Title:

Exploring genome-wide variation in *Leishmania donovani*

Authors & affiliations:

Tim Downing [1], Hideo Imamura [1,2], Saskia Decuypere [2], Suman Rijal [3], Shyam Sundar [4], Christiane Hertz-Fowler [1,5], Jean-Claude Dujardin [2], Matt Berriman [1].

[1] Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK.
[2] Institute of Tropical Medicine, Antwerp, Belgium.
[3] B.P. Koirala Institute of Health Sciences, Dharan, Nepal.
[4] Kala-Azar Medical Research Center, Dept. of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India.
[5] Centre for Genomic Research, University of Liverpool, UK.

Abstract:

Over half a million people each year are afflicted by visceral leishmaniasis (kala-azar), a tropical disease caused by the Trypanosomid *Leishmania donovani*. The parasite's severe clinical virulence and capacity to resist drug treatment necessitate a diverse combination of synergistic molecular approaches in order to elucidate the key genetic and metabolic changes occuring during disease manifestation and drug pressure. Using Illumina sequencing technology, we have so far resequenced the genomes of 17 *L. donovani* strains with differing phenotypes isolated from separate patients in reference hospitals in the Indian subcontinent. These 2 Indian and 15 Nepalese samples were evaluated in terms of their resistance to SSG (stibogluconate) drug treatment and the various treatment outcomes of the patients. Genetic variability among the strains encompassed differences in SNP sites, structural variation and whole chromosome number. With regard to SNP diversity, we identified 7,641 variable sites, including 858 in coding regions – of which 526 cause changes in the protein sequence. 1717 SNPs were uniformly heterozygous in all samples: this included 121 nonsynonymous mutations, providing evidence of distinctive functional consequences of different mutations. In addition, to determine *L. donovani*-specific SNPs that were segregating at high frequencies, the other published *Leishmania* genomes (*L. infantum, major, mexicana* and *braziliensis*) were aligned to the *L. donovani* reference genome. 154,816 differences fixed between species were observed in the *L. infantum* genome (93.4%) orthologously aligning to *L. donovani*, illustrating the limited extent of variation in the Indian and Nepalese strains: this was more acute in coding regions. Of the 730 SNPs with elevated derived allele frequencies (> 0.5) in the *L. donovani* samples, majority (60%) of those in coding regions lead to amino acid substitutions. These mutations represent candidates for further investigation of the phenotypic effects of genetic variation, with the prospect of assessing metabolomic data in the framework of genomic diversity.

For more details on the projects, see www.leishrisk.net/kaladrug and www.leishrisk.net/gemini.