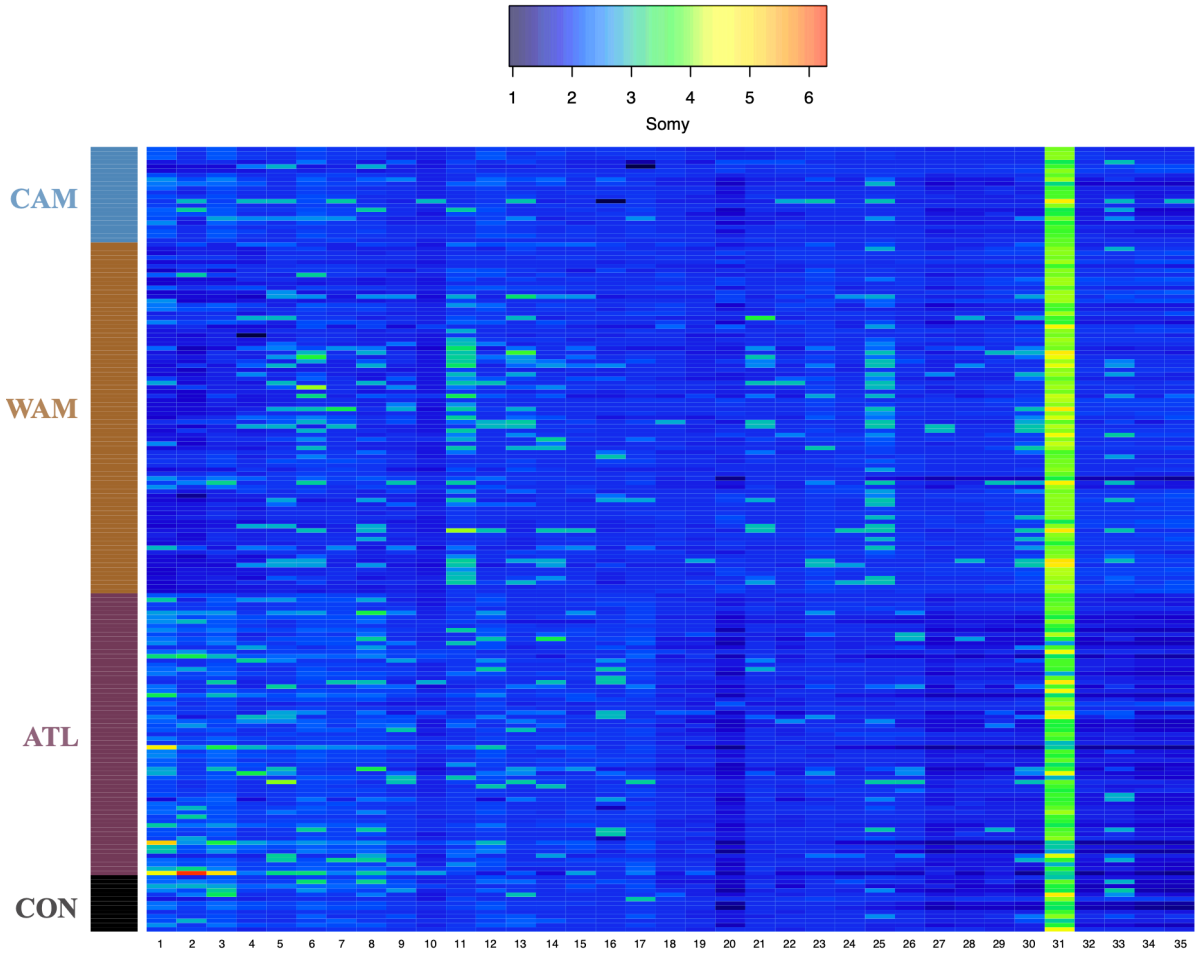
Supp. Fig 2 A) The quality of sequence alignments against the M2904 reference was investigated by characterizing the accessible genome, i.e. genomic regions of sufficient mapping quality (>25), base quality (>25) and read depth (>5). Isolates contained a median 17.8k accessible genomic regions, altogether spanning a median 29.96 Mb (i.e. 91.5% of the haploid genome). Three isolates (55_AVS, ELIZETE, NMT_LTCP_14417_P) were removed because of an aberrantly low coverage of accessible regions (10.2-10.8 Mb) compared to the other isolates; in silico multi-locus sequencing analysis (MLSA) revealed that these isolates were Leishmania amazonensis (55_AVS, ELIZETE) and Leishmania infantum (NMT_LTCP_14417_P) (results not shown). Three other isolates identified as L. braziliensis (4450, GC08, ORO13) were also removed because of low median coverages (9x-14x) and fragmented callable genomes. **B)** Inspection of the homozygous and heterozygous SNP counts of the remaining 244 L. Viannia isolates. **C)** Phylogenetic network of the 244 *L.* (*Viannia*) isolates based on 828,314 bi-allelic SNPs.

823



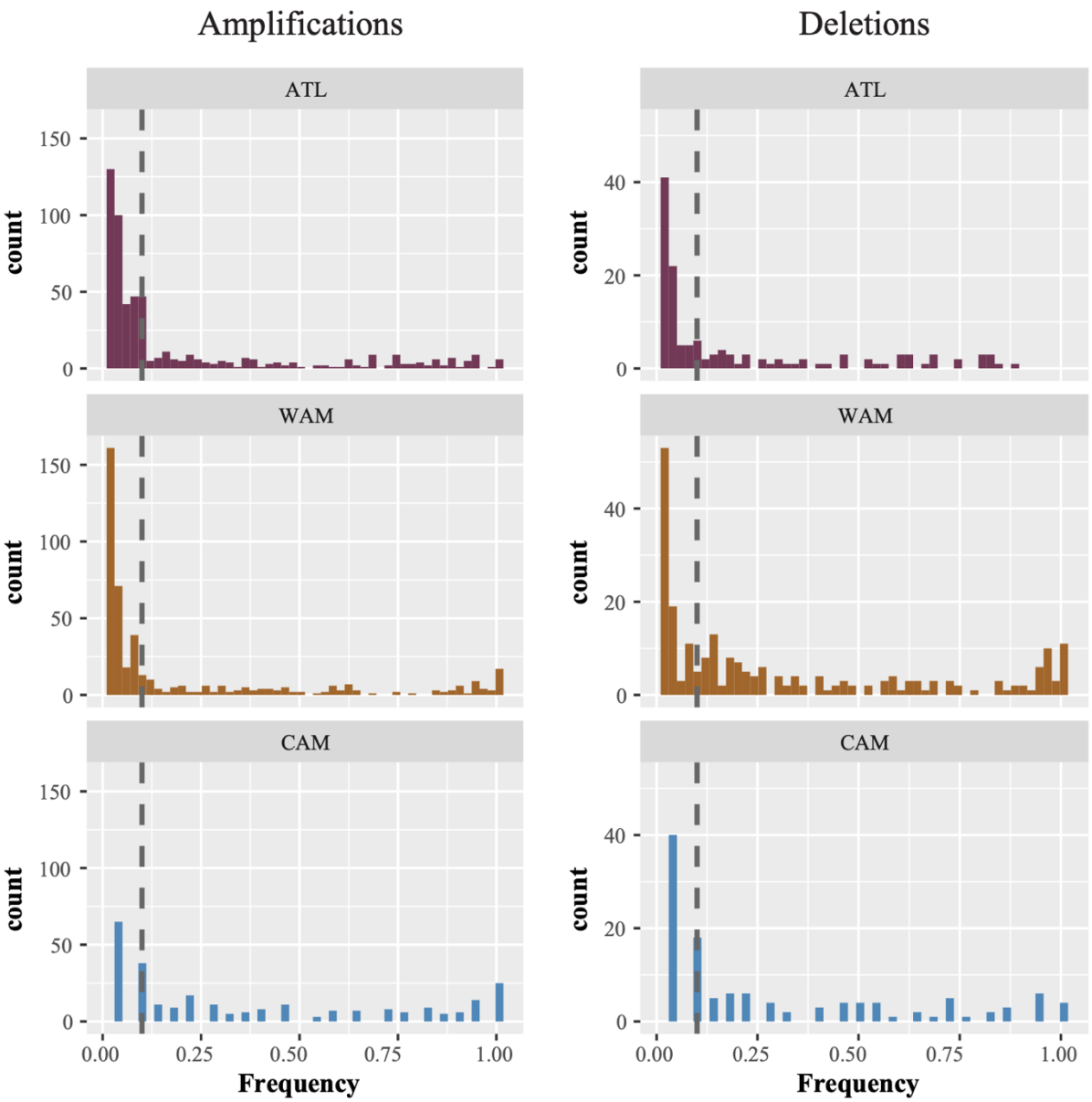
Supp. Fig 3 Population allele frequency spectra for *L. braziliensis* ecotypes L1 **(A)**, L2 **(B)**, L3 **(C)** and *L. peruviana* **(D)**.

824

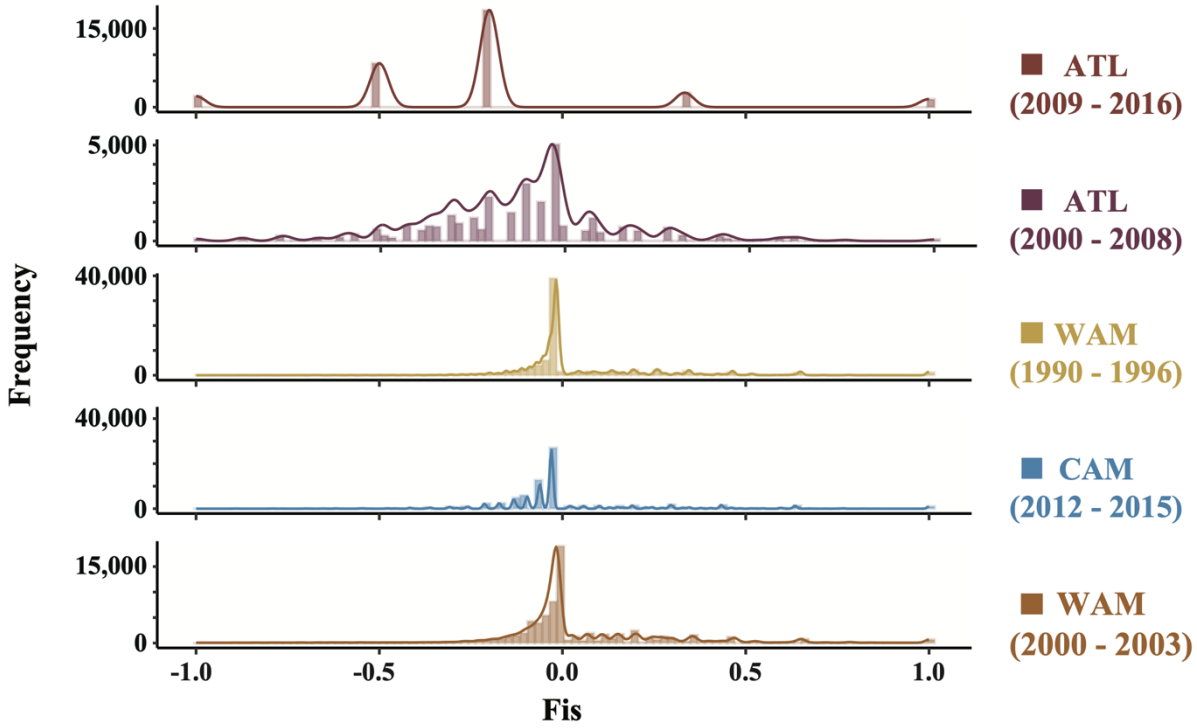


Supp. Fig 4 Variation in chromosome copy numbers across 182 *L. braziliensis* L1 genomes. Each row in the heatmap corresponds to an isolate, which were clustered per major parasite population as inferred by ADMIXTURE and fineSTRUCTURE. CAM= Central Amazon; WAM= West Amazon; ATL= Atlantic; CON= conglomerate.

825



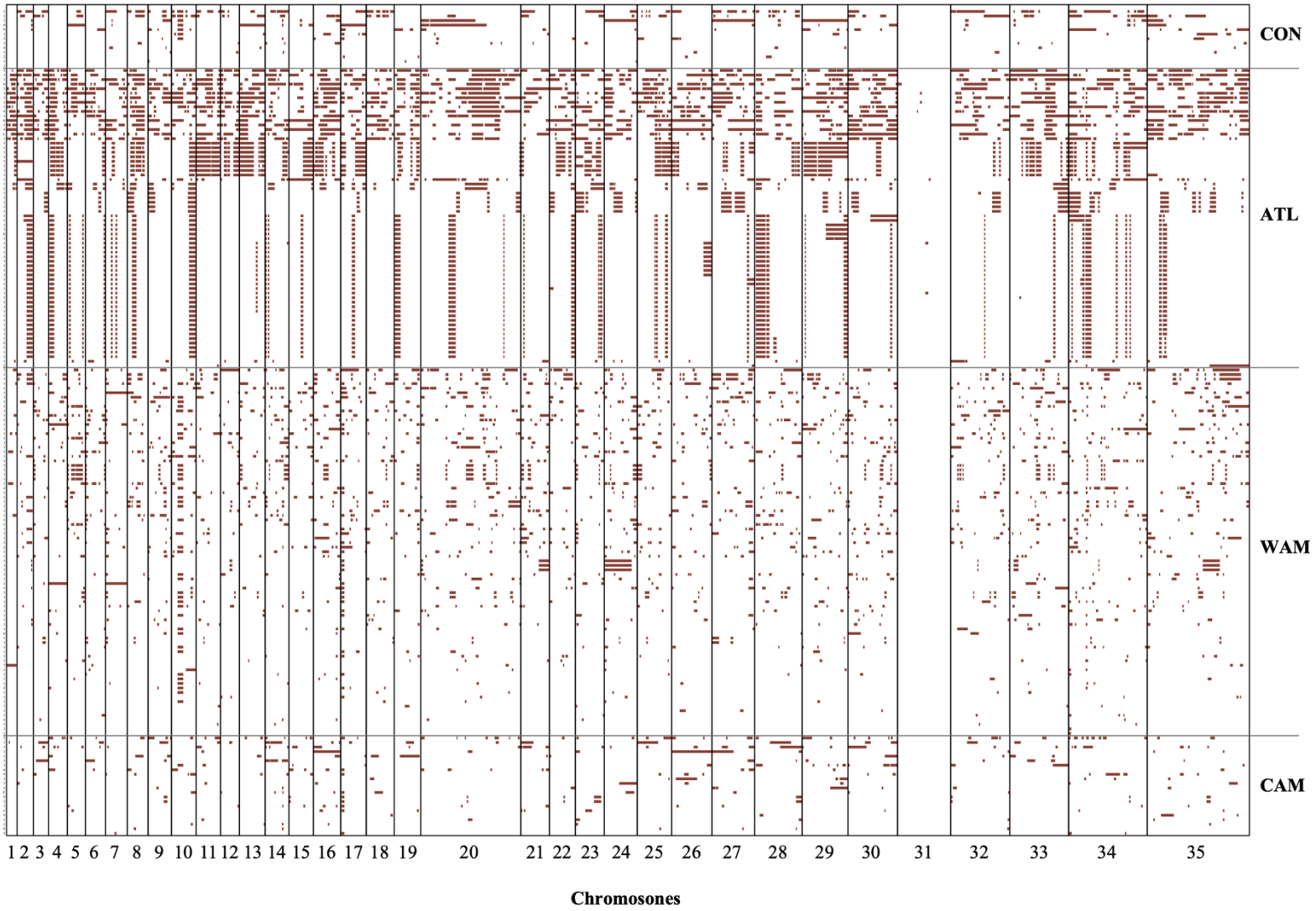
826 **Supp. Fig 5** Frequency distribution of amplifications occurring in the three major L1 populations. (left) Amplifications. (right) deletions.



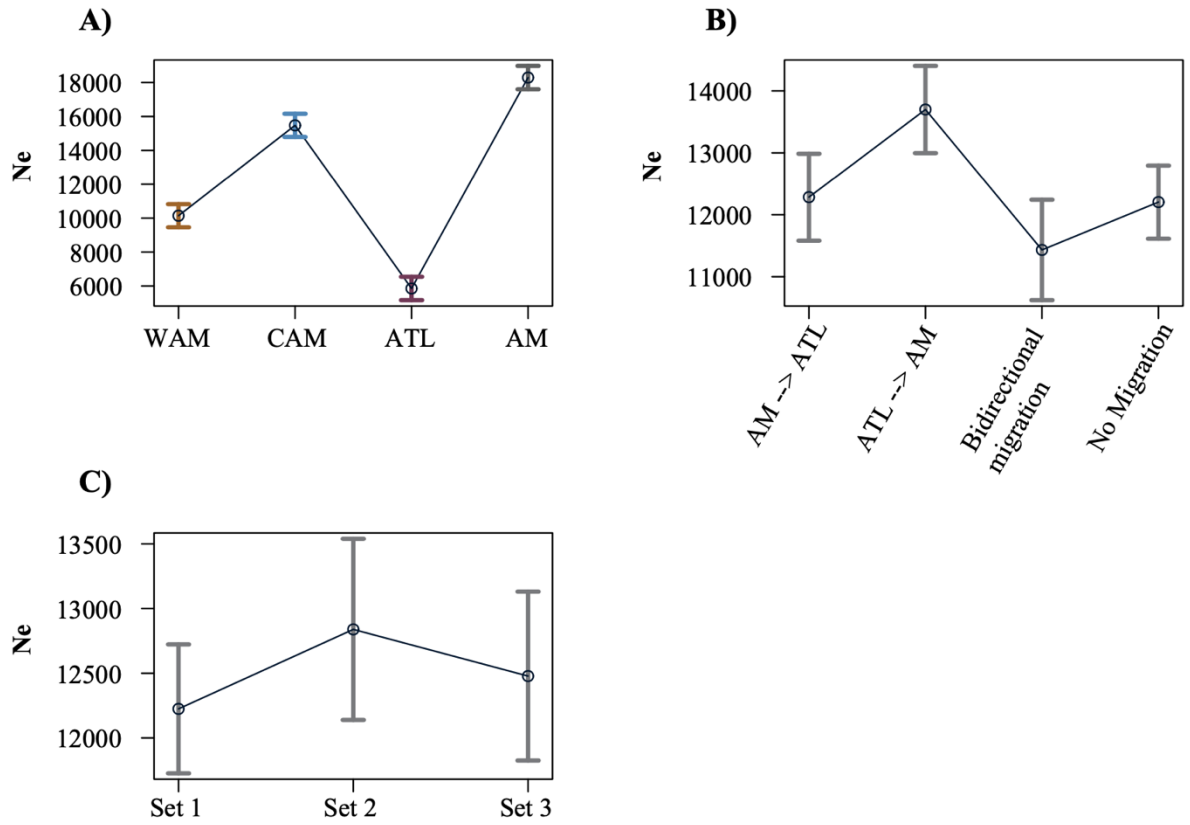
Supp. Fig 6 Inbreeding coefficient (Fis) distributions for each of the major L1 populations, taking into account spatio-temporal Wahlund effects.

827

828

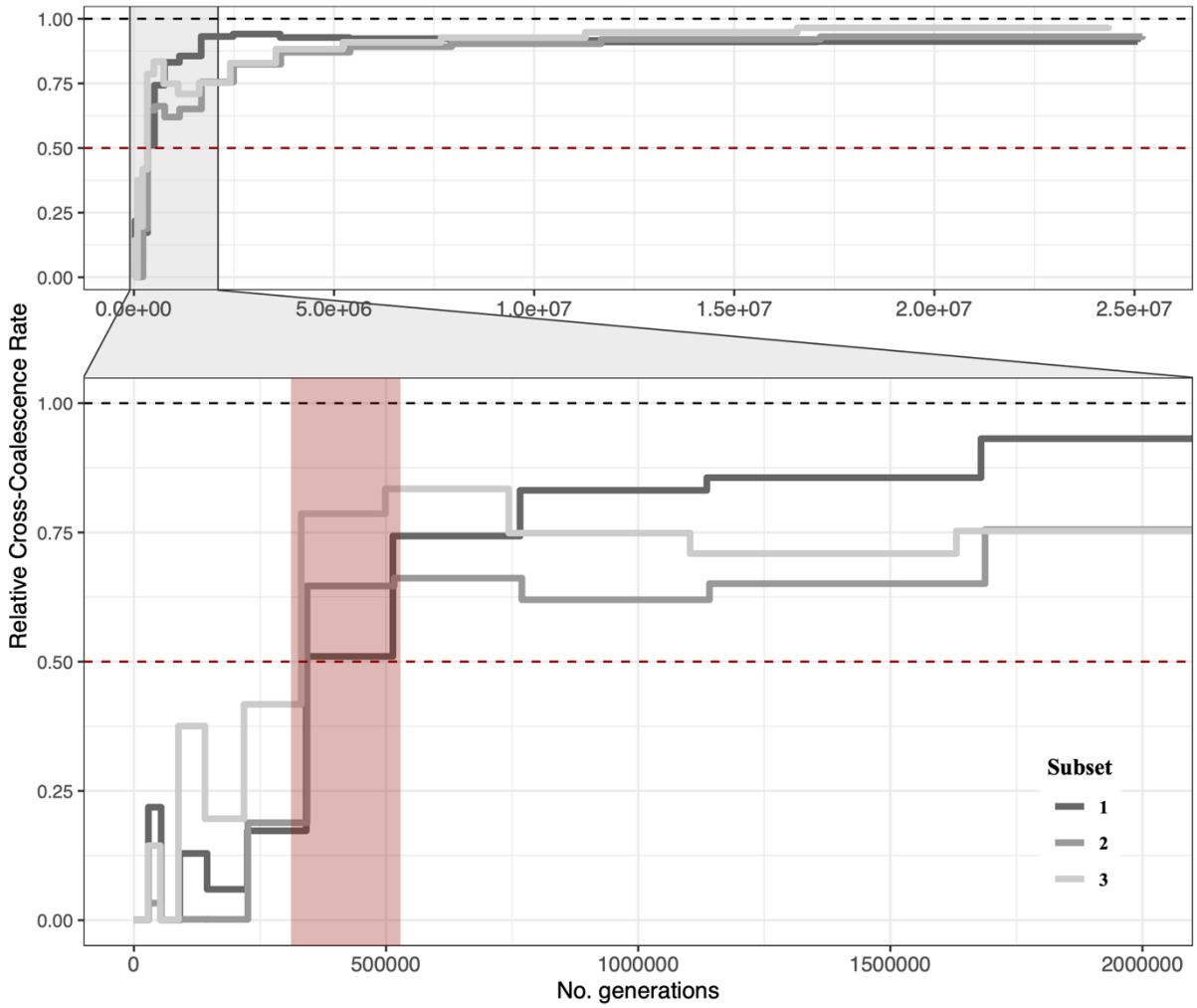


Supp. Fig 7 Loss-of-heterozygosity pattern distribution across the genome in all of the 244 isolates.



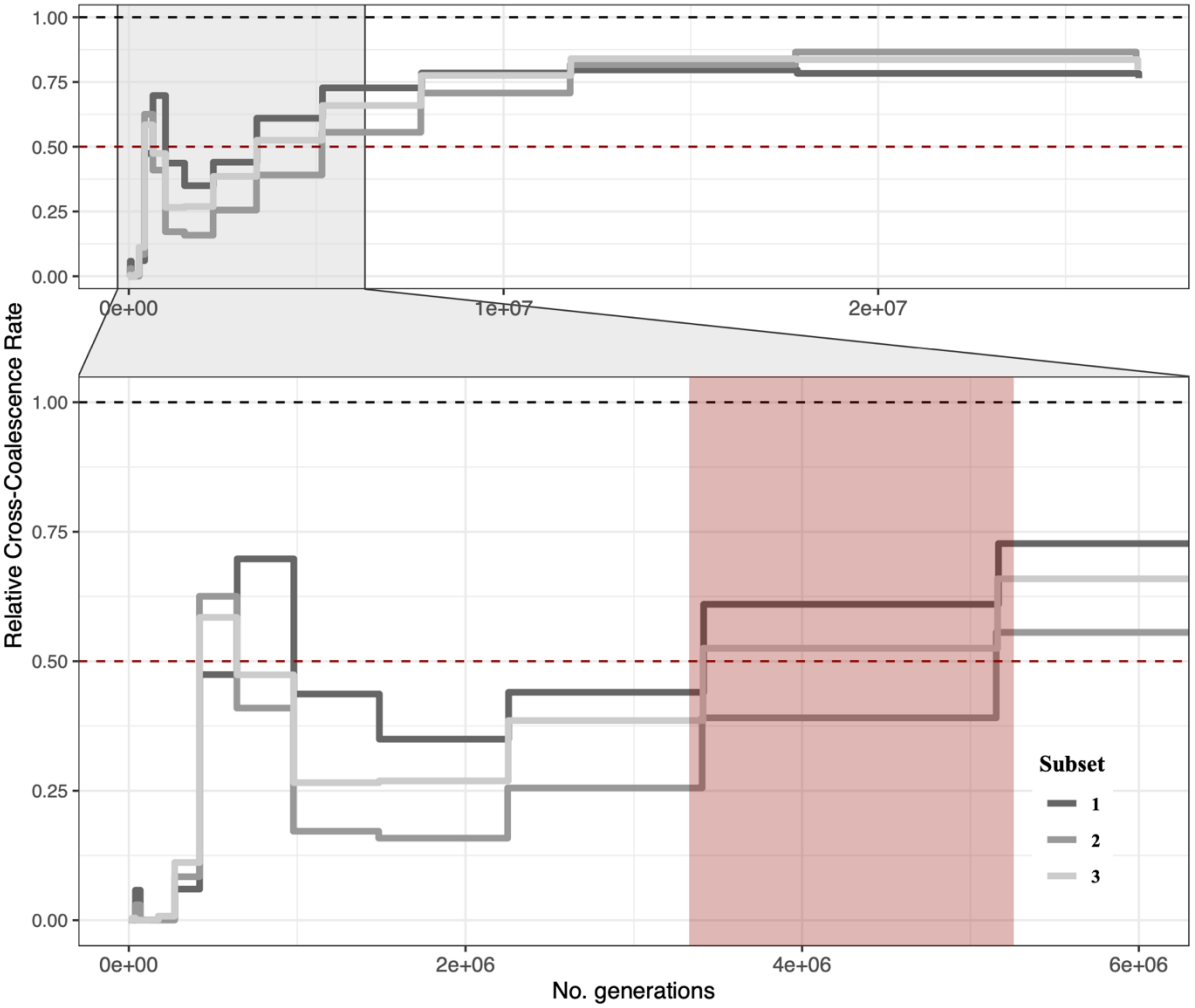
Supp. Fig 8 Main effect plots for Ne between the different L1 populations **(A),** migration model **(B)** and independent sample subsets **(C)**.

829



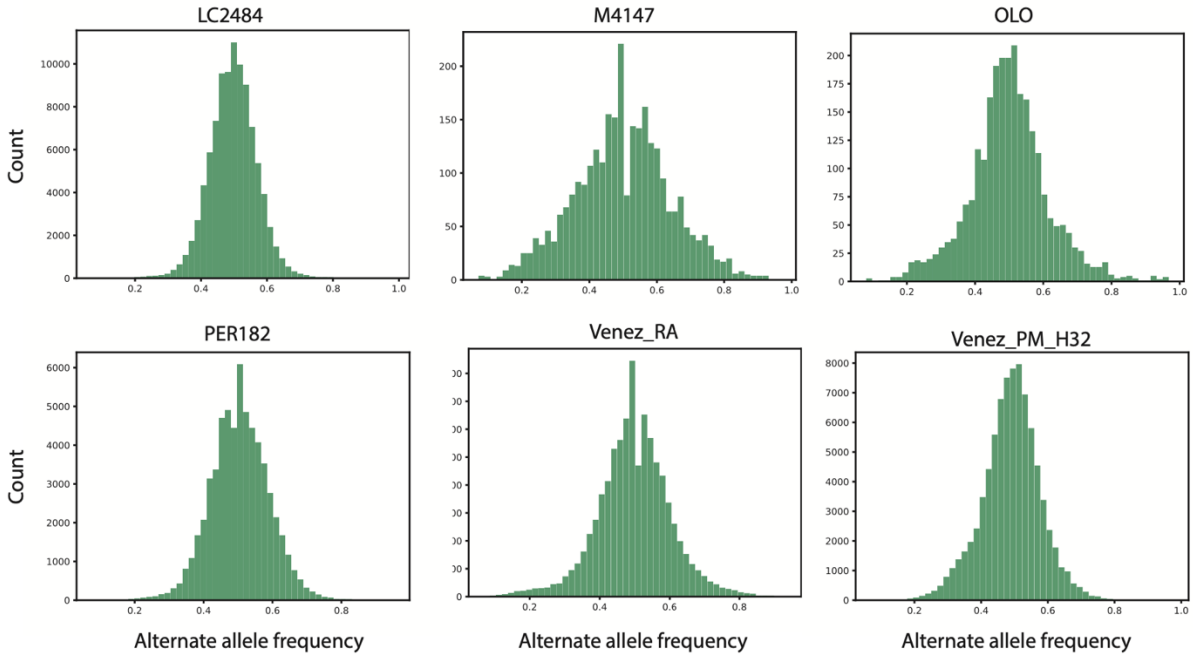
Supp. Fig 9 Relative cross-coalescence rate (rCCR) between the two Amazonian populations (WAM, CAM). Gray lines indicate the three independent runs with different sample subsets. The red box indicates the period where the rCCR first fell to 0.5, a proposed estimate for the split time between the two populations (4).

830

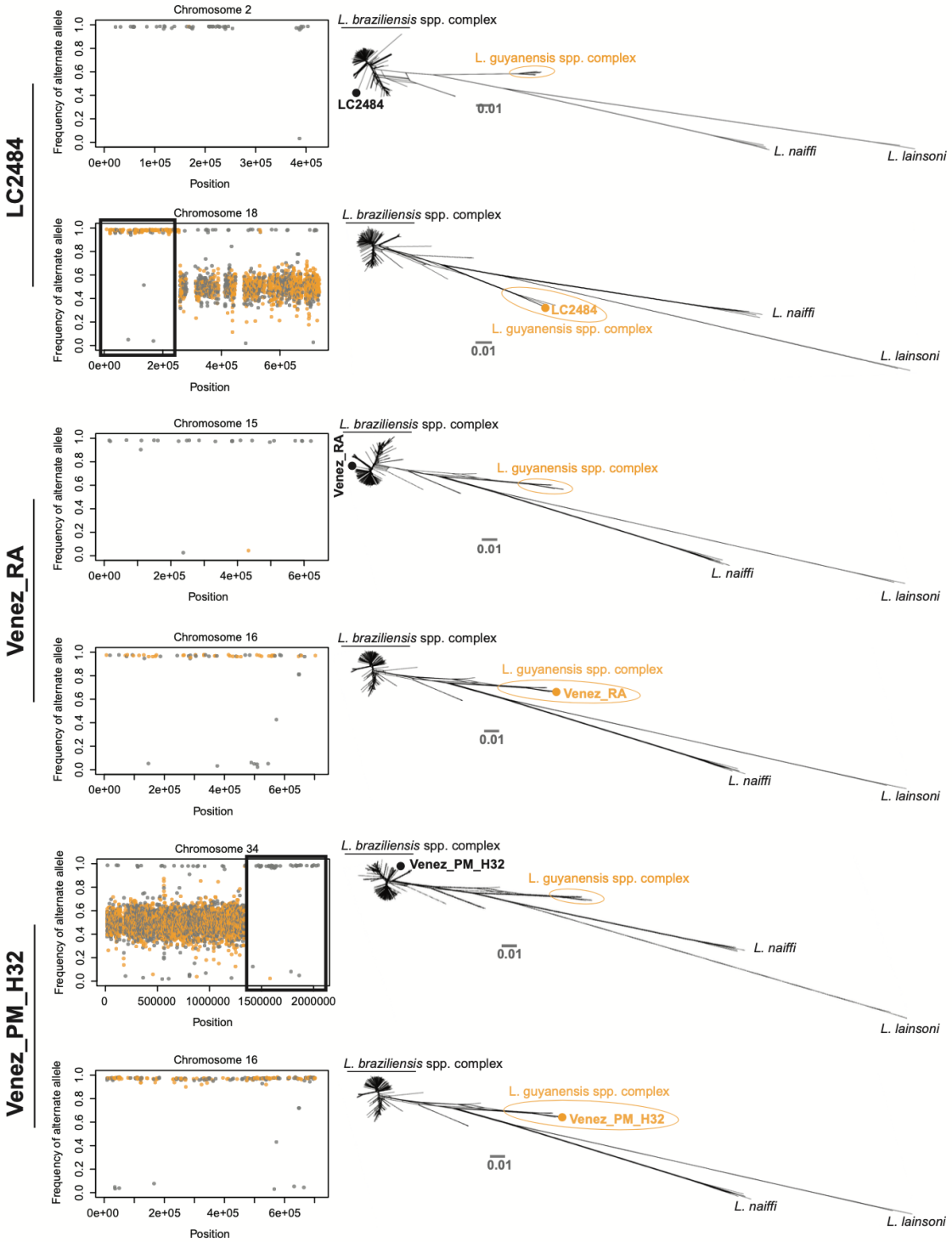


Supp. Fig 10 Relative cross-coalescence rate (rCCR) between the Amazonian (WAM, CAM) and Atlantic (ATL) populations. Gray lines indicate the three independent runs with different sample subsets. The red box indicates the period where the rCCR first fell to 0.5, a proposed estimate for the split time between the two populations (4).

831

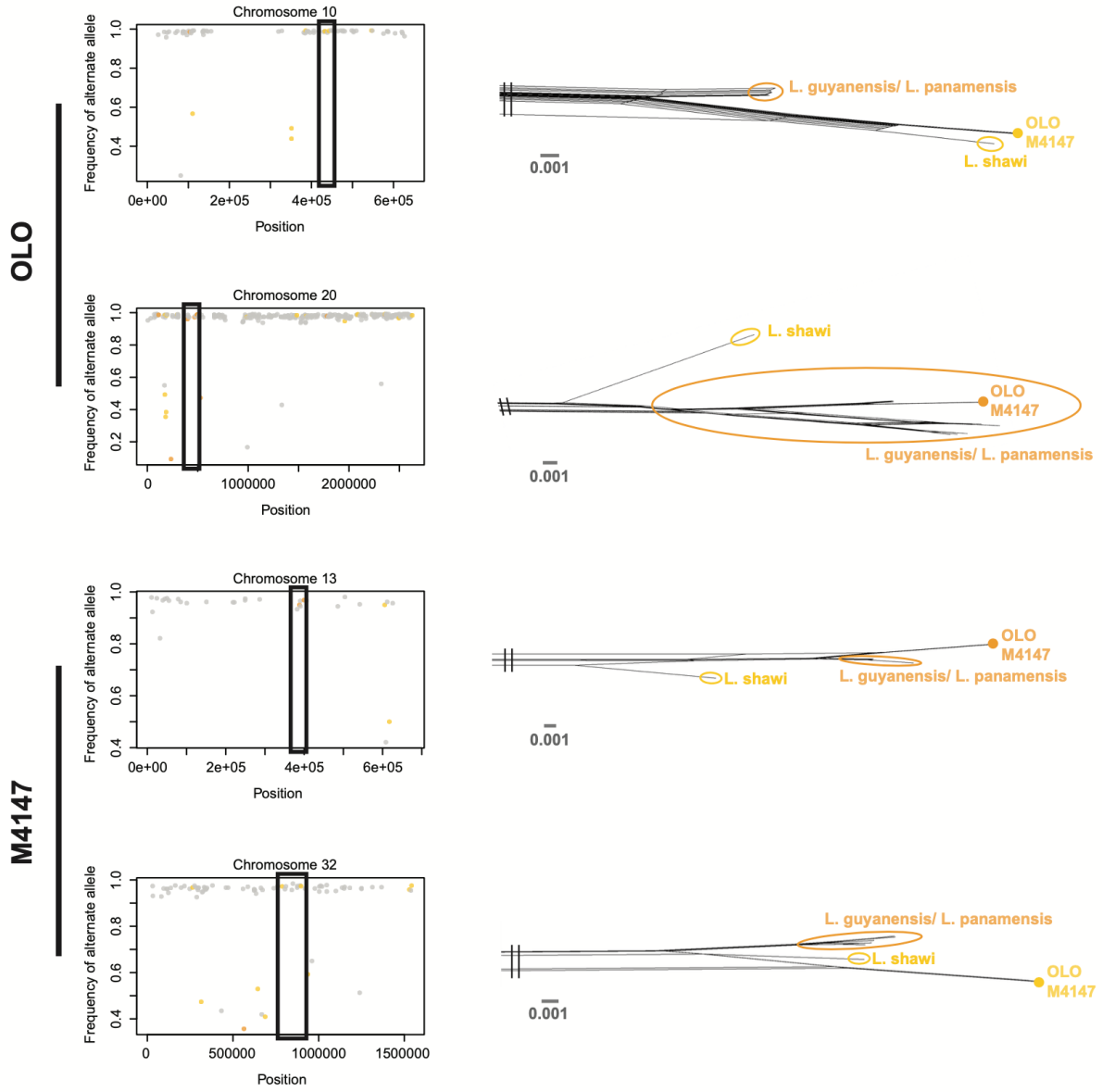


832 **Supp. Fig 11** Alternate allele frequency distribution at heterozygous sites for putative hybrid isolates. Each of the isolates show the expected unimodal distribution around 0.5 assuming diploidy.



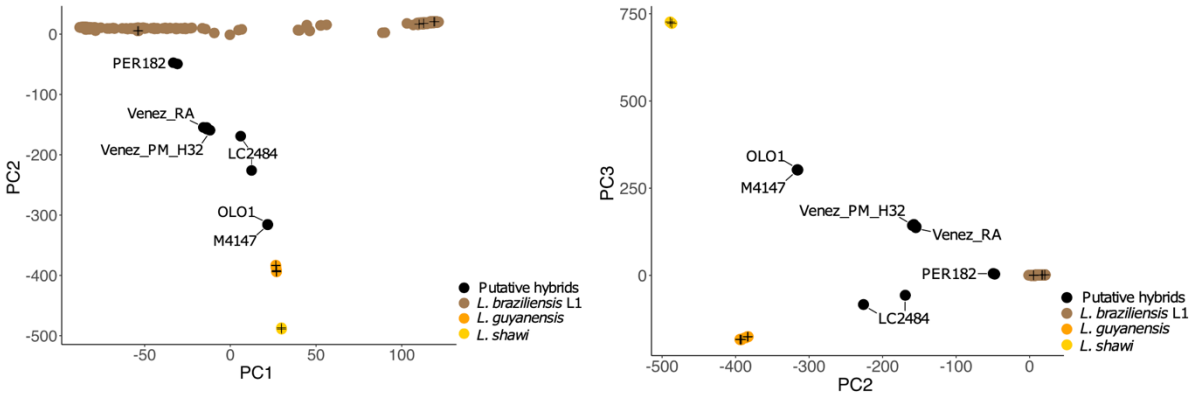
Supp. Fig 12 Interspecific hybrids between *L. braziliensis* and *L. guyanensis*. (Left) chromosome specific frequency distributions of the alternate allele that are partially or almost entirely homozygous. Orange dots represent SNPs that are unique for *L. guyanensis*. (Right) Phylogenetic networks based on SNPs called in the presented chromosome or boxed regions (i.e. homozygous regions).

833



Supp. Fig 13 Intraspecific hybrids within the *L. guyanensis* species complex. (Left) chromosome specific frequency distributions of the alternate allele that are partially or almost entirely homozygous. Orange dots represent SNPs that are unique for *L. guyanensis/L. panamensis*, yellow dots represent SNPs unique to *L. shawi*. (Right) Phylogenetic networks based on SNPs called in the presented chromosome or boxed regions (i.e. homozygous regions).

834



Supp. Fig 14 PCA-based ancestry estimation of putative hybrid isolates assuming three putative parental lineages: *L. braziliensis* L1 (brown*), L. guyanensis/L. panamensis* (orange) and *L. shawi* (yellow). Plus-signs indicate isolates of the different ancestral lineages (three per lineage) that were used as controls.

REFERENCES

1. M. C. Boité, I. L. Mauricio, M. A. Miles, E. Cupolillo, New insights on taxonomy, phylogeny and population genetics of Leishmania (Viannia) parasites based on multilocus sequence analysis. *PLoS Negl. Trop. Dis.* **6**, e1888 (2012).

2. S. Coughlan, *et al.*, Leishmania naiffi and Leishmania guyanensis reference genomes highlight genome structure and gene evolution in the Viannia subgenus. *R Soc Open Sci* **5**, 172212 (2018).

3. G. Van Der Auwera, *et al.*, Evaluation of four single-locus markers for leishmania species discrimination by sequencing. *J. Clin. Microbiol.* **52**, 1098–1104 (2014).

4. S. Schiffels, K. Wang, MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).