RESEARCH   **Open Access**

Leishmania donovani populations in Eastern Sudan: temporal structuring and a link between human and canine transmission

Rania Baleela[1,3*], Martin S Llewellyn[1,4], Sinead Fitzpatrick[1], Katrin Kuhls[2,5], Gabriele Schönian[2], Michael A Miles[1] and Isabel L Mauricio[1,6]

Abstract

Background: Visceral leishmaniasis (VL), caused by the members of the *Leishmania donovani* complex, has been responsible for devastating VL epidemics in the Sudan. Multilocus microsatellite and sequence typing studies can provide valuable insights into the molecular epidemiology of leishmaniasis, when applied at local scales. Here we present population genetic data for a large panel of strains and clones collected in endemic Sudan between 1993 and 2001.

Methods: Genetic diversity was evaluated at fourteen microsatellite markers and eleven nuclear sequence loci across 124 strains and clones.

Results: Microsatellite data defined six genetic subpopulations with which the nuclear sequence data were broadly congruent. Pairwise estimates of F_{ST} (microsatellite) and K_{ST} (sequence) indicated small but significant shifts among the allelic repertoires of circulating strains year on year. Furthermore, we noted the co-occurrence of human and canine *L. donovani* strains in three of the six clusters defined. Finally, we identified widespread deficit in heterozygosity in all four years tested but strong deviation from inter-locus linkage equilibrium in two years.

Conclusions: Significant genetic diversity is present among *L. donovani* in Sudan, and minor population structuring between years is characteristic of entrenched, endemic disease transmission. Seasonality in vector abundance and transmission may, to an extent, explain the shallow temporal clines in allelic frequency that we observed. Genetically similar canine and human strains highlight the role of dogs as important local reservoirs of visceral leishmaniasis.

Background

Visceral leishmaniasis (VL) is caused by parasites of the *Leishmania donovani* complex. The *L. donovani* complex is distributed throughout Asia, North Africa, Latin America and Southern Europe, affecting mostly vulnerable and neglected populations. Infection is spread via the bite of haematophagous phlebotomine sand fly species, while the role of non-human reservoir hosts varies from region to region [1,2]. The most important endemic foci in terms of prevalence, morbidity and mortality are located in India, Sudan and Brazil. Leishmaniasis is likely to have been endemic to Sudan since antiquity (e.g. Zink et al. [3]). Epidemic outbreaks are periodically reported (e. g. Dereure et al. [4]) with high mortality (e.g. Seaman et al. [5]). Recent surveys of disease burden still show consistently high infection and mortality rates in Eastern Sudan, with up to 16% of all deaths attributed to VL regionally [6]. Infection rates in Sudan are thought to be seasonal, linked to moisture and sand fly abundance [7].

Molecular studies, such as the analysis of the ribosomal DNA internal transcribed spacer (ITS) [8,9], multilocus sequence typing (MLST) [10,11] and multilocus microsatellite typing (MLMT) [12], have shown that VL in Sudan, and the contiguous focus in Ethiopia, is caused by one to two genetic groups of *L. donovani*, distinct from *L. infantum* and other *L. donovani* genetic groups. Nevertheless, unlike on the Indian subcontinent, where

* Correspondence: rania.baleela@uofk.edu
[1]Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, UK
[3]Current address: Department of Zoology, Faculty of Science, University of Khartoum, PO Box 321, Khartoum, Sudan
Full list of author information is available at the end of the article

an emergent epidemic clone seems responsible for most cases, there is significant genetic diversity within Sudanese *L. donovani* [12-14]. More recently, MLMT typing of Sudanese *L. donovani* has focussed on the role genetic recombination might have in influencing local patterns of population genetic diversity [15]. Genetic recombination in the field and laboratory is increasingly reported within and between *Leishmania* species, with important consequences in terms of vector compatibility and the spread of drug resistance [16-19]. Several studies based on MLMT have used widespread homozygosity within populations as a proxy for inbreeding in *Leishmania*, in the face of widespread linkage disequilibrium and irrespective of whether parasites undergo 'classic' (Mendelian) gametic sex [15,20,21],

In the current study we evaluated the genetic diversity of *L. donovani* in Sudan using MLST and MLMT markers in parallel, with special focus on longitudinal patterns of parasite genetic diversity in the hyperendemic village Barbar El Fugara of the Atbara River Region and around it, 1993–2001. We successfully incriminated dogs as important reservoirs of *L. donovani* locally, by comparisons to local strains isolated from patients. Furthermore, we were able to show significant, but minor, subdivision between *L. donovani* isolated from different years based on MLMT and MLST, which we discuss in the light of VL epidemiology. We found evidence for excess homozygosity across all populations and associated linkage disequilibrium, but based on the available data we are unable to attribute this pattern of diversity to either genetic exchange (inbreeding) or gene conversion.

Methods
Ethical statement
Sampling in Barbar El Fugara was approved by both the Federal and Gedarif State Ministries of Health and by the Faculty of Medicine, Khartoum University. Informed consent was obtained from the district authorities and from the village committee as well as from all the adults who participated in the study. For younger children the consent was obtained from their parents. Other samples included in this analysis were archival or reference strains.

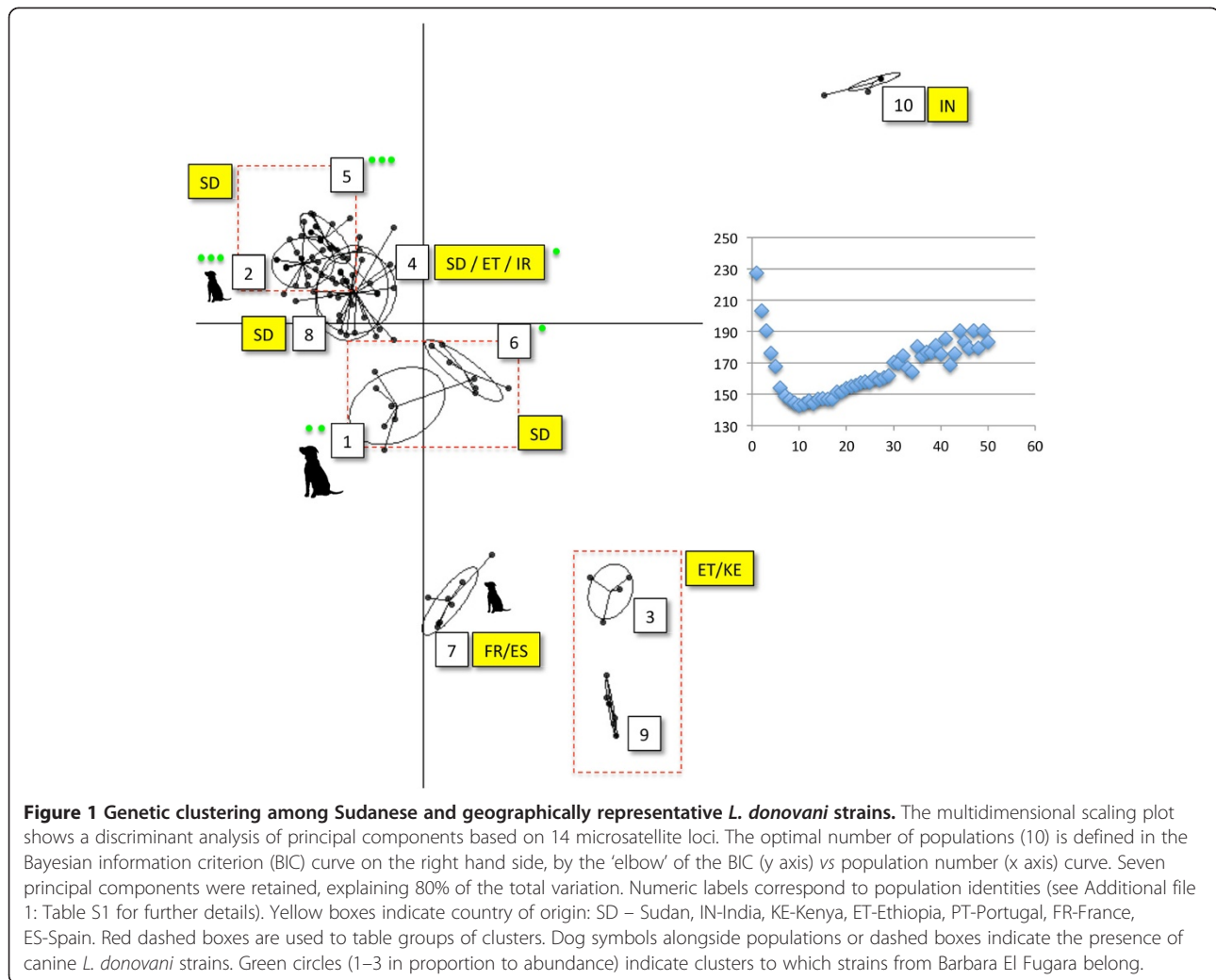### Strains, reference strains and clones
A panel of 124 *L. donovani* strains and clones was assembled (Additional file 1: Table S1). Twenty-three strains were biologically cloned (cultures founded from a single organism - one to four clones per strain) on solid media in 3.5 cm Petri dishes incubated at 24°C, using a protocol adapted from Yeo et al. [22]. All but one sample selected for cloning originated from the Atbara River region, and our aim was to facilitate the identification of local hybrids among contemporary circulating strains. As such, cloning

was undertaken to eliminate the possibility that heterozygous microsatellite loci or SNPs were the result of mixed infections and clones are indicated in Additional file 1: Table S1. Most strains were collected in the Atbara River region of Eastern Sudan in and around the village of Barbar El Fugara. Sudanese strains were collected from human and canine hosts over an eight-year period (1993–2001). However, further Sudanese samples prior to and after this period were also included for reference. In addition, a geographically representative selection of strains collected from Europe, East Africa and the Middle East was included for comparison. A subset of those strains sampled from Barbar El Fugara has been analysed previously (Additional file 1: Table S1) via MLMT but with different markers to those employed here [15].

Multilocus microsatellite typing and analysis
Fragment length analysis of 14 microsatellite markers was undertaken as previously described in Kuhls et al., [12] with the exception of locus CS19 which failed to amplify in our study. Positive controls (HU3 and DD8) and negative controls (i.e. reactions lacking DNA) were included in each set of PCR amplifications and subsequent analyses to ensure compatibility across data sets. To define *a posteriori* the number of putative populations in the data set using a non-parametric (free from Hardy-Weinberg constraints) approach, we employed a *K*-means clustering algorithm, implemented in adegenet [23]. As such, the 'true' number of populations can be defined by reference to the Bayesian Information Criterion (BIC), which reaches a minimum when the best-supported assignment of individuals to the appropriate number of clusters is approached. In practice, this number is selected at the 'elbow' of the BIC curve (Figure 1). The relationship between these clusters and the individuals within them was evaluated via a discriminant analysis of principal components (DAPC) [24]. We chose to retain the number of principal components (PCs) that represented the first 80% of the total variation in the data set. DAPC results are presented as multidimensional scaling plots in Figure 1. Individual level sample clustering was defined via a neighbour-joining tree based on pairwise distances between multilocus genotypes MLGs [evaluated using D_{AS} (1 – proportion of shared alleles at all loci/*n*)] calculated in MICROSAT ([25].

Population-level analyses of microsatellite data were undertaken exclusively on Sudanese *L. donovani* based on populations defined *a priori* by year 1993, 1997, 1998, and 2001 (Additional file 1: Table S1). First we undertook to estimate the level of gene flow between years in Arlequin v3.5 using F_{ST} (equivalent to Weir and Cockerman's 1984 estimator (θ_w) [26]) and tested this value for significance using a non-parametric random permutation procedure [27]. Secondly, we linearised these values as in Slatkin,

Figure 1 Genetic clustering among Sudanese and geographically representative *L. donovani* strains. The multidimensional scaling plot shows a discriminant analysis of principal components based on 14 microsatellite loci. The optimal number of populations (10) is defined in the Bayesian information criterion (BIC) curve on the right hand side, by the 'elbow' of the BIC (y axis) *vs* population number (x axis) curve. Seven principal components were retained, explaining 80% of the total variation. Numeric labels correspond to population identities (see Additional file 1: Table S1 for further details). Yellow boxes indicate country of origin: SD – Sudan, IN-India, KE-Kenya, ET-Ethiopia, PT-Portugal, FR-France, ES-Spain. Red dashed boxes are used to table groups of clusters. Dog symbols alongside populations or dashed boxes indicate the presence of canine *L. donovani* strains. Green circles (1–3 in proportion to abundance) indicate clusters to which strains from Barbara El Fugara belong.

1995, to facilitate direct comparison between values for population pairs [28]. Finally, we calculated population-specific statistics by year: sample size corrected allelic richness (A_r) in FSTAT 2.9.3.2 [29] and F_{IS} (an index of the distribution of heterozygosity within and between individuals), per locus per population, also in FSTAT 2.9.3.2. Tests for population-level deviation from Hardy-Weinberg allele frequencies were calculated in Arlequin v3.5 and associated significance levels for p values derived after sequential Bonferroni correction to minimise the likelihood of Type 1 errors. Linkage disequilibrium was defined via the Index of Association and calculated exclusively from biological clones in two populations.

Multilocus sequence typing and analysis

Direct DNA sequences were generated from the PCR amplification products of eleven MLST targets. Targets included four housekeeping genes previously identified as suitable markers: *asat-Ch24, asat-Ch35, fh-Ch24, gp63-mspC*) [10,30], and seven new targets developed in the current study. The new targets include a house-keeping gene *cytb5R-Ch22II* (LinJ.22.0590), 4 hypothetical protein-coding genes (*LinJ.01.0010, LinJ.28.0190, LinJ.34.0550* and *LinJ.36.1190*) and 2 pseudogenes (*LinJ.11.0280* and *LinJ.36.0350*). PCR reactions to amplify these targets were undertaken in a final volume of 25 μl, comprising: 2.5 μl 10× buffer, 0.05 mM MgCl₂, 2.5 μl 0.8 mM dNTPs, 12.5 pmol of each primer, 1.25 U *Taq* polymerase and 25 ng genomic DNA. Amplification conditions were: 30 cycles at 95°C for 1 min, 54–62°C (dependent on primer (Additional file 2: Table S2)) for 1.5 min and 72°C for 1.5 min, with a final extension at 72°C for 10 min. Amplification of targets *LinJ.01.0010, LinJ.11.0280, LinJ.34.0550, LinJ.36.0350* and *LinJ.36.1190* required 10% DMSO. Direct sequencing reactions were performed with internal primers, BigDye™ terminator cycle sequencing V3.1 kits (ABI PRISM® Applied Biosystems) and analysed in an ABI PRISM™ 3730 DNA sequencer (Applied Biosystems), according to the manufacturer's instructions. Sequences
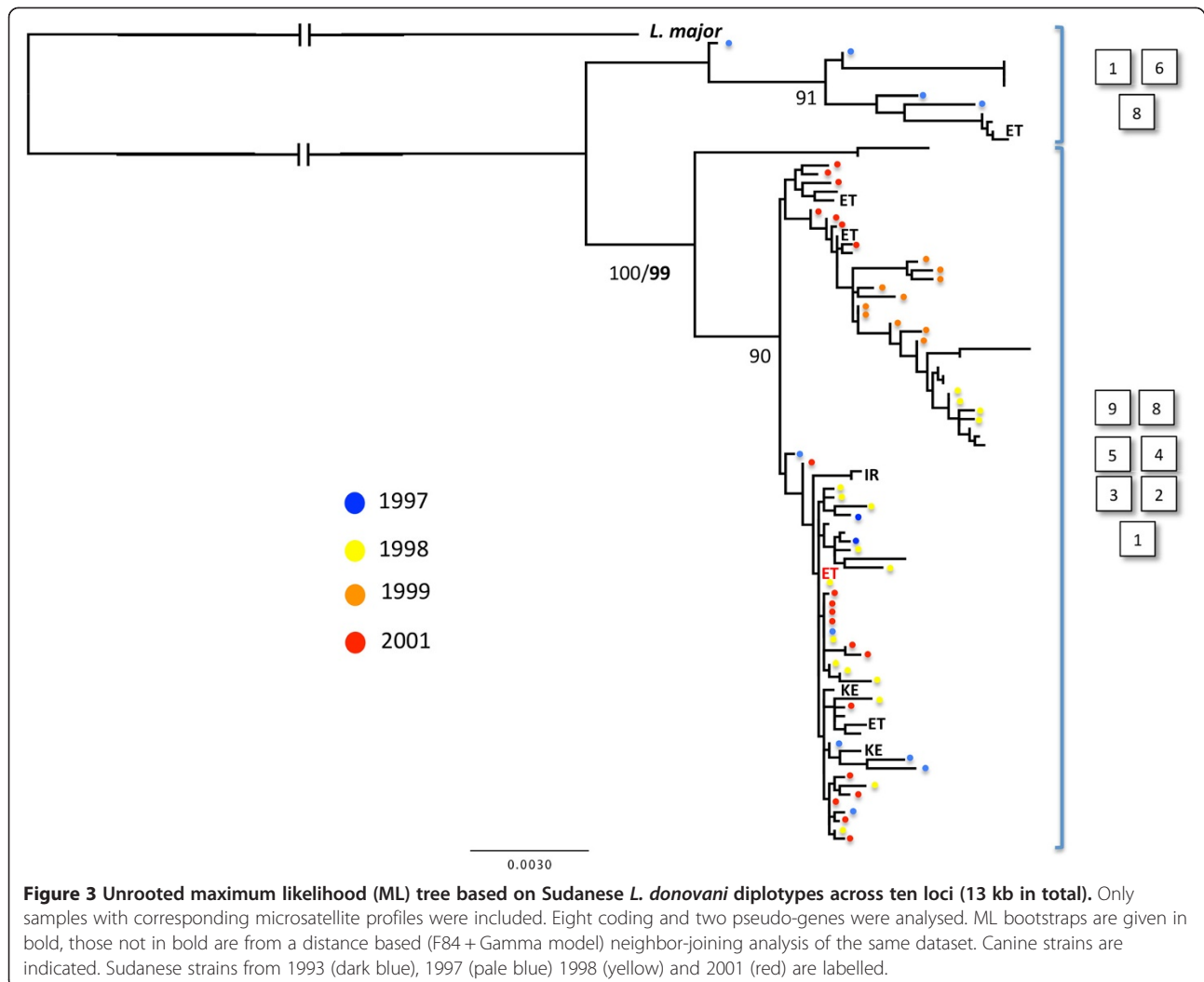
were inspected and edited visually in Chromas Lite (Copyright © 2005 Technelysium Pty Ltd) and assembled using ClustalW [31] in BioEdit v 5.0.6 [32]. Haplotype phases were reconstructed using the software PHASE v. 2.1.1 [33]. MLST data were analyzed as two separately concatenated target haplotype datasets: coding and hypothetical (10995 bp in total) and non-coding pseudogenes (2173 bp in total). Genome sequences of the reference strains *L. major* MHOM/IL/80/Friedlin and *L. infantum* MCAN/ES/98/LLM-877 were obtained from www.genedb.org. Haplotypes for both coding and non-coding sequence datasets were scanned for mosaic breakpoints in RDP [34]. Sequences for each gene were submitted to Genbank (Accession numbers: FR775540.1-FR775754.1, FR796277.1-FR846 362.2, HE648217.1-HE648270.1).

Phylogenetic and population genetic analysis of various sequence and population sets were undertaken. Phylogenies were inferred using Maximum-Likelihood (ML) implemented in PhyML (4 substitution rate categories) [35]. The best-fit model of nucleotide substitution was selected from 88 models and its significance evaluated according to the Akaike Information Criterion (AIC) in MODELTEST 1.0 [36]. Bootstrap support for clade topologies was estimated following the generation of 100 pseudo-replicate datasets. Neighbour-joining phylogenies were also inferred from the same alignments, and bootstrap support over 1000 pseudo-replicates on congruent clades added to the ML phylogeny. The resultant trees are shown in Figures 2 and 3.

Population genetic differentiation between strains from years 1997, 1998, 1999 and 2001 was calculated from
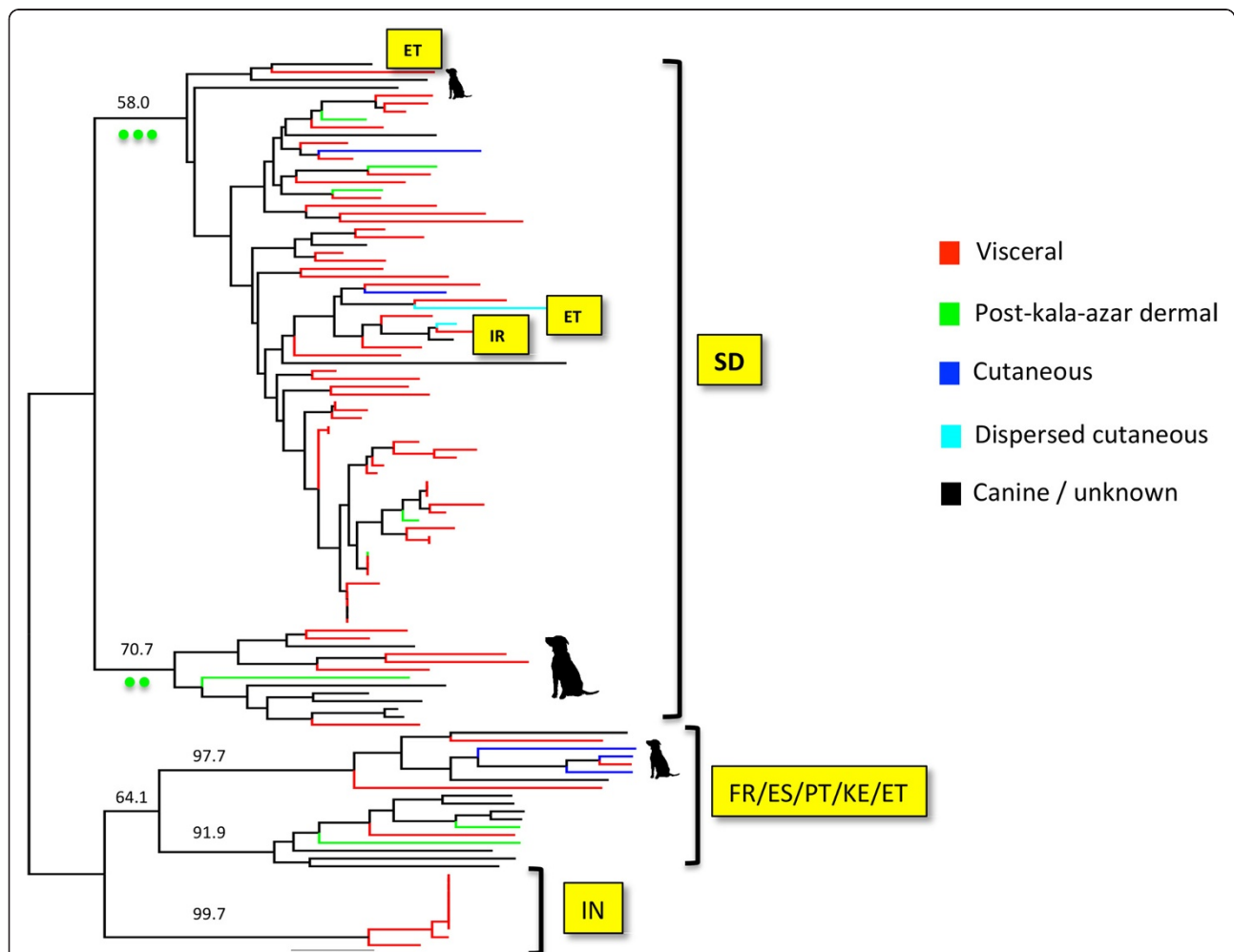
Figure 2 Maximum likelihood (ML) tree based on phased and concatenated *L. donovani* haplotypes across eight coding loci (10.9 kb in total). The ML substitution model adopted was the General Time Reversible plus Gamma with Invariable sites. ML bootstrap support is given in bold. Non-bold bootstraps are derived from the distance based (F84 + Gamma model) neighbor-joining tree from the same dataset. Non-Sudanese strains are identified by country code: SD – Sudan, IN-India, KE-Kenya, ET-Ethiopia, PT-Portugal, FR-France, ES-Spain. Sudanese strains from 1997 (blue), 1998 (yellow), 1999 (orange) and 2001 (red) are labelled. Population codes on the left hand site correspond to those identified via microsatellites (Figure 1).
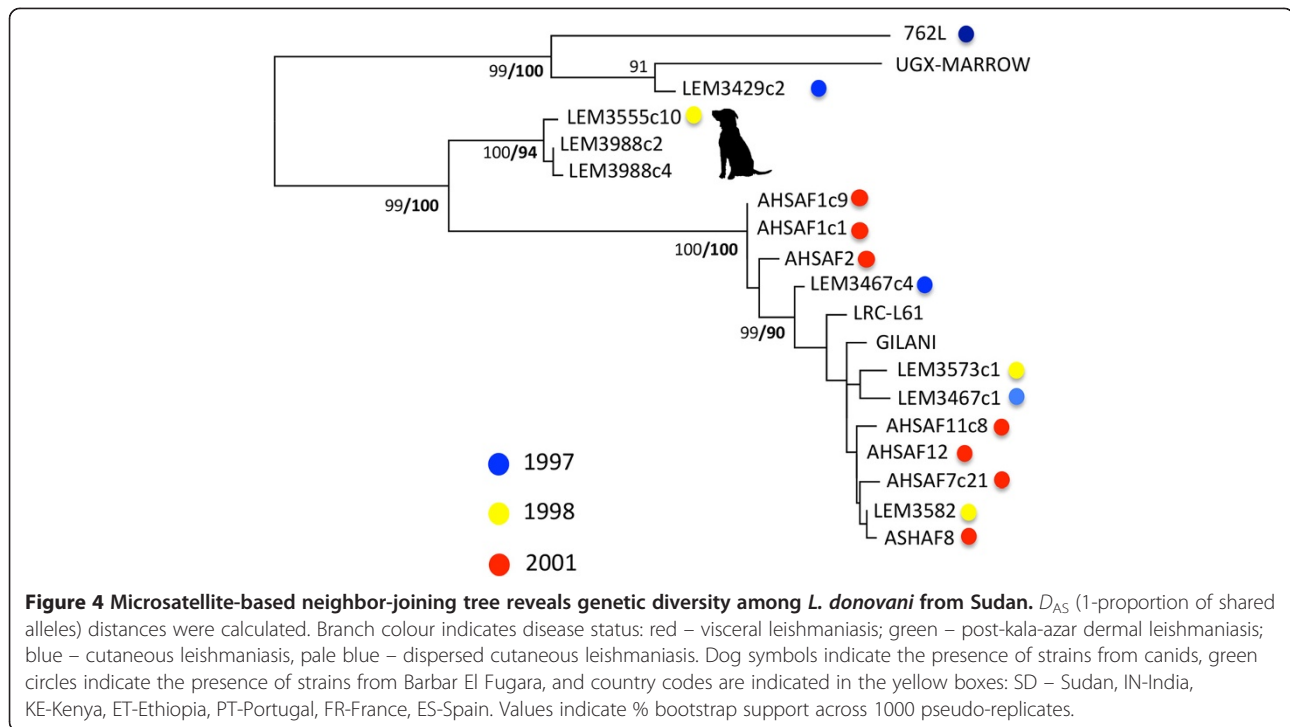
Figure 3 Unrooted maximum likelihood (ML) tree based on Sudanese *L. donovani* diplotypes across ten loci (13 kb in total). Only samples with corresponding microsatellite profiles were included. Eight coding and two pseudo-genes were analysed. ML bootstraps are given in bold, those not in bold are from a distance based (F84 + Gamma model) neighbor-joining analysis of the same dataset. Canine strains are indicated. Sudanese strains from 1993 (dark blue), 1997 (pale blue) 1998 (yellow) and 2001 (red) are labelled.

sequence haplotypes in DNAsp [37] using K_{ST}. K_{ST} compares the expected number of nucleotide differences between a pair of sequences within one population with a pair taken across all populations [38]. Statistical significance for observed differentiation was inferred via 10,000 random permutations.

Results

Microsatellite and DNA sequence data were generated across 124 strains and clones. Twenty-two strains were genotyped using both marker classes. Ninety-two were typed using microsatellite markers only, ten with MLST markers only. K-means clustering of the 112 microsatellite profiles revealed ten populations (Figure 1). The majority of Sudanese strains fell into six related populations (2,5,8,4,1,6), while strains from Europe (7) and the Indian subcontinent (10) were clear outliers. A subset of strains from Ethiopia and Iran shared some genetic affinity with those from Sudan. Some other strains from East Africa (ET/KE - 3) were, by contrast, highly divergent.

Importantly, canine *L. donovani* from Barbar El Fugara, Sudan, were found alongside strains from humans in four populations (1,2.5,6) across several different sampling years (1997, 1998, 1999, 2000, Additional file 1: Table S1). Clustering observed in Figure 1 was supported by the topology of the neighbour joining tree in Figure 4. Furthermore, no clustering of distinct disease outcome (cutaneous, visceral, post-kala-azar dermal or diffuse cutaneous leishmaniasis, respectively, CL, VL, PKDL or DCL) was observed (Figure 4).

For population level analyses of the microsatellite dataset we focused on only four well represented years from which samples were available: 1993, 1997, 1998 & 2001. Samples from the latter three time points were isolated uniquely from the village of Barbar El Fugara in the hyper-endemic Atbara River region of Eastern Sudan. Estimates for allelic richness over this time period show few radical differences between populations year-on-year (Table 1). Similarly, values for the inbreeding co-efficient F_{IS} were consistently positive over all loci over the years

Figure 4 Microsatellite-based neighbor-joining tree reveals genetic diversity among *L. donovani* from Sudan. D_{AS} (1-proportion of shared alleles) distances were calculated. Branch colour indicates disease status: red – visceral leishmaniasis; green – post-kala-azar dermal leishmaniasis; blue – cutaneous leishmaniasis, pale blue – dispersed cutaneous leishmaniasis. Dog symbols indicate the presence of strains from canids, green circles indicate the presence of strains from Barbar El Fugara, and country codes are indicated in the yellow boxes: SD – Sudan, IN-India, KE-Kenya, ET-Ethiopia, PT-Portugal, FR-France, ES-Spain. Values indicate % bootstrap support across 1000 pseudo-replicates.

(Table 1). The index of association, a measure of linkage disequilibrium, was calculated only for years for which ≥8 biological clones were available (1997 and 2001). In both cases, the null hypothesis of random inter-locus associations was strongly rejected (Table 1, *P* <0.0001).

As well as calculating population diversity statistics within each time point, we were interested in examining the extent of parasite population differentiation longitudinally. We therefore calculated values for pairwise population F_{ST} and tested for associated significance using a permutation test. Significant structure was detected between all population pairs except years 1997 and 1998 (Table 2). Interestingly, once values for F_{ST} had been linearised using Slatkin's correction, the extent of subdivision between time points increased in proportion to the time elapsed between sample collections dates (Table 3).

In parallel to our analysis of microsatellite fragment size data, we also undertook analysis of DNA sequence data derived from a representative group of 34 strains and clones (including genomic reference strains). Both

coding and non-coding regions were scanned for evidence of mosaic breakpoints that might be associated with homologous recombination, which would also potentially disrupt phylogenetic signal in subsequent trees, but no evidence for such events was uncovered. The ML topology derived from coding loci revealed substantial genetic diversity, but little bootstrap support, as one might expect between samples from the same species across a restricted area (Sudan/East Africa, Figure 2). The only clearly divergent clade contained samples classified as population 6, 8 and 1 based on microsatellite typing, as well as a single strain from Ethiopia. Correspondence between sample year and tree topology was limited (Figure 2). As a second approach we examined only those strains for which we had microsatellite data. To improve resolution, we concatenated both coding and non-coding genes and constructed an ML tree from unphased sequence haplotypes (Figure 4). In this case there was a closer match between microsatellite and sequence data. Strains from MLMT-defined populations 6 and 1 were outliers with respect to other Sudanese strains.

Table 1 Year on year population genetic statistics for Sudanese *L. donovani* populations

Year	N	Ar	F_{IS}	%LociHd	%LociHe	Ia
1993	11	2.643 ± 0.289	0.201 ± 0.150	8.3	8.3	ND
1997	11	3.071 ± 0.355	0.596 ± 0.073	46.1	0	ND
1998	14	3.144 ± 0.339	0.479 ± 0.107	46.1	0	2.95 < 0.001
2001	27	2.430 ± 0.291	0.275 ± 0.172	0.33	0.25	2.03 < 0.001

Index of association only calculated for clones.

Table 2 Pairwise F_{ST} between years suggests incremental shifts in allelic frequencies in Sudanese populations *L. donovani*

	1993	1997	1998	2001
1993		*0.000**	*0.000**	*0.000**
1997	**0.152**		*0.563*	*0.003**
1998	**0.213**	0.008		*0.003**
2001	**0.297**	**0.097**	**0.074**	

In bold: statistically significant values of pairwise F_{ST}. In italics, upper left triangle: p values. *represents statistically significant p values.

Table 4 Genetic differentiation (K_{ST}) between years based on concatenated coding sequence data

	1997	1998	1999
1998	0.115 (0.0090**)		
1999	0.2077 (0.000**)	0.47319 (0.000**)	
2001	0.1079 (0.0050**)	0.03672 (0.0380*)	0.37466 (0.000**)

P-values (in parentheses) are made with reference to 10000 random permutations. Statistical significance: *P < 0.05 and **P < 0.001.

A notable exception was strain 762 L, which grouped differently between the two sets of markers.

For the sequence data, population genetic analyses were undertaken to explore patterns of diversification across years in samples only from Barbar El Fugara. Among those samples sequenced, the distribution of available data per year was marginally different. Years 1997, 1998, 1999 and 2001 were compared. Pairwise permutations tests suggested significant K_{ST} for most population pairs, except 1998/2001, where significance was marginal (*p* =0.0380) (Table 4).

Discussion

Visceral leishmaniasis in Sudan is a major and on-going public health problem [6]. Molecular epidemiological studies like ours can have a significant role in guiding and informing public health professionals. Our first key observation in this context is that dogs and humans in the region share similar strains. PCR-based and parasitological approaches have already identified dogs as important carriers of *L. donovani* in Sudan [1,4], although circumstantial evidence also points to other truly sylvatic hosts (e.g. [39]). Our high-resolution genetic data clearly demonstrate sharing of parasites between dogs and humans. Previous work on a limited number of the same strains from the same area suggested the possible presence of distinct human and canine transmission cycles [15], however, all three clusters containing canine hosts also contained humans (Figure 1).

The stability of genetic diversity in parasite populations in space is frequently used to infer patterns of regional and global parasite spread (e.g. [40]). Temporal variation in parasite populations can also be highly informative, especially pre- and post- large scale treatment

Table 3 Pairwise F_{ST} linearised with Slatkin correction

	1993	1997	1998
1993			
1997	**0.17997**		
1998	**0.27118**	0.00844	
2001	**0.42205**	**0.1072**	**0.07969**

In bold: statistically significant values of pairwise F_{ST}.

interventions (e.g. [41,42]. The majority of samples we analysed came from an outbreak first reported in 1996 [4]. Given that high rates of infection still occur in the same region today, it is not clear whether 'outbreak' successfully described the diseases' local status [6]. Both our sequence and microsatellite data from different years suggest incremental changes in allelic composition (although, like in an earlier study, no subdivision is detected between years 1997 and 1998 [15]). Mutational instability of highly variable microsatellite markers could play a role. However, it is not clear over what timescale such changes might be expected to happen. In *Trypanosoma cruzi* discrete typing unit I, a related trypanosomatid, two samples taken 20 years apart from the same geographic focus can be identical at 48 microsatellite loci [43]. There are multiple examples in the current dataset where temporally separated strains are closely related to each other. Population 4, for example contains samples from 1967 and 2001. It is, thus, more likely that population processes, such as immigration, founder events and bottlenecks, define the differences between years. However, the shallow clines in allelic composition we observe in the data are not reminiscent of intense serial reductions in parasite population size. Interpopulation variation is perhaps more consistent with seasonal changes in infection intensity.

As well as the defining patterns of parasite genetic diversification in the Atbara River region, a secondary goal was to evaluate evidence for genetic exchange among Sudanese *L. donovani* strains. Like previous authors, we were able to detect reduced heterozygosity in the populations studied [15]. However, unlike other authors, we are reluctant to interpret our data as evidence for genetic exchange [15,20,21]. Analysis of cloned *L. donovani* from two populations revealed strong evidence for predominant clonality, despite consistently high values for F_{IS}. Furthermore, sequence data showed no evidence for the mosaics that one might expect to accompany recombination. It is important to state that, although our data do not confirm the presence of genetic exchange, we cannot rule out the occurrence of some recombination/inbreeding, as suggested by Rougeron et al. [15]. We note that extensive genomic-level hybridisation was recently detected among a population of *Leishmania*

infantum in Turkey [16], while microsatellite data based on the same strains detected no such phenomenon [44].

Conclusion

Rapid, low cost, high-resolution genotyping strategies have an important role in elucidating the molecular epidemiology of visceral leishmaniasis, especially where the burden of the disease is felt the most. Population genomic studies of *Leishmania* have now demonstrated the power that such approaches have to reveal the extent and mechanism of genetic exchange in natural populations [16]. It has become apparent that genetic exchange is not a rare event but a feature of natural populations of several *Leishmania* species [16,20,21]; experimental crosses in sand flies suggest Mendelian segregation [17]. Although proof of genetic exchange was not evident among the Sudanese populations analysed here, excess homozygosity occurred in conjunction with LD, and this was interpreted by others as inbreeding [20]. Furthermore, some foci of human and canine *L. donovani* transmission were coincident with overlapping *L. donovani* genotypes, indicating that dogs may have an important role in sustaining human VL in Sudan, which deserves further investigation.

Additional files

> **Additional file 1: Table S1.** Panel of clones and strains analysed in this study.
>
> **Additional file 2: Table S2.** Primers and annealing temperature for PCR amplification of the new MLST targets.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
RB, SF, KK, and MSL analysed the data. GS, MAM, and ILM, conceived the experiments. RB MSL MAM and ILM wrote the paper. All authors read and approved the final manuscript.

Author details
[1]Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT London, UK. [2]Institut für Mikrobiologie und Hygiene, Charité Universitätsmedizin, Berlin, Germany. [3]Current address: Department of Zoology, Faculty of Science, University of Khartoum, PO Box 321, Khartoum, Sudan. [4]Current address: Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, University of Wales, Bangor, Deiniol Road, Bangor, Gwynedd LL57 2UW, UK. [5]Current address: Division of Molecular Biotechnology and Functional Genetics, Technical University of Applied Sciences Wildau, Hochschulring 1, 15745 Wildau, Germany. [6]Current address: Instituto de Higiene e Medicina Tropical/Unidade de Parasitologia e Microbiologia Médicas, UEI Parasitologia Médica, Rua da Junqueira, 100, 1349-008 Lisbon, Portugal.

References

1. Hassan MM, Osman OF, El-Raba'a FM, Schallig HD, Elnaiem DE: **Role of the domestic dog as a reservoir host of *Leishmania donovani* in eastern Sudan.** *Parasit Vectors* 2009, **2**:26.
2. Singh N, Mishra J, Singh R, Singh S: **Animal reservoirs of visceral leishmaniasis in India.** *J Parasitol* 2013, **99**:64–67.
3. Zink AR, Spigelman M, Schraut B, Greenblatt CL, Nerlich AG, Donoghue HD: **Leishmaniasis in Ancient Egypt and Upper Nubia.** *Emerg Infect Dis* 2006, **12**:1616–1617.
4. Dereure J, El-Safi SH, Bucheton B, Boni M, Kheir MM, Davoust B, Pratlong F, Feugier E, Lambert M, Dessein A, Dedet JP: **Visceral leishmaniasis in eastern Sudan: parasite identification in humans and dogs; host-parasite relationships.** *Microbes Infect* 2003, **5**:1103–1108.
5. Seaman J, Mercer AJ, Sondorp E: **The epidemic of visceral leishmaniasis in western Upper Nile, southern Sudan: course and impact from 1984 to 1994.** *Int J Epidemiol* 1996, **25**:862–871.
6. Mueller YK, Nackers F, Ahmed KA, Boelaert M, Djoumessi JC, Eltigani R, Gorashi HA, Hammam O, Ritmeijer K, Salih N, Worku D, Etard JF, Chappuis F: **Burden of visceral leishmaniasis in villages of eastern Gedaref State, Sudan: an exhaustive cross-sectional survey.** *PLoS Negl Trop Dis* 2012, **6**:e1872.
7. Zijlstra EE, el-Hassan AM: **Leishmaniasis in Sudan: Visceral leishmaniasis.** *Trans R Soc Trop Med Hyg* 2001, **95**(Suppl 1):S27–S58.
8. Kuhls K, Mauricio IL, Pratlong F, Presber W, Schonian G: **Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex.** *Microbes Infect* 2005, **7**:1224–1234.
9. Mauricio IL, Howard MK, Stothard JR, Miles MA: **Genetic diversity in the *Leishmania donovani* complex.** *Parasitology* 1999, **119**:237–246.
10. Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, Zemanova E, Dedet JP, Lukes J, Miles MA: **Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD).** *Int J Parasitol* 2006, **36**:757–769.
11. Zemanova E, Jirku M, Mauricio IL, Horak A, Miles MA, Lukes J: **The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing.** *Int J Parasitol* 2007, **37**:149–160.
12. Kuhls K, Keilonat L, Ochsenreither S, Schaar M, Schweynoch C, Presber W, Schönian G: **Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis.** *Microbes Infect* 2007, **9**:334–343.
13. Alam MZ, Kuhls K, Schweynoch C, Sundar S, Rijal S, Shamsuzzaman AK, Raju BV, Salotra P, Dujardin JC, Schönian G: **Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of *Leishmania donovani* strains in the Indian subcontinent.** *Infect Genet Evol* 2009, **9**:24–31.
14. Gelanew T, Kuhls K, Hurissa Z, Weldegebreal T, Hailu W, Kassahun A, Abebe T, Hailu A, Schönian G: **Inference of population structure of *Leishmania donovani* strains isolated from different Ethiopian visceral leishmaniasis endemic areas.** *PLoS Negl Trop Dis* 2010, **4**:e889.
15. Rougeron V, De Meeûs T, Hide M, Le Falher G, Bucheton B, Dereure J, El-Safi SH, Dessein A, Bañuls AL: **Multifaceted population structure and reproductive strategy in *Leishmania donovani* complex in one Sudanese village.** *PLoS Negl Trop Dis* 2011, **5**:e1448.
16. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M, Svobodova M, Volf P, Berriman M, Cotton JA, Smith DF: **Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated *Leishmania* population.** *PLoS Genet* 2014, **10**:e1004092.
17. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, Dobson DE, Beverley SM, Sacks DL: **Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector.** *Science* 2009, **324**:265–268.
18. Volf P, Benkova I, Myskova J, Sadlova J, Campino L, Ravel C: **Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids.** *Int J Parasitol* 2007, **37**:589–593.
19. Sadlova J, Yeo M, Seblova V, Lewis MD, Mauricio I, Volf P, Miles MA: **Visualisation of *Leishmania donovani* fluorescent hybrids during early stage development in the sand fly vector.** *PLoS One* 2011, **6**:e19851.

20. Rougeron V, De Meeus T, Hide M, Waleckx E, Bermudez H, Arevalo J, Llanos-Cuentas A, Dujardin JC, De Doncker S, Le Ray D, Ayala FJ, Bañuls AL: Extreme inbreeding in *Leishmania braziliensis*. *Proc Natl Acad Sci U S A* 2009, **106**:10224–10229.

21. Rougeron V, Banuls AL, Carme B, Simon S, Couppié P, Nacher M, Hide M, De Meeûs T: Reproductive strategies and population structure in *Leishmania*: substantial amount of sex in *Leishmania Viannia guyanensis*. *Mol Ecol* 2011, **20**:3116–3127.

22. Yeo M, Lewis MD, Carrasco HJ, Acosta N, Llewellyn M, da Silva Valente SA, de Costa Valente V, de Arias AR, Miles MA: Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium. *Int J Parasitol* 2007, **37**:111–120.

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

24. Jombart T, Devillard S, Balloux F: Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 2010, **11**:94.

25. Minch E, Ruíz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L: *MICROSAT - the Microsatellite Distance Program*. Stanford: Stanford University Press; 1995.

26. Weir B, Cockerham C: Estimating F-statistics for the analysis of population structure. *Evolution* 1984, **38**:1358–1370.

27. Excoffier L, Lischer HE: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010, **10**:564–567.

28. Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995, **139**:457–462.

29. Goudet J: FSTAT version 1.2: a computer program to calculate F-statistics. *J Heredity* 1995, **86**:485–486.

30. Mauricio IL, Gaunt MW, Stothard JR, Miles MA: Glycoprotein 63 (gp63) genes show gene conversion and reveal the evolution of Old World *Leishmania*. *Int J Parasitol* 2007, **37**:565–576.

31. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994, **22**:4673–4680.

32. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999, **41**:95–98.

33. Stephens M, Smith N, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001, **68**:978–989.

34. Martin D, Williamson C, Posada D: RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 2005, **21**:260–262.

35. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010, **59**:307–321.

36. Posada D, Crandall KA: MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998, **14**:817–818.

37. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**:1451–1452.

38. Hudson RR, Boos DD, Kaplan NL: A statistical test for detecting geographic subdivision. *Mol Biol Evol* 1992, **9**:138–151.

39. Elnaiem DA, Hassan MM, Maingon R, Nureldin GH, Mekawi AM, Miles M, Ward RD: The Egyptian mongoose, *Herpestes ichneumon*, is a possible reservoir host of visceral leishmaniasis in eastern Sudan. *Parasitology* 2001, **122**:531–536.

40. Lukes J, Mauricio IL, Schonian G, Dujardin JC, Soteriadou K, Dedet JP, Kuhls K, Tintaya KW, Jirků M, Chocholová E, Haralambous C, Pratlong F, Obrník M, Horák A, Ayala FJ, Miles MA: Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc Natl Acad Sci U S A* 2007, **104**:9375–9380.

41. Gosi P, Lanteri CA, Tyner SD, Se Y, Lon C, Spring M, Char M, Sea D, Sriwichai S, Surasri S, Wongarunkochakorn S, Pidtana K, Walsh DS, Fukuda MM, Manning J, Saunders DL, Bethell D: Evaluation of parasite subpopulations and genetic diversity of the msp1, msp2 and glurp genes during and following artesunate monotherapy treatment of Plasmodium falciparum malaria in Western Cambodia. *Malar J* 2013, **12**:403.

42. Norton AJ, Gower CM, Lamberton PH, Webster BL, Lwambo NJ, Blair L, Fenwick A, Webster JP: Genetic consequences of mass human chemotherapy for *Schistosoma mansoni*: population structure pre- and post-praziquantel treatment in Tanzania. *Am J Trop Med Hyg* 2010, **83**:951–957.

43. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, Vargas J, Torrico F, Diosque P, Valente V, Valente SA, Gaunt MW: Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathog* 2009, **5**:e1000410.

44. Gouzelou E, Haralambous C, Amro A, Mentis A, Pratlong F, Dedet JP, Votypka J, Volf P, Toz SO, Kuhls K, Schönian G, Soteriadou K: Multilocus microsatellite typing (MLMT) of strains from Turkey and Cyprus reveals a novel monophyletic *L. donovani sensu lato* group. *PLoS Negl Trop Dis* 2012, **6**:e1507.