INSTITUT PASTEUR

Microbes and Infection 9 (2007) 334–343

www.elsevier.com/locate/micinf

Original article

Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis

Katrin Kuhls*, Lyvia Keilonat, Sebastian Ochsenreither, Matthias Schaar, Carola Schweynoch, Wolfgang Presber, Gabriele Schönian

Institute of Microbiology and Hygiene (Parasitology), Charité Universitätsmedizin Berlin, Dorotheenstr. 96, 10117 Berlin, Germany

Received 27 October 2006; accepted 1 December 2006
Available online 9 January 2007

Abstract

Multilocus enzyme electrophoresis (MLEE) is the gold standard for taxonomy and strain typing of *Leishmania*, but has some limitations. An alternative reliable and fast genotyping method for addressing population genetic and key epidemiological questions, is multilocus microsatellite typing (MLMT). MLMT using 15 markers was applied to 91 strains of *L. donovani*, *L. archibaldi*, *L. infantum* and *L. chagasi* from major endemic regions of visceral leishmaniasis. Population structures were inferred by combination of Bayesian model-based and distance-based approaches. Six main genetically distinct populations were identified: (1) *L. infantum*/*L. chagasi* MON-1 and (2) *L. infantum*/*L. chagasi* non-MON-1, both Mediterranean region/South America; (3) *L. donovani* (MON-18), *L. archibaldi* (MON-82), *L. infantum* (MON-30, 81) and (4) *L. donovani* (MON-31, 274), *L. archibaldi* (MON-82, 257, 258), *L. infantum* (MON-267), both Sudan/Ethiopia; (5) *L. donovani* MON-2, India; (6) *L. donovani* (MON-36, 37, 38), Kenya and India. Substructures according to place and time of strain isolation were detected. The VL populations seem to be predominantly clonal with a high level of inbreeding. Allelic diversity was highest in the Mediterranean region, intermediate in Africa and lowest in India. MLMT provides a powerful tool for global taxonomic, population genetic and epidemiological studies of the *L. donovani* complex.

Keywords: *Leishmania donovani*; *Leishmania archibaldi*; *Leishmania infantum*; Visceral leishmaniasis; Kala-Azar; Population genetics; Sudan; Microsatellites; Genotyping; Epidemiology

1. Introduction

Species of the *Leishmania donovani* complex are causative agents of visceral leishmaniasis (VL) and post-Kala-Azar dermal leishmaniasis (PKDL) affecting approximately half a million people annually. The current taxonomy and strain identification is based on multilocus enzyme analysis (MLEE) [1,2]. Several regions are endemic for VL with different species linked to the respective geographical locations. *Leishmania donovani* Laveran and Mesnil, 1903 (*sensu stricto*) is endemic in three regions: (1) Indian subcontinent; (2) China; (3) Sudan, Ethiopia, Kenya. *Leishmania archibaldi* Castellani and Chalmers, 1919 is known from Sudan/Ethiopia. *Leishmania infantum* Nicolle, 1908 (Old World)/*L. chagasi* Cunha and Chagas, 1937 (New World) is distributed over four areas: (1) the Mediterranean basin to Iran, (2) China, (3) Sudan, Ethiopia, (4) Central and South America. *L. chagasi* is considered a synonym for *L. infantum* [1,3]. *L. archibaldi* was originally applied to strains from Sudan and later (1936), synonymized with *L. donovani*. In 1963 it was raised to subspecies or even species level, and applied to all East African parasites. In 1990, a single zymodeme (MON-82) was assigned to *L. archibaldi* [1], two other zymodemes were described 11 years later [2]. The taxonomic

* Corresponding author. Tel.: +49 30 450524028; fax: +49 30 450524902.
E-mail address: katrin.kuhls@charite.de (K. Kuhls).

1286-4579/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.micinf.2006.12.009

position of *L. archibaldi* was, however, not clearly established by MLEE, since the differentiation between *L. donovani*, *L. infantum* and *L. archibaldi* is based on differences in a single enzyme (glutamate oxaloacetate transaminase; GOT).

MLEE is laborious, slow and depends on mass parasite culturing. It has several other drawbacks, e.g. synonymous nucleotide substitutions may not be observed, distinct allozymes can have coincident mobilities, influence of posttranslational modifications, and restricted discriminatory power. Genotype-based methods are an alternative. Recent studies analysing rDNA internal transcribed spacers [4,5], *gp63* [6,7], mini-exon [4], cysteine proteinase B [7], PCR-fingerprinting [8], SCAR [9], mitochondrial cytochrome oxidase II [10] led to results conflicting the isoenzyme-based taxonomy, especially with regard to East African *L. archibaldi* and *L. infantum*. Most of them analyzed only a few strains of different geographic origin or strains from a single origin. Only a few studies included representative sets of strains from the main endemic regions to provide a more complex picture of the *L. donovani* complex [5,6,11]. Neither MLEE nor the genetic approaches, with the exception of kDNA-based markers, were however, able to discriminate closely related strains [12].

Highly polymorphic and co-dominant microsatellite markers have proven very useful for population studies. There are only few reports on microsatellite markers for *Leishmania* [13–15]. With the exception of a recent study on *L. tropica* [15], no report has applied a population-genetic approach. Because of the high polymorphism of microsatellite sequences, new markers have to be developed for nearly every *Leishmania* species. Different panels of microsatellite markers have been reported recently for the *L. donovani* complex. Bulle et al. [13] found 10 microsatellites for *L. infantum* on three discrete loci; sequencing was, however required for analysing variation in repeat numbers. Polymorphic microsatellite markers for the *L. donovani* complex suitable for fragment size analysis have only been developed very recently [14,16].

In this study, for the first time a population genetic approach was conducted for 91 strains of the *L. donovani* complex using 15 microsatellite markers. Sample sets were representative for the main endemic regions as East Africa, India, and the Mediterranean region. In order to understand the population structure of the *L. donovani* complex the degree of polymorphism between and within each endemic region and the possible existence of subpopulations have been assessed.

2. Materials and methods

2.1. Parasite cultures and DNA extraction

Sources, designation and geographical origins of the 91 *Leishmania* strains studied are listed in Table 1. Three of the main endemic regions (East Africa, India, Mediterranean region) were represented by a number of strains, suitable for population genetics studies. We included all zymodemes predominating in these regions and, where possible, all known zymodemes, as in the case of Sudan. All parasites were cultivated as described previously, and DNA was isolated using proteinase K-phenol/chloroform extraction [8].

2.2. PCR amplification assays of the microsatellite markers

Thirteen of the 15 microsatellite markers (Table 2) used (Li41-56, Li46-67, Li22-35, Li23-41, Li45-24, Li71-33, Li71-5/2, Li71-7, TubCA, Lm4TA, TubCA, LIST7031 and LIST7039) have been described previously [14,16]. Markers CS19 (fw 5′-TGC AGC AGA TGA GAA GGA AA/rev 5′-GTG AGA AGG CAG GGA TTC AA, $T_{an} = 56\,^{\circ}C$) and CS20 (fw 5′-CGT TGG CTG TTG ATT GTG-TA/rev 5′-GCG TGG CAA TCT CCT CAT T, $T_{an} = 56\,^{\circ}C$) were isolated from a genomic library based on an Indian strain of *L. donovani* (MHOM/IN/01/BHU20140). PCR reactions and amplifications were performed as described elsewhere [14,16]. PCR conditions for markers CS19 and CS20 ($T_{an} = 56\,^{\circ}C$) were the same as for the markers described by Ochsenreither et al. [16]. Amplification products were tested by electrophoresis in 1.2% agarose gels in 1× TBE-buffer.

2.3. Microsatellite electrophoretic analysis

Fragments containing single microsatellites were either analyzed by MetaPhor agarose gel, PAGE or capillary electrophoresis. All three methods are able to detect differences of a single dinucleotide repeat and yielded reproducible and comparable results [16]. Four percent MetaPhor agarose gels (BioWhittaker Molecular Applications, USA) were used to screen the strains for polymorphisms [16]. For PAGE 6–15 µl of the PCR product were mixed with loading buffer and run under non-denaturating conditions on 12% polyacrylamide gels at 1 kV for 6 h. Gels were silver-stained and dried [9]. For high throughput studies, fluorescence labeled PCR products were analyzed using the fragment analysis tool of the CEQ 8000 automated genetic analysis system (Beckman Coulter, USA) [15,16].

2.4. Data analysis

Population structure was investigated using the STRUCTURE software [17], which applies a Bayesian model-based clustering approach. The following parameters were used: burn-in period of 50 000 iterations, probability estimates based on 500 000 Markov Chain Monte Carlo iterations and 10 runs for each K (number of populations), to quantify the amount of variation of the likelihood for each K. The most appropriate number of populations was determined by comparing log-likelihood values for K between 1 and 16. The log-likelihood values were compared in a diagram. At the plateau (maximum) of the derived Gaussian graph, the value of K captures the major structure of the populations. In addition, ΔK based on the rate of change in the log probability of data between successive K values [18] has been calculated. Ancestral source populations were identfied by decreasing the number of K.

Microsatellite genetic distances were analyzed using the MICROSAT program [19]. They were calculated for the numbers of repeats within a locus using the measure D_{AS} (proportion of shared alleles) which calculates multilocus pairwise distance measurements as $1 - s/n$, where n is the number of loci compared and s the total number of shared alleles at all loci. This metric follows the infinite allele model (IAM). Confidence intervals for D_{AS} were calculated by bootstrapping over loci (1000 replications). Neighbor-joining trees were constructed in PHYLIP, version 3.6 [20].

Microsatellite data were analyzed with respect to diversity of alleles (A), expected (H_e, gene diversity) and observed (H_o) heterozygosity, and the inbreeding coefficient F_{is} applying GENEPOP 3.4 [21] and GDA (http://lewis.eeb.uconn.edu/lewishome). Indices of association (I_A) were calculated to test each population for clonality and recombination using MULTILOCUS [22]. Genetic differentiation and gene flow were assessed by calculating F_{st} values (IAM) [23] with corresponding p-values using the MSA software [24].

3. Results

3.1. Microsatellite loci

We applied 15 microsatellite markers previously developed for the *L. donovani* complex, for the analysis of 91 strains of this species complex. Strains were obtained from different endemic regions, East Africa (32 strains: *L. donovani*, *L. archibaldi*, *L. infantum*), India (21 strains: *L. donovani*), China (4 strains: *L. donovani*, *L. infantum*), the Mediterranean region

Table 1
Designation and characteristics of *Leishmania* strains used in this study

Taxa[a]	Laboratory code (project code)	WHO code	Origin	Zymodeme	Pathology[b]	Population assignment[c]
L. infantum (Old World)	INF-01[d]	MHOM/TN/80/IPT1	Tunisia	MON-1	VL	5
	INF-39 (LG1)	MHOM/FR/78/LEM75	France	MON-1	VL	5
	INF-40 (LG2)	MHOM/FR/95/LPN114	France	MON-1	VL	5
	INF-41 (LG3)	MHOM/ES/93/PM1	Spain	MON-1	VL	5
	INF-42 (LG4)	MHOM/FR/97/LSL29	France	MON-1	CL	5
	INF-43 (LG5)	MHOM/ES/86/BCN16	Spain	MON-1	CL	5
	INF-44 (LG6)	MHOM/PT/00/IMT260	Portugal	MON-1	CL	5
	INF-09 (LG132)	MCAN/PT/93/IMT193	Portugal	MON-1	CanL	5
	INF-02	MHOM/CN/54/Peking	China	MON-1	VL	7
	INF-05	MHOM/CN/78/D2	China	LON-49	VL	7
	INF-67	MHOM/CN/80/StrainA	China	MON-34	n.d.	7
	INF-03	MHOM/FR/62/LRC-L47	France	n.d.	VL	6
	INF-04	MHOM/ES/87/Lombardi	Spain	MON-24	CL	6
	INF-35	MCAN/FR/87/RM1	France	MON-108	VL	5
	INF-37 (LG19)	MHOM/ES/88/LLM175	Spain	MON-198	VL	6
	INF-45 (LG7)	MHOM/FR/96/LEM3249	France	MON-29	CL	6
	INF-46 (LG8)	MHOM/ES/91/LEM2298	Spain	MON-183	VL	6
	INF-48 (LG15)	MHOM/MT/85/BUCK	Malta	MON-78	CL	7
	INF-47 (LG14)	MHOM/FR/80/LEM189	France	MON-11	CL	6
	INF-32	MCAN/ES/86/LEM935	Spain	MON-77	VL	5
	INF-55 (LG20)	MHOM/ES/92/LLM373	Spain	MON-199	VL	6
	INF-56 (LG21)	MHOM/IT/94/ISS1036	Italy	MON-228	VL	6
	INF-57 (LG22)	MHOM/IT/93/ISS800	Italy/Sicily	MON-188	VL	7/8
	INF-10	MCAN/TR/96/EP16	Turkey	n.d.	CanL	5
	INF-11	MHOM/TR/94/EP3	Turkey	n.d.	n.d.	5
	INF-12	MCAN/IL/94/LRC-L639	Israel	n.d.	CanL	5
	INF-13	MCAN/IL/96/LRC-L685	Israel	n.d.	CanL	5
	INF-38 (LG23)	MHOM/SD/97/LEM3472	Sudan	MON-267	PKDL	4
	INF-33 (LG18)	MHOM/SD/62/3S	Sudan	MON-81	VL	3
	INF-58	MHOM/SD/93/452BM	Sudan	MON-30	PKDL	3
	INF-59 (LG12)	MHOM/SD/82/GILANI	Sudan	MON-30	VL	3
	INF-65	MHOM/SD/93/38-UMK	Sudan	MON-30	VL	3
	INF-61	MHOM/SD/93/762L	Sudan	MON-30	VL	3
	INF-62	MHOM/SD/93/597-2	Sudan	MON-30	PKDL	3
	INF-63	MHOM/SD/92/51-band	Sudan	MON-30	VL	3
	INF-60	MHOM/SD/93/45-UMK	Sudan	MON-30	VL	3
	INF-64	MHOM/SD/93/597LN	Sudan	MON-30	PKDL	3
L. chagasi (New World)	CHA-01[d]	MHOM/BR/74/PP75	Brazil	MON-1	VL	5
	CHA-02	MHOM/PA/79/WR317	Panama	n.d.	CL	6
	CHA-03	MHOM/BR/85/M9702	Brazil	n.d.	VL	5
	CHA-04	MHOM/BR/??/Edmael	Brazil	n.d.	n.d.	5
	CHA-05	MHOM/CR/9?/LVCR	Costa Rica	n.d.	n.d.	6

Table 1 (*continued*)

Taxa[a]	Laboratory code (project code)	WHO code	Origin	Zymodeme	Pathology[b]	Population assignment[c]
L. donovani (Old World)	DON-11	MHOM/CN/00/Wangjie1	China	MON-35	VL	8
	DON-38 (LG13)	MHOM/ET/00/HUSSEN	Ethiopia	MON-31	VL	4
	DON-06	MHOM/SD/75/LV139	Sudan	n.d.	CL	3
	DON-08	MHOM/SD/68/1S	Sudan	n.d.	VL	3
	DON-17	MHOM/SD/93/9S	Sudan	MON-18	VL	3
	DON-24 (LG76)	MHOM/ET/67/HU3	Ethiopia	MON-18	VL	3
	DON-23	MHOM/SD/93/338	Sudan	MON-18	PKDL	3
	DON-49 (LG17)	MCAN/SD/00/LEM3946	Sudan	MON-274	VL	4
	DON-02	MHOM/KE/83/NLB189	Kenya	MON-37	PKDL	2
	DON-03	MHOM/KE/84/NLB218	Kenya	n.d.	PKDL	2
	DON-04	MHOM/KE/85/NLB323	Kenya	MON-37	VL	2
	DON-66	MHOM/KE/55/LRC-L53	Kenya	n.d.	n.d.	2
	DON-67	MHOM/KE/??/LRC-L445	Kenya	n.d.	n.d.	2
	DON-78 (LG65)	MHOM/KE/73/MRC74	Kenya	LON-51	n.d.	2
	DON-79 (LG170)	IMAR/KE/62/LRC-L57	Kenya	MON-37	n.d.	2
	DON-83 (LG67)	MHOM/KE/54/LRC-L53	Kenya	MON-36	n.d.	2
	DON-01[d]	MHOM/IN/80/DD8	India	MON-2	VL	1
	DON-39 (LG9)	MHOM/IN/00/DEVI	India	MON-2	VL	1
	DON-40 (LG10)	MHOM/IN/96/THAK35	India	MON-2	n.d.	1
	DON-45	MHOM/IN/01/BHU20140	India	n.d.	VL	1
	DON-10	MHOM/IN/71/LRC-L51a	India	n.d.	VL	2
	DON-48 (LG16)	MHOM/IN/54/SC23	India	MON-38	VL	2
	DON-51	MHOM/IN/02/BHU1	India	n.d.	VL	1
	DON-52	MHOM/IN/02/BHU2	India	n.d.	VL	1
	DON-53	MHOM/IN/02/BHU3	India	n.d.	VL	1
	DON-54	MHOM/IN/02/BHU4	India	n.d.	VL	1
	DON-55	MHOM/IN/02/BHU5	India	n.d.	VL	1
	DON-56	MHOM/IN/02/BHU6	India	n.d.	VL	1
	DON-57	MHOM/IN/02/BHU7	India	n.d.	VL	1
	DON-58	MHOM/IN/02/BHU8	India	n.d.	VL	1
	DON-59	MHOM/IN/02/BHU9	India	n.d.	VL	1
	DON-60	MHOM/IN/02/BHU11	India	n.d.	VL	1
	DON-61	MHOM/IN/02/BHU12	India	n.d.	VL	1
	DON-62	MHOM/IN/02/BHU13	India	n.d.	VL	1
	DON-63	MHOM/IN/02/BHU15	India	n.d.	VL	1
	DON-64	MHOM/IN/02/BHU17	India	n.d.	VL	1
	DON-65	MHOM/IN/02/BHU20	India	n.d.	VL	1
	DON-29	MHOM/IL/98/LRC-L740	Israel	n.d.	n.d.	8
	DON-80	MCAN/IQ/81/SUKKAR2	Iraq	LON-43	CanL	8
	DON-82	MHOM/IQ/77/BUMM3	Iraq	LON-43	n.d.	8
	DON-81	MHOM/SA/81/Jeddah KA	Saudi Arabia	MON-31	n.d.	4
	DON-84	MHOM/SU/84/MARZ-KRIM	Soviet Union	MON-73	n.d.	7
L. archibaldi (Old World)	ARC-21	MHOM/SD/93/35-band	Sudan	MON-82	VL	3
	ARC-43 (LG11)	MHOM/ET/72/GEBRE1	Ethiopia	MON-82	VL	3
	ARC-14	MHOM/SD/93/AEB	Sudan	MON-82	VL	3
	ARC-07	MHOM/SD/62/LRC-L61	Sudan	MON-82	n.d.	3
	ARC-13	MHOM/SD/93/GE	Sudan	MON-82	VL	4
	ARC-46 (LG24)	MHOM/SD/97/LEM3429	Sudan	MON-257	VL	4
	ARC-47 (LG25)	MHOM/SD/97/LEM3463	Sudan	MON-258	VL	4

[a] Strain identification according to isoenzyme analysis.

[b] VL, visceral leishmaniasis; CL, cutaneouse leishmaniasis; PKDL, post-Kala-Azar dermal leishmaniasis.

[c] Population assignment according to STRUCTURE results for $K = 8$: Population 1, *L. donovani* India 1; population 2, *L. donovani* Kenya/India 2; population 3, *L. donovani* Sudan/Ethiopia 1; population 4, *L. donovani* Sudan/Ethiopia 2; population 5, *L. infantum*/*L. chagasi* MON-1 Mediterranean region/South America; population 6, *L. infantum*/*L. chagasi* non-MON-1 Mediterranean region/South America; population 7, *L. infantum* China, Malta, Sicily + *L. donovani* Soviet Union; population 8, *L. donovani* Iraq, Israel and China.

[d] WHO reference strain.

and South America (22 strains: *L. infantum*) and the Middle East (7 strains: *L. donovani*, *L. infantum*) (Table 1).

Table 2 shows the microsatellite type, repeat and fragment size arrays and numbers of allelic variants for each marker. The number of allelic variants ranged between 4 and 20, with a mean of 11.5. Seventy-five of the 91 strains tested showed unique genotypes. The observed heterozygosity (H_o) varied between 0.022 and 0.253, showing that all microsatellite loci were at least partially heterozygous. The expected heterozygosity (H_e), a measure of genetic diversity, was

Table 2
Characterization of the 15 microsatellite markers used for population analysis of the *Leishmania donovani* complex (91 strains)

No.	Marker	*n*	Repeat array	Fragment size array (bp)	*A*	H_e	H_o	F_{is}
1	Lm2TG	91	TG 9–28	110–148	18	0.740	0.099	0.867
2	TubCA	91	CA 8–17	78–96	9	0.762	0.033	0.957
3	Lm4TA	91	TA 6–16	67–87	9	0.842	0.110	0.870
4	Li 41-56 (B)	91	CA 7–17	84–104	9	0.778	0.253	0.677
5	Li 46-67 (C)	91	CA 6–9	74–80	4	0.630	0.154	0.757
6	Li 22-35 (E)	91	CA 5–28 (36[a])	78–124 (140[a])	18	0.868	0.154	0.824
7	Li 23-41 (F)	91	GT 6–32	65–117	20	0.898	0.176	0.805
8	Li 45-24 (G)	91	CA 7–20	89–115	13	0.850	0.110	0.871
9	Li 71-33 (P)	91	TG 6–27	95–137	15	0.779	0.176	0.775
10	Li 71-5/2 (Q)	91	CA 7–10	106–112	4	0.598	0.121	0.799
11	Li 71-7 (R)	91	CA 8–22	90–118	8	0.807	0.022	0.973
12	CS19	91	AC 13–26	94–120	8	0.828	0.099	0.881
13	CS20	90	TG 17–32	81–111	12	0.809	0.222	0.726
14	LIST7031	91	CA 7–30	103–149	11	0.803	0.143	0.823
15	LIST7039	91	CA 8–29	193–235	14	0.865	0.143	0.836
Overall		90.9			11.5	0.790	0.134	0.831

A, number of alleles; *n*, sample size; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{is}, inbreeding coefficient.
[a] Exceptual value found only for one strain.

between 0.598 and 0.898, and always much higher than the mean observed heterozygosity (0.134). Inbreeding coefficients differed between 0.677 and 0.973 (mean 0.831) indicating inbreeding within loci and the predominance of homozygotes in the investigated strains.

The strains were grouped according to the main endemic regions, East Africa (Sudan, Ethiopia, Kenya: 32 strains), India (21 strains), Mediterranean region (24 strains) to compare the number of alleles in different main endemic regions. Strains from the Middle East, China and South America have not been included due to their small number. In order to compare groups of identical size we used a resampling method [15]. The number of alleles was highest (96 ± 0.4) in the Mediterranean region, followed by Africa (57 ± 5.4) and lowest for the Indian strains (29).

3.2. Inference of population structure of the L. donovani complex

We used a Bayesian model-based clustering algorithm, implemented in STRUCTURE to infer the *L. donovani* population structure. It identifies genetically distinct populations on the basis of allele frequencies. Genetic clusters are then constructed from a collection of individual genotypes, estimating for each strain the fraction of its genotype that belongs to each cluster. Increasing values of K (2–16) were used to identify ancestral source populations (Fig. 1). At $K = 2$ the first split divided *L. infantum*/*L. chagasi* (Mediterranean region, China, South America) from *L. donovani*/*L. archibaldi*/*L. infantum* (East Africa, India). Strains from the Middle East, China and the Soviet Union had memberships in both clusters. The successive splits in the populations could be followed by increasing the number of K. All clusters were well defined and stable at $K = 8$, and corresponded generally to the geographical origin. Calculation of ΔK confirmed the presence of eight populations. The six main populations were: (1) *L. infantum* MON-1; (2) *L. infantum* non-MON-1; (3) Sudan/Ethiopia 1; (4) Sudan/Ethiopia 2 and the single *L. donovani* strain from Saudi Arabia; (5) India 1; (6) Kenya + India 2. The remaining two populations consisted of individual strains of different origin: (7) *L. donovani* China + Middle East; and (8) *L. infantum* China + Malta + Sicily. Only minor changes were observed with those single strains when K was further increased to 9 and 10.

D_{AS} values were calculated for the 91 strains and a neighbor-joining tree was constructed from the resulting distance matrix (Fig. 2A) using a distance-based analysis method. The same six major clusters were detected as with STRUCTURE analysis. Minor differences concerned the single *L. donovani* strains from the Middle East, China, the Soviet Union and *L. infantum* from Sicily and Malta. Midpoint rooting led to a separation of *L. infantum* from *L. donovani* strains. Strains of zymodemes MON-108 and MON-77 and the three strains from China were assigned to the *L. infantum* MON-1 population and two *L. chagasi* strains with unknown zymodeme to the non-MON-1 population. The strain from Malta had a basal position in the *L. infantum* cluster. The *L. donovani* cluster fell into three main subclusters: India 1, Sudan/Ethiopia with two further subgroups, and Kenya + India 2, all identical with the corresponding STRUCTURE populations. *L. infantum* from Sicily and *L. donovani* from Israel formed a basal separate branch in the *L. donovani* cluster at an intermediate position between the two main clusters. When single strains have been excluded because of their limited use for population analysis, most of the branches were well supported by bootstrap analysis (Fig. 2B).

3.3. Population characteristics: geographical differentiation

Those ‘‘populations’’ represented by only 1–2 strains were excluded from further population analyses. A significant

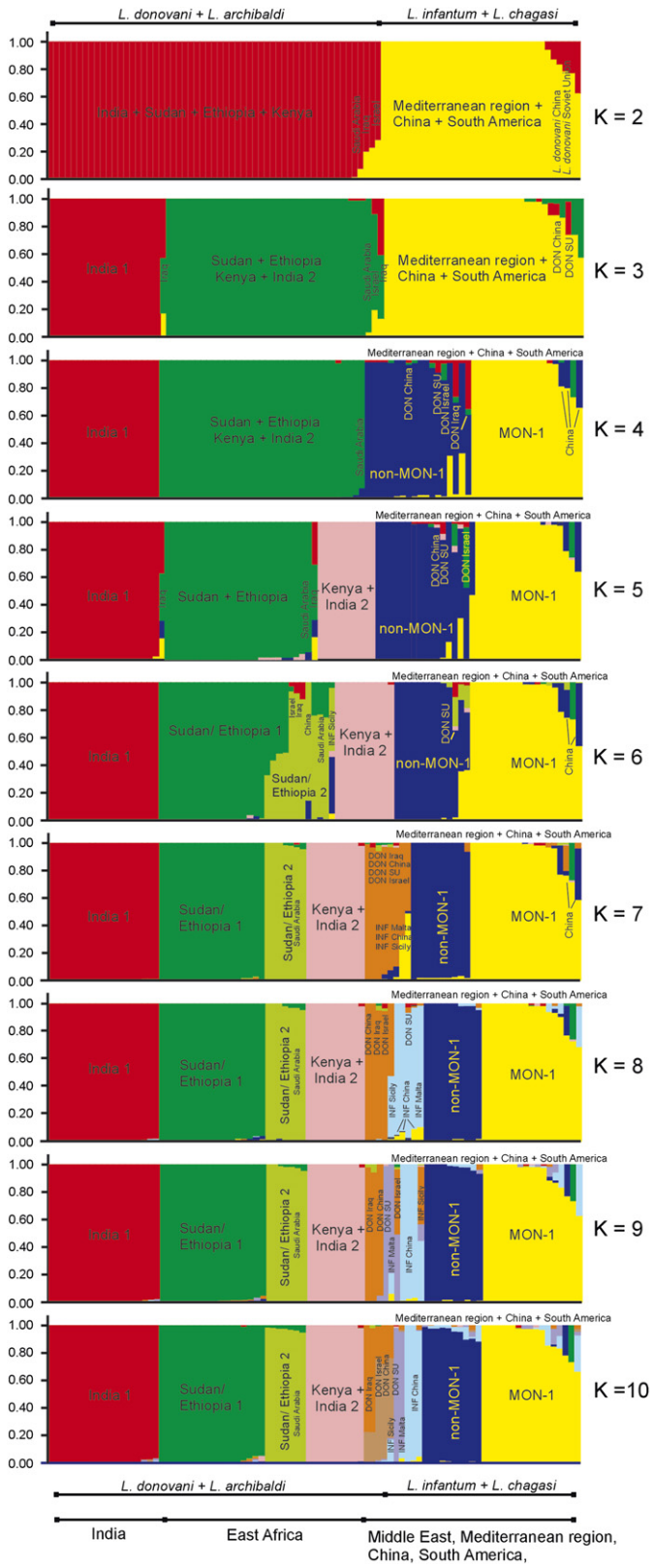

Fig. 1. Population structure of the *Leishmania donovani* complex as inferred by analysis with STRUCTURE based on 15 microsatellite markers. Each of the 91 strains is represented by a single line divided into K colors, where K is the number of populations assumed. Each color represents one population, and the length of the colored segment shows the estimated proportion of membership of the strain in that population.

amount of structure in the data for the *Leishmania donovani* complex was shown by the model-based data analysis and F_{st} values. All F_{st} values (Table 3) are significant and >0.25 indicating very great genetic differentiation between the populations.

The genetic diversity within each population was measured using MNA, P, H_o and H_e (Table 4). The mean number of alleles (MNA) over a range of loci, which is considered an indicator of genetic variation within a population, ranged between 1.1 for India 1 and 4.4 for non-MON-1. Also P, the proportion of polymorphic loci, was lowest for India 1 (0.067) and highest for non-MON-1 (1.0). These differences are even more relevant, as the number of strains investigated for India 1 was almost twice the number for non-MON-1. The values for H_e differed significantly between 0.019 for India 1 and 0.619 for non-MON-1. The inbreeding coefficient F_{is} ranged from 0.254 for India 1 to 0.907 for MON-1, reflecting the great difference between H_e and H_o in the MON-1 population and suggesting a very high degree of inbreeding in this population. All the above measures indicated that the India 1 population was least diverse and the non-MON-1 and MON-1 populations were most diverse. The Sudan/Ethiopia 2 population is also very heterogenous, especially in view of the high diversity measures despite the fewer strains tested. Homozygous alleles predominated in India 1 and MON-1 whereas the highest degree of heterozygosity was observed for Sudan/Ethiopia 1. Multilocus linkage associations were highly significant for five populations pointing to a largely clonal reproduction within these populations.

4. Discussion

The aim of this study was to address key epidemiological questions concerning the *L. donovani* complex, as the extent of genetic isolation and gene flow between different populations, by analyzing variation of microsatellite loci. Model- and distance-based methods have been applied for microsatellite data analysis. The model-based method implemented in STRUCTURE has advantages over distance-based data evaluation because its algorithm uses patterns of allele frequencies to identify distinct subpopulations and determine fractions of the genotype belonging to each subpopulation. It was shown to accurately infer individual ancestries, to be more appropriate for characterization of population structure and to provide information on population relationships and history [17]. Both model- and distance-based algorithms identified the same main populations and thus also supported those clusters in distance trees for which bootstrap values were not sufficiently significant.

MLMT identified six major geographical populations with a high degree of genetic isolation. The major split observed between strains from the Mediterranean region and the New World, and strains from East Africa and India corroborates the existence of two species/subspecies within *L. donovani sensu lato*, namely *L. donovani* (*sensu stricto*) and *L. infantum*, as proposed by previous taxonomic studies [1,7,10,25].

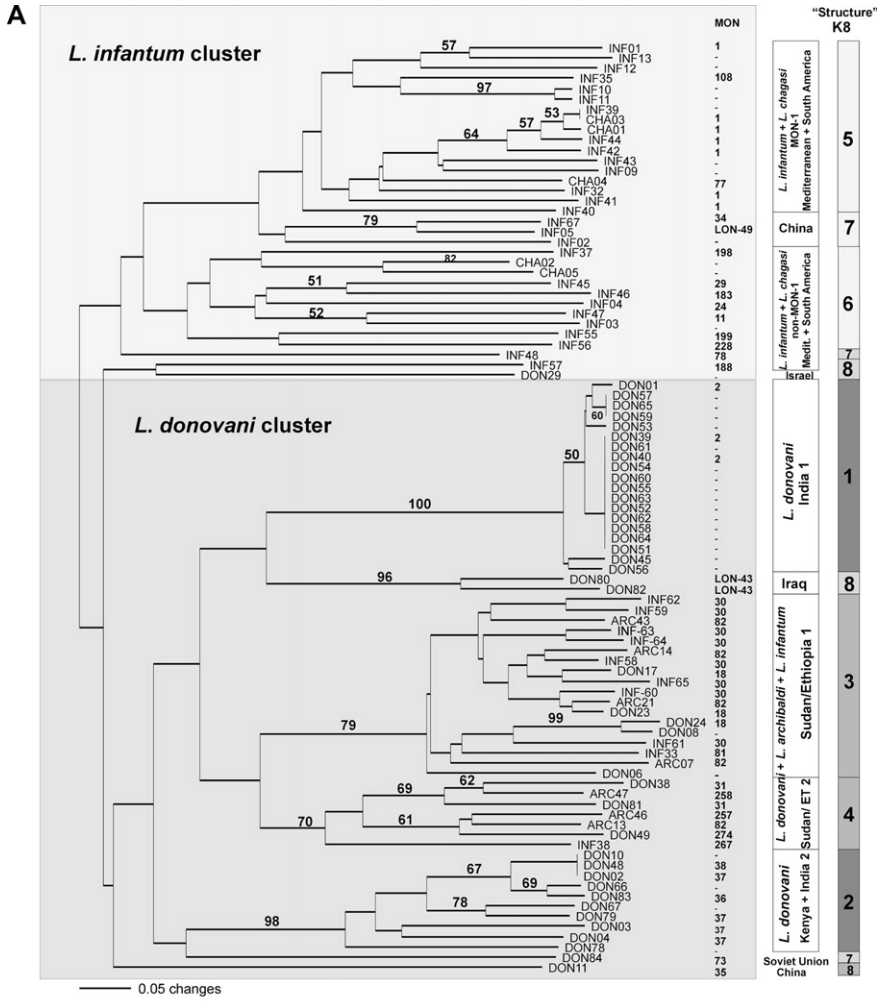
A
L. infantum cluster
L. donovani cluster
MON
"Structure" K8
INF01
INF13
INF12
INF35
INF10
INF11
INF39
CHA03
CHA01
INF44
INF42
INF43
INF09
CHA04
INF32
INF41
INF40
INF67
INF05
INF02
INF37
CHA02
CHA05
INF45
INF46
INF04
INF47
INF03
INF55
INF56
INF48
INF57
DON29
DON01
DON57
DON65
DON59
DON53
DON39
DON61
DON40
DON54
DON60
DON55
DON63
DON52
DON62
DON58
DON64
DON51
DON45
DON56
DON80
DON82
INF62
INF59
ARC43
INF-63
INF-64
ARC14
INF58
DON17
INF65
INF-60
ARC21
DON23
DON24
DON08
INF61
INF33
ARC07
DON06
DON38
ARC47
DON81
ARC46
ARC13
DON49
INF38
DON10
DON48
DON02
DON66
DON83
DON67
DON79
DON03
DON04
DON78
DON84
DON11
L. infantum + L. chagasi MON-1 Mediterranean + South America
China
L. infantum + L. chagasi MON-1 Medit. + South America
Israel
L. donovani India 1
Iraq
L. donovani + L. archibaldi + L. infantum Sudan/Ethiopia 1
L. donovani Sudan/ ET 2
L. donovani Kenya + India 2
Soviet Union China
0.05 changes
B
predominantly anthroponotic
L. donovani MON-31, 274
L. archibaldi MON-82, 257, 258
L. infantum MON-267
SUDAN 2
predominantly anthroponotic
L. donovani MON-18
L. archibaldi MON-82
L. infantum MON-30, 81
SUDAN 1
anthroponotic
L. donovani MON-2
INDIA 1
L. donovani MON-36, 37, 38
KENYA+ INDIA 2
anthroponotic
China
L. infantum/ L. chagasi MON-1, (77, 108)
MEDITERRANEAN SOUTH AMERICA
L. infantum/ L. chagasi non-MON-1
MEDITERRANEAN SOUTH AMERICA
zoonotic
zoonotic

Table 3
F_{st} values and corresponding p-values for the main six populations as assumed by STRUCTURE

F_{st}-values	India 1	Kenya + India 2	Sudan/ ET 1	Sudan/ ET 2	MON-1	Non-MON-1
India 1	0	0.8496	0.7523	0.8338	0.7474	0.7431
Kenya + India 2	0.0001	0	0.6130	0.5381	0.5671	0.4928
Sudan/Ethiopia 1	0.0001	0.0001	0	0.4077	0.5597	0.4674
Sudan/Ethiopia 2	0.0001	0.0006	0.0001	0	0.5141	0.3657
MON-1	0.0001	0.0001	0.0001	0.0001	0	0.3309
Non-MON-1	0.0001	0.0001	0.0001	0.0001	0.0001	0

Populations consisting of only 1–3 strains have been omitted. F_{st} values are in the upper triangle, p-values in the lower triangle.

The *L. infantum* cluster was further subdivided into two populations, one comprising all MON-1 strains and a second strains of non-MON-1 zymodemes. The MON-1 group contained isolates not typed by MLEE, as well as two strains assigned to zymodemes MON-77 and MON-108 that are closely related to MON-1 [2]. The hypothesis of recent introduction of this parasite into China, possibly along the silk-road [10], was supported since three Chinese strains grouped next to the MON-1 population. *L. chagasi*, the New World synonym of *L. infantum* [1,3,5,6], was present in both MON-1 and non-MON-1 populations.

Within the MON-1 group, strains from North Africa, Turkey and Israel were separated from the majority of strains from Southern Europe, except the MON-108 strain from France. More strains from North Africa, Southern Europe, Middle East and Asia should be analyzed to test for the presence of distinct, well-supported subgroups within the MON-1 and the non-MON-1 clusters. Strains from Malta (INF-48) and Sicily (INF-57) shared alleles from different clusters and could not be assigned unequivocally to a specific population. They were placed between *L. infantum* and *L. donovani* which, in the case of the strain from Malta, corresponds with previous studies [6,7]. Whether strains from these geographical regions might represent an intermediate between *L. infantum* and *L. donovani* needs to be investigated with more strains.

Four genetically isolated populations were differentiated among *L. donovani* strains from East Africa and India. Strains from India were assigned to two different clusters with high bootstrap support. The majority of these strains formed population India 1, whereas two other strains were grouped with Kenyan strains. The relationship between the populations India 1 and Kenya + India 2 could not be resolved with confidence. Interestingly, MLMT also detected two genetically distinct populations in Sudan/Ethiopia. All strains with zymodemes MON-30 and MON-81 (“*L. infantum*”), MON-18 (*L. donovani*) plus some with MON-82 (“*L. archibaldi*”) were found in population Sudan/Ethiopia 1. Zymodemes MON-257, MON-258 and a single MON-82 (“*archibaldi*”), MON-267 (“*L. infantum*”), as well as MON-274 and MON-31 (*L. donovani*) were concentrated in the population Sudan/Ethiopia 2. According to MLMT, Sudan/Ethiopia 1 and 2 are sister groups supported by 80% and 77%, respectively, whereas the branch combining both showed 52% bootstrap support (Fig. 2B).

Considerable variation was observed for the microsatellite profiles of strains within the Sudan/Ethiopia 1 and 2 populations. Both are subdivided into subclusters obviously associated with the year of isolation. Most strains in the first subcluster of Sudan/Ethiopia 1 were isolated between 1992 and 1993. The second subcluster contains mainly older strains collected between 1962 and 1975. The majority of the strains in population Sudan/Ethiopia 2 were isolated between 1997 and 2000. This might reflect different consecutive outbreaks or be due to instability of microsatellites. Long term cultivation studies (authors’ unpublished data) and a previous study on *L. tropica* [15] showed however, that microsatellite markers in *Leishmania* are stable over time.

The East African strains assigned to three different species by MLEE, namely *L. infantum*, *L. archibaldi* and *L. donovani*, were scattered among the two well-supported Sudan/Ethiopia 1 and 2 populations. Neither an independent *L. archibaldi* cluster nor an East African *L. infantum* cluster were detectable, which agrees with numerous genetic analyses [5–7,9,10]. It has been suggested that the misleading results from MLEE are attributable to convergent evolution in the GOT gene [11]. The consequence is that the species *L. archibaldi* is invalid and that *L. infantum* does not exist in East Africa.

All but two strains from India formed the highly uniform India 1 cluster representing the zymodeme MON-2. Twelve of the 19 strains had identical MLMT profiles, other strains differed in only one locus. Most of these strains were isolated in Bihar, a highly endemic focus of VL with an emerging refractoriness to standard antimonial treatment [26]. Whether the remarkable homogeneity of India 1 is related to an epidemic spread of a single clone or to a recent bottleneck remains to be established.

The existence of genetically isolated populations is supported by differences in biology and clinical behaviour of

Fig. 2. (A) Neighbor-joining tree (midpoint rooted) inferred from the D_{AS} distances calculated for the data of 15 microsatellite markers and 91 strains with MICROSAT, and PHYLIP softwares. The numbers above the branches indicate the percentage with which a given branch is supported in 1000 bootstrap replications. Only values above 50% are shown. Strain names are based on isoenzyme identification. Geographical origins of the populations found are shown. These populations have been compared with populations inferred by STRUCTURE for $K = 8$. (B) Neighbor-joining tree (unrooted) inferred from the D_{AS} distances calculated for the data of 15 microsatellite markers with MICROSAT and PHYLIP software. Populations represented by only single strains (Iraq, DON-80, DON-82; Israel, DON-29; Saudi Arabia, DON-81; Soviet Union, DON-84; China, DON-11; Malta, INF-48; Sicily, INF-57) have been excluded. The numbers above the branches indicate the percentage with which a given branch is supported in 1000 bootstrap replications. Six main populations basically correlating with geographical origin could be recognized. Species (based on MLEE identification) found in each population as well as the zymodemes and reservoir host are shown. Microsatellite markers place all strains from Sudan/Ethiopia, regardless of whether they were typed by MLEE as *L. donovani*, *L. archibaldi* or *L. infantum*, into the same two groups. There is no correlation between these populations and MLEE identification.

Table 4
Characterization of the main populations found for the *L. donovani* complex

Group	Population	Origin	Species[a]	*N*	*P*	MNA	H_o	H_e	F_{is}	I_A
1	India 1	India (Bihar)	*L. donovani*	19	0.067	1.133	0.014	0.019	0.254	–
2	Kenya + India 2	Kenya, India	*L. donovani*	10	0.533	2.200	0.047	0.288	0.845	2.132[b]
3	Sudan 1	Sudan, Ethiopia	*L. donovani*, *L. archibaldi*, *L. infantum*	18	0.733	2.667	0.304	0.376	0.196	0.757[b]
4	Sudan 2	Sudan, Ethiopia	*L. donovani*, *L. archibaldi*, *L. infantum*	6	0.800	2.267	0.267	0.439	0.416	1.144[b]
5	MON-1	Spain, Portugal, France, Turkey, Israel, Tunisia, Brazil	*L. infantum*, *L. chagasi*	20	0.933	4.267	0.043	0.454	0.907	1.022[b]
6	Non-MON-1	Spain, France, Italy, China, Panama, Costa Rica	*L. infantum*, *L. chagasi*	10	1.000	4.400	0.180	0.619	0.720	0.536[c]

Populations consisting of only 1–3 strains have been excluded.
[a] Species identification according to isoenzyme analysis; *N*, number of strains; *P*, proportion of polymorphic loci; MNA, mean number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{is}, inbreeding coefficient; I_A, index of association.
[b] $p < 0.001$.
[c] $p = 0.019$.

the *L. donovani* strains. VL due to *L. infantum* is zoonotic, whereas infections by Indian and, perhaps, Kenyan [27] *L. donovani* are anthroponotic. Sudanese *L. donovani* is considered to be, at least in part, zoonotic [25] with dogs as the putative reservoir host [28]. In India, PKDL develops in 10–20% of the patients mostly 2–3 years after treatment, in Sudan in 50–60% of the patients within weeks or months and in Kenya up to 30 years after cure [27,29]. The existence of distinct populations in East Africa might be linked to the transmission by different vectors, *P. martini* in Kenya and *P. orientalis* in Sudan/Ethiopia. There was no correlation between detected MLMT subgroups and clinical manifestations as VL, CL, PKDL.

MLMT revealed a predominantly clonal population genetic structure of VL parasites with high levels of inbreeding especially for the MON-1 strains, despite evidence of recombination based on single strains, the recent detection of hybrids and mosaic genome structures [30]. This is of important epidemiological significance for the spread of drug refractoriness.

Most authors advocate a spread of *Leishmania* parasites "out of Africa" [2,10,15,25]. However, the higher number of alleles for the Mediterranean *L. infantum* would admit another interpretation. A global population genetic study for the reconstruction of the history of the *L. donovani* complex should include appropriate sampling in other endemic areas in the Middle East, Asia, and North Africa in order to identify possible intermediate genotypes.

Microsatellite repeats are located on multiple unlinked loci, the markers are co-dominant and all possible allele combinations can be detected. Microsatellite typing is, however influenced by the number of loci tested, since simple repeats are prone to homoplasy due to their nature of evolution and mutation. For an optimal analysis, it was recommended to use at least 12–15 highly variable loci and 15–20 individuals per hypothesized population. In this study, 21–33 strains per endemic area (Mediterranean basin, India and East Africa) were investigated with 15 microsatellite markers. Increasing the number of markers to 24 led to identical major populations, all measures of diversity were comparable and the relationships among the populations reflected by F_{st} values were nearly identical (data not shown).

MLMT can be used directly on biological material without cultivation of the parasite (authors' unpublished data) and high-throughput analyses are possible. Microsatellite profiles can be stored in databases and compared between different laboratories. The discriminatory power of MLMT is higher than that of other genetic markers used for typing of *Leishmania* and comparable to that of kDNA markers [12]. In this study, highly polymorphic and less discriminatory microsatellite markers were combined to elucidate population structures within the entire *L. donovani* complex. This allowed not only identification of six major geographical populations but also detection of substructures according to the place or time of strain isolation. With the MLMT approach presented herein, a powerful tool is now available for global taxonomic, epidemiological and population genetic studies of the *L. donovani* complex.

Acknowledgements

We would like to thank Anna Braune and Oliver Radtke for technical assistance and Renate Rebenstorff for her help with parasite cultivation. We are very grateful to Francine Pratlong and Jean-Pierre Dedet from the Laboratoire de Parasitologie et Centre National de Référence des Leishmania, Montpellier, Isabel Mauricio from London School of Hygiene and Tropical Medicine, Nel Kroon and Henk Schallig from the Royal Tropical Institute Amsterdam, Lee Schnur from the Kuvin Centre for the Study of Tropical and Infectious Diseases Hadassah Medical School Jerusalem, Syamal Roy from the Indian Institute of Chemical Biology Calcutta and Shyam Sundar from the Banaras Hindu University, Varanasi for providing *Leishmania* strains and DNA samples. The research was supported by a grant (QLK2-CT-2001-01810; http://www.leishdomus.org) from the European Union.

References

[1] J.A. Rioux, G. Lanotte, E. Serres, F. Pratlong, P. Bastien, J. Perieres, Taxonomy of *Leishmania*, use of isozymes. Suggestions for a new classification, Ann. Parasitol. Hum. Comp. 65 (1990) 111–125.

[2] F. Pratlong, J. Dereure, B. Bucheton, S. El-Safi, A. Dessein, G. Lanotte, J.-P. Dedet, Sudan: the possible original focus of visceral leishmaniasis, Parasitology 122 (2001) 599–605.
[3] I. Mauricio, J.R. Stothard, M.A. Miles, The strange case of *Leishmania chagasi*, Parasitol. Today 16 (2000) 188–189.
[4] I. Mauricio, J.R. Stothard, M.A. Miles, *Leishmania donovani* complex: genotyping with the ribosomal internal transcribed spacer and the mini-exon, Parasitology 128 (2004) 1–5.
[5] K. Kuhls, I.L. Mauricio, F. Pratlong, W. Presber, G. Schönian, Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex, Microb. Infect. 7 (2005) 1224–1234.
[6] I.L. Mauricio, M.W. Gaunt, J.R. Stothard, M.A. Miles, Genetic typing and phylogeny of the *Leishmania donovani* complex by restriction analysis of PCR amplified gp63 intergenic regions, Parasitology 122 (2001) 393–403.
[7] K.W. Quispe Tintaya, X. Jing, J.-P. Dedet, S. Rijal, X. De Bolle, J.C. Dujardin, Antigen genes for molecular epidemiology of leishmaniasis: polymorphism of cysteine proteinase B and surface metalloprotease glycoprotein 63 in the *Leishmania donovani* complex, J. Infect. Dis. 189 (2004) 1035–1043.
[8] G. Schönian, C. Schweynoch, K. Zlateva, L. Oskam, N. Kroon, Y. Gräser, W. Presber, Identification and determination of the relationships of species and strains within the genus *Leishmania* using single primers in the polymerase chain reaction, Mol. Biochem. Parasitol. 77 (1996) 19–29.
[9] S. Lewin, G. Schönian, N. El Tai, L. Oskam, P. Bastien, W. Presber, Strain typing in *Leishmania donovani* by using sequence-confirmed amplified region analysis (SCAR), Int. J. Parasitol. 32 (2002) 1267–1276.
[10] M.E. Ibrahim, D.C. Barker, The origin and evolution of the *Leishmania donovani* complex as inferred from a mitochondrial cytochrome oxidase II gene sequence, Inf. Gen. Evol. 1 (2001) 61–68.
[11] M.B. Jamjoom, R.W. Ashford, P.A. Bates, M.L. Chance, S.J. Kemp, P.C. Watts, H.A. Noyes, *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 (2004) 1–11.
[12] Y. Botilde, T. Laurent, W. Quispe Tintaya, C. Chicharro, C. Cañavate, I. Cruz, K. Kuhls, G. Schönian, J.-C. Dujardin, Comparison of molecular markers for strain typing of *Leishmania infantum*, Inf. Gen. Evol. 6 (2006) 440–446.
[13] B. Bulle, L. Millon, J.M. Bart, M. Gállego, F. Gambarelli, M. Portús, L. Schnur, C.L. Jaffe, S. Fernandez-Barredo, J.M. Alunda, R. Piarroux, Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis, J. Clin. Microbiol. 40 (2002) 3391–3397.
[14] M.B. Jamjoom, R.W. Ashford, P.A. Bates, S.J. Kemp, H.A. Noyes, Towards a standard battery of microsatellite markers for the analysis of the *Leishmania donovani* complex, Ann. Trop. Med. Parasitol. 96 (2002) 265–270.
[15] J.M. Schwenkenbecher, T. Wirth, L.F. Schnur, C.L. Jaffe, H. Schallig, A. Al-Jawabreh, O. Hamarsheh, K. Azmi, F. Pratlong, G. Schönian, Microsatellite analysis reveals genetic structure of *Leishmania tropica*, Int. J. Parasitol. 36 (2006) 237–246.
[16] S. Ochsenreither, K. Kuhls, M. Schaar, W. Presber, G. Schönian, Multilocus microsatellite typing as a new tool for the discrimination of *Leishmania infantum* MON-1 strains, J. Clin. Microbiol. 44 (2006) 495–503.
[17] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[18] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[19] E. Minch, A. Ruiz-Linares, D. Goldstein, M. Feldman, L.L. Cavalli-Sforza, MICROSAT, The Microsatellite Distance Program, Stanford University Press, Stanford, CA, 1995.
[20] J. Felsenstein, PHYLIP (Phylogeny inference package) v. 3.6 Distributed by the author, Department of Genetics, University of Washington, Seattle, 2004.
[21] F. Roussett, M. Raymont, Testing heterozygote excess and deficieny, Genetics 140 (1995) 1413–1419.
[22] G.M. Agapow, A. Burt, Indices of multilocus linkage disequilibrium, Mol. Ecol. Notes 1 (2001) 101–102.
[23] B.S. Weir, C.C. Cockerman, Estimating F statistics for the analysis of population structure, Evolution 38 (1984) 1358–1370.
[24] D. Dieringer, C. Schlötterer, Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite sets, Mol. Ecol. Notes 3 (2002) 167–169.
[25] R.W. Ashford, The leishmaniases as emerging and reemerging zoonoses, Int. J. Parasitol. 30 (2000) 1269–1281.
[26] S. Sundar, D.K. More, M.K. Singh, V.P. Singh, S. Sharma, A. Makharia, P.C. Kumar, H.W. Murray, Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic, Clin. Infect. Dis. 31 (2000) 1104–1107.
[27] Y.B. Mebrahtu, P.G. Lawyer, H. Pamba, D. Koech, P.V. Perkins, C.R. Roberts, J.B. Were, L.D. Hendricks, Biochemical characterization and zymodeme classification of *Leishmania* isolates from patients, vectors, and reservoir hosts in Kenya, Am. J. Trop. Med. Hyg. 47 (1992) 852–892.
[28] J. Dereure, S. El-Safi, B. Bucheton, M. Boni, M.M. Kheir, B. Davoust, F. Pratlong, E. Feugier, M. Lambert, A. Dessein, J.-P. Dedet, Visceral leishmaniasis in eastern Sudan: parasite identification in humans and dogs: host-parasite relationships, Microbes Infect. 5 (2003) 1103–1108.
[29] E.E. Zijlstra, A.M. Musa, E.A.G. Khalil, I.M. El-Hassan, A.M. El-Hassan, Post-kala-azar dermal leishmaniasis, Lancet Infect. Dis. 3 (2003) 87–98.
[30] I. Mauricio, M. Yeo, M. Baghaei, D. Doto, F. Pratlong, E. Zemanová, J.-P. Dedet, J. Lukeš, M.A. Miles, Towards multilocus sequence typing in the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD), Int. J. Parasitol. 36 (2006) 757–769.