Infection, Genetics and Evolution 12 (2012) 149–159

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the *Leishmania donovani* species complex

Tim Downing [a,*], Olivia Stark [b,*], Manu Vanaerschot [c], Hideo Imamura [c], Mandy Sanders [a], Saskia Decuypere [c], Simonne de Doncker [c], Ilse Maes [c], Suman Rijal [d], Shyam Sundar [e], Jean-Claude Dujardin [c], Matthew Berriman [a], Gabriele Schönian [b]

[a] *Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK*
[b] *Institut für Mikrobiologie und Hygiene, Charité Universitätsmedizin Berlin, Berlin, Germany*
[c] *Unit of Molecular Parasitology, Department of Parasitology, Institute of Tropical Medicine, Antwerp, Belgium*
[d] *B.P. Koirala Institute of Health Sciences, Ghopa, Dharan, Nepal*
[e] *Institute of Medical Sciences, Banaras Hindu University, Varanasi, India*

ARTICLE INFO

Article history:
Received 29 June 2011
Received in revised form 28 October 2011
Accepted 10 November 2011
Available online 20 November 2011

Keywords:
Leishmania infantum
Visceral leishmaniasis
Diversity
Population markers
Population genetics
Drug resistance

ABSTRACT

The species of the *Leishmania donovani* species complex cause visceral leishmaniasis, a debilitating infectious disease transmitted by sandflies. Understanding molecular changes associated with population structure in these parasites can help unravel their epidemiology and spread in humans. In this study, we used a panel of standard microsatellite loci and genome-wide SNPs to investigate population-level diversity in *L. donovani* strains recently isolated from a small geographic area spanning India, Bihar and Nepal, and compared their variation to that found in diverse strains of the *L. donovani* complex isolates from Europe, Africa and Asia. Microsatellites and SNPs could clearly resolve the phylogenetic relationships of the strains between continents, and microsatellite phylogenies indicated that certain older Indian strains were closely related to African strains. In the context of the anti-malaria spraying campaigns in the 1960s, this was consistent with a pattern of episodic population size contractions and clonal expansions in these parasites that was supported by population history simulations. In sharp contrast to the low resolution provided by microsatellites, SNPs retained a much more fine-scale resolution of population-level variability to the extent that they identified four different lineages from the same region one of which was more closely related to African and European strains than to Indian or Nepalese ones. Joining results of *in vitro* testing the antimonial drug sensitivity with the phylogenetic signals from the SNP data highlighted protein-level mutations revealing a distinct drug-resistant group of Nepalese and Indian *L. donovani*. This study demonstrates the power of genomic data for exploring parasite population structure. Furthermore, markers defining different genetic groups have been discovered that could potentially be applied to investigate drug resistance in clinical *Leishmania* strains.

1. Introduction

An estimated 12 million people are currently infected by *Leishmania* parasites, including the visceral (VL), cutaneous (CL) and mucocutaneous forms of the disease (www.who.int/leishmaniasis/). The species of the *Leishmania donovani* complex cause the most severe type of leishmaniasis (VL) in tropical and sub-tropical regions – though asymptomatic infections are common (Rijal et al., 2010). Unlike the other *Leishmania* parasites, those causing VL disseminate to internal organs and are responsible for 50,000 deaths and the loss of over 2.3 million disability-adjusted life years annually (WHO Expert Committee on the Control of the Leishmaniases, 2010). The number of people in India, Nepal and Bangladesh at risk of infection is 190 million and the annual volume of cases in India is over 100,000 – mostly occur in the north-western Indian state of Bihar, adjacent to the Terai region of Nepal, where VL is also endemic (Sundar et al., 2008). This shows the scale of the disease burden and the challenge in combating VL as detailed in the first regional programme of VL elimination in the Indian subcontinent by 2015 (WHO, 2005).

As a result, improved molecular tools are vital for monitoring the epidemiology and diversity of strains in the *L. donovani* species complex. A wide variety of approaches has been applied, including those using multilocus enzyme electrophoresis (Rioux et al., 1990),

Abbreviations: MLMT, multi-locus microsatellite typing ; VL, visceral leishmaniasis; CL, cutaneous leishmaniasis; PKDL, post kala-azar dermal leishmaniasis; SSG, sodium stibogluconate.

* Corresponding authors.
E-mail addresses: Tim.Downing@sanger.ac.uk (T. Downing), Olivia.Stark@charite.de (O. Stark).

1567-1348 © 2011 Elsevier B.V. Open access under CC BY license.
doi:10.1016/j.meegid.2011.11.005

sequencing of ribosomal loci (Kuhls et al., 2005), microsatellite typing (for example, Ochsenreither et al., 2006), gene sequences (for example, Mauricio et al., 2006), random amplification of polymorphic DNA (Botilde et al., 2006), amplified fragment length polymorphisms (for example, Kumar et al., 2010) and kinetoplastid minicircle DNA (kDNA) restriction fragment length polymorphisms (RFLP; for example, Laurent et al., 2007). Among this multitude of approaches used to assess diversity, kDNA RFLP and multi-locus microsatellite typing (MLMT) have proven to be highly discriminatory for typing *L. donovani* species complex parasites (Botilde et al., 2006). Although kDNA diversity can be informative, the variability in both experimental replication (Bhattarai et al., 2010) and total DNA yield (Downing et al., 2011) limit the effectiveness of this approach. Nonetheless, the hypothesis of multiple origins for drug resistance in these parasites stems from kDNA variation in Nepalese *L. donovani* (Laurent et al., 2007), and was supported by genome-wide variability (Downing et al., 2011).

MLMT has been used extensively for typing strains of the *L. donovani* complex. An initial study investigating variation in 15 microsatellite markers for 91 strains of the *L. donovani* complex from different VL foci in the Mediterranean Basin, East Africa and the Bihar state of India highlighted significant differentiation between the continental groups (Kuhls et al., 2007). However, in contrast to strains from East Africa and the Mediterranean that appeared to be highly variable, Bihari strains had little microsatellite profile diversity. Improved sampling of more and newly isolated strains from Bihar, Nepal and Bangladesh with MLMT suggested that *L. donovani* strains in the Indian subcontinent showed genetic homogeneity regardless of geographical origin, clinical manifestation, and whether they presented *in vitro* or *in vivo* susceptibility to antimonial drugs (Alam et al., 2009). This result was reflected in the poor resolution of MLMT as well as also PCR-RFLP targeting kDNA and genomic coding sequences to clearly determine the genetic relationships of Nepalese strains (Bhattarai et al., 2010). In contrast, wider MLMT analysis of *L. infantum* strains from Europe and South America ($N = 406$), and of East African *L. donovani* ($N = 123$) further supported their considerable diversity (Gelanew et al., 2010; Kuhls et al., 2011).

With ongoing improvements in sequencing technology, genome-wide SNP typing represents a powerful alternative approach for differentiating parasite strains (Mardis, 2011). In this study, we sought to elucidate the variation in the *L. donovani* complex firstly within a population and secondly between species using SNP and microsatellite genotyping, while also comparing the power of each marker type. While both methods were effective at discriminating isolates over long geographic distances, informing on the population history of *Leishmania* strains, microsatellites lacked sufficient power to resolve diversity in closely related strains of Nepalese and Indian *L. donovani*. In contrast, genome-wide SNP variation provided new evolutionary insights into the ongoing diversification of this phenotypically variable set of strains, potential links between the genotypes of the strains with *in vitro* sodium stibogluconate (SSG) resistance, as well as identifying protein sequence mutations that may underlie the phenotypic differences.

2. Material and methods

2.1. Sample collection

To compare the discriminatory power of microsatellite and SNP typing for unravelling diversity within a set of genetically homogeneous *L. donovani*, 25 clinical isolates taken from a small focus with endemic VL in the Indian subcontinent were examined by assaying their microsatellite and SNP genotypes (Supplementary Fig. 1). These parasites were isolated n the Terai area of Nepal (23) and the nearby Indian state of Bihar (two). Each isolate was independently taken from unique VL patients (with the exception of two strains from BPK173; Rijal et al., 2010). The genomes of 17 of these have been scrutinised (Downing et al., 2011) and so eight of the Nepalese strains represent entirely new samples assessed using both SNPs and microsatellites (Table 1). DNA was isolated and phenotype responses were assessed for *in vitro* susceptibility to SSG relative to the Nepalese reference genome (BPK282/0cl04, see Supplementary data) for lines not already assessed (Rijal et al., 2007; Downing et al., 2011). In total, nine clinical samples were classed as resistant and 15 as sensitive (the phenotype of BPK077/0cl5 was not determined).

Genetic variation between strains of the *L. donovani* complex from different continents was studied using both marker types to frame diversity observed with the Terai-Bihar region. Five additional *L. donovani* strains originally taken in Sri Lanka (L60b), Kenya (LRC-L53, NLB218), Sudan (597LN), Ethiopia (GEBRE1), as well as three *L. infantum* from China (D2) and France (LPN114, LRC-L47) were assessed. All of these caused VL or PKDL except L60b, whose pathology was cutaneous (CL): this strain was closely related to other Sri Lankan ones causing CL (Alam et al., 2009). Genome, enzyme and microsatellite markers have been previously used to classify these eight samples (Kuhls et al., 2005, 2007; Mauricio et al., 2006; Lukes et al., 2007; Zemanová et al., 2007; Kuhls et al., 2008), leading to their usage here as informative representatives of genetic variation in the wider *L. donovani* complex. Ethiopian GEBRE1 and the Kenyan strains LRC-L53 and NLB218 are representatives of East African diversity, though NLB218 may be more divergent within this group (Alam et al., 2009). Among the *L. infantum* samples, French strain LPN114 represents Mediterranean strains belonging to zymodeme MON-1. Strains LRC-L47 (France) and D2 (China) are part of the non-MON-1 group, though D2 has a more divergent genetic profile (Kuhls et al., 2005, 2007).

2.2. Microsatellite profiling

In light of the extensive research already performed on microsatellite variability in *L. donovani*, a combination of new and existing microsatellite data yielded a total 193 strains of the *L. donovani* complex that were investigated here. Analysis of variation in tandem repeat numbers at 15 unlinked microsatellites (Supplementary Table 1) was completed for 41 newly typed strains from Bangladesh, the Bihar state of India and Terai region of Nepal (Supplementary Table 2). Their microsatellite profiles were compared to those of 152 strains previously studied of which 25 came from African and European foci endemic for VL (Ochsenreither et al., 2006; Kuhls et al., 2007; Alam et al., 2009; Bhattarai et al., 2010). Of the 193 strains, 168 strains were from the Indian subcontinent, 17 *L. donovani* strains from Sudan (7), Kenya (7), and Ethiopia (3), and eight genetically diverse strains of *L. infantum* were from France (3), China (3), Spain (1) and Tunisia (1). The strains of *L. infantum* represented the MON-1 and different non-MON-1 zymodemes (Ochsenreither et al., 2006; Kuhls et al., 2007).

PCR amplification and determination of the DNA fragment sizes was performed as described elsewhere (Ochsenreither et al., 2006; Kuhls et al., 2007). Ten of the 15 microsatellite markers were invariant in the previously published strains from the Indian subcontinent (Ochsenreither et al., 2006; Kuhls et al., 2007; Alam et al., 2009) and, consequently, five markers (Lm4TA, TubCA, B_Li41-56, F_Li23-41 and CS20) were amplified in the 41 newly typed strains (Supplementary Table 1).

2.3. SNP genotyping

SNP variation was examined in the 33 strains (23 Nepalese, two Indian and eight globally diverse *L. donovani* complex strains) out