International Journal for Parasitology 49 (2019) 555–567

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

A novel multilocus sequence typing scheme identifying genetic diversity amongst *Leishmania donovani* isolates from a genetically homogeneous population in the Indian subcontinent ☆

Sultana Shahana Banu [a,b,c,d], Wieland Meyer [b,c], Kennio Ferreira-Paim [b,c,e], Qinning Wang [a], Katrin Kuhls [f], Elisa Cupolillo [g], Gabriele Schönian [h], Rogan Lee [a,b,*]

[a] *Parasitology Department, Centre for Infectious Diseases and Microbiology Laboratory Services (CIDMLS), Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, Westmead, Sydney, NSW, Australia*
[b] *Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Faculty of Medicine and Health, Sydney Medical School, Westmead Clinical School, Marie Bashir Institute for Emerging Infectious Diseases and Biosecurity, Westmead Hospital (Research and Education Network), The University of Sydney, Sydney, NSW, Australia*
[c] *Westmead Institute for Medical Research, Westmead, Sydney, NSW, Australia*
[d] *Directorate General of Health Services (DGHS), Ministry of Health and Family Welfare (MOHFW), Dhaka, Bangladesh*
[e] *Department of Microbiology, Federal University of Triangulo Mineiro, Uberaba, Brazil*
[f] *Division of Molecular Biotechnology and Functional Genomics, Technical University of Applied Sciences Wildau, Wildau, Germany*
[g] *Laboratory on Leishmaniasis Research, Oswaldo Cruz Institute – Fiocruz, Rio de Janeiro, Brazil*
[h] *Institute for Microbiology and Hygiene CC05, Charité University Medicine Berlin, Berlin, Germany*

ARTICLE INFO

Article history:
Received 15 August 2018
Received in revised form 20 February 2019
Accepted 23 February 2019
Available online 18 May 2019

Keywords:
Leishmania donovani
Genetic diversity
Multilocus sequence typing
Indian subcontinent
Bangladesh

ABSTRACT

In the Indian subcontinent, infection with *Leishmania donovani* can cause fatal visceral leishmaniasis. Genetic variation in *L. donovani* is believed to occur rapidly from environmental changes and through selective drug pressures, thereby allowing continued disease occurrence in this region. All previous molecular markers that are commonly in use multilocus microsatellite typing and multilocus sequence typing, were monomorphic in *L. donovani* originating from the Indian subcontinent (with only a few exceptions) and hence are not suitable for this region. An multilocus sequence typing scheme consisting of a new set of seven housekeeping genes was developed in this study, based on recent findings from whole genome sequencing data. This new scheme was used to assess the genetic diversity amongst 22 autochthonous *L. donovani* isolates from Bangladesh. Nineteen additional isolates of the *L. donovani* complex (including sequences of *L. donovani* reference strain BPK282A1) from other countries were included for comparison. By using restriction fragment length polymorphism of the internal transcribed spacer 1 region (ITS1-RFLP) and ITS1 sequencing, all Bangladeshi isolates were confirmed to be *L. donovani*. Population genetic analyses of 41 isolates using the seven new MLST loci clearly separated *L. donovani* from *Leishmania infantum*. With this multilocus sequence typing scheme, seven genotypes were identified amongst Bangladeshi *L. donovani* isolates, and these isolates were found to be phylogenetically different compared with those from India, Nepal, Iraq and Africa. This novel multilocus sequence typing approach can detect intra- and inter-species variations within the *L. donovani* complex, but most importantly these molecular markers can be applied to resolve the phylogenetically very homogeneous *L. donovani* strains from the Indian subcontinent. Four of these markers were found suitable to differentiate strains originating from Bangladesh, with marker *A2P* being the most discriminative one.

1. Introduction

Visceral leishmaniasis (VL) is a 'vector-borne disease' caused mainly by species of the *Leishmania donovani* complex, which is comprised of *L. donovani* in the eastern hemisphere, and *Leishmania infantum* in both eastern and western hemispheres (Kuhls et al., 2005; Lukes et al., 2007; Leblois et al., 2011). These parasites spread to organs, mainly those of the reticuloendothelial system

☆ Note: Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers: MK211285 – MK211306 for ITS1; KX155845 – KX155886 for *RSP*; KX155887 – KX155928 for *FAD*; KX155929 – KX155970 for *STK*; KX155971 – KX156012 for *A2P*; KX156013 – KX156054 for *UBE1*; KX156055 – KX156096 for *MTF*; and KX156097 – KX156138 for *SF3B1*.

* Corresponding author at: Parasitology Department, Centre for Infectious Diseases and Microbiology Laboratory Services (CIDMLS), ICPMR, Westmead Hospital, Westmead, Sydney, NSW, Australia. Fax: +61 2 9893 8659.
E-mail address: rogan.lee@health.nsw.gov.au (R. Lee).

https://doi.org/10.1016/j.ijpara.2019.02.010
0020-7519/

and the disease is fatal if left untreated (Murray et al., 2005). More than 90% of VL cases are reported from just eight countries (in order of highest case numbers): India, Bangladesh, Sudan, Brazil, Ethiopia, South Sudan, Iraq and Nepal (Alvar et al., 2012). The vast majority of cases occur in the Indian subcontinent, mainly from remote areas of rural India, Bangladesh, and Nepal (Alvar et al., 2012). Although Bangladesh has reached the elimination target in 90% of the endemic sub-districts, 16 of those still report 1.06 to 18.25 VL cases per 10,000 people per year (Chowdhury et al., 2014; WHO, 2015). The persistence of VL in some endemic foci, even after extensive control interventions, might be due to genetic variants of *L. donovani* that are refractory to current treatment regimens. These pockets of infection may give rise to new outbreaks and spread to other areas. In addition, new variants may also be imported from neighbouring India to adjacent districts of Bangladesh with population movement across the border. A recent study shows that genetic mutations and associated structural changes have occurred in *L. donovani* isolates from India and Nepal, therefore indicating that this parasite can persist by adapting rapidly to the changing environment and drug pressures (Imamura et al., 2016). Investigations to detect genetic diversity within *L. donovani* population in Bangladesh are now urgently required.

Leishmania donovani in the Indian subcontinent was found to be genetically extremely homogeneous by multilocus microsatellite typing (MLMT) (Alam et al., 2009). The first steps towards revealing heterogeneity in the Indian subcontinent population were two studies applying genome-wide single nucleotide polymorphism (SNP) analysis and whole-genome-sequencing (WGS) (Downing et al., 2011, 2012). In contrast to the limited resolving power of MLMT within the Indian subcontinent, significant genetic diversity and extensive structural variations were observed within Indian and Nepalese isolates. A recent study applying WGS on 204 *L. donovani* strains from the Indian subcontinent (98 strains from Nepal, 98 from India, and only eight from Bangladesh) described three divergent genetic lineages of *L. donovani* in this region, with a core population of 191 strains that further subdivided into nine sub-populations. These results pointed to a significant spatial and temporal structure of the parasite in the Indian subcontinent and their complex evolution, despite the extremely low level of diversity in comparison to other geographical areas (Imamura et al., 2016). A previous Indian study detected 15 genotypes of *L. donovani* by restriction fragment length polymorphism of the kinetoplast deoxyribonucleic acid (kDNA-RFLP) analysis and eight genotypes by other markers (one microsatellite, internal transcribed spacer 1 (ITS1), hydrophilic acylated surface protein b and cysteine proteinase b) indicating the presence of diverse genotypes of the parasite in the country (Srivastava et al., 2011). Furthermore, *Leptomonas* co-infection with VL was described in another study, which revealed a genetic association of Indian *L. donovani* isolates with *Leptomonas* spp. (Singh et al., 2013). In addition, detection of *Leishmania tropica* in India (Sharma et al., 2005; Kumar et al., 2007) and *Leishmania major* in Nepal (Kumar et al., 2008) demonstrates the presence of other *Leishmania* spp. in the Indian subcontinent. Hybridization or recombination events could occur between different species of *Leishmania* (Ravel et al., 2006) in this region and result in marked genetic variation of *L. donovani* with the potential of evolving into more virulent strains (Ravel et al., 2006; Nolder et al., 2007; Chargui et al., 2009; Odiwuor et al., 2011; Gelanew et al., 2014). Genetic exchange in *L. donovani* and spread of mutations by recombination were recently described in Indian and Nepalese *L. donovani* isolates (Imamura et al., 2016). However, the eight Bangladeshi isolates included in this earlier study did not show remarkable diversity, which might be due to the small sample size.

Among various methods used for typing *L. donovani* complex species and strains, multilocus sequence typing (MLST) can be highly discriminatory with reproducible and universally comparable results, and therefore is expected to be the future 'gold standard' (Mauricio et al., 2006; Zemanova et al., 2007; El Baidouri et al., 2013; Akhoundi et al., 2016; Pérez-Losada et al., 2017). This method is able to discriminate at the inter- and intra-species level across the *Leishmania* genus (Miles et al., 2009; Akhoundi et al., 2016), including closely related *Leishmania* strains, which has been shown for both Old (Mauricio et al., 2006; Zemanova et al., 2007; El Baidouri et al., 2013; Zhang et al., 2013; Gelanew et al., 2014) and New World *Leishmania* spp. (Tsukayama et al., 2009; Boite et al., 2012; Marlow et al., 2014; Marco et al., 2015).

On the other hand, molecular markers, previously used in MLST for population genetic studies on the *L. donovani* complex, did not show sufficient discriminatory power; neither Bangladeshi nor Nepalese isolates were included in these earlier studies (Mauricio et al., 2006; Zemanova et al., 2007; El Baidouri et al., 2013; Gelanew et al., 2014; Akhoundi et al., 2016). These studies did include a few Indian strains, all of which belonged to the zymodeme MON-2 and showed identical genotypes with the majority of genetic markers used (Mauricio et al., 2006; Zemanova et al., 2007; El Baidouri et al., 2013). Although these studies recommend MLST as the method of choice, there is no ideal MLST panel of molecular markers able to discriminate between and within species of the *L. donovani* complex, especially those from the highly homogeneous group of the Indian subcontinent. Since WGS and the respective data analysis is not readily available in a resource-poor setting, an improved MLST method using seven new housekeeping genes from the *L. donovani* genome was applied in the current study to analyse genetic diversity and phylogenetic relationships of *L. donovani* isolates from Bangladesh. It was expected that these new molecular markers could be deposited in common accessible databases for future identification-based projects.

2. Materials and methods

2.1. Study populations, isolation of parasites and DNA extraction

Clinical specimens were obtained from VL patients attended Surya Kanta Kala-azar Research Centre (SKKRC), Mymensingh, Bangladesh, for diagnosis and treatment from May 2013 to May 2014. Patients were suspected of having VL based on the clinical features of prolonged fever for 2 weeks or more, hepatosplenomegaly, anaemia, weight loss, weakness, and cachexia (Kumar and Nylen, 2012; van Griensven and Diro, 2012). Approximately 3–4 ml of venous blood was collected in a Vacutainer® containing lithium-heparin (BD, Franklin Lakes, NJ, USA) from each clinically suspected patient for sero-diagnosis by a rK39 immunochromatographic test (rK39 ICT) and microscopy. The clinically suspected and serologically positive patients were admitted into SKKRC for confirmative diagnosis by parasite detection from blood buffy coat and splenic aspirates by microscopy and culture. The SKKRC is the only Government run VL research centre in Bangladesh where splenic aspiration is routinely performed to isolate parasites for research purposes. This research centre provides facilities for diagnosis and treatment of VL patients under the National VL Elimination Program. Routine investigations were carried out to exclude other possible causes of hepatosplenomegaly, fever and anaemia. All cases were screened for malaria using a rapid test kit (SD BIOLINE, Malaria Ag P.f/Pan, Standard Diagnostics, Korea). Individuals presenting with the previously mentioned clinical features and showing positive reactions to the Widal and Mantoux tests, were not included in this study. Complete blood counts, haemoglobin levels, bleeding and clotting times were investigated, and severely anaemic patients were given a blood transfusion.

Only patients who were assessed as clinically fit for splenic aspiration were chosen for this procedure. An experienced physician was designated to collect splenic aspirates by using a standard universal procedure (Kager, 1992). A few drops of aspirate were smeared on glass slides for microscopic examination and the rest was inoculated into Nicolle-Novy-McNeal (NNN) media (Lumsden and McMillan, 1996) with the antibiotics penicillin and gentamycin (Square Pharmaceuticals Ltd, Dhaka, Bangladesh) and incubated at 24–26 °C for at least 1 month until proved negative for parasite growth. All patients were treated with liposomal amphotericin B (AmBisome) (Gilead, Foster, CA, USA), the anti-*Leishmania* treatment, under the National VL Elimination Program.

The cultured parasites from NNN media were harvested in DMEM (Sigma–Aldrich, Chemie GmbH, St. Louis, USA) supplemented with 20% heat-inactivated FCS (HyClone Laboratories, Inc., South Logan, Utah, USA) and 4 mM L-glutamine (Sigma–Aldrich, St Louis, USA) at 24–26 °C, and underwent several passages before being stored frozen in liquid nitrogen at the Pathology Department, Faculty of Veterinary Medicine, Bangladesh Agricultural University (BAU), Mymensingh. Liquid phases from all positive cultures in NNN media were stored at −80 °C. All frozen cultures were transported on dry ice to Australia for molecular testing.

DNA was extracted from frozen cultures (22 Bangladeshi and one Sudanese isolates) using the DNA mini kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions and stored at −20 °C until use. DNA only from uncontaminated cultures of *Leishmania* spp. were analysed for MLST in this study. For this purpose, extracted DNA that showed expected clean bands in the gel after the Qiagen HotStar *Taq* PCR, and had quality sequences after sequencing, was used for analyses.

Bangladeshi isolates were named sequentially as follows for laboratory identification (laboratory code): two letters representing the names of sub-district and district of their origin (e.g. TM: T = Trishal sub-district, M = Mymensingh district), two to three digits indicating the serial number of the isolate, two letters (BD) abbreviated the country, four digits representing the year of isolation followed by two letters that abbreviated the name of the collector. All variables were separated by a slash except between geographical location and serial number, which were separated by a hyphen (e.g. TM-29/BD/2014/SB). World Health Organization (WHO) codes were also assigned for 22 Bangladeshi and one Sudanese (CHW/SD/2010/MW) *Leishmania* isolates according to WHO recommendations (Supplementary Table S1) (WHO, 2010). This Sudanese isolate originated from a Sudanese child who developed VL in Australia. It was included in this study for comparison.

2.2. Species identification of the isolates

The species identification of the 22 new Bangladeshi *Leishmania* isolates was performed by RFLP analysis of the ITS1 region (ITS1-RFLP) as described elsewhere (Schonian et al., 2003). The amplified ITS1 region was also sequenced. The obtained ITS1 sequences from all Bangladeshi *Leishmania* isolates were aligned against ITS1 sequences of *L. donovani* available in the National Center for Biotechnological Information (NCBI) nucleotide database (GenBank). All sequences generated from Bangladeshi isolates were deposited in GenBank. The Sudanese isolate was previously identified as *L. donovani* based on ITS1-RFLP (Schonian et al., 2003) at St. Vincent Hospital, Darlinghurst, New South Wales, Australia. Species of the reference isolates included in this study had previously been identified by isoenzymatic analysis using multilocus enzyme electrophoresis (MLEE) in the WHO reference centers of Montpellier (MON), France, London (LON), UK, and the Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil. The ITS1 and ITS2 regions of these reference strains were also sequenced and studied by MLMT (Kuhls et al., 2005, Kuhls et al., 2007) (Supplementary Table S1).

2.3. Parasite isolates analysed for MLST

Forty-one isolates of the *L. donovani* complex analysed in this study included 22 new isolates from Bangladesh, which were obtained from autochthonous VL patients attending the SKKRC and one Sudanese isolate received from the Institute of Clinical Pathology and Medical Research (ICPMR), Australia. Reference isolates ($n = 17$) were received as DNA samples. Eleven of those originating from India, Sudan, Ethiopia, Kenya, Iraq, Tunisia, France, and Costa Rica were provided by the Institute of Microbiology and Hygiene, Charite University, Berlin, Germany. The other six DNA samples from India, Sudan, Ethiopia, Portugal, and Brazil were provided by Fiocruz, Rio de Janeiro, Brazil. The genome sequence of the Nepalese *L. donovani* reference strain BPK282A1 (retrieved from GenBank, assembly accession: GCA_000227135.2 and made available by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, http://www.genome.jp/kegg-bin/show_organism?org=ldo) was used as the master sequence in this study.

Bangladeshi *Leishmania* isolates were obtained from five districts by uncloned culturing into NNN media. Uncloned cultures were specifically used in this study with the intention of capturing the diversity present within natural *L. donovani* populations in Bangladesh (Bhattacharyya et al., 2013). Of the total identified isolates used in this study, 34 isolates were *L. donovani*, four were *L. infantum* of the Old World and three were *L. infantum* of the New World. Information regarding the parasite isolates is available in the Supplementary Table S1.

2.4. Selection of target loci

A panel of seven single copy housekeeping genes was selected based on the observed sequence diversity within *L. donovani* strains of the Indian subcontinent, detected by whole-genome SNP analysis of isolates from India and Nepal (Downing et al., 2012). These were genes of rhomboid-like serine protease (*RSP*), fatty acid/sphingolipid δ-4 desaturase (*FAD*), serine/threonine-protein kinase (*STK*), 5′ a2rel-related protein (*A2P*), ubiquitin-activating enzyme e1 (*UBE1*), mannosyltransferase (*MTF*), and splicing factor 3B subunit1 (*SF3B1*). These genes are located on seven different chromosomes of the *L. donovani* genome (Downing et al., 2012). The search for candidate genes was performed on the whole genome sequences of *L. donovani* established by the KEGG database. The target genes are distributed across the *L. donovani* genome with orthologues present throughout the genomes of other *Leishmania* spp. such as *L. infantum*, *L. major*, *Leishmania mexicana*, *Leishmania panamensis*, *Leishmania peruviana*, and *Leishmania braziliensis* (Downing et al., 2011). The homology of the selected gene sequences was also checked with other *Leishmania* spp. using the Basic Local Alignment Search Tool (BLAST) at https://blast.ncbi.nlm.nih.gov/Blast.cgi to use these targets for future typing-based studies for a broader range of *Leishmania* spp. The designated genes were not under positive selection as previously shown (Downing et al., 2011). The selected loci with their chromosomal locations are shown in Table 1.

2.5. Primer design

The respective primers were designed using the genome sequence of the strain BPK282A1. The genome sequence of *L. donovani* was chosen, as this parasite species is the major cause of VL in the Indian subcontinent. Using the web-based Custom Primers-Oligoperfect software (https://tools.thermofisher.com/content.cfm?pageid=9716&icid=fr-oligo-6?CID=fl-oligoperfect), seven

Table 1
Information of multilocus sequence typing loci and PCR conditions used in the current study.

Gene/locus	Chromosome location	Primer name [a]	Primer sequences (5′−3′)	Genome position of Primer sequences	Amplification conditions (used in HotStar *Taq* PCR system)	Amplicon size (bp)	Size of aligned fragment analysed (bp)	Analysed sequence fragment start (5′) and end (3′) points	Genome position of analysed fragment (5′-3′)
Rhomboid like serine protease (*RSP*)	4	RSPF	CGAGGAAGAGGATCATGACG	337,457–337,476	95 °C for 15 min; followed by 30 cycles of 94 °C for 1 min; annealing at 55 °C for 1 min; extension at 72 °C for 90 s; final extension at 72 °C for 5 min and at 4 °C on hold	500	410	5′-GCCTGGGCC..	337,515–337,918
		RSPR	CATGAGCAGCAGCACTAGCA	337,956–337,937				..AGGCGTGCA-3′	
Fatty acid/ sphingolipid δ-4 desaturase (*FAD*)	26	FADF	AGACACAGAGTGGGAACTGG	610,391–610,410		500	344	5′-CCGAGGTAG..	610,480–610,817
		FADR	GTGATGCTCAGCGTGGTAAG	610,871–610,890				..CGACCTTTA-3′	
Serine/threonine-protein kinase (*STK*)	29	STKF	CGGCCTTTTTGAAAGAGAGA	544,865–544,884		576	472	5′-GCGCCTGTG..	544,931–545,392
		STKR	CAGAAAACGAAAAGGCGAAG	545,421–545,440				..CTGCTTTCT-3′	
5′ a2rel-related protein (*A2P*)	22	A2PF	ATGCTCCACGCCATGTTAAG	302,887–302,906		600	459	5′-CGGTGGC-G..	302,940–303,386
		A2PR	ACGACAGCGTCCACGAAAG	303,486–303,468				.. GTGGCGCCC-3′	
86Ubiquitin-activating enzyme e1 (*UBE1*)	23	UBE1F	TCTTGAGCGCAATCATGAAC	254,790–254,809		483	388	5′-AGCTGCACC..	254,837–255,206
		UBE1R	TCACGTACTGCTCCAACCTG	255,272–255,253				..CGATGCGTT-3′	
Mannosyltransferase (*MTF*)	31	MTFF	ATCTTCATGCCCTCCTCGTA	930,604–930,623		500	378	5′-CGCGCTCC..	93,0686–931,045
		MTFR	TCGAGCTTCAAGTGGGAGAG	931,084–931,103				..GAACGCGG-3′	
Splicing factor 3B subunit1 (*SF3B1*)	28	SF3B1F	GTCATCGCCAAGAAAATTCG	1,023,180–1,023,199		450	329	5′-CAGATTCGC..	102,3252–102,3567
		SF3B1R	ATTCTTAGCCAGCAGCTCCA	1,023,629–1,023,610				..GCTTTGAGG-3′	

[a] Primers are named based on the abbreviations of the respective gene; the letters F and R are suffixed after the abbreviated gene names to denote forward and reverse primers, respectively.

pairs of primers were constructed based on DNA sequences surrounding the protein coding SNPs detected in the target regions (Downing et al., 2012). The newly created primers were 19–20 bp in length with the melting temperatures ranging between 58C and 62C and the product sizes ranging between 450 bp and 600 bp. To minimize cross-priming with other organisms or unrelated genes, BLAST searches were performed for these primers against GenBank. A range of different primer conditions was tested, with the best amplification results being achieved with the primers listed in Table 1.

2.6. PCR amplification of MLST markers

The PCR mix was prepared using the Qiagen HotStar *Taq* PCR system (QIAGEN, Hilden, Germany). A separate reaction tube was used for each of the seven amplicons. Each reaction contained 25 μl of 2X pre-optimized Multiplex PCR buffer containing 3 mM $MgCl_2$, HotStar *Taq* DNA polymerase and a mixture of dNTPs (10 mM of each dNTP), 0.2 μM each forward and reverse primers (Table 1), 5 μl of template DNA (final concentration of approximately 5 ng/μl) and molecular grade water added to a final volume of 50 μl. The PCR amplification conditions are described in Table 1. Double amplification using the same amplification settings was performed in the samples, when expected product bands were not clearly visible in the gel after the first PCR amplification.

PCR products were visualized in 2% agarose gels containing GelRed™ nucleic acid stain (Biotium, USA) in 0.25X TBE (Tris-Borate, EDTA, pH 8.0) buffer (Sigma, USA).

2.7. DNA sequencing

The PCR products of the seven MLST loci were commercially sequenced by Macrogen Inc. (Seoul, South Korea) for 40 isolates (excluding BPK282A1) using the primers described in Table 1. The forward and reverse sequences were manually edited and assembled to generate consensus sequences using the software ChromasPro v7.4.1 (Technelysium Pty Ltd, Australia). All sequences from the seven MLST loci were deposited in GenBank.

2.8. Allele and sequence type determination

The consensus sequences from each of the seven MLST loci of the 40 isolates were aligned against the reference strain BPK282A1 using the Clustal W2 algorithm (https://www.ebi.ac.uk/Tools/msa/clustalw2/) (Thompson et al., 1994) in the BioEdit v7.2.5 software (Ibis Biosciences, CA, USA). The aligned sequences were trimmed from both ends to the same length using the GeneDoc v2.7 software (NRBSC) (http://genedoc.software.informer.com/2.7/). From these sequences, the unambiguous sites of SNPs were determined.

The allele types (ATs) were assigned based on the difference in SNP position for each of the seven loci. A seven-digit allelic profile was then produced as a sequence type (ST) for each isolate. Allele numbers and STs were calculated based on the genotypic diversity according to the allelic profiles using MLSTest v1.0.1.23 software (Tomasini et al., 2013).

2.9. Genetic diversity determination

The extent of DNA polymorphisms was determined by analysing discriminatory parameters, such as the number of polymorphic sites (S), number of alleles, discriminatory power (DP), nucleotide diversity (π), haplotype diversity (H_d), number of mutations, mutation rate per sequence (θ), and average nucleotide differences per sequence (k). These parameters were calculated using DNAsp v5.10.01 software (Librado and Rozas, 2009; Rozas, 2009). Tajima's D, Fu & Li's D^*, Fu &Li's F^*, and Fu's F, tests for neutrality based on individual and concatenated sequences were also calculated using this software.

2.10. Phylogenetic analysis

Phylogenetic analysis was performed using each single locus, the concatenated sequences of the seven MLST loci from all 40 isolates and the reference strain BPK282A1. The Bayesian Information Criterion (BIC) was executed to determine the optimal molecular evolutionary model for the concatenated sequences using jModelTest software 2.1.7 (Guindon and Gascuel, 2003; Darriba et al., 2012). The Hasegawa-Kishino-Yano model (HKY) with gamma distribution (G) was selected and used in the phylogenetic analysis applying the respective parameters: fA: 0.1846, fC: 0.3114, fG: 0.3011, fT: 0.2029, ti/tv: 1.5355, and gamma: 0.0420. A phylogenetic tree was constructed using the maximum likelihood method with the nearest-neighbor interchange (NNI) algorithm based on the HKY model (Hasegawa et al., 1985) with 1000 bootstrap replicates in MEGA 6.0 software (Tamura et al., 2013).

2.11. Genetic differentiation and recombination

The genetic differentiation of pre-defined *Leishmania* populations (e.g. *L. donovani* versus *L. infantum*) and the distribution of *L. donovani* isolates according to their continent of origin (either Asia or Africa) was performed by applying the MLST allelic profiles in the Principal Component Analysis (PCA) using the *Adegenet* v2.0.1 package for statistical software R v3.2.4 (https://www.R-project.org/). To test if there was any genetic differentiation between the pre-defined populations, the *Fst* test was calculated using the *Hierfstat* package in R applying the Monte-Carlo test.

To investigate the possibility of recombination present amongst *L. donovani* and *L. infantum* populations, a separate split decomposition analysis was executed for isolates of each species using the Neighbor-net method implemented in SplitsTree4 v4.14.3 software (Huson and Bryant, 2006). The SplitsTree analyses were performed using alignments of both the single locus sequences (intragenic) and the concatenated dataset of all seven loci (intergenic) of 34 *L. donovani* isolates derived from Asia (Indian subcontinent and Iraq) and Africa, and of seven *L. infantum* isolates recovered from South America and Europe. The second-best model selected in jModelTest software 2.1.7 was used in this analysis. The Generalized Time Reversible (GTR + I) model was applied in the analysis of all *L. donovani* isolates using the following parameters: fA: 0.1847, fC: 0.3182, fG: 0.3033, fT: 0.1938, and p-inv: 0.9420. The Pair-wise Homoplasy Index (PHI) test available in SplitsTree4, was performed separately on individual gene sequences (intragenic) and on the concatenated dataset (intergenic) from isolates of two major groups to check if there was any statistically significant evidence for recombination within isolates (Huson and Bryant, 2006).

2.12. Ethics statement

This study was conducted with the human ethics approval from the Bangladesh Medical Research Council (BMRC: # BMRC/NREC/20BMRC/NREC/2010-2013/655(1-10)) and from the Human Research Ethics Committee, at Western Sydney Local Health District, Australia (LNR/13/WMEAD/173). Informed written consent was obtained from each participant from Bangladesh before collecting personal details, medical records, socio-demographic data, and biological specimens after the need for collecting splenic aspirates for research had been clearly explained. Informed written consent was also taken from parents or legal guardians of those less than 18 years old as proxy informants. The objectives of the

study and the collection procedure for each specimen with their associated health risks, if any, were explained clearly in the native language, Bengali, before obtaining consent by signature or thumb impression on the consent form. The Sudanese strain was obtained from the ICPMR, Westmead Hospital, Australia. Data on samples and isolates were coded and anonymously analysed. All Bangladeshi VL patients included in this study were treated with liposomal amphotericin B under the National VL Elimination Program.

2.13. Data accessibility

All aligned sequences used in this manuscript have been deposited in Mendeley Data (https://data.mendeley.com/) and are available at DOI: https://doi.org//10.17632/tcn8dph4dn.1.

3. Results

3.1. Species identified for Bangladeshi Leishmania isolates

Species of 22 Bangladeshi *Leishmania* isolates were identified as *L. donovani* by the ITS1-RFLP. The ITS1 sequences of Bangladeshi *Leishmania* isolates were found 100% identical to the available ITS1 sequences of *L. donovani* from the Indian subcontinent belonging to the zymodemes MON-2 and MON-38 (sequence type H as in Kuhls et al. (2005)) (Kuhls et al., 2005), some from Kenya belonging to zymodeme MON-37 (MHOM/KE/---/KENO49, MHOM/KE/1975/H9, MHOM/KE/1983/NBL189, MHOM/KE/1985/NLB323, MHOM/KE/1984/NLB218), and from Sri Lanka (MON-37) (MHOM/LK/2002/L60c, MHOM/LK/2002/L60b). All submitted ITS1 sequences obtained from Bangladeshi *Leishmania* isolates are available under the following GenBank accession numbers: MK211285 – MK211306 (Supplementary Table S1).

3.2. PCR amplification and DNA sequencing

Locus-specific PCR generated single products of the expected sizes for each of the seven MLST loci as in Table 1, listed from all 40 isolates (excluding the reference strain BPK282A1).

All submitted sequences are available under the following GenBank accession numbers: KX155845 – KX155886 for *RSP*; KX155887 – KX155928 for *FAD*; KX155929 – KX155970 for *STK*; KX155971 – KX156012 for *A2P*; KX156013 – KX156054 for *UBE1*; KX156055 – KX156096 for *MTF*; and KX156097 – KX156138 for *SF3B1* (Supplementary Table S1). The alignments are available in Mendeley Data at: DOI: https://doi.org//10.17632/tcn8dph4dn.1.

3.3. SNP analysis and nucleotide diversity

The number of substitutions, deletions and insertions found between isolates varied from gene to gene. Gene fragments showed segregating sites with *SF3B1* having the least (15 sites) and *MTF* having the most polymorphic regions (73 sites) considering the whole sample set of 41 strains including BPK282A1. Only two alternative bases were detected at each polymorphic site due to substitution. The details of DNA polymorphisms found in the seven MLST loci obtained from all 41 strains are shown in Table 2. The highest discriminatory power and haplotype diversity are observed in the *A2P* gene followed by the *MTF* gene (Table 2). Table 3 shows DNA polymorphisms in the seven MLST loci obtained from 22 Bangladeshi *L. donovani* isolates. The *A2P* gene shows the highest discriminatory power across Bangladeshi *L. donovani* followed by *SF3B1* gene. Tajima's *D*, Fu & Li's D^*, Fu & Li's F^*, and Fu's *Fs* for all individual genes and for the concatenated sequences of the whole dataset of 41 isolates and of 22 Bangladeshi isolates separately show negative values (Tables 2 and 3). Significant Tajima's *D* value ($P < 0.05$) is observed for variable regions of

Table 2
DNA polymorphisms found in the seven multilocus sequence typing loci obtained from 40 isolates including Bangladeshi strains and sequences from the reference strain BPK282A1 of the *Leishmania donovani* complex.

Locus	Length (bp)	*S*	No. alleles	*DP*	π	*k*	H_d	No. mutation	θ	F_D	F_F	F_S	*D*	*P* value
RSP	410	19	9	0.69	0.00941	3.782	0.635	20	4.67449	1.66523	1.05821	4.788	−0.62905	>0.05
FAD	344	24	11	0.636	0.00619	2.093	0.6	24	5.609	−4.27617	−4.19159	0.192	−2.10588	<0.05
STK	472	33	10	0.604	0.00758	3.4878	0.571	33	7.71291	−4.3929	−4.18886	2.105	−1.89255	<0.05
A2P	459	66	11	0.827	0.01095	4.86	0.818	67	15.65954	−5.40376	−5.19898	0.51	−2.48182	<0.01
UBE1	388	26	12	0.495	0.00512	1.87439	0.467	27	6.31056	−4.50238	−4.47999	−2.649	−2.38807	<0.01
MTF	378	73	16	0.699	0.01667	5.95	0.655	76	17.76306	−4.68655	−4.61376	−0.493	−2.40497	<0.01
SF3B1	329	15	15	0.621	0.00381	1.185	0.39	15	3.50587	−1.91509	−2.33182	−1.665	−2.1005	<0.05
Concatenated	2780	256	25	0.929	0.00868	23.232	0.909	262	61.2358	−4.58756	−4.48516	0.799	−2.30523	<0.01

S, number of polymorphic sites; *DP*, discriminatory power; π, nucleotide diversity; *k*, average number of nucleotide differences per sequence; *Hd*, haplotype diversity; θ, mutation rate per sequence; *D*, F_D, F_F, and F_S, Tajima's D, Fu and Li's D*, Fu and Li's F*, and Fu's Fs, respectively.
See Table 1 for full names of loci.

Table 3
DNA polymorphisms found in the seven multi-locus sequence typing loci obtained from 22 Bangladeshi *Leishmania donovani* isolates.

Locus	Length (bp)	*S*	No. alleles	*DP*	π	*k*	H_d	*n mutation*	θ	F_D	F_F	F_S	*D*	*P* value
RSP	410	0	1	0.083	0	0	0	0	0	0	0	0	0	–
FAD	344	0	1	0.083	0	0	0	0	0	0	0	0	0	–
STK	472	0	1	0.087	0	0	0	0	0	0	0	0	0	–
A2P	459	8	5	0.672	0.00251	1.121	0.641	8	2.195	−2.73508	−2.79317	−0.699	−1.60561	>0.05
UBE1	388	3	2	0.17	0.00049	0.182	0.091	3	0.549	−2.10924	−2.23805	−0.112	−1.51481	>0.05
MTF	378	1	2	0.17	0.00025	0.091	0.091	1	0.274	−1.57469	−1.67803	−0.957	−1.1624	>0.05
SF3B1	329	2	4	0.324	0.00058	0.182	0.177	2	0.549	−2.10924	−2.23805	−1.974	−1.51481	>0.05
Concatenated	2780	13	7	0.739	0.00058	1.576	0.684	13	3.566	−3.28526	−3.37241	−0.777	−1.97688	<0.05

S, number of polymorphic sites; *DP*, discriminatory power; π, nucleotide diversity; *k*, average number of nucleotide differences per sequence; *Hd*, haplotype diversity; θ, mutation rate per sequence; *D*, F_D, F_F, and F_S, Tajima's D, Fu and Li's D*, Fu and Li's F*, and Fu's Fs, respectively.
See Table 1 for full names of loci.

the individual (except *RSP*) and concatenated gene sequences across isolates of the whole dataset, indicating significant evidence for purifying selection for the MLST loci (Table 2). Although Tajima's *D* value for the concatenated sequences of 22 Bangladeshi isolates was significant ($P < 0.05$), it was not significant ($P > 0.05$) for individual gene sequences across Bangladeshi isolates when analysed separately (Table 3).

3.4. MLST analysis

The genetic diversity was measured by determining the number of alleles per locus. The SNPs generated nine, 11, 10, 11, 12, 16, and 15 alleles in *RSP*, *FAD*, *STK*, *A2P*, *UBE1*, *MTF*, and *SF3B1* loci, respectively (for the whole sample set of 41 strains including BPK282A1). From the combination of the allelic profiles, 25 STs were identified in isolates originating from 12 different countries (Supplementary Table S1). Amongst 25 STs, 18 STs were for *L. donovani* and seven STs were for *L. infantum*. Within these 18 STs for *L. donovani*, half were from the Indian subcontinent. In 22 Bangladeshi isolates, the SNP diversity generated five, two, two, and four allele types in *A2P*, *UBE1*, *MTF*, and *SF3B1* genes, respectively, and seven STs were identified (Supplementary Table S1). Using the concatenated dataset, 10 isolates were identified as ST5 and another seven were ST4. The remaining five isolates had unique STs. ST4 and ST7 differed only by one nucleotide at the *SF3B1* locus. The most polymorphic locus for Bangladeshi isolates was *A2P* with five alleles. The locus *MTF* with 16 alleles showed the highest polymorphism within the entire sample set studied (Supplementary Table S1).

3.5. Phylogenetic analysis and genetic variability

Phylogenetic analysis (using the maximum likelihood method with NNI algorithm) of the individual locus sequences showed that only the *A2P* gene could distinguish *L. donovani* from *L. infantum*, although bootstrap support was low (51%) (Supplementary Figs. S1–S4). However, phylogenetic analysis of their concatenated sequences could clearly differentiate *L. donovani* from *L. infantum* with 99% bootstrap agreement, using the MLST scheme proposed herein (Fig. 1). Due to different levels of intraspecific variation observed in both species, isolates were sub-grouped with varying bootstrap support. Within the *L. donovani* populations, two main groups (Asian and African) were observed with low Bootstrap agreement (Fig. 1). The Asian group contained the isolates from countries of the Indian subcontinent (Bangladesh, India and Nepal) and Iraq, and the African group was composed of the isolates from countries in Africa. The Iraqi isolate SUKKAR2 positioned between and linked the isolates from the Indian subcontinent and Africa (Fig. 1).

In the Asian group (Fig. 1), the Bangladeshi *L. donovani* isolates grouped into a large cluster except one (TM-29), which was more diverse than other Bangladeshi isolates. This large Bangladeshi cluster could be further split into two sub-clusters, although bootstrap support was low. Each sub-cluster contained isolates of different STs. Two identical Indian *L. donovani* isolates (THAK35 and BHU17) were placed within one of these two Bangladeshi subclusters but grouped separately from Bangladeshi *L. donovani* isolates (Fig. 1). Bangladeshi isolates originated from five endemic districts such as Mymensingh, Tangail, Jamalpur, Pabna and Sirajganj with Mymensingh district having the highest number of diverse genotypes (five of seven genotypes). The geographical distribution of seven genotypes detected in 22 Bangladeshi *L. donovani* isolates is shown in Fig. 2. All Bangladeshi and Indian *L. donovani* isolates differed from the Nepalese strain BPK282A1 with high bootstrap agreement, while their difference from the Iraqi strain SUKKAR2 showed low bootstrap support. The bootstrap values above 90% were used to make definite inferences or interpretations in this study and only bootstrap values above 80% are presented in the phylogenetic tree (Fig. 1).

Within the African group (Fig. 1), isolates were subdivided into two clusters: Cluster I contained two Kenyan isolates (NLB218A and MRC74) and one Sudanese *L. donovani* strain (CHW). Cluster II included three Sudanese (1S, SO3, and GILANI) and two Ethiopian (HU3 and GEBRE1) *L. donovani* isolates. All isolates belonging to *L. infantum* showed a scattered distribution (Fig. 1).

Two species (*L. donovani* and *L. infantum*) of the *L. donovani* complex were also clearly differentiated by the PCA ($P = 0.002$), using the MLST scheme proposed herein (Fig. 3A). Furthermore, two major *L. donovani* populations (e.g. Asian and African) were evidently separated in the PCA ($P = 0.002$) using the MLST allelic profiles (Fig. 3B).

The SplitsTree analysis performed within each of the species identified in the maximum likelihood tree (NNI algorithm), again identified two main clusters (Asian and African) for *L. donovani* (Fig. 4). The Asian cluster included a group of 22 Bangladeshi, two Indian (BHU17 and THAK35) and one Nepalese (BPK282A1) isolates. The African cluster composed of eight diverse isolates from countries in Africa. As observed in phylogenetic analysis (Fig. 1), the Iraqi strain SUKKAR2 was also found to form a link between isolates from the Indian subcontinent and Africa in the SplitsTree analysis. The overall distribution of *L. donovani* isolates from Asian and African countries showed consistency with their continent of origin (Figs. 1, 3B and 4).

The SplitsTree network showed that the presence of reticulating parallelograms was more frequent within the African group than that within the Indian subcontinent group (Fig. 4). Results of the PHI test did not show evidence for recombination in any of the loci on single gene analysis across isolates from both groups (PHI > 0.05). However, results of the PHI test analysis using the concatenated data from both groups showed that significant reticulation events (PHI = 0.03) occurred between isolates in the second group represented by African isolates (Fig. 4). Similar findings were also observed amongst seven *L. infantum* isolates, where PHI test analysis showed a statistically significant reticulation event (PHI = 0.007) for the concatenated dataset (data not shown).

4. Discussion

Although *L. donovani* in the Indian subcontinent was previously detected as genetically homogeneous (Alam et al., 2009), genetic variation was recently observed among *L. donovani* isolates from the region (Downing et al., 2011, 2012; Imamura et al., 2016). A better method that is readily applied in resource limited countries and has higher discriminatory power is therefore needed to identify diverse genotypes within closely related *L. donovani* strains from the Indian subcontinent. Differentiation of genotypes within closely related *L. donovani* strains will assist in understanding the clinical and epidemiological aspects of VL control in a certain region.

Several methods have been used for discriminating *L. donovani* complex species and strains, all of which were reviewed by Akhoundi et al. (2016). Except for MLEE, MLMT, and MLST, most of these techniques are not suitable for discriminating *Leishmania* parasites at the strain level (Schonian et al., 2011; Almeida and Araujo, 2013). Most microsatellite markers are species-specific in *Leishmania* and cannot be used across species or species complexes. Therefore, MLMT is not very appropriate for inferring phylogeny between different *Leishmania* spp. (Schonian et al., 2011). In addition, MLMT has lower discriminatory power for differentiating *L. donovani* isolates from India and Nepal compared with genome-wide SNP analysis (Downing et al., 2012). Although MLEE is still the 'gold standard' for *Leishmania* spp. identification, it shows poor

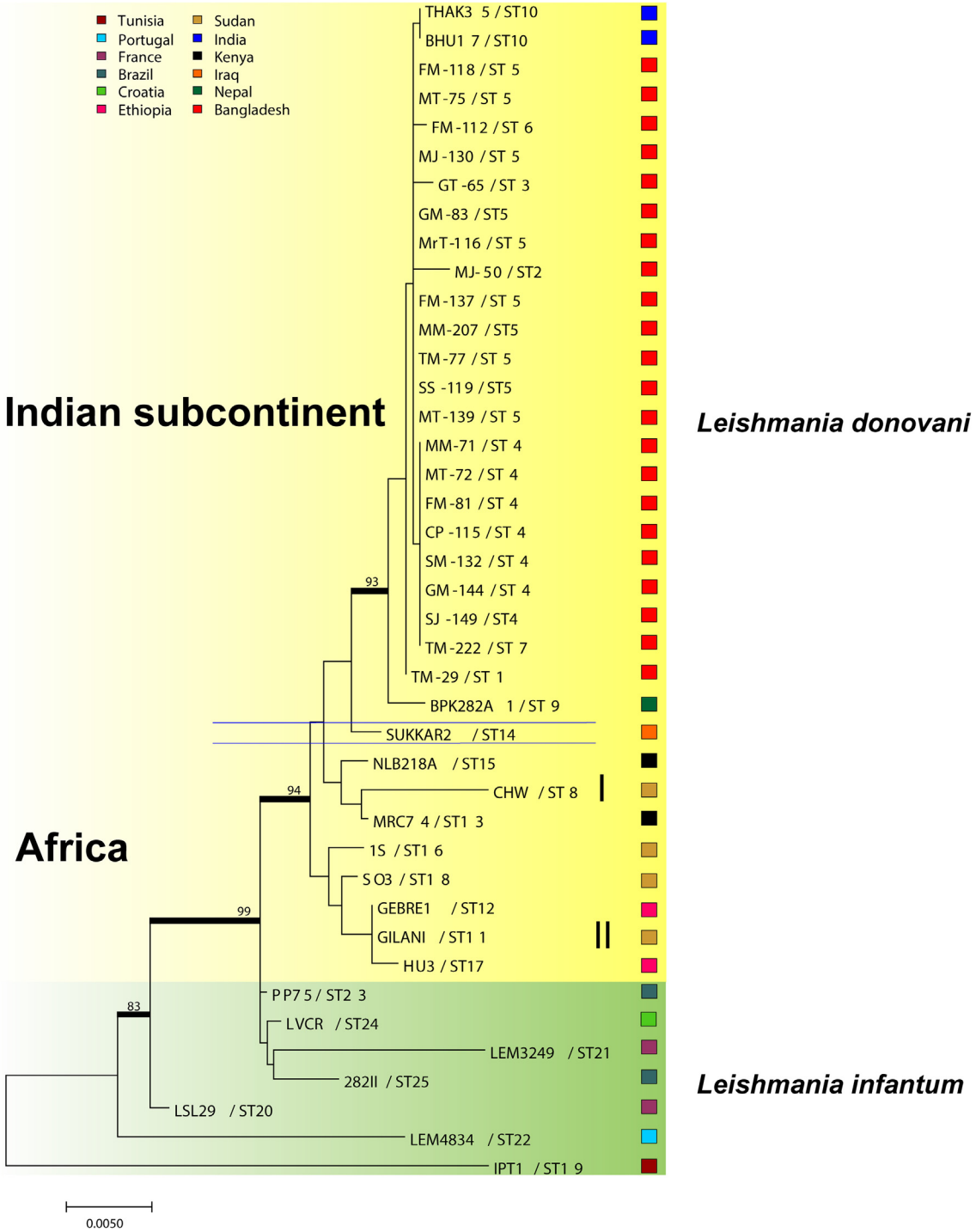

Fig. 1. A mid-point rooted maximum likelihood tree with the highest log likelihood (-5729.7437) constructed from the concatenated sequences of *RSP*, *FAD*, *STK*, *A2P*, *UBE1*, *MTF*, and *SF3B1* genes from 41 isolates of the *Leishmania donovani* complex with 1000 bootstrap replicates. Two blue lines on either side of SUKKAR2 separate the African isolates from those of the Indian subcontinent (Asia) and signify the geographical positioning the SUKKAR2. Cluster I contained two Kenyan and one Sudanese *L. donovani* isolates and Cluster II included three Sudanese and two Ethiopian *L. donovani* isolates. Bootstrap values greater than 80% are shown above the branches. The analysis involved a total of 2780 positions in the final dataset. Bar = 0.005 and represents nucleotide substitutions per site. Different colours indicate geographical origins of the isolates.

efficiency in evaluating genotypes and lower sensitivity in discriminating closely related strains (Ochsenreither et al., 2006; Schonian et al., 2011). On the other hand, MLST is a powerful genotyping approach, having advantages over WGS for which MLST is more accessible to resource poor countries. In MLST, a number of specific housekeeping genes are sequenced whereas in WGS, DNA sequence from the whole parasite is generated. A parasite genome is much larger than a bacterium or a virus and the mega data produced by WGS will subsequently require large electronic storage capacity. In addition, experts with advanced training on computer procedures and bioinformatics are essential. These experts will use high end computers requiring a continual and reliable electricity supply (Bersch, 2011). A resource poor country will lack these facilities. Management and analysis of MLST data is easier and simpler.

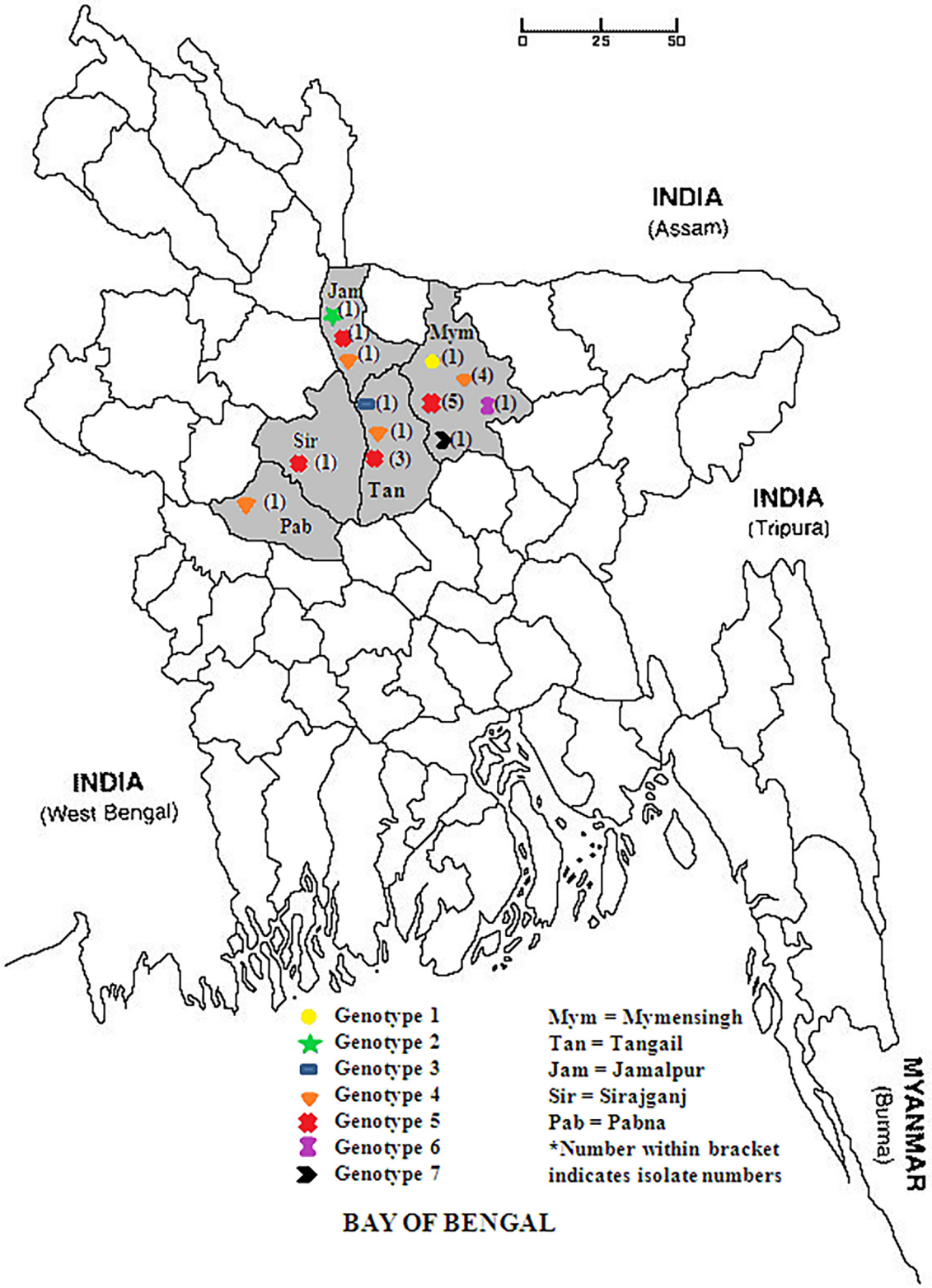

Fig. 2. Map showing the geographical distribution of the Bangladeshi *Leishmania donovani* isolates and their genotypes in five endemic districts.

Besides, an effective MLST system has discriminatory power comparable to that of WGS and can provide the same genetic data (Lytsy et al., 2017). With advances in genomic technologies, MLST is currently being used with WGS as an integrated method (wgMLST) to study population dynamics, pathogenicity, and molecular evolution of pathogens (Lytsy et al., 2017; Pérez-Losada et al., 2017). Furthermore, MLST can be chosen as a better option for a population genetic study on a small sample size with a limited budget (cost effective) (Pérez-Losada et al., 2017). However, there are no specific studies addressing the need for a MLST marker panel to specifically distinguish extremely homogeneous *L. donovani* strains from the Indian subcontinent. Although the

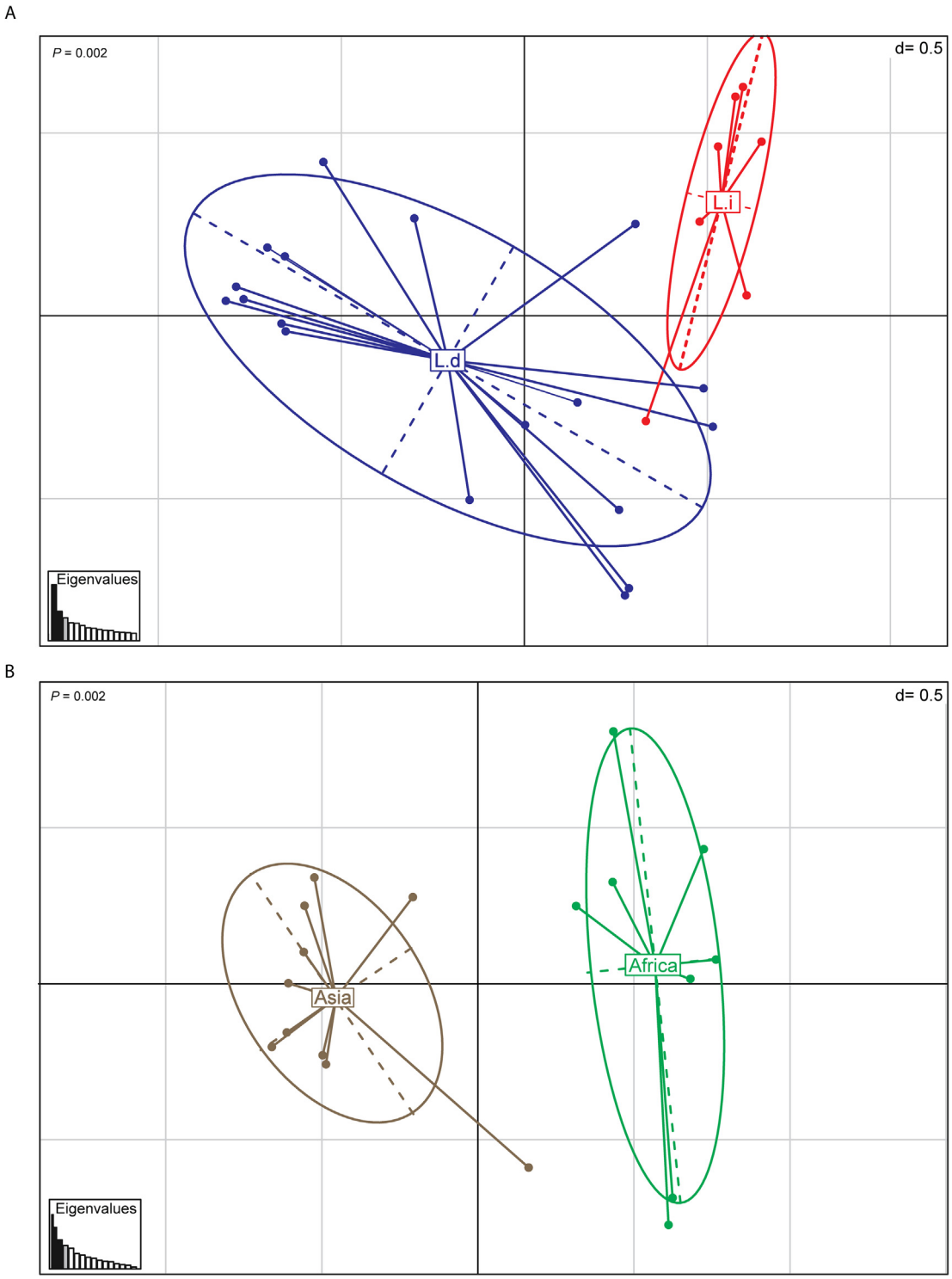

Fig. 3. Principal Component Analysis of the *Leishmania donovani* complex multilocus sequence typing allelic profile showing the differences between the species (A) and the distributions of *L. donovani* according to continent of origin (B). Dots represent sequence types (STs) linked by coloured lines to form clusters, proportional in size to the number of isolates (25 isolates in A and 18 isolates in B). The two main clusters separating the *L. donovani* (L. d) from *L. infantum* (L. i) populations are highlighted in blue (L. d) and red (L. i), respectively. Within the *L. donovani* species, the samples recovered from Asia (Indian subcontinent and Iraq) and Africa are highlighted in brown and green, respectively. Eigenvalues highlighting the two first components are represented in the bar plot. The *P* value calculated using the Monte-Carlo test is described in the upper corner.

WGS method has currently been recognized as more informative, less time consuming, and convenient for a large sample size (Downing et al., 2011, 2012; Pérez-Losada et al., 2017), the technique is not widely accessible in the most affected developing countries as e.g. Bangladesh and Nepal. Therefore the nucleotide sequence-based method, MLST, was applied in the current study to assess the genetic diversity and phylogenetic relationships between *L. donovani* isolates recovered from Bangladesh (Supplementary Table S2).

Housekeeping genes with high and variable rates of sequence diversity within and between species or genera are the ideal targets for characterization of microorganisms including *Leishmania*

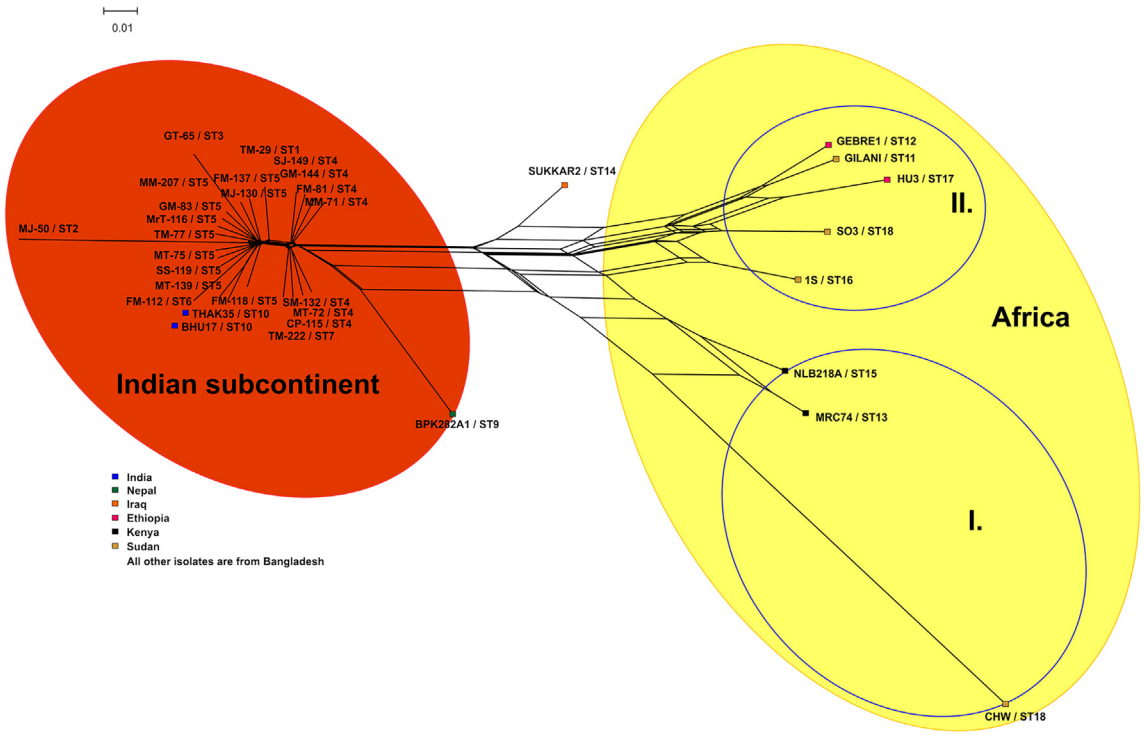

Fig. 4. SplitsTree decomposition network of the concatenated DNA sequences of seven genetic loci for 34 *Leishmania donovani* isolates derived from Asia (Bangladesh, India, Nepal, and Iraq) and Africa (Sudan, Kenya, and Ethiopia). This network was built based on split decomposition using the Neighbor-net method. Cluster I contained two Kenyan and one Sudanese *L. donovani* isolates; Cluster II included three Sudanese and two Ethiopian *L. donovani* isolates. Bar = 0.1 and represents nucleotide substitutions per site. The Pairwise Homoplasy Index value showed evidence for recombination in the African and one Iraqi isolates ($P < 0.05$).

(Bennasar et al., 2010; Weirather et al., 2011; Boite et al., 2012; Marco et al., 2015). Moreover, single copy protein coding genes are preferred because further analysis by cloning or other allele-specific analysis can be avoided (Schonian et al., 2011). A number of housekeeping genes were investigated by MLST for taxonomic/-phylogenetic analyses of *L. donovani* complex species and strains, but they showed poor discriminatory power (Mauricio et al., 2006; Zemanova et al., 2007; El Baidouri et al., 2013; Akhoundi et al., 2016). Although other markers such as the genes of cytochrome oxidase II (Ibrahim and Barker, 2001), glycoprotein 63 (Mauricio et al., 2007), cysteine protease (Hide et al., 2007), and cytochrome b (Luyo-Acero et al., 2004) could distinguish between *Leishmania* spp. and complexes, they were not able to differentiate between strains of *L. donovani*, especially from the Indian subcontinent (Downing et al., 2011). In the current study, we selected a new set of seven single copy housekeeping genes from *L. donovani* based on the presence of significant SNP diversity across the genomic data and their distribution on different chromosomes (Downing et al., 2011). These seven enzyme-coding genes play a significant role in *Leishmania* metabolism and in the pathogenesis of leishmaniasis (Denny et al., 2004; Urban, 2006; Besteiro et al., 2007; McConville and Naderer, 2011). Furthermore, there are several other advantages of the MLST proposed herein. Firstly, the targeted housekeeping loci are located on different chromosomes and thus are unlinked. Secondly, the primer set selected can amplify all target regions under the same amplification conditions, including the same annealing temperature, and can therefore save significant time and reduce laboratory costs. This is an important advantage over other studies where different annealing temperatures were used to amplify different targets (Mauricio et al., 2006; Zemanova et al., 2007). Thirdly, these target genes are present in most of the medically relevant *Leishmania* spp. and could be applied to genotype a wider range of *Leishmania* spp.

The herein described MLST scheme revealed significant genetic polymorphisms in the *L. donovani* complex at the intra- and inter-species levels. The targeted loci were highly diverse, possessing between nine and 16 alleles across the 41 *L. donovani* and *L. infantum* isolates. Population genetic analysis also demonstrated that *L. donovani* isolates from Bangladesh, India and Nepal were distinct (Figs. 1 and 4). This agrees with a recent study where eight Bangladeshi *L. donovani* isolates formed a distinct population from Indian and Nepalese *L. donovani* isolates (Imamura et al., 2016). Therefore, it is necessary to evaluate the impact of genetic variants detected in Bangladeshi isolates on the epidemiology, pathogenicity and therapeutic response of VL in the country. Future studies should include more isolates from the Indian subcontinent, encompassing a broader period of sampling, as this would increase the chances of finding more variation of genotypes among *Leishmania* isolates.

In the current study, the locus *A2P* showed the highest diversity within 22 Bangladeshi isolates, while the *MTF* locus had the highest diversity within the entire study population, reflecting the utility of these two markers for discriminating *L. donovani* complex species and strains. The *MTF* gene is situated on chromosome 31, which is reported to be tetraploid and to be a cause of genetic variation (Downing et al., 2011). Further evaluation of this single genetic marker on a broader range of *Leishmania* spp. could show whether it alone is sufficiently diverse to distinguish within and between *Leishmania* spp.

The differentiation between the two species within the *L. donovani* complex was clearly evidenced by the phylogenetic analysis of the concatenated dataset (Fig. 1), and by the PCA using the allelic profiles (Fig. 3A). This distinction between the species was in agreement with other MLST studies (Mauricio et al., 2006; Zemanova et al., 2007). Intraspecific variation was also observed when *L. donovani* isolates were analyzed according to their continent of origin by the PCA. They grouped either by Asian or African

continents (Fig. 3B) as observed in previous studies (Mauricio et al., 2006; Zemanova et al., 2007). In addition, a tendency to cluster and the absence of recombination observed within Bangladeshi isolates might indicate a clonal model of reproduction. Similar findings were noticed in a study where no recombination was detected in Bangladeshi *L. donovani* isolates, even though a low level of genetic diversity was found amongst them (Imamura et al., 2016). Alternatively, this could be due to the limited geographical scale of the study. In this same study (Imamura et al., 2016), eight Bangladeshi *L. donovani* isolates which originated from only the Mymensingh district were analyzed, and in the current study 22 Bangladeshi *L. donovani* isolates were derived from only five endemic districts (Mymensingh, Tangail, Jamalpur, Sirajganj, and Pabna) located in the northwestern parts of the country. These districts share borders with each other and some with neighboring India (Fig. 2). Despite population movement across the border with the highly endemic state of West Bengal in India, our results suggest a clonal population structure within Bangladeshi isolates. Expanding the study beyond borders using the proposed MLST scheme is, therefore, expected to contribute to a better understanding of the genetic population structure of *L. donovani* within the Indian subcontinent and globally.

A previous study using microsatellites shows genetic homogeneity among 85 *L. donovani* isolates from Bangladesh, India and Nepal, with identical MLMT genotypes mostly of the 1a type (Alam et al., 2009). The new MLST scheme proposed herein classified 22 new Bangladeshi isolates of *L. donovani* into seven STs. Interestingly, the reference *L. donovani* strains BPK282A1 from Nepal, and THAK35 and BHU17 from India, also shared the predominating MLMT profile 1a in the same previous study (Alam et al., 2009), but presented different STs in this study. This comparison suggests that the target genes selected for the new MLST scheme are more effective at discriminating strains of *L. donovani* from the Indian subcontinent than the known microsatellite markers for the DON complex.

This new MLST scheme possesses sufficient discriminatory power to detect genetic heterogeneity within and between species of the *L. donovani* complex, irrespective of their geographical origin. The novel seven molecular markers will be useful when investigating the emergence of new parasite variants, for tracking their spread and for monitoring the development of resistance to anti-*Leishmania* drugs. Furthermore, the *A2P* and *MTF* genes have the greatest diversity compared with the other loci used in this study, suggesting that sequences of these two genes could be particularly useful for discriminating species and strains within the *L. donovani* complex. This however needs to be validated through larger and more geographically diverse sampling.

The sequences generated herein, employed in a new MLST scheme, are available in GenBank and can be used by others to build a standardized approach to gaining insights into the evolutionary ecology and molecular epidemiology of this pathogen at a local and a global scale. One big advantage of this study was that it could use numerous whole genome sequences published for *L. donovani* from the Indian subcontinent which became available only recently and helped to identify discriminatory MLST markers for the Bangladeshi strains. In future, if there is interest to proceed with MLST, the different markers published by different authors should be analysed in terms of their discriminatory power for the strains of interest.

Acknowledgements

The study was conducted in collaboration with the Communicable Disease Control Unit of the Directorate General of health Services (DGHS), Bangladesh. This study was supported by the Centre for Infectious Diseases and Microbiology (CIDM) Public Health fund, an Endeavour Foundation scholarship (ID: 3039_2012) and Top-up scholarship from the Sydney Medical School, Australia, to Sultana Shahana Banu. Dr. Progga Nanda Nath, Surya Kanta Kala-azar Research Centre (SKKRC), did splenic aspiration in all patients. Help given by staff of SKKRC and the Microbiology Department of Mymensingh Medical College (MMC), Bangladesh, is acknowledged for collection, culturing and storage of specimens. Equipment and facilities provided by the Japan International Cooperation Agency (JICA) at SKKRC were used. Prof. Abu Hadi Noor Ali Khan ensured space was available to store parasite isolates in liquid nitrogen at the Pathology Department, Faculty of Veterinary Medicine, Bangladesh Agricultural University (BAU), Mymensingh, during the field work. Cooperation from staff of a PC4 laboratory in the Institute of Clinical Pathology and Medical Research (ICPMR), Australia for sub-culturing Leishmania is acknowledged. Institute of Microbiology and Hygiene, Charite University, Berlin, Germany and Fiocruz, Rio de Janeiro, Brazil, are highly appreciated for providing the reference strains necessary for this study. Kennio Ferreira-Paim was supported by a Coordination for the Improvement of Higher Education Personnel (CAPES) Science without Borders visiting fellow (N° 9313133) from Brazil.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2019.02.010.

References

Akhoundi, M., Kuhls, K., Cannet, A., Votypka, J., Marty, P., Delaunay, P., Sereno, D., 2016. A historical overview of the classification, evolution, and dispersion of *Leishmania* Parasites and Sandflies. PLoS Negl. Trop. Dis. 10, e0004349.

Alam, M.Z., Kuhls, K., Schweynoch, C., Sundar, S., Rijal, S., Shamsuzzaman, A.K., Raju, B.V., Salotra, P., Dujardin, J.C., Schonian, G., 2009. Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of *Leishmania donovani* strains in the Indian subcontinent. Infect. Genet. Evol. 9, 24–31.

Almeida, L.A., Araujo, R., 2013. Highlights on molecular identification of closely related species. Infect. Genet. Evol. 13, 67–75.

Alvar, J., Velez, I.D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., Den Boer, M., WHOLC Team, 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7, e35671.

Bennasar, A., Mulet, M., Lalucat, J., Garcia-Valdes, E., 2010. PseudoMLSA: a database for multigenic sequence analysis of *Pseudomonas* species. BMC Microbiol. 10, 118.

Bersch, C. 2011. Unplugged: the code of life. Medical Laboratory Observer, May, 40-43.

Besteiro, S., Williams, R.A., Coombs, G.H., Mottram, J.C., 2007. Protein turnover and differentiation in *Leishmania*. Int. J. Parasitol. 37, 1063–1075.

Bhattacharyya, T., Boelaert, M., Miles, M.A., 2013. Comparison of visceral leishmaniasis diagnostic antigens in African and Asian *Leishmania donovani* reveals extensive diversity and region-specific polymorphisms. PLoS Negl. Trop Dis. 7, e2057.

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

Chargui, N., Amro, A., Haouas, N., Schonian, G., Babba, H., Schmidt, S., Ravel, C., Lefebvre, M., Bastien, P., Chaker, E., Aoun, K., Zribi, M., Kuhls, K., 2009. Population structure of Tunisian *Leishmania infantum* and evidence for the existence of hybrids and gene flow between genetically different populations. Int. J. Parasitol. 39, 801–811.

Chowdhury, R., Mondal, D., Chowdhury, V., Faria, S., Alvar, J., Nabi, S.G., Boelaert, M., Dash, A.P., 2014. How far are we from visceral leishmaniasis elimination in Bangladesh? An assessment of epidemiological surveillance data. PLoS Negl. Trop. Dis. 8, e3020.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

Denny, P.W., Goulding, D., Ferguson, M.A., Smith, D.F., 2004. Sphingolipid-free *Leishmania* are defective in membrane trafficking, differentiation and infectivity. Mol. Microbiol. 52, 313–327.

Downing, T., Imamura, H., Decuypere, S., Clark, T.G., Coombs, G.H., Cotton, J.A., Hilley, J.D., De Doncker, S., Maes, I., Mottram, J.C., Quail, M.A., Rijal, S., Sanders, M., Schonian, G., Stark, O., Sundar, S., Vanaerschot, M., Hertz-Fowler, C., Dujardin, J.C., Berriman, M., 2011. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res. 21, 2143–2156.

Downing, T., Stark, O., Vanaerschot, M., Imamura, H., Sanders, M., Decuypere, S., De Doncker, S., Maes, I., Rijal, S., Sundar, S., Dujardin, J.C., Berriman, M., Schonian, G., 2012. Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the *Leishmania donovani* species complex. Infect. Genet. Evol. 12, 149–159.

El Baidouri, F., Diancourt, L., Berry, V., Chevenet, F., Pratlong, F., Marty, P., Ravel, C., 2013. Genetic structure and evolution of the *Leishmania* genus in Africa and Eurasia: what does MLSA tell us. PLoS Negl. Trop. Dis. 7, e2255.

Gelanew, T., Hailu, A., Schonian, G., Lewis, M.D., Miles, M.A., Yeo, M., 2014. Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of *Leishmania donovani*. Int. J. Parasitol. 44, 751–757.

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

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Hide, M., Bras-Goncalves, R., Banuls, A.L., 2007. Specific cpb copies within the *Leishmania donovani* complex: evolutionary interpretations and potential clinical implications in humans. Parasitology 134, 379–389.

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

Ibrahim, M.E., Barker, D.C., 2001. The origin and evolution of the *Leishmania donovani* complex as inferred from a mitochondrial cytochrome oxidase II gene sequence. Infect. Genet. Evol. 1, 61–68.

Imamura, H., Downing, T., Van Den Broeck, F., Sanders, M.J., Rijal, S., Sundar, S., Mannaert, A., Vanaerschot, M., Berg, M., De Muylder, G., Dumetz, F., Cuypers, B., Maes, I., Domagalska, M., Decuypere, S., Rai, K., Uranw, S., Bhattarai, N.R., Khanal, B., Prajapati, V.K., Sharma, S., Stark, O., Schonian, G., De Koning, H.P., Settimo, L., Vanhollebeke, B., Roy, S., Ostyn, B., Boelaert, M., Maes, L., Berriman, M., Dujardin, J.C., Cotton, J.A., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5.

Kager, P.A., 1992. Splenic aspiration in visceral leishmaniasis. Rev. Soc. Bras. Med. Trop. 25, 217–223.

Kuhls, K., Keilonat, L., Ochsenreither, S., Schaar, M., Schweynoch, C., Presber, W., Schonian, G., 2007. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. Microbes Infect. 9, 334–343.

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.

Kumar, R., Ansari, N.A., Avninder, S., Ramesh, V., Salotra, P., 2008. Cutaneous leishmaniasis in Nepal: *Leishmania major* as a cause. Trans. R. Soc. Trop. Med. Hyg. 102, 202–203.

Kumar, R., Bumb, R.A., Ansari, N.A., Mehta, R.D., Salotra, P., 2007. Cutaneous leishmaniasis caused by *Leishmania tropica* in Bikaner, India: parasite identification and characterization using molecular and immunologic tools. Am. J. Trop. Med. Hyg. 76, 896–901.

Kumar, R., Nylen, S., 2012. Immunobiology of visceral leishmaniasis. Front. Immunol. 3, 251.

Leblois, R., Kuhls, K., Francois, O., Schonian, G., Wirth, T., 2011. Guns, germs and dogs: on the origin of *Leishmania chagasi*. Infect. Genet. Evol. 11, 1091–1095.

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

Lukes, J., Mauricio, I.L., Schonian, G., Dujardin, J.C., Soteriadou, K., Dedet, J.P., Kuhls, K., Tintaya, K.W., Jirku, M., Chocholova, E., Haralambous, C., Pratlong, F., Obornik, M., Horak, A., Ayala, F.J., Miles, M.A., 2007. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. Proc. Natl. Acad. Sci. USA 104, 9375–9380.

Lumsden, W.H.R., McMillan, S., 1996. Practical medical microbiology: Protozoa. In: Collee, J.G., Fraser, A.G., Marmion, B.P., Simmons, A. (Eds.), Mackie and McCartney's Practical Medical Microbiology. 14th ed. Churchill Livingstone, New York, pp. 739–740.

Luyo-Acero, G.E., Uezato, H., Oshiro, M., Takei, K., Kariya, K., Katakura, K., Gomez-Landires, E., Hashiguchi, Y., Nonaka, S., 2004. Sequence variation of the cytochrome b gene of various human infecting members of the genus *Leishmania* and their phylogeny. Parasitology 128, 483–491.

Lytsy, B., Engstrand, L., Gustafsson, A., Kaden, R., 2017. Time to review the gold standard for genotyping vancomycin-resistant enterococci in epidemiology: Comparing whole-genome sequencing with PFGE and MLST in three suspected outbreaks in Sweden during 2013–2015. Infect. Genet. Evol. 54, 74–80.

Marco, J.D., Barroso, P.A., Locatelli, F.M., Cajal, S.P., Hoyos, C.L., Nevot, M.C., Lauthier, J.J., Tomasini, N., Juarez, M., Estevez, J.O., Korenaga, M., Nasser, J.R., Hashiguchi, Y., Ruybal, P., 2015. Multilocus sequence typing approach for a broader range of species of *Leishmania* genus: describing parasite diversity in Argentina. Infect. Genet. Evol. 30, 308–317.

Marlow, M.A., Boite, M.C., Ferreira, G.E., Steindel, M., Cupolillo, E., 2014. Multilocus sequence analysis for *Leishmania braziliensis* outbreak investigation. PLoS Negl. Trop. Dis. 8, e2695.

Mauricio, I.L., Gaunt, M.W., Stothard, J.R., Miles, M.A., 2007. Glycoprotein 63 (gp63) genes show gene conversion and reveal the evolution of Old World *Leishmania*. Int. J. Parasitol. 37, 565–576.

Mauricio, I.L., Yeo, M., Baghaei, M., Doto, D., Pratlong, F., Zemanova, E., Dedet, J.P., Lukes, J., Miles, M.A., 2006. Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int. J. Parasitol. 36, 757–769.

Mcconville, M.J., Naderer, T., 2011. Metabolic pathways required for the intracellular survival of *Leishmania*. Annu. Rev. Microbiol. 65, 543–561.

Miles, M.A., Llewellyn, M.S., Lewis, M.D., Yeo, M., Baleela, R., Fitzpatrick, S., Gaunt, M.W., Mauricio, I.L., 2009. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. Parasitology 136, 1509–1528.

Murray, H.W., Berman, J.D., Davies, C.R., Saravia, N.G., 2005. Advances in leishmaniasis. Lancet 366, 1561–1577.

Nolder, D., Roncal, N., Davies, C.R., Llanos-Cuentas, A., Miles, M.A., 2007. Multiple hybrid genotypes of *Leishmania (viannia)* in a focus of mucocutaneous Leishmaniasis. Am. J. Trop. Med. Hyg. 76, 573–578.

Ochsenreither, S., Kuhls, K., Schaar, M., Presber, W., Schonian, G., 2006. Multilocus microsatellite typing as a new tool for discrimination of *Leishmania infantum* MON-1 strains. J. Clin. Microbiol. 44, 495–503.

Odiwuor, S., De Doncker, S., Maes, I., Dujardin, J.C., Van Der Auwera, G., 2011. Natural *Leishmania donovani*/*Leishmania aethiopica* hybrids identified from Ethiopia. Infect. Genet. Evol. 11, 2113–2118.

Pérez-Losada, M., Arenas, M., Castro-Nallar, E., 2017. Multilocus sequence typing of pathogens. Genet. Evol. Infect. Dis. 2, 383–404.

Ravel, C., Cortes, S., Pratlong, F., Morio, F., Dedet, J.P., Campino, L., 2006. First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major*. Int. J. Parasitol. 36, 1383–1388.

Rozas, J., 2009. DNA sequence polymorphism analysis using DnaSP. Methods Mol. Biol. 537, 337–350.

Schonian, G., Kuhls, K., Mauricio, I.L., 2011. Molecular approaches for a better understanding of the epidemiology and population genetics of *Leishmania*. Parasitology 138, 405–425.

Schonian, G., Nasereddin, A., Dinse, N., Schweynoch, C., Schallig, H.D., Presber, W., Jaffe, C.L., 2003. PCR diagnosis and characterization of *Leishmania* in local and imported clinical samples. Diagn. Microbiol. Infect. Dis. 47, 349–358.

Sharma, N.L., Mahajan, V.K., Kanga, A., Sood, A., Katoch, V.M., Mauricio, I., Singh, C. D., Parwan, U.C., Sharma, V.K., Sharma, R.C., 2005. Localized cutaneous leishmaniasis due to *Leishmania donovani* and *Leishmania tropica*: preliminary findings of the study of 161 new cases from a new endemic focus in himachal pradesh, India. Am. J. Trop. Med. Hyg. 72, 819–824.

Singh, N., Chikara, S., Sundar, S., 2013. SOLiD sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm *leptomonas* co-infection and raise some key questions. PLoS One 8, e55738.

Srivastava, P., Singh, T., Sundar, S., 2011. Genetic heterogeneity in clinical isolates of *Leishmania donovani* from India. J. Clin. Microbiol. 49, 3687–3690.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S., Diosque, P., 2013. MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. Infect. Genet. Evol. 20, 188–196.

Tsukayama, P., Lucas, C., Bacon, D.J., 2009. Typing of four genetic loci discriminates among closely related species of New World *Leishmania*. Int. J. Parasitol. 39, 355–362.

Urban, S., 2006. Rhomboid proteins: conserved membrane proteases with divergent biological functions. Genes Dev. 20, 3054–3068.

Van Griensven, J., Diro, E., 2012. Visceral leishmaniasis. Infect. Dis. Clin. North Am. 26, 309–322.

Weirather, J.L., Jeronimo, S.M., Gautam, S., Sundar, S., Kang, M., Kurtz, M.A., Haque, R., Schriefer, A., Talhari, S., Carvalho, E.M., Donelson, J.E., Wilson, M.E., 2011. Serial quantitative PCR assay for detection, species discrimination, and quantification of *Leishmania* spp. in human samples. J. Clin. Microbiol. 49, 3892–3904.

WHO, 2010. Control of the Leishmaniases. Tech. Rep. Ser. 20.

WHO, 2015. Kala-Azar elimination programme: report of a WHO consultation of partners, Geneva, Switzerland, 10–11 February.

Zemanova, E., Jirku, M., Mauricio, I.L., Horak, A., Miles, M.A., Lukes, J., 2007. The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. Int. J. Parasitol. 37, 149–160.

Zhang, C.Y., Lu, X.J., Du, X.Q., Jian, J., Shu, L., Ma, Y., 2013. Phylogenetic and evolutionary analysis of Chinese *Leishmania* isolates based on multilocus sequence typing. PLoS One 8, e63124.