Multifaceted Population Structure and Reproductive Strategy in Leishmania donovani Complex in One Sudanese Village

Virginie Rougeron1*3, Thierry De Meeûs2,3, Mallorie Hide1, Georges Le Falher4, Bruno Bucheton2, Jacques Dereure1,5, Sayda H. El-Safi6, Alain Dessein7, Anne-Laure Bañuls1

1 Laboratoire MIVEGEC (UMR IRD 224-CNRS 5290-Université Montpellier 1), Montpellier, France, 2 UMR 177 IRD-CIRAD, Centre International de Recherche-Développement sur l’Elevage en zone Subhumide (CIRDES), 01 BP 454, Bobo-Dioulasso, Burkina-Faso, 3 CNRS, Délégation Languedoc-Roussillon, Montpellier, France, 4 Centre Hospitalier, Gui de Chauliac, Montpellier, France, 5 Centre Hospitalier Universitaire Service de Parasitologie - Mycologie, Montpellier, France, 6 Department of Microbiology and Parasitology, Faculty of Medicine, Khartoum University, Khartoum, Sudan, 7 Université de la Méditerranée - INSERM UMR 906, Marseille, France

Abstract

Leishmania species of the subgenus Leishmania and especially L. donovani are responsible for a large proportion of visceral leishmaniasis cases. The debate on the mode of reproduction and population structure of Leishmania parasites remains opened. It has been suggested that Leishmania parasites could alternate different modes of reproduction, more particularly clonality and frequent recombinations either between related individuals (endogamy) or between unrelated individuals (outcrossing) within strongly isolated subpopulations. To determine whether this assumption is generalized to other species, a population genetics analysis within Leishmania donovani complex strains was conducted within a single village. The results suggest that a mixed-mating reproduction system exists, an important heterogeneity of subsamples and the coexistence of several genetic entities in Sudanese L. donovani. Indeed, results showed significant genetic differentiation between the three taxa (L. donovani, L. infantum and L. archibaldi) and between the human or canine strains of such taxa, suggesting that there may be different imbricated transmission cycles involving either dogs or humans. Results also are in agreement with an almost strict specificity of L. donovani stricto sensu to human hosts. This empirical study demonstrates the complexity of population structure in the genus Leishmania and the need to pursue such kind of analyses at the smallest possible spatio-temporal and ecological scales.

Introduction

Leishmaniases are worldwide vector-borne diseases of humans and domestic animals, caused by protozoan parasites of the genus Leishmania. These parasitic infections are a serious public health problem, with about 350 million persons at risk and 2,357,000 new cases per year [1]. The genus Leishmania totals approximately 20 described species causing human infections (reviewed in [2]) with a wide variety of clinical symptoms: cutaneous, visceral, mucocutaneous, mucosal and post-kala-azar dermal (PKDL) leishmaniases. Visceral leishmaniasis is the most severe form of the disease, which can be lethal if it goes untreated. It is the most widespread leishmaniasis form, especially in India, Bangladesh, Nepal, Sudan, Ethiopia and Brazil [1,3,4]. In this study, we focused on human and canine samples collected in Sudan, where visceral leishmaniasis is endemic in the eastern and southern parts of the country and has claimed the lives of thousands of people [5].

Visceral leishmaniasis is mainly caused by species from the Leishmania donovani complex [6]. Multilocus enzyme electrophoresis (MLEE) studies generated the description of three different species in this complex: L. donovani in the Old World, L. infantum in the Old World and the New World (also named L. chagasi there), and L. archibaldi in Sudan and Ethiopia [7,8]. In Sudan, the taxonomic status of these three species has been challenged using several different molecular markers, such as random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and microsatellites [9,10]. On the basis of both sequencing and microsatellite analysis, Jamjoom et al. proposed that Leishmania donovani sensu lato was the only cause of visceral leishmaniasis in East Africa (the three species falling in one clade), including Sudan [11]. Lukes et al. [12], by a multifactorial genetic analysis that includes DNA sequences of protein-coding genes as well as noncoding segments, microsatellites, restriction-fragment length polymorphisms, and randomly amplified polymorphic DNAs, suggested that Leishmania infantum and L. donovani were the only recognized species of the L. donovani complex [12]. It was even recently suggested that the only valid name is L. donovani [13].
Leishmaniases are a serious public health problem, especially in developing countries, caused by Leishmania parasites and transmitted by sandfly bites. More information is needed on the population biology of these pathogens for diagnostic and epidemiological inquiries and for drug and vaccine elaboration. For studies dealing with the population genetics, exploring the genetic patterns of such organisms at microgeographic scales is fundamental. In this context, we made a population genetic study, based on 20 microsatellite loci, on 61 strains of Leishmania donovani complex collected in a Sudanese village, Babar El Fugara, during the epidemic of 1996–2000. Results showed that considering the whole sample as a single population was not adequate because of the coexistence of several genetic entities and a genetic differentiation between the human or canine strains. In addition, our findings suggested that clonality may have a strong impact on the L. donovani complex, unlike other Leishmania species. This study demonstrates the need to pursue population genetics studies in Leishmania species from sampling designs that control maximum possible confounding factors and to elaborate such kinds of analyses at the smallest possible spatio-temporal and ecological scales.

Materials and Methods

Study site, parasites, cultures and DNA extraction
A census of the village population was conducted by Bucheton et al. [24], making personal and clinical data available. From 1997 to 2000, 61 isolates of Leishmania donovani complex were collected and then cultured. We obtained the samples for this study from the “the French National Reference Center of Leishmania”, under the agreement of Dr. Alain Dessein.

The 61 strains from Sudan were isolated from dogs (ten strains) and humans (51 strains) and characterized using the MLEE technique by Dereure et al. [25]. Thirty-three strains were identified as L. donovani, 17 strains as L. infantum and 11 strains as L. archibaldi (see supplementary data Table S1). Promastigotes were cultured at 26°C by weekly subpassages in RPMI 1640 medium, buffered with 25 mM HEPES, 2 mM NaHCO3 and supplemented with 20% heat-inactivated fetal calf serum, 2 mM glucose, 100 U/ml penicillin and 100 µg/ml streptomycin. Cultures were harvested by centrifugation and stored at −80°C until DNA extraction. Genomic DNA was extracted using the DNeasy Blood and Tissues Kit (Qiagen, Courtaboeuf, France), following the manufacturer’s recommendations.

Statistical analysis
Data were processed through Create V 1.1 [27] to convert the data for different usage. We mainly analysed data with Fstat Version 2.9.3.2 software (Goudet 2002, updated from Goudet [28]), which computes estimates and tests the significance of the following population genetics parameters. Genetic polymorphism was measured by the number of alleles per locus (N_a) and by Nei’s unbiased estimate of genetic diversity within subsamples H. [29]. We estimated Wright’s F statistics [30] with Weir and Cockerman’s method [31]: F_IS measures the relative inbreeding of individuals due to the local non-random union of gametes in each subpopulation, and F_ST measures the relative inbreeding in subpopulations attributable to the subdivision of the total population into subpopulation of limited size. F_ST thus also measures genetic differentiation between subpopulations. F_IS ranges between −1 and 1: a negative value corresponds to an excess of heterozygotes, a positive value to heterozygote deficiency; 0 is expected under panmixia. The significance of the departure from 0 was tested by 10,000 randomisations of alleles within subpopulations (to test random mating) and individuals across subsamples (for differentiation). The statistic used for random mating (Hardy-Weinberg Equilibrium) testing was simply Weir
and Cockerham’s estimator \( f (F_{IS} \text{ and } F_{ST}) \). For the genetic differentiation test, we used the log likelihood ratio \( G \)-based test of Goudet et al. [32] summed over all loci. Confidence intervals were estimated by bootstrapping over loci or jack-knifing over populations with Fstat as described in De Meeës et al. [33].

Genetic diversity, as measured by Nei’s \( H_s \), can lower the maximum possible value for \( F_{ST} \). According to classical formulation (e.g. [34]) \( F_{ST} = (Q_s - Q_T) / (1 - Q_T) \), where \( Q_s \) is the probability to sample twice the same allele in a subpopulation and \( Q_T \) is the probability to sample twice the same allele in different subpopulations. If a population was totally subdivided, then the probability to sample twice the same allele in two different subpopulations should be null and thus \( F_{ST} \) should be equal to the probability to sample twice the same allele in a subpopulation \( Q_s \). \( H_s \), being the probability to sample two alleles that are different hence \( Q_s = 1 - H_s \). The maximum possible value for \( F_{ST} \) in a sample with a given \( H_s \) can thus be estimated as \( 1 - H_s \) and a corrected version of \( F_{ST} \) as \( F_{ST}^* = F_{ST} / (1 - H_s) \) [33,35].

Data were heterogeneous regarding Leishmania species (as recognized by MLEE typing), year of sampling and host species. To assess the possible contribution of these factors to genetic partitioning (Wahlund effect), we compared \( F_{IS} \) obtained with four different sampling strategies. The first sampling strategy considered each Leishmania species-year of sampling-host species combinations as different subsamples (14 subsamples, “All separated” strategy). The second strategy ignored the Leishmania species distinction (six subsamples, “Species fused” strategy). The third strategy ignored the year of sampling (six subsamples, “Years fused” strategy) and the fourth one ignored the host species (10 subsamples, “Hosts fused” strategy). For significant difference testing, we undertook planned pairwise Wilcoxon signed rank tests between “All separated” and each of the other three strategies ordered as above with sequential Bonferroni correction (multiplying the \( P \)-values by 3, 2 and 1, respectively). Unilateral (“All separated” has a smaller \( F_{IS} \) than the other three strategies) Wilcoxon signed rank tests were undertaken under R [36]. Differentiation between the relevant units controlled for the other factors were then undertaken with paired subsample differentiation tests (\( F_{ST} \) estimation and \( G \)-based randomisation test). When two values were obtained for the same type of differentiation (e.g. differentiation between \( L. \) archibaldi and \( L. \) infantum in 1997 and 1998), these values were combined with an unweighted mean for \( F_{ST} \) (e.g. over years) and Stouffer’s \( Z \)-test (Whitlock, 2005) for \( P \)-values as recommended [37].

Linkage disequilibrium between pairs of loci (non-random association of alleles at different loci) was assessed with a randomisation test (genotypes at two loci are associated at random a number of times) using Fstat software Version 2.9.3.2 software (Goudet 2002, updated from Goudet [28]). The statistic used was the log likelihood ratio \( G \) summed over all subpopulations, known to be more powerful than other combinatorial procedures [37]. Because there are as many tests as locus pairs tested (here 15 × 14/
distribution identified by BAPS and noted population, we ran the program 50 times in order to obtain the software uses stochastic optimization to infer the posterior mode of been successfully applied to other parasites [21,39,40]. The BAPS This software was used to detect possible Wahlund effects and has within populations (admixture) through a Bayesian analysis [38].

\[ F_{IS_C} = 0.05. \] Thus we used the unilateral ("greater") exact binomial test to check if there was significantly more than 5% significant tests in the 190 tests series under \( R \) [36].

The BAPS version 5.1 software identifies a hidden structure within populations (admixture) through a Bayesian analysis [30]. This software was used to detect possible Wahlund effects and has successfully applied to other parasites [21,39,40]. The BAPS software uses stochastic optimization to infer the posterior mode of genetic structure. To obtain the best distribution of the entire population, we ran the program 50 times in order to obtain the right number of clusters. The same approach has been applied within \( L. donovani \) 1997, \( L. donovani \) 1998 and \( L. infantum \) from humans for which enough individuals were available. Each of the three samples was submitted to a clustering exploration by BAPS with a maximum number of clusters set (19, 13 and 12, respectively, these values corresponding to the number of individuals in each sample). \( F_{IS} \) was recalculated in each best distribution identified by BAPS and noted \( F_{IS_C} \). Then, for the three samples corresponding to the three species of \( Leishmania \), the \( F_{IS_C} \) was compared with the initial \( F_{IS} \) using a unilateral Wilcoxon signed-rank test for paired data (with the software \( R \)), the pairing units being the 20 loci. If \( F_{IS_C} \) is significantly lower than \( F_{IS} \), it is probable that the initial subsamples were composed of several genetically distinct entities (e.g. geographical microstructure or subpopulations).

Since we got the data's prevalence from Dereure et al. study [25], the prevalences were compared for each \( Leishmania \) species between humans and dogs (50 human strains and 20 dog strains), and the significance was tested using an exact Fisher test under the software \( R \) [36].

A Neighbor-Joining (NJ) tree [41] was constructed out of a Cavalli-Sforza and Edwards genetic distance matrix [42]. The robustness of tree topology was obtained by bootstrap resampling of loci, with 500 replications per set. We used PHYLIP software (version 3.5c; J. Felsenstein, Department of Genetics, University of Washington, Seattle, 1993) and the tree was edited using TreeDyn software [43].

Simulations where also handled with Easypop 2.0.1 (Balloux 2006, updated from Balloux 2001 [44]) to find possible sets of parameters fitting our observations.

Ethical statement

The approval for human strain study was obtained by both the federal and state Ministries of Health and by the Faculty of Medicine of Khartoum. Approval of the project to be performed was also approved for each field visit by the village committee, which included elected delegates from all ethnic groups and as well elected citizens. Since an important proportion of the population in Barbar El Fugara was illiterate, oral informed consent was obtained after the aim of the study was explained to study participants in their own language by a translator. For child participants, oral consent was obtained from their parents. The verbal consent was also obtained in the presence of the ethnic group leader, who eventually provide more explanations if required. After verbal informed consent obtained from the patient, the clinician recorded it on a written form.

Results

We obtained clear electrophoregrams for all genotypes at all 20 loci investigated, with only one or two alleles per strain at each locus, which excludes events of aneuploidy (for which we would have also expected individuals with no alleles, three or four alleles). The genotypes obtained are presented in supplemental Table S1. The data showed a low level of genetic diversity, with an average number of alleles per locus of 4.29±1.74, ranging from 2 (LIST7026, LIST7028 and LIST7030 to 8 (LIST7025) and a mean genetic diversity \( H_S = 0.475±0.148 \) (Table 1).

Phylogenetic analysis and genetic differentiation

The dendrogram, based on 20 polymorphic microsatellite loci, represented in Figure 1 underlined two main clusters. Cluster A (36% bootstraps) regroups strains from \( L. infantum \) and \( L. donovani \). Cluster B (sustained by a bootstrap of 32%) corresponded to \( L. archibaldi \) taxon and three \( L. infantum \) from dogs. It has to be noticed that other studies have observed, using microsatellite method, small bootstrap for large clusters and important bootstrap values only for small clusters for \( L. braziliensis \) [43] and \( L. infantum \) [46].

\( F_{IS} \) comparisons between “All separated” strategy and the three others gave significant differences, as illustrated in Figure 2, meaning each factor, \( Leishmania \) species, year of sampling and host species in order of importance, displays a significant signature on the apportioning of genetic information. Consequently, each \( Leishmania \) species of each year and each host species must be considered as separate subsamples. It has to be noticed that the significant results we obtained cannot come from an insufficient number of samples. Indeed, the significant differences evidenced are statistically valid and ignoring it might lead to overlook important ecological processes currently involved in the population biology of these \( Leishmania \) “lineages”. Moreover, these differentiations were confirmed by paired subsample differentiation tests, as indicated in Table 2. All \( Leishmania \) species are genetically different. Species differentiation seems very pronounced between \( L. donovani \) and \( L. archibaldi \) (\( F_{S1} \sim 0.767 \)) and smaller for the two other pairs (\( F_{S1} \sim 0.2–0.3 \)) (Table 2). Temporal differentiation seems only to affect \( L. donovani \) in humans. Considering the host origin, a weak and marginally non-significant differentiation is found between human and dog strains for \( L. archibaldi \) while a strong differentiation seems to affect \( L. infantum \) strains between the two host species (Table 2).

Clinical forms (visceral versus PKDL in humans, see Supplementary Table S1) could only be compared for \( L. donovani \) in 1997 and 1998 where no differentiation could be evidenced (\( F_{IS} \sim 0, \) \( P \)-value >0.4 in both cases). Consequently, clinical forms were not considered further in our analyses.

Prevalence comparisons

The data’s prevalence from Dereure et al. [25] was compared for each \( Leishmania \) species between humans and dogs (50 human strains and 20 dog strains). The results, presented in Table 3, show that \( L. donovani \) is clearly found in humans rather than in dogs (\( P \)-value = 0.001), that \( L. infantum \) displays a tendency to infect dogs more often (\( P \)-value = 0.04), while the difference is not significant for \( L. archibaldi \) (\( P \)-value = 0.2). If Bonferroni adjusted, only \( L. donovani \) test stays significant (\( P \)-value = 0.003).

Linkage disequilibrium study

This analysis was undertaken over all the data but considering each \( Leishmania \) species, year of sampling and host species combination as a distinct subsample. This provided 19 locus pairs out of 190 tests in significant linkage. This is far above the 5% expected under the null hypothesis (\( P \)-value = 0.0001). These significant tests involved 18 of the 20 loci. Within each \( Leishmania \) species, small subsample sizes limited the power of the test. For \( L. archibaldi \) (very small subsamples of four and seven individuals in dogs and human hosts respectively) only five tests out of 190 were
significant (P-value = 1). In *L. donovani* 22 tests were significant (P-value = 0.0003) and in *L. infantum* 19 tests were significant (P-value = 0.0034). There is thus a global linkage at a genome-wide scale in the three *Leishmania* species populations.

**Genetic diversity and heterozygote deficiency within *Leishmania* species**

For each *Leishmania* species, a global and highly significant heterozygote deficit, highly variable across loci, was observed (Figure 3). These heterozygote deficits significantly decrease (P-values < 0.005) in the best partitions found by BAPS for the two species for which such analyses could be done (*L. donovani* and *L. infantum*) (Table 4 and Figure 4). Simulations, undertaken using the software EasyPop, provided patterns convergent with the pattern observed for some parameter sets only for very high clonal rates (minimum $c = 0.99$) and strong Wahlund effects (pooling one representative of each strongly isolated subpopulation into one subsample). Nevertheless, in each of these simulations, fairly numerous multilocus genotypes (MLGs) appeared, in contrast to the real data, where on the whole data set only two MLGs (2 observations of two samples presented the same multilocus genotypes) were observed. Consequently, something else is occurring. Finally, using the NJ Tree pattern of Figure 1, keeping only *L. donovani* strains belonging to most homogeneous clusters (no leaf longer than 0.1, see Figure 1) and subdividing it into subclusters belonging to the same year indeed produced lower $F_{IS} \approx 0.27$, but still with a very strong variance across loci (ranging from −0.1 to 0.7), no significant linkage disequilibrium and a reasonable proportion of MLGs (one repeated twice and a second repeated three times) but very small subsample sizes. It has to be noticed, that the global same topology of the NJ tree using Cavalli-Sforza distances has been obtained using shared allele distances.
Discussion

Despite the latest studies in this area, the debate on population structure and *Leishmania* reproductive mode is far from being settled and therefore deserves further investigation. Recent publications on different *Leishmania* species and in different environments seriously challenge the view that the species of the genus should display a predominantly clonal genetic signature because of important homogeneity levels and rarity of MLGs [17,18,19,20,21]. As suggested for *L. braziliensis* [21], these parasites could alternate different modes of reproduction: clonality in both vertebrate host and insect vector and sexual recombination (similar to other kinetoplastid parasites, such as *Trypanosoma brucei* [47], or other Trypanosomatidae such as *Crithidia bombi* [48]) between genetically related cells (endogamy) resulting in high levels of inbreeding. Most of these studies also revealed strong heterogeneities within *Leishmania* subsamples that probably results from Wahlund effects (mixture of differentiated true populations), and also the Minimum Evolution tree using either Cavalli Sforza heterogeneities within levels of inbreeding. Most of these studies also revealed strong between genetically related cells (endogamy) resulting in high s.l. [47], or other Trypanosomatidae such as *Crithidia bombi* [48]) (similar to other kinetoplastid parasites, such as *L. braziliensis* [21], these parasites could alternate different modes of reproduction: clonality in both vertebrate host and insect vector and sexual recombination (similar to other kinetoplastid parasites, such as *Trypanosoma brucei* [47], or other Trypanosomatidae such as *Crithidia bombi* [48]) between genetically related cells (endogamy) resulting in high levels of inbreeding. Most of these studies also revealed strong heterogeneities within *Leishmania* subsamples that probably results from Wahlund effects (mixture of differentiated true populations), because strains were collected at too large spatial and/or temporal scales. To prevent such possible biases, we selected a sample of *L. donovani*, collected at a village scale, reducing the risk of hidden substructuring.

In this Sudanese village, the validity of the distinction between *L. donovani* sensu stricto, *L. archibaldi* and *L. infantum*, be it a true species, a subspecies or any other taxonomic level, is supported by our results, in contradiction with recent papers [9,11,13,49]. As shown here, ignoring such delimitations dangerously biases genetic data interpretation. It remains that taxonomic distinction based on isoenzymes does not seem very clear as can be seen from Figure 1 and it would be worth trying other kind of markers as MultiLocus Sequencing Typing or MultiLocus Sequencing Analysis [50] to clarify this issue.

Another significant subdivision arose between dogs and human hosts, primarily regarding *L. infantum* and to a much lesser extent *L. archibaldi*. Gene flow (gene flow) appears much reduced between dogs and human hosts for *L. infantum* and two different kinds of cycles must be present here, involving probably different vector’s species and reservoirs. For *L. archibaldi* the difference is much less obvious but may be as a result of modest sampling sizes. In *L. donovani*, the greater specificity of strains to human (Table 3) and the resulting reduced number of strains found in dog did not allow for such testing. Nevertheless, the single *L. donovani* strain (LEM3785) genotyped from a dog did not show any originality as regard to its human counterparts (Figure 1). This apparent strong specificity for human hosts would mean, at least for *L. donovani*, that dogs are not a significant reservoir for these pathogens, in agreement with the anthropomonic feature of this species.

Time also appeared as a significant subdividing factor but only for *L. donovani*. The simplest interpretation being that, as patient once diagnosed are treated, a drop in subpopulation size may occur in the following year, thus leading to a genetic differentiation as a result of a bottleneck or of the replacement of empty places by other strains. Genetic diversities being not significantly different between 1998 and 1999 (Wilcoxon signed rank test, *P*-value = 0.27), the second hypothesis appears more likely.
Failing to consider all the above factors as relevant resulted in a very odd $F_{IS}$ distribution as illustrated by Supplementary Figure S1.

Our data, and especially the NJTree approach, also suggest that hybridization between the different taxa is not impossible, though rare enough to prevent homogenization, but frequent enough to enhance heterogeneity within each cluster that could be defined.

An interesting point to notice is the absence of genetic differentiation obtained between $L. donovani$ clinical forms (visceral leishmaniasis and PKDL, $F_{ST}$=0, $P$-value>0 in 1997 and 1998). Indeed, this result could suggest that the development of PKDL in treated patients is more likely linked to host’s factors than to parasite’s factors. This potential association between PKDL and host has already been suggested by Blackwell J.M.’s team. Indeed, results of this study proposed a genetic association between the polymorphism at IFNGR1 and the susceptibility of patients after treatments to PKDL (and not to visceral leishmaniasis) [51].

Regarding the reproductive strategy and population structure of these parasites, further studies should focus on the effect of individual hosts to detail the respective contribution of population differentiation as well as clonal, endogamic and outcrossing modes of reproduction in the genotypic distribution of these parasites. Nonetheless, clonality does not totally explain the strong variance across $F_{IS}$ loci, that displayed a much wider range than what was observed for the much more homozygous $L. braziliensis$ [21]. Our simulation approach suggested that obtaining the $F_{IS}$ and its variance across loci with very few MLGs, as in the real data, was impossible to achieve. The existence of a strongly structured hierarchical meta-population, with for instance the individual hosts playing the role of micro-populations for the parasites, in combination with occasional gene flow between different genetically distant entities (species hybridizations) and/or different cycle types (zoonotic vs. anthroponotic), could explain the pattern observed on our microsatellite loci. However, as previously said, this requires further investigation. We cannot exclude the possible disturbing role played by gene conversion known to occur in Leishmania [52] though we do not favour much such a hypothesis. If gene conversion is a genome wide process in Leishmania (genomic conversion) we would have expected a much more homogeneous homozygosity across loci than what was observed. Some loci are indeed almost always homozygous while some others display substantial amounts of heterozygosity (Figure 3). If gene conversion is site specific, we would expect it to preferentially affect coding sequences and its surrounding more than non coding zones. A glance at the localisation in the chromosome of markers did not suggest that microsatellite markers situated closer to coding sequences were more prone to display positive $F_{IS}$ than the other microsatellites. Moreover, even if the correlation between species is good, it can be seen that it is not perfect and that some loci with $F_{IS}$~1 in one species can display a fairly lower $F_{IS}$ in another. This does not strongly support the site specific DNA conversion hypothesis. But here again, further studies would be worth being undertaken on that issue.

Null alleles are often encountered in population genetics studies. They may be frequent in allozymes [53,54] and in DNA markers such as microsatellites [55,56,57]. In our data, no blank has ever been observed in the genotypes (no missing data, i.e. all individuals were amplified at all loci), which, given the high homozygosity encountered (increased probability of blank homozygotes), makes the null allele explanation very unlikely.

Rarity of MLGs, variable but globally positive $F_{IS}$ and strong heterogeneity within subsamples seem to be the rule for $L. donovani$ as such a pattern was already reported in Eastern Africa [19]. Such results suggest the existence of strongly differentiated hidden entities. A different pattern was found in $L. donovani$ from the Indian subcontinent [58] where all loci appeared weakly polymorphic, dominated by a single MLG with a few variants at one locus and, in spatially and temporally homogeneous subsamples no deviation from panmixia. Just as if this subcontinent had been colonised by one of the entities we are dealing with Africa.

The village Babar El Fugara is characterized by an epidemic context, with the occurrence of several epidemic episodes.
genetic diversity revealed by our results is not due to the arrival of a new variant but more likely was already present. Indeed, during this epidemic, all the population have been exposed to the disease and only $\frac{1}{4}$ develop visceral leishmaniasis. This observation means that the majority of the population is probably asymptomatic and contagious for the transmission [24]. In this context, this suggested the need to pursue research in order to identify which reservoir could be involved in the maintenance of the diversity and the transmission cycles (vectors or mammal reservoirs).

To conclude on this population genetics study within the *L. donovani* complex, it clearly appears that considering the whole sample as a single population was not adequate. In addition, our findings suggested that clonality may have a stronger impact on the *L. donovani* complex than on *L. braziliensis*. It also suggested that exploring the possible strong impact of the host individual (sandy or mammal hosts) was worth trying and indeed represents a too often neglected factor in *Leishmania* population studies in particular and in pathogenic microbes in general [22,33,59,60,61]. These results demonstrate the need to pursue population genetics studies in *Leishmania* species from sampling designs that control maximum possible confounding factors. These parasites indeed seem to be subdivided at very narrow spatio-temporal and ecological (host) scales.

**Supporting Information**

**Figure S1** $F_{IS}$ for each of the loci in the entire population of *L. donovani* complex. There is a large heterozygote deficiency at each locus. (THF)

**References**


11. Jamjoom MB, Ashford RW, Bates PA, Chance ML, Kemp SJ, et al. (2004) *Leishmania* donovani is the only cause of visceral leishmaniasis in East Africa; previous descriptions of *L. infantum* and “*L. archibaldi*” from this region are a consequence of convergent evolution in the isoenzyme data. Parasitology 129: 399–409.


13. Schonian G, Mauricio I, Cagolillo E (2010) Is it time to revise the nomenclature for providing the samples. The authors acknowledge a Dessen’s team for the strains and patient data collection. We also thank the French National Reference Center of *Leishmania* for providing the samples. The authors acknowledge the assistance of Dr. Linda Norup for the correction in English of the manuscript and the IFR-119 platform from Montpellier, France.

**Author Contributions**

Conceived and designed the experiments: VR TDM A-LB. Performed the experiments: VR. Analyzed the data: VR TDM A-LB. Contributed reagents/materials/analysis tools: VR TDM A-LB MH BB. Wrote the paper: VR TDM A-LB. Samples providers or collectors: BB AD SHE-S JD GLF.

**Checklist S1** STROBE checklist. Checklist of items included in this population genetic study. (PDF)

**Table S1** Description of data set and microsatellite genotypes. Each sample is detailed by sample code, species attribution by MLEE [25], host, clinical forms (VL for Visceral Leishmaniasis and PKDL for PostKala azar Dermatite Leishmaniasis) and year of collection, and microsatellite genotypes obtained at each locus. (XLS)

**Acknowledgments**

The authors acknowledge F. Prugnolle, F. Renaud and M. Choisy for helpful discussions and their assistance in the results. We acknowledge A. Dessen’s team for the strains and patient data collection. We also thank the French National Reference Center of *Leishmania* for providing the samples. The authors acknowledge the assistance of Dr. Linda Northup for the correction in English of the manuscript and the IFR-119 platform from Montpellier, France.

**References**


