The dynamics of population genomic variation in Leishmania donovani clinical strains

Tim Downing1, Hideo Imamura1,2, Saskia Decuyper2, Suman Rijal3, Shyam Sundar3, Christiane Hertz-Fowler1,3, Jean-Claude Dujardin5, Matt Berriman1.

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

Contact: tim.downing@sanger.ac.uk

Introduction

Leishmania donovani is a principle cause of visceral leishmaniasis (kala-azar), which has a global incidence rate of over 500k cases per year. This tropical protozoan parasite is developing resistance to the limited repertoire of available drugs and so genetic drug target discovery could instigate the creation of novel therapeutics.

In this study we by initially produced a complete L. donovani genome from an isolate cloned from a patient.

Subsequently, we sequenced the genomes of 17 samples from India (2) and Nepal (15) strains to preliminary explore their population genomics.

Combining this data with the other complete Leishmania genomes, phenotype information relating to both the patients’ case history and the biochemical characteristics of L. donovani as well as the genomes of additional strains in production is expected will key genetic changes associated with drug resistance and visceralisation.

Comparing Leishmania genomes

In order to determine the L. donovani-specific SNP alleles, the published L. infantum genome was conservatively aligned to the L. donovani reference.

This identified conserved orthologous gene regions that were iteratively expanded for each chromosome using homology identification and alignment programs. This method will also be used to predict alleles for sites in other complete Leishmania genomes: major, mexicana and braziliensis.

Functional SNP discovery

An average of nearly 28 million uniquely mapping paired-end reads were generated for each of the 17 strains. These were mapped to the new L. donovani reference genome with a mean coverage of over 68-fold.

In order to avoid identifying errors due to single base mistakes, repetitive gene arrays or paralogous regions as SNPs, candidate SNP sites were examined for quality, coverage levels and local sequence uniqueness. A subset of SNPs (289) were genotyped in 36 strains, including African strains, using the Sequenom iPLEX platform.

Creating a reference genome

Paired-end 454 FLX reads were assembled into de novo contigs for L. donovani reference sample BPK282A1.

Illumina

This initial set of contigs was enhanced by the addition of high-coverage Illumina reads using a series of programs to remove gaps (IMAGE), clarify repetitive regions (Velvet) and reduce nucleotide errors (ICORN).

Reference genome

These contigs were aligned, ordered and oriented (ABACAS). Ssah2 was used to map this data to the L. infantum genome so that the high-quality L. donovani genome ascertained could be used as a reference for sequence from other strains.

Annotating the reference genome

The conserved nature of the coding sequences between L. infantum and L. donovani was illustrated by the successful transfer of 8,254 gene models from a total of 8,395 using the tools RATT (Otto et al. 2010), Artemis and ACT (Carver et al. 2008).

Using a conservative SNP screening approach, 3,549 candidate sites were segregating within the 17 samples. This includes 616 in coding regions, of which 396 cause amino acid changes.

A McDonald-Kreitman test comparing the relative rate of nonsynonymous to synonymous mutations within the 17 strains (1.8) to those fixed between L. infantum and L. donovani (1.1) indicated that nonsynonymous substitutions were subject to purifying selection (p < 10^{-6}).

We ascertained nonsynonymous SNPs in the set of 17 in genes relating to:

- amastins (17) receptors (10)
- kinases (15) protein transport (11)
- protein transfer (24)

Additionally, we identified donovani-specific protein-level mutations that were variable in at least half of the strains (52).

- Future tests for adaptive evolution will link genomic & metabolomic data.
- At least 5 of the 17 strains show resistance to sodium stibo-gluconate (SSG) and a total of 7 are associated with treatment failure in patients.
- The strategy outlined here will be augmented by the continued sequencing of strains with increasing heterogeneity.

An ACT screenshot (below) of genes on chromosome 12 in L. infantum (top) and L. donovani (bottom) with orthologous BLAST hits (red), genes (green, yellow and orange) on the forward and reverse direction of each species’ chromosome.

Contact: tim.dowling@sanger.ac.uk

Website: www.leishrisk.net/kaladrug

Grant number: EC-FP7-222895