Available online at www.sciencedirect.com
ScienceDirect
International Journal for Parasitology 37 (2007) 149–160
www.elsevier.com/locate/ijpara

The *Leishmania donovani* complex: Genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing

Eva Zemanová [a], Milan Jirků [a], Isabel L. Mauricio [b], Aleš Horák [a], Michael A. Miles [b], Julius Lukeš [a,*]

[a] *Biology Centre, Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, České Budějovice, (Budweis), Czech Republic*
[b] *Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK*

Received 24 July 2006; accepted 15 August 2006

Abstract

Flagellates of the *Leishmania donovani* complex are causative agents of human cutaneous and visceral leishmaniasis. The complex is comprised of *L. donovani*, *Leishmania infantum* and *Leishmania archibaldi*, although the latter is not now considered to be a valid species. Morphological distinction of *Leishmania* species is impractical, so biochemical, immunological and DNA-based criteria were introduced. Multilocus enzyme electrophoresis (MLEE) is the present gold standard. We have sequenced the genes encoding five metabolic enzymes used for MLEE, both to resolve the DNA diversity underlying isoenzyme mobility differences and to explore the potential of these targets for higher resolution PCR-based multilocus sequence typing. The genes sequenced were *isocitrate dehydrogenase*, *malic enzyme*, *mannose phosphate isomerase*, *glucose-6-phosphate dehydrogenase*, and *fumarate hydratase*, for 17 strains of *L. infantum*, seven strains of *L. donovani*, and three strains of *L. archibaldi.* Protein mobilities predicted from amino acid sequences did not always accord precisely with reported MLEE profiles. A high number of heterozygous sites was detected. Heterozygosity was particularly frequent in some strains and indirectly supported the presence of genetic exchange in *Leishmania*. Phylogenetic analysis of a concatenated alignment based on a total of 263 kb protein-coding sequences showed strong correlation of genotype with geographical origin. Europe and Africa appear to represent independent evolutionary centres.
© 2006 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

Keywords: *Leishmania donovani* complex; Zymodeme; Multilocus sequence typing; *Leishmania*; Phylogenetic network

1. Introduction

Flagellates of the genus *Leishmania* are the causative agents of the leishmaniases, human diseases with symptoms ranging from the localised skin lesions of cutaneous leishmaniasis (CL), to the destructive metastatic nasopharyngeal lesions of mucocutaneous leishmaniasis (MCL) and to progressive systemic visceral leishmaniasis (VL), which is fatal if left untreated (Lainson and Shaw, 1987). Visceral leishmaniasis is caused by the *Leishmania donovani* complex, comprised of the species *L. donovani*, *Leishmania infantum* and *Leishmania archibaldi* (Pratlong et al., 2001). However, *L. archibaldi* is now considered to be encompassed within *L. donovani* and not valid as a separate species (Lewin et al., 2002; Jamjoon et al., 2004; Mauricio et al., 2004; Kuhls et al., 2005).

During the last decade, genetic diversity within the *L. donovani* complex has been the subject of intense interest. A number of approaches have been applied, including: restriction fragment length polymorphisms (RFLPs); restriction of kinetoplast DNA minicircles; analysis of

* Corresponding author. Tel.: +420 38 7775416; fax: +420 38 5310388.
E-mail address: jula@paru.cas.cz (J. Lukeš).

0020-7519/$30.00 © 2006 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijpara.2006.08.008

telomeric repeats; random amplified polymorphic DNA (RAPD); microsatellite analysis and sequencing of targets such as the ribosomal internal transcribed spacers, the miniexon locus and a few protein-coding genes (Noyes et al., 1998; Breniere et al., 1999; El-Tai et al., 2000; Chiurillo et al., 2001; Jamjoom et al., 2002; Castilho et al., 2003; Gangneux et al., 2003; Zemanová et al., 2004; Kuhls et al., 2005). So far, the data have supported the affinity between *L. donovani* and *L. infantum* and indicated that genetic diversity is more pronounced in *L. donovani*, compared with more homogeneous European *L. infantum* (Mauricio et al., 1999, 2000; Jamjoon et al., 2004; Kuhls et al., 2005).

Multilocus enzyme electrophoresis (MLEE), is currently regarded as the ''gold standard'' for identification of *Leishmania* strains and has been widely used since its introduction in the 1980s (Rioux et al., 1990). Distinct combinations of isoenzyme mobilities for up to 15 enzymes have been assigned zymodeme numbers (MON-1–MON-274). Our aim was first, to determine the DNA sequence diversity for five of the genes encoding enzymes used in MLEE and second, to explore the potential of these data for PCR-based multilocus sequence typing (MLST). According to available *L. major* and *L. infantum* genome sequences the five enzyme genes that we selected (see Section 2 for EC codes) are nuclear, single or double-copy and unlinked, (all in http://www.genedb.org/; Ivens et al., 2005). Such housekeeping genes have the advantage of reliability and resolving power, as they are rarely subject to concerted evolution (Senchina et al., 2003), do not display high intragenic polymorphism and can be unambiguously aligned.

Selection of genes encoding enzymes used in MLEE therefore not only resolves genetic diversity at the DNA level but connects resultant observations to the substantial existing MLEE dataset. With the exception of occasional single gene phylogenies, these targets have not been applied in phylogenetics, partially because of difficulties in identifying strictly orthologous genes. However, identification of orthologous genes is now achievable through the *Leishmania* genome sequencing projects. Our results provide data for both MLST and for phylogeny reconstruction.

Here, we present results for DNA sequence diversity of isocitrate dehydrogenase (*icd*), cytosolic malic enzyme (*me*), mannose phosphate isomerase (*mpi*), glucose-6-phosphate dehydrogenase (*gpi*), and fumarate hydratase (*fh*) in the *L. donovani* complex. Results for five other genes (*asat*, *gpi*, *nh1*, *nh2*, and *pgd*) have been presented elsewhere (Mauricio et al., 2006).

2. Materials and methods

2.1. Parasites

Strains of the *L. donovani* complex and *Leishmania major*, *Leishmania tropica* and *Leishmania gerbilli* are listed in Table 1. Table 1A gives strains for which all five enzymes were sequenced and Table 1B strains of the *L. donovani* complex for which only one or two genes, encoding enzymes with distinct electrophoretic mobilities, were sequenced.

2.2. PCR amplification

Primers were designed from GeneDB data of the *L. major* and *L. infantum* genome sequencing projects http://www.genedb.org/ and, for *g6pdh*, from the *L. major* sequence CT005271 in GenBank/EMBL. If possible, primers were designed to span from start to stop codon and amplify the entire coding region of the gene. For *fh* the primers were designed in the intergenic region. For *me* the forward primer was designed 70 bp downstream because of the high GC content in the first part of the gene. Amplification reactions were in 25 µl volumes containing 10 pmol of each primer, reaction buffer (100 mM Tris–HCl, pH 8.8; 500 mM KCl, 1% Triton X-100; 15 mM $MgCl_2$), 0.25 mM deoxyribonucleotide triphosphate (dNTPs), 0.5 U *Taq* polymerase and 10–20 ng DNA. Amplification conditions were: 96 °C for 5 min, followed by 30 cycles with 96 °C for 1 min, 55–58 °C for 1 min and 72 °C for 90 s, with a final extension at 72 °C for 10 min.

2.3. Gene sequencing

Databases for *L. major* and *L. infantum*- and for *Trypanosoma brucei* and *Trypanosoma cruzi* were searched for orthologues of genes encoding enzymes used in MLEE (Ivens et al., 2005). The following sequences were used to design forward (Fw) and reverse (Rv) primers:

(i) *icd*, encoding isocitrate dehydrogenase (EC 1.1.1.42., ICD) – *L. major* (LmjF10.0290), *L. infantum* (LinJ10.0200) at chromosome 10. Fw 5′ ATG TTC CGC CAT GTT TCG GC 3′; Rv 5′ TTA CGC GCT CAT CGC CTT 3′; annealing temperature 55 °C.

(ii) *me*, encoding cytosolic NADP-malic enzyme (EC 1.1.1.40., ME) – *L. major* (LmjF24.0770), *L. infantum* (LinJ24.0440) at chromosome 24. Fw 5′ CGC AAC CGC TTC ACC AAT AAG GGC 3′; Rv 5′ CAA CTC CTT CTC CAG GTA GTA GTT 3′; annealing temperature 50 °C.

(iii) *mpi*, encoding mannose phosphate isomerase (EC 5.3.1.8., MPI) – *L. major* (LmjF32.1580), *L. infantum* (LinJ32.1600) at chromosome 32. Fw 5′ ATG TCT GAG CTC GTA AAG CT 3′; Rv 5′ CTA CCT GTC GCT CAA GTC 3′; annealing temperature 55 °C.

(iv) *g6pdh*, encoding glucose-6-phosphate dehydrogenase (EC 1.1.1.49., G6PDH) – *L major* (CT005271) at chromosome 34. Fw 5′ ATG TCG GAA GAG CAG TCT 3′; Rv 5′ TCA CAG CTT ATT CGA GGG AA 3′; annealing temperature 50 °C.

Table 1
WHO code, isoenzyme identification (as MON zymodeme number), country of origin and clinical manifestation in humans (or the non-human host) for strains of the *Leishmania donovani* complex, *Leishmania major*, *Leishmania gerbilli*, *Leishmania tropica*, and *Leishmania infantum*

Name	Strain	Zymodeme	Country	Type of infection or host
(A), All five genes sequenced				
LEM75	MHOM/FR/1978/LEM75	MON 1	France	Visceral
LPN114	MHOM/FR/1995/LPN114	MON 1	France	Visceral
PM1	MHOM/ES/1993/PM1	MON 1	Spain	Visceral
LSL29	MHOM/FR/1997/LSL29	MON 1	France	Cutaneous
BCN16	MHOM/ES/1986/BCN16	MON 1	Spain	Cutaneous
IMT260	MHOM/PT/2000/IMT260	MON 1	Portugal	Cutaneous
LEM3249	MHOM/FR/1996/LEM3249	MON 29	France	Cutaneous
LEM2298	MHOM/ES/1991/LEM2298	MON 183	Spain	Visceral
DEVI	MHOM/IN/0000/DEVI	MON 2	India	Visceral
THAK35	MHOM/IN/1996/THAK35	MON 2	India	Visceral
GEBRE1	MHOM/ET/1972/GEBRE 1	MON 82	Ethiopia	Visceral
GILANI	MHOM/SD/1982/GILANI	MON 30	Sudan	Visceral
HUSSEN	MHOM/ET/0000/HUSSEN	LON 42 = MON31	Ethiopia	Visceral
LEM189	MHOM/FR/1980/LEM189	MON 11	France	Visceral
BUCK	MHOM/MT/1985/BUCK	MON 78	Malta	Visceral
SC23	MHOM/IN/54/SC23	MON 38	India	Visceral
LEM3946	MCAN/SD/2000/LEM3946	MON 274	Sudan	Visceral
3S	MHOM/SD/62/3S	MON 81	Sudan	Visceral
LLM175	MHOM/ES/88/LLM175	MON 198	Spain	Visceral
LLM373	MHOM/ES/92/LLM373	MON 199	Spain	Visceral
ISS1036	MHOM/IT/94/ISS1036	MON 228	Italy	Visceral
ISS800	MHOM/IT/93/ISS800	MON 188	Italy	Visceral
LEM3472	MHOM/SD/97/LEM3472	MON 267	Sudan	Visceral
LEM3429	MHOM/SD/97/LEM3429	MON 257	Sudan	Visceral
LEM3463	MHOM/SD/97/LEM3463	MON 258	Sudan	Visceral
DD8	MHOM/IN/80/DD8	MON 2	India	Visceral
RRR-B	RRR-B	?	Greece	?
L. major	MPSA/JO/87/JPs-G	MON 103	Jordan	*Psammomys obesus*
L. tropica	MHOM/SU/74/K 27	MON 60	Sudan	?
L. gerbilli	MRHO/CN/60/Gerbilli	MON 22	China	*Rhombomys opinus*
L. infantum-clone JPCM5	MCAN/ES/98/LLM-877	?	Spain	?
(B), One or two genes encoding enzymes with distinct electrophoretic mobilities				
GH6 (*me*)	MHOM/GR/2001/GH6	MON 98	Greece	Visceral
GD8 (*me*)	MCAN/GR/2001/GD8	MON 98	Greece	Visceral
GH15 (*me*)	MHOM/GR/2003/GH15	MON 98	Greece	Visceral
LEM935 (*g6pdh*)	MCAM/ES/86/LEM935	MON 77	Spain	Visceral
KA-Jeddah (*mpi*)	MHOM/SA/81/KA-Jeddah	LON 42 = MON31	Saudi Arabia	Visceral
VL29 (*mpi*)	MHOM/SA/87/VL29	LON 42 = MON31	Saudi Arabia	Visceral
GH16 (*me*)	MHOM/GR/2003/GH16	MON 98	Greece	Visceral
GH18 (*me*)	MHOM/GR/2003/GH18	MON 98	Greece	Visceral
GH20 (*me*)	MHOM/GR/2003/GH20	MON 98	Greece	Visceral
GD17 (*me*)	MHOM/GR/2004/GD17	MON 98	Greece	Visceral
LLM580 (*g6pdh*)	MHOM/ES/96/LLM580	MON 253	Spain	Visceral
LLM1026 (*g6pdh*)	MHOM/ES/2001/LLM1026	MON 105	Spain	Visceral
ISS510 (*fh; mpi*)	MHOM/IT/90/ISS510	MON 136	Italy	Visceral
ADDIS164 (*mpi*)	MHOM/ET/84/ADDIS164	MON 83	Ethiopia	Visceral

(v) *fh*, encoding fumarate hydratase (EC 4.2.1.2., FH) – *L major* (LmjF29.1960) at chromosome 29 – Fw 5′ AGC GTC TTG TGT TTC CCA 3′; Rv 3′ GAG CCC GTG TAA GGA GGC 3′; annealing temperature 58 °C.

PCR products of expected size were purified either from the gel using QIAquick gel extraction kits (Qiagen) or directly from the PCR with GenElute PCRClean-up Kit (Sigma) and sequenced directly by cycle sequencing using either the BigDye™ Big Dye Terminator Cycle Sequencing V3.1 (Perkin-Elmer) or CEQ chemistry (Beckman Inc.). Several internal primers were designed based on resolved 5 and 3 sequences. Both strands were sequenced, so that each region was available as at least two independent reads. The determined sequences were deposited in GenBank™ under the following Accession Nos.: DQ142922–DQ142950 (*g6pdh*), DQ142951–DQ142978 (*icd*), DQ142979–DQ143006 (*mpi*), DQ143007–DQ143037 (*me*) and DQ449802–DQ449829 (*fh*).

2.4. Nucleotide sequence analysis

Partial sequences were assembled manually using SeqMan in DNAStar and all consensus sequences were generated in BioEdit (Hall, 1999). No gaps or insertions were detected in sequences obtained from strains of the *L. donovani* complex. Heterozygosity was considered to be present when direct sequencing of PCR products yielded the same split peak at the same site in both strands.

2.5. Protein sequence analysis

Molecular weight, overall charge, charge variation according to theoretical isoelectric point (p*I*) and predictions of post-translational modifications for each protein sequence, were determined through http://www.expasy.org or EditSeq in DNAStar or Memo program.

2.6. Phylogenetic analysis

The MLEE frequencies were transformed into Cavalli–Sforza's chord measure distances (Cavalli-Sforza and Edwards, 1967) and visualized with the Fitch–Margoliash tree building method, with Gendist and Fitch software, respectively, of the PHYLIP 3.6 package (Felsenstein, 1989).

Maximum likelihood trees were calculated from the concatenated nucleotide (nt) dataset under the TN93+I model of evolution using PHYML 2.4.4 (Guindon and Gascuel, 2003). We performed the hierarchical likelihood ratio test (Goldman, 1993) and the Akaike Information Criterion search as implemented in Modeltest 3.75 (Posada and Crandall, 1998), according to which the more complex models, with gamma distribution, were selected. However, our dataset contains few variable characters and the resulting parameter of gamma distribution was very small (0.0188). Therefore, we decided to choose only one category for variable sites and invariants for the maximum likelihood (ML) analyses. ML and maximum parsimony (MP) bootstrap support values were calculated by PHYML 2.4.4 and PAUP*4.0b10 (Swofford, 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland) from 1,000 replications. Bayesian posterior probabilities were assessed under the GTR+Γ_4+I model with MrBayes 3.0b4 software (Bayesian inference – BI; Huelsenbeck and Ronquist, 2001), where the Markov chain was set to 2×10^6 generations, every 100th tree was sampled and the first 10^5 generations were omitted from phylogeny reconstruction.

2.7. Statistical parsimony network

The statistical parsimony network was constructed using TCS 1.21 software (Clement et al., 2000). This software implements the coalescent based method of sequence analysis (Templeton et al., 1992), where the maximum number of differences among the "haplotypes" resulting from single substitution is estimated with the statistical confidence of 95% (Posada and Crandall, 2001) and further transformed to the parsimony network. Heterozygous positions were coded as ambiguous characters so that the software considered them as missing data.

3. Results

Sequences of *icd*, *me*, *mpi*, *g6pdh*, and *fh* from the *L. major* genome were used to design conserved primers spanning the entire coding regions. After optimization of conditions, PCRs amplified single products of expected size from total DNA of various *Leishmania* species. The five genes could be also amplified from reference strains of Old World species outside the *L. donovani* complex, (*Leishmania aethiopica*, *Leishmania arabica*, *L. gerbilli*, *L. major*, and *L. tropica*), except that *me* and *fh* primers yielded no product with *L. arabica* (data not shown).

Genetic diversity in the *L. donovani* complex was assessed using 15 European strains of *L. infantum*, two Sudanese strains of *L. infantum*, seven *L. donovani* strains originating from Europe, India and Sudan, and three strains of North African origin originally designated *L. archibaldi*, with *L. major*, *L. tropica*, and *L. gerbilli* as representatives outside the complex (Table 1). Sequence analysis is described separately for each of the five genes. Only two alternative bases were found at each polymorphic site across all strains. Discrepancies between predicted amino acid (aa) sequences and enzyme electrophoretic profiles (Table 2) were followed up by searching for modifications such as glycosylations, methylations, and acetylations; various prediction programs were applied but failed to predict with high level of certainty (data not shown).

3.1. Isocitrate dehydrogenase (*icd*)

The genomes of *L. major* and *L. infantum* contain two *icd* genes, located on chromosomes 10 and 33, which differ in nt and aa sequences by 41.9% and 11%, respectively. Since the size predicted for both *icd* genes is very similar (1,308 nt versus 1,278 nt), we first analysed *icd* from chromosome 33 by sequencing the entire gene from six isolates, namely *L. infantum* BCN16 and LLM373, *L. donovani* THAK35 and LEM3946, *L. archibaldi* LEM3463 and *L. major*, representing the known mobility variants for the ICD proteins (Pratlong et al., 2001). There were extensive nt and aa differences, yet these did not correspond with MLEE analyses (data not shown). Moreover, reported MLEE diversity for ICD is extremely limited (Rioux et al., 1990). We therefore concluded that the MLEE data are derived from the *icd* gene on chromosome 10.

Consequently, the *icd* gene on chromosome 10 was sequenced for the 27 strains of the *L. donovani* complex (Table 2). Silent mutations were detected at six positions (0.46%), without leading to a change in aa. There was no *icd* diversity among MON-1, MON-183, MON-198, and MON-199 strains (LEM75, LPN114, PM1, LSL29,

Table 2
Polymorphisms within the *Leishmania donovani* complex for all five genes used

Gene	Strain	MON	MLEE	Amino acid/position				Nucleotide								DIPLOTYPE	p*I*	MMass
icd	10 Strains	1, 183, 198, 199	100					C	G	T	T	A	C			1	8.31	48474.41
	LEM3249, LEM189, BUCK, ISS1036, ISS800	29, 11, 78, 228, 188	100					C	G	T	C	A	C			2	8.31	48474.41
	DEVI	2	100					C	G	T	T/C	C	C			3/4	8.31	48474.41
	THAK35, SC23, DD8	2, 38	100					C	G	T	C	C	C			4	8.31	48474.41
	HUSSEN+LEM3463	31, 258	100					T	G	C	C	C	C			5	8.31	48474.41
	GILANI+3S	30, 81	100					C	A	T	C	A	C			6	8.31	48474.41
	GEBRE1	82	100					C	A	T	T	A	C			7	8.31	48474.41
	LEM3946	274	100					T	G/A	C/T	C	A/C	C/A			8	8.31	48474.41
	LEM3472	267	100					C	A	T	C	A	A			9	8.31	48474.41
	LEM3429	257	100					C/T	G/A	C/T	C	A/C	A			10	8.31	48474.41
				59	133	330	563											
me	LEM75, LPN114, PM1, LSL29, BCN16, IMT260	1	100	V	I	V	E	C	T	T	C	C	C	A	C	1	7.01	60480.47
	LEM3249,LEM2298, LEM189, LLM175, RRR-B	29, 183, 11, 198,	100	V	I	V	E	C	G	T	C	C	C	A	C	2	7.01	60480.47
	GH6, GD8, GH15, GH16, GH18, GH20, GD17	98	**90**	V	I	V	E	C	G	T	C	C	C	A	C	2	7.01	60480.47
	ISS1036	228	100	V	I	V	**E/D**	C	G	T	C	C	C	A	C	2/3	7.01	60466.44 or 60480.47
	DEVI, THAK35, DD8	2	100	V	S	I	E	C	G	G	C	C	C	A	T	4	7.01	60468.41
	GEBRE1	82	100	V	S	I	E	C	G	G	A	C	C	A	C	5	7.01	60468.41
	GILANI	30	100	V	S	I	**E/D**	C	G	G	A	C	C	A	C	5/6	7.01	60454.39 or 60468.41
	HUSSEN, SC23, LEM3472, LEM3429, LEM3463	31, 38, 267, 257, 258	100	V	S	I	E	C	G	G	C	C	C	A	C	7	7.01	60468.41
	BUCK	78	100	V	I	I	E	C	G	G	C	C	C	C	T	8	7.01	60494.50
	LEM3946	274	100	V	S	I	**E/D**	C	G	G	C	G	C	A	C	9/10	7.01	60454.39 or 60468.41
	3S	81	100	V	S	I	E	C	G	G	A	G	C	A	C	11	7.01	60468.41
	LLM373	199	100	**M**	S	V	E	C	G	G	C	C	C	C	C	12	7.01	60486.45
	ISS800	188	100	V	S	I	E	T	G	G	C	C	T	A	C	13	7.01	60468.41
				131	213	277												
mpi	16 strains	1, 29, 183, 82, 30, 11, 78, 81, 198, 199, 228	100	K	K	S		C	G	C	G					1	5.75	46533.18
	ISS510	136	**100/127**	K	**K/E**	S		C	G/A	C	G					1/2	5.56 or 5.75	46534.12 or 46533.18
	ISS800	188	**127**	K	**E**	S		C	G	C	G					2	5.56	46534.12
	DEVI,THAK35,DD8	2	100	K	K	S		T	G	T	G					3	5.75	46533.18
	HUSSEN, LEM3946, KA-Jeddah, VL29, ADDIS164	31, 274, 83	**110**	K	K	S		T	G	T	G					3	5.75	46533.18
	LEM3472, LEM3429, LEM3463	267, 257, 258	**110**	**K/R**	K	S		T	G	T	G					3/4	5.75	46651.20 or 46533.18
	SC23	38	100	K	K	**S/I**		C	G	T	C					5/6	5.75	46559.26 or 46533.18
	RRR-B	?	?	K	K	S		C/T	G	C/T	G					7	5.75	46533.18

(*continued on next page*)

Table 2 (*continued*)

Gene	Strain	MON	MLEE	Amino acid/position						Nucleotide								TYPE	p*I*	Mmass
				120	135	301														
g6pdh	LEM75, LPN114, PM1, BCN16, IMT260, ISS1036, RRR-B	1, 228	100	K	E	S				T	C	T	C	C	A	C	G	1	5.81	63299.02
	LLM1026	105	**105**	K	E	S				T	C	T	C	C	A	C	G	1	5.81	63299.02
	LSL29+LEM2298	1, 183	100	**K/M**	E	S				T	C	T	C	C	A	C	G	1/2	5.73 or 5.81	63302.04 or 63299.02
	LLM373+LEM935	199, 77	**102**	**K/M**	E	S				T	C	T	C	C	A	C	G	1/2	5.73 or 5.81	63302.04 or 63299.02
	LLM580	253	**95**	K	**K**	S				T	C	T	C	C	A	C	G	3	5.98	63298.08
	BUCK+ISS800	78, 188	100	K	E	S				C	C	T	C	C	A	C	G	4	5.81	63299.02
	LEM3249	29	**105**	**K/M**	E	S				C	C	T	C	C	A	C	G	4/5	5.73 or 5.81	63302.04 or 63299.02
	LEM189+LLM175	11, 198	**105**	**M**	E	S				C	C	T	C	C	A	C	G	5	5.73	63302.04
	DEVI,THAK35, DD8	2	100	K	E	S				C	C	C	C	C	A	T	G	5	5.81	63299.02
	GEBRE1	82	100	K	E	S				C	G	C	T	T	G	C	G	7	5.81	63299.02
	GILANI, HUSSEN, LEM3946, 3S, LEM3472, LEM3429, LEM3463	30, 31, 274, 81, 267, 257, 258	100	K	E	S				C	G	C	T	T	A	C	G	8	5.81	63299.02
	SC23	38	100	K	E	**N**				C	C	C	C	C	A	C	A	9	5.81	63326.04
				16	346	358	390	444	499											
fh	18 strains	1, 29, 183, 2, 31, 11, 274 198, 199, 228, 267, 257	100	I	C	K	D	T	T	G	G	G	G					1	6.43	62622.33
	PM1	1	100	I	C	K	D	T	**T/M**	G	G	G	G					1/2	6.43	62652.42 or 62622.33
	BCN16	1	100	I	C	K	D	T	T	G	G	G	C/G					1/3	6.43	62622.33
	GEBRE1, GILANI, 3S	82, 30, 81	100	I	C	K	D	T	T	A	G	G	G					4	6.43	62622.33
	BUCK	78	**110**	**I/M**	**C/F**	K	D	T	**T/M**	G	G	G	G					5	6.43	62622.33 or 62714.49
	ISS510	136	**110**	I	C	**Q**	D	T	T	G	G	G	G					6	6.33	62622.29
	SC23	38	100	I	C	K	D	T	T	G	A	G	G					7	6.43	62622.33
	ISS800	188	**105**	I	C	K	D	**M**	T	G	G	G	G					8	6.43	62632.56
	LEM3463	258	**93**	I	C	K	**G**	T	T	G	G	G/A	G					9/10	6.54	62564.29

Previously described multilocus enzyme electrophoresis (MLEE) phenotype (expressed as mobility of their protein in arbitrary units with "100" being the most prevalent mobility), amino acid differences and silent nucleotide differences are shown, along with the positions of alternative amino acids. Sites of particular interest are in bold. See deposited sequences for additional detail. In the case of putative heterozygotes, theoretical isoelectric point (p*I*) and molecular mass (MMass) have been calculated for all possible amino acid combinations.

BCN16, IMT260, LEM2298, LLM175, LLM373, and RRR-B). Seven homozygous and three heterozygous genotypes were identified. There were four heterozygous sites in LEM3946 and LEM3429, for which we did not determine the haplotypes (phase), yet we speculate that strains LEM3472 and LEM3463/HUSSEN may represent candidate parental strains (see also Mauricio et al., 2006).

3.2. Malic enzyme (me)

The 1,716 nt *me* gene was sequenced from 33 strains of the *L. donovani* complex and from three other *Leishmania* species, representing 23 MON zymodemes. There were 12 single nucleotide polymorhisms (SNPs, 0.73%) of which eight were silent (Table 2) and four changed the aa sequence. The sequences of all six MON-1 strains were identical, whereas the non-MON-1 European *L. infantum* strains differed at several sites (Table 2). The base change in LLM373 changed the aa, from V to M at aa 59, but did not alter the molecular mass (MMass) or p*I*. Three MONs (30, 228, and 274) exhibited the same alternative amino acids, E/D, at heterozygous sites in the same position (aa 563, Table 2).

With the exception of MON-98, all MONs, are reported to be monomorphic in electrophoretic mobility of ME (Rioux et al., 1990). We included seven recently isolated Greek MON-98 *L. infantum* strains, which had the distinct MON-98 MLEE phenotype (Ch. Haralambous, personal communication). However, the MON-98 amino acid sequence, which was identical for all seven strains, did not explain the ME phenotype on MLEE (Table 2).

3.3. Mannose phosphate isomerase (mpi)

The 1,266 nt *mpi* sequence was identical for 16 of 31 strains (LEM75, LPN114, PM1, LSL29, BCN16, IMT260, LEM3249, LEM2298, GEBRE1, GILANI, LEM189, BUCK, 3S, LLM175, LLM373, and ISS1036), which included all the MON-1 strains (Table 2). There were seven SNPs (0.55%) among the remaining 15 strains; four SNPs were silent and three were associated with amino acid changes. Three SNPs at heterozygous sites gave alternative aa sequences, two in single strains (ISO510, K/E, at aa 213 and SC23, S/I, at aa 277) and a third occurred in several strains (K/R, aa 131,Table 2). One of same SNPs also gave the second homozygous site (ISS800, K–E, aa 213).

The mobilities 100, 127, and 100/127 were clearly explicable in terms of p*I* (Table 2). Eight strains in our dataset represented mobility 110, three of which had K/R alternative aa at aa 131, which did not alter p*I* (Table 2). However, the other five strains with mobility 110 had the same genotype as strains with mobility 100. Thus, MPI mobility 110 was not explicable in terms of the observed genotypes.

3.4. Glucose-6-phosphate dehydrogenase (g6pdh)

The 1,689 nt *g6pdh* gene was sequenced in 31 strains of the *L. donovani* complex. There were 11 SNPs (=0.65%), eight were silent and their distribution was associated with the geographical origin of the strains (Table 2).

Four (95, 100, 102, and 105) of five known G6PDH electrophoretic mobilities (Pratlong et al., 2001) were represented among the strains analysed. The same alternative aa, K/M, at a heterozygous site at aa 120, occurred for the three electrophoretic mobilities 100, 102, and 105. In other strains with electrophoretic mobilities 100 or 105, K or M were present at homozygous sites. Therefore, the aa differences observed in G6PDH are not entirely consistent with the electrophoretic mobilities. Note, however, that the three different electrophoretic mobilities concerned, 100, 102, and 105, are very similar, so presumably vulnerable to equivocal resolution. For strain SC23, there was a change from S to N at aa 301 but no change in mobility was predicted or seen.

3.5. Fumarate hydratase (fh)

Four different phenotypes, ranging from mobility 93 to 110 have been described for FH (Rioux et al., 1990; Pratlong et al., 2001). Two *fh* genes of *L. major* are 1,650 nt, at chromosome 24, and 1707 nt, at chromosome 29, diverging 44% and 39% at nt and aa levels, respectively (Ivens et al., 2005). We initially sequenced the gene from chromosome 24 in seven strains originating from India, Ukraine, Malta, Tunisia and Sudan. However, only five SNPs were detected and all were silent.

Accordingly, we sequenced the *fh* gene from chromosome 29. The sequence was identical in 18 of 28 strains (LEM75, LPN114, LSL29, IMT260, LEM3249, LEM2298, DEVI, THAK35, HUSSEN, LEM189, LEM3946, LLM175, LLM373, ISS1036, LEM3472, LEM3429, DD8, and RRR-B). However, of the five genes studied *fh* of chromosome 29 displayed the greatest aa sequence diversity. Six sites were involved, three of the changes were: K–Q at aa 358, D–G at aa 390, and T–M at aa 444 and the remaining three gave alternative amino acids at heterozygous sites (I/M at aa 16, C/F at aa 346, and T/M at aa 499). Silent mutations occurred at four additional sites, with SNPs overall at 0.58% of sites. Such genetic diversity was to be expected from the described range of MLEE profiles. Furthermore, the four electrophoretic mobilities (93, 100, 105, and 110) were explicable by the observed aa diversity. Nevertheless, the limits of the MLEE approach and greater resolution of MLST were clearly seen from the shared 110 mobility of strains BUCK and ISS510, which have four aa differences.

3.6. Phylogenetic analysis

The complete sequences for *icd*, *me*, *mpi*, *g6pdh*, and *fh* from 27 strains of the *L. donovani* complex were used for phylogenetic inference of their relationships, using ML and MP networks. The trees were rooted with *L. tropica* and *L. major* (Fig. 1). The few differences in sequences gave poor resolution for single gene trees. In the case of aa sequences, even concatenated alignments

did not contain sufficient differences for robust trees (data not shown). We therefore confined ML analysis to concatenated alignment of nt sequences of all five genes. The consensus tree showed 100% bootstrap support for the monophyly of the *L. donovani* group (Fig. 1) and, at low bootstrap values, the monophyly of the 14 European *L. infantum* strains. Within the European clade BUCK and ISS800 apparently differed from the bulk of closely related strains isolated from patients in France, Spain, and Portugal.

Other branches also reflected geographical origin of *Leishmania* strains. The Ethiopian and Sudanese strains comprised a common cluster, although not strongly supported. Within this cluster, one of the two African strains of *L. infantum* was identical to *L. archibaldi* strain LEM3429. The other African *L. infantum* 3S grouped with *L. archibaldi* GEBRE1 and *L. donovani* GILANI. African strains assigned to *L. infantum*, *L. donovani* and *L. archibaldi* therefore show no such species-specific grouping. Three *L. donovani* strains of Indian origin (DEVI, THAK35, and DD8) were virtually identical but the position of the Indian *L. donovani* strain SC23 was unresolved (Fig. 1).

To compare MLEE and sequence data, using 15 enzyme mobilities described for the strains used here, we generated a MLEE-based tree, which suffers, however, from generally low bootstrap support (Fig. 2). The MLEE tree was not congruent with the sequence-based tree, in that the MLEE tree split the European strains into two branches. The MLEE-based dataset returned the three *L. archibaldi* strains as a monophyletic group separated from the other strains of African origin (Fig. 2), whilst there was no support for the monophyly of *L. archibaldi* in the sequence-based tree (Fig. 1).

For additional insight, we analysed the data using the statistical parsimony algorithm. To simplify presentation of results, heterozygous positions were considered as ambiguous. The program recognised 22 different haplotypes (13 in Europe, seven in Africa and two in India). This number is an underestimate due to the exclusion of heterozygotes, most of which are of African origin. The concatenated network revealed a strong correlation between molecular data and geographical origin (Fig. 3). The number of mutational steps between Indian and European strains, and between Indian and African strains was comparable. Within the continents, however, geographical

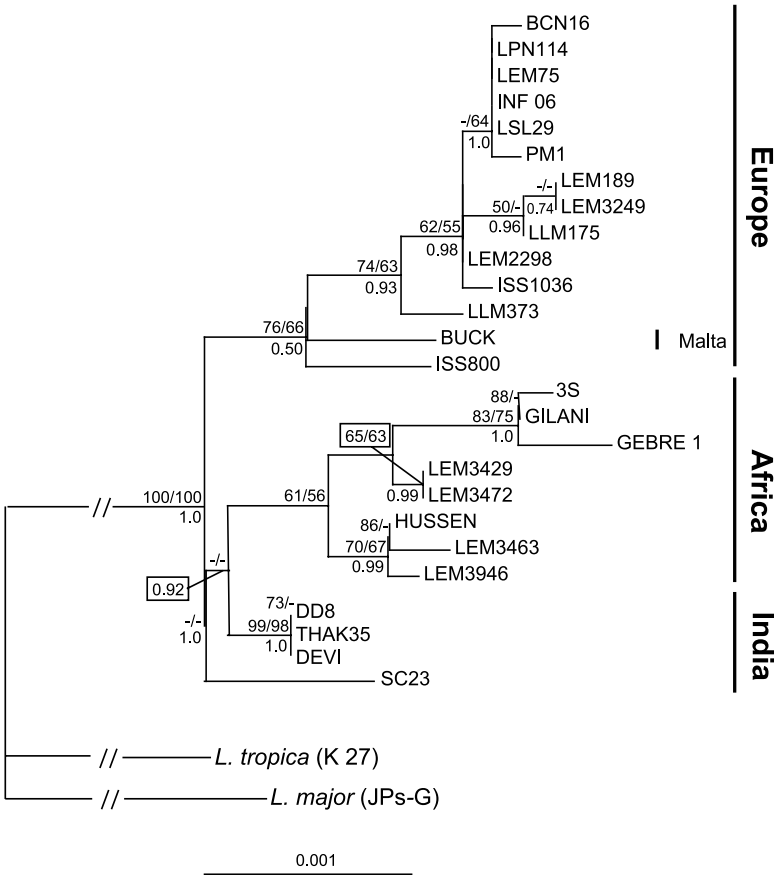

Fig. 1. Phylogenetic analysis of the *Leishmania donovani* complex based on sequences of the *icd*, *me*, *mpi*, *g6pdh*, and *fh* genes. The maximum likelihood tree (log lk = −12,552.86874) was constructed from a concatenated dataset (7,614 characters, 161 parsimony-informative) under the TN93+Γ_8 model of evolution as implemented in Phyml 2.4.4 (Guindon and Gascuel, 2003), the proportion invariant (0.283) was estimated from the dataset. Numbers above branches represent bootstrap support (ML/MP, see text for details). Numbers below branches stand for the bayesian posterior probability. The geographical origin of the strains is indicated to the right.

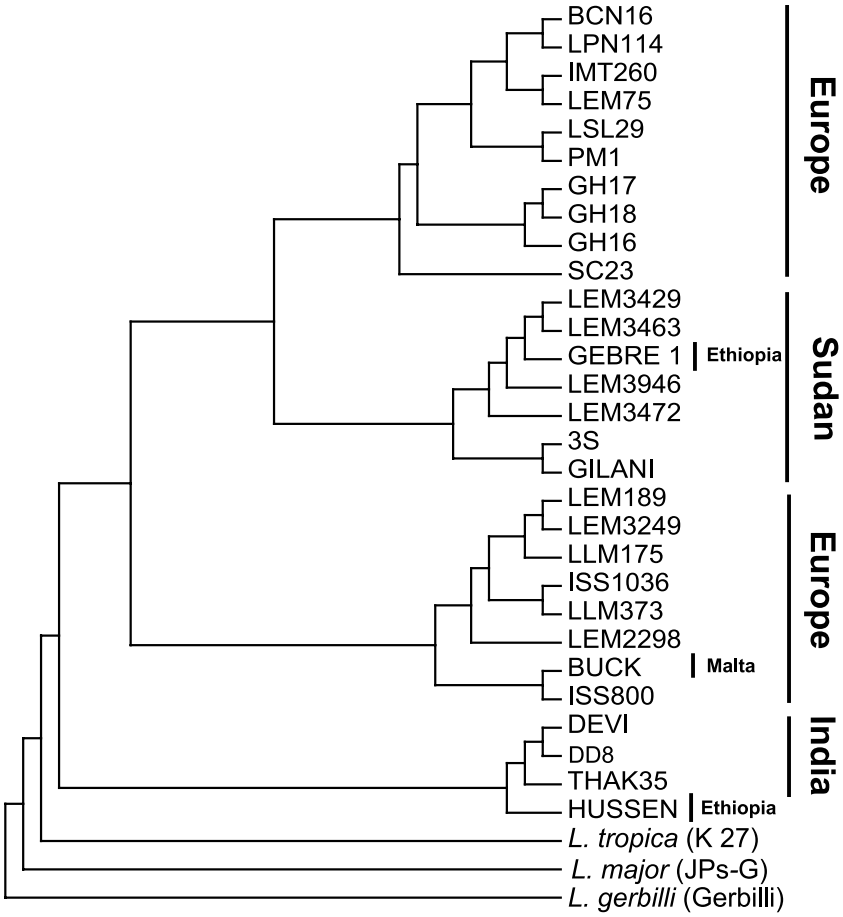

Fig. 2. Phylogenetic analysis of the *Leishmania donovani* complex based on the enzyme profiles of 15 enzymes established by the MLEE-analysis. The geographical origin of the strains is indicated to the right.

correlation was lost: Sudanese and Ethiopian strains intermingled, as did Spanish, French, and Italian strains. As a reflection of its geographical origin, the strain from Malta (BUCK) occupied an intermediate position between European and non-European strains. Only one representative of the Indian group, namely SC23, was separated from the other strains from the Indian subcontinent (Fig. 3).

4. Discussion

The aim of this study was to examine the genetic diversity within five genes encoding enzymes used for MLEE of *Leishmania*, with a view to encouraging the development of MLST as a high resolution alternative to MLEE. The identification of *Leishmania* is still largely based on MLEE. However, MLEE fails to detect nucleotide substitutions that do not give rise to aa changes and aa substitutions that do not alter electrophoretic mobility, and it does not distinguish similar electrophoretic mobilities that are dependent on distinct genotypes. Furthermore, post-translational modification may change electrophoretic mobility, for example, by methylation of lysine and arginine. A number of DNA-based analyses of *Leishmania* have been compared with the MLEE-based typing, some producing equivalent conclusions (Tibayrenc et al., 1993; Brisse et al., 2000; Telleria et al., 2004) others showing significant incongruence (Kuhls et al., 2005; Mauricio et al., 2006; this work). Protein-coding genes, including those encoding enzymes, are promising targets for the identification of *Leishmania* (Greenblatt et al., 2002; Castilho et al., 2003; Jamjoon et al., 2004; Tintaya et al., 2004). Markers that evolve more rapidly than enzyme genes, such as surface antigen genes, non-transcribed regions or microsatellites, may give supplementary insight into gene flow among intraspecific populations (Bulle et al., 2002; Tintaya et al., 2004; Kuhls et al., 2005).

Herein, we focused on sequencing *icd*, *me*, *mpi*, *g6pdh*, and *fh*. Complete sequences were obtained for at least 27 strains of the *L. donovani* complex and for *L. tropica*, *L. gerbilli*, and *L. major*. The *Leishmania* strains were selected to represent named species, designated MON zymodemes, geographical origins and clinical presentations. The gene sequences were between 1.0 and 2.0 kb, with less than 1% of polymorphic sites.

The enzyme ICD is monomorphic on MLEE and as expected no *icd* polymorphisms were found that would yield amino acid changes. Similarly, for ME, among the strains that we examined, only MON-98 had distinct

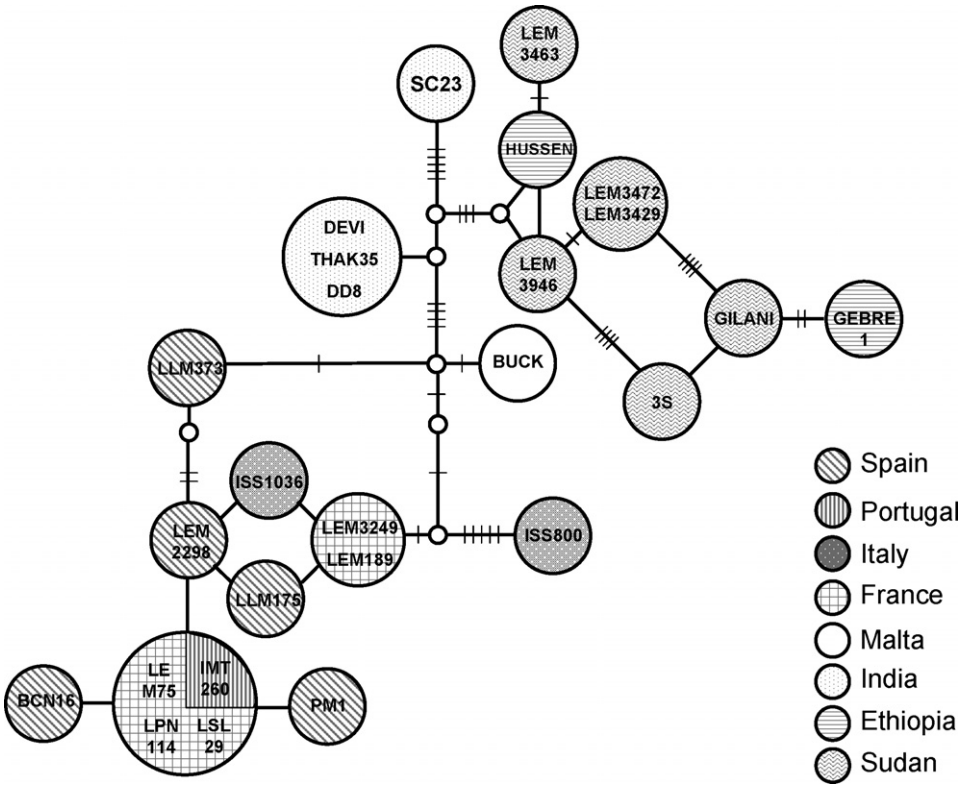

Fig. 3. Statistical parsimony network constructed from the concatenated dataset of *icd*, *me*, *mpi*, *g6pdh*, and *fh* genes with TCS 1.21 software (Clement et al., 2000). Large circles represent *Leishmania* haplotypes with shading indicating their geographical origin. Small circles represent hypothetical ancestors and commas indicate mutational steps between individual haplotypes.

MLEE profiles (Rioux et al., 1990; Pratlong et al., 2001, 2004). Nevertheless, notable hidden genetic diversity was encountered in *me*, including different aa sequences, M at aa 59 and E/D at aa 563. Consistent with the MLEE data, since the differences in molecular mass are negligible and the p*I* remains unaltered, no differences in enzyme electrophoretic mobilities have been detected. However, we were unable to find DNA sequence polymorphisms to explain the observed MLEE phenotype for the MON-98 strains from Greece.

Three of four MPI mobilities, 100, 127, and 100/127, were explicable by *mpi* genetic diversity, but not phenotype 110. Some of the relevant strains are heterozygous at aa 131, yet other strains are not. The discrepancy between mobility 110 and genotype may be explicable by post-translational modification.

The correlation between MLEE for G6PDH and sequence analysis of *g6pdh* is less obvious. One genotype, K/M at aa 120, was associated with three different MLEE mobilities. However, it is notable that the three different mobilities concerned, 100, 102, and 105, were extremely close and therefore vulnerable to equivocal interpretation, although slight differences in mobility caused by post-translational modification cannot be ruled out. An almost perfect match was found between FH electrophoretic mobilities on MLEE and sequence diversity of *fh*. Furthermore, *fh* gave additional discrimination over MLEE, making this target highly suitable for MLST. Also one methylation site was observed due to the mutation in aa in LEM3463.

Overall, these data support the potential of MLST to provide a high resolution alternative to MLEE. However, some apparent discrepancies between MLEE groups and sequence data need to be explained, in particular the failure of sequence data to separate MON-98. It is possible that a post-translational modification, such as glycosylation, protein–protein interaction or an artificial phenotype, may be involved. It is conceivable but unlikely that the *me* analyzed does not encode the ME seen in MLEE.

DNA sequencing revealed a high number of heterozygotes, and these were present in different genes (Table 2). Heterozygosity at several sites within a gene and at several loci can be a consequence of genetic exchange; mutation is a less likely explanation (Mauricio et al., 2006). The abundance of heterozygotes provides support for the presence of genetic exchange in *Leishmania*, which is much debated (Gibson and Stevens, 1999; Gaunt et al., 2003). Heterozygosity was particularly prominent in two strains, LEM3946 and LEM3429, as also noted for LEM3946 with other MLST targets, which suggested genetic exchange between distant genetic groups (Mauricio et al., 2006). Both strains seem to bear traces of relatively recent recombination events and the putative parents may be candidates for genetic hybridization experiments.

Considering its scale (263 kb of new protein-coding sequences), along with the complementary analysis by Mauricio et al. (2006), this is one of the first extensive attempts to examine the intraspecific diversity of protein-coding genes in protozoa. The concatenated tree (Fig. 1) is based on ~7.7 kb of protein-coding sequence from each strain and despite the lack of resolution in several nodes, important conclusions can be drawn. Since bifurcating trees may not optimally represent the intraspecific gene evolution, we also applied the statistical parsimony network to the concatenated dataset. Both methods showed similar genetic distances between Indian and African clusters and between Indian and European clusters.

As in several recent studies (Jamjoon et al., 2004; Kuhls et al., 2005), the *L. donovani*, *L. infantum*, and *L. archibaldi* strains of African origin do not form monophyletic groups but are intermingled in the tree and network. Geographical origin seems to govern the position rather than designated species, zymodeme or pathogenicity. Regardless of species, the Indian and Sudanese strains are closely related to strains originating from the same region. All 14 *L. infantum* strains isolated from patients in Europe cluster together, to the exclusion of two *L. infantum* strains of Sudanese origin. Europe and Africa may represent independent evolutionary centres. There is no support for monophyly of *L. archibaldi*.

From the potential MLST targets examined here it is clear that *fh* is particularly promising, with high resolution and clear compatibility with MLEE. The other four enzyme genes are potentially exploitable for MLST, irrespective of precise compatibility with MLEE phenotype, although the resolving power of *icd* is low for the *L. donovani* complex. It is not a prerequisite that MLST targets must correspond with MLEE targets and therefore the more diverse *icd* on chromosome 33 warrants further investigation. Sequences of five other enzyme genes have already yielded more than seven alleles each for *gpi*, *nh1*, and *pgd*, five for *nh2* and four for *asat* (Mauricio et al., 2006).

To make MLST a reality, these targets need to be applied to a wider range of *Leishmania* species or alternative targets selected, for example for the subgenus *Viannia*. Ten or more genes could be applied in a traditional MLST format of 700 bp regions or in an extended MLST format to include the entire gene sequences, although this would be more laborious. The SNPs described for *nh1*, *nh2*, and *asat* are within 700 and 700 bp regions within *gpi* and *pgd* could be selected to cover most of the genetic diversity. It will be necessary to deposit *Leishmania* MLST data in a publicly accessible database.

It is clear the MLST provides higher resolution than MLEE. MLST is also applicable to fewer parasites and to biopsies. Furthermore, MLST overcomes the problem of comparability between locally generated data. Nevertheless, the development of high resolution, more efficient alternatives to MLST cannot be ignored.

Acknowledgements

We thank Jean-Pierre Dedet and Francine Pratlong (University of Montpellier, Montpellier), Gabriele Schonian and Katrin Kuhls (Humboldt University, Berlin), Ketty Soteriadou and Christos Haralambous (Hellenic Pasteur Institute, Athens), and Carmen Chicharro Gonzalo and Israel Cruz Mata (National Centre of Tropical Medicine, Madrid) for providing cell cultures and/or DNA samples of different *Leishmania* strains used in this study. This work was supported by grants from the European Union (QLK2-CT-2001-01810), the Grant Agency of the Czech Academy of Sciences (Z60220518) and the Ministry of Education of the Czech Republic (6007665801 and 2B06129).

References

Breniere, S.F., Telleria, J., Bosseno, M.F., Buitrago, R., Bastrenta, B., Cuny, G., Banuls, A.L., Brewster, S., Barker, D.C., 1999. Polymerase chain reaction-based identification of New World *Leishmania* species complexes by specific kDNA probes. Acta Trop. 73, 283–293.

Brisse, S., Barnabé, C., Tibayrenc, M., 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. Int. J. Parasitol. 30, 35–44.

Bulle, B., Millon, L., Bart, J.-M., Gállego, M., Gambarelli, F., Portús, M., Schnur, L., Jaffe, C.L., Fernandez-Barredo, S., Alunda, J.M., Piarroux, R., 2002. Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis. J. Clin. Microbiol. 40, 3391–3397.

Castilho, T.M., Shaw, J.J., Floeter-Winter, L.M., 2003. New PCR assay using glucose-6-phosphate dehydrogenase for identification of *Leishmania species*. J. Clin. Microbiol. 41, 540–546.

Cavalli-Sforza, L.L., Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19, 233.

Chiurillo, M.A., Sachdeva, M., Dole, V.S., Yepes, Y., Miliani, E., Vazquez Rojas, A., Crisante, G., Guevara, P., Anez, N., Madhubala, R., Ramirez, J.L., 2001. Detection of *Leishmania* causing visceral leishmaniasis in the Old and New Worlds by a polymerase chain reaction assay based on telomeric sequences. Am. J. Trop. Med. Hyg. 65, 673, 582.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.

El-Tai, N.O., Osman, O.F., El Fari, M., Presber, W., Schonian, G., 2000. Genetic heterogeneity of ribosomal internal transcribed spacer in clinical samples of *Leishmania donovani* spotted on filter paper as revealed by single-strand conformation polymorphisms and sequencing. Trans. R. Soc. Trop. Med. Hyg. 94, 575–579.

Felsenstein, J., 1989. PHYLIP – phylogeny inference package (version 3.2). Cladistics 5, 164–166.

Gangneux, J.P., Menotti, J., Lorenzo, F., Sarfati, C., Blanche, H., Bui, H., Pratlong, F., Garin, Y.J.F., Derouin, F., 2003. Prospective value of PCR amplification and sequencing for diagnosis and typing of Old World *Leishmania* infections in an area of non-endemicity. J. Clin. Microbiol. 41, 1419–1422.

Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Solis Mena, S., Veazey, P., Miles, G.A.J., Acosta, N., Rojas de Arias, A., Miles, M.A., 2003. Mechanism of genetic exchange in American Trypanosomes. Nature 421, 936–939.

Gibson, W.C., Stevens, J.R., 1999. Genetic exchange in the Trypanosomatidae. Adv. Parasitol. 43, 1–46.

Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36, 182–198.

Greenblatt, C.L., Schnur, L.F., Bar-Gal, G.K., Ermolaev, H., Peleg, N., Barrett, M.P., 2002. Polymorphism among alleles of the 6-phosphogluconate dehydrogenase gene from *Leishmania major* and *Leishmania tropica*. Mol. Biochem. Parasitol. 125, 185–188.

Guindon, S., Gascuel, O., 2003. Simple, fast and accurate algorithm to estimate large phylogenies by maximal likelihood. Syst. Biol. 52, 696–704.

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

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Ivens, A., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., et al., 2005. The genome of the kinetoplastid parasite *Leishmania major*. Science 309, 436–442.

Jamjoom, M.B., Ashford, R.W., Bates, P.A., Kemp, S.J., Noyes, H.A., 2002. Towards a standard battery of microsatellite markers for the analysis of the *Leishmania donovani* complex. Ann. Trop. Med. Parasitol. 96, 265–270.

Jamjoon, M.B., Ashford, R.W., Bates, P.A., Chance, M.L., Kemp, S.J., Watts, P.C., Noyes, H.A., 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, 1–11.

Kuhls, K., Mauricio, I.L., Pratlong, F., Presber, W., Schonian, G., 2005. Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex. Microbes Infect. 7, 1224–1234.

Lainson, R., Shaw, J.J., 1987. Evolution, classification and geographical distribution. In: Peters, W., Killick-Kendrick, R. (Eds.), The Leishmaniases in Biology and Medicine. Academic Press, London, pp. 1–120.

Lewin, S., Schonian, G., El Tai, N., Oskam, L., Bastien, P., Presber, W., 2002. Strain typing in *Leishmania donovani* by using sequence-confirmed amplified region analysis. Int. J. Parasitol. 32, 1267–1276.

Mauricio, I.L., Howard, M.K., Stothard, J.R., Miles, M.A., 1999. Genomic diversity in the *Leishmania donovani* complex. Parasitology 119, 237–246.

Mauricio, I.L., Stothard, J.R., Miles, M.A., 2000. The strange case of *Leishmania chagasi*. Parasitol. Today 16, 188–189.

Mauricio, I.L., Stothard, J.R., Miles, M.A., 2004. *Leishmania donovani* complex: genotyping with the ribosomal internal transcribed spaces and mini-exon. Parasitology 128, 1–5.

Mauricio, I.L., Yeo, M., Baghaei, M., Doto, D., Silk, R., Pratlong, F., Zemanová, E., Dedet, J.-P., Lukeš, J., Miles, M.A., 2006. Towards multi locus sequence typing in the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GP1, NH1, NH2, PGD). Int. J. Parasitol. 36, 757–769.

Noyes, H.A., Reyburn, H., Bailey, J.V., Smith, D., 1998. A nested-PCR-based schizodeme method for identifying *Leishmania* kinetoplast minicircle classes directly from clinical samples and its application to the study of the epidemiology of *Leishmania tropica* in Pakistan. J. Clin. Microbiol. 36, 2877–2881.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Posada, D., Crandall, K.A., 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. 16, 37–45.

Pratlong, F., Dereure, J., Bucheton, B., El-Safi, S., Dessein, A., Lanotte, G., Dedet, J.P., 2001. Sudan: the possible original focus of visceral leishmaniasis. Parasitology 122, 599–605.

Pratlong, F., Rioux, J.-A., Marty, P., Faraut-Gambarelli, F., Dereure, J., Lanotte, G., Dedet, J.-P., 2004. Isoenzymatic analysis of 712 strains of *Leishmania infantum* in the south of France and relationship of enzymatic polymorphism to clinical and epidemiological features. J. Clin. Microbiol. 42, 4077–4082.

Rioux, J.-A., Lanotte, G., Serres, E., Pratlong, F., Bastien, P., Perieres, J., 1990. Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. Ann. Parastol. Hum. Comp. 65, 111–125.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J.K., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., Wendel, J.F., 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossipium*. Mol. Biol. Evol. 20, 633–643.

Telleria, J., Barnabé, C., Hide, M., Banuls, A.-L., Tibayrenc, M., 2004. Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 137, 133–141.

Templeton, A.R., Crandall, K.A., Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132, 619–633.

Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D., Ayala, F.J., 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Natl. Acad. Sci. USA 90, 1335–1339.

Tintaya, K.W.Q., Ying, X., Dedet, J.-P., Rijal, S., De Bolle, X., Dujardin, J.-C., 2004. 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, 1035–1043.

Zemanová, E., Jirků, M., Mauricio, I.L., Miles, M.A., Lukeš, J., 2004. Genetic polymorphism within the *Leishmania donovani* complex: correlation with geographical origin. Am. J. Trop. Med. Hyg. 70, 613–617.