Title:

A global study of whole genome and metabolome diversity in natural *Leishmania donovani* populations

Authors & affiliations:

S. Decuypere[1], H. Imamura[1,3], R. t'Kindt [1,4], A. Jankevics[2,6], T. Downing[3], R. A. Scheltema[6] C. Hertz-Fowler[3,7], D. Watson[4], S. Rijal[5], G.H. Coombs[4], M. Berriman[3], R. Breitling[2,6], JC Dujardin[1]

1 Institute of Tropical Medicine Antwerp, Belgium; **2** University of Glasgow, Glasgow, UK; **3** Wellcome Trust Sanger Institute, Cambridge, UK; **4** University of Strathclyde, Glasgow, UK; **5** B.P. Koirala Institute of Health Sciences, Dharan, Nepal; **6** University of Groningen, Groningen, the Netherlands; **7** Centre for Genomic Research, University of Liverpool, Liverpool, UK.

Abstract:

Characterizing the diversity of pathogen populations is a crucial step towards understanding the clinical polymorphism of infectious diseases. The aim of our GeMInI study is to document and integrate data on the genomic and metabolic diversity of natural pathogen populations. We are applying new sequencing technologies and ultra-high mass accuracy mass-spectrometry to comparatively analyse the whole genome and metabolome of multiple strains of a single pathogen species, *Leishmania donovani*, the causative agent of visceral leishmaniasis.

Comparative genomics of *L. donovani* strains first required the construction of a *de novo* reference genome for this species. The genomic diversity was assessed among clones derived from 20 Nepalese clinical isolates and three types of genomic diversity elements were surveyed: (i) single nucleotide polymorphisms (SNPs), (ii) copy number variations (CNVs) and (iii) chromosome number (ploidy) differences. SNPs were correlated with observed clinical phenotypes and previously observed population structure. Several large scale CNVs were also identified and the most prominent genomic diversity was found at the level of chromosome number and ploidy, both of which appeared to be strain-specific.

Comparative metabolomics of *L. donovani* strains required the optimisation of a specific protocol suitable for comprehensive *Leishmania* metabolite profiling. Our current protocol detects the relative abundance of 340 metabolites including approximately 20% of the predicted core *Leishmania* metabolome and a large number of lipids.

Unsupervised clustering and principal component analysis clearly distinguishes the drug-sensitive and drug-resistant strains, yielding a total of 100 metabolites differing more than three-fold between the two phenotypes. Many of these differences are in specific areas of lipid metabolism, suggesting that the membrane composition of the drug-resistant parasites is extensively modified.

In the next stage of the study, we will integrate the whole genome and metabolome diversity data and investigate their relationship with clinical polymorphisms.