GENOME WATCH

Sequencing parasite populations

Alejandro Sanchez-Flores

This month's Genome Watch highlights how a population study, in conjunction with a reference genome, can identify the evolutionary features that contribute to drug resistance in a protozoan parasite.

Faster and cheaper technologies for sequencing whole genomes have greatly increased the availability of complete genome sequences for a range of organisms. Furthermore, the adoption of certain standards for the quality of sequence data[1] has meant that the genome sequences in public databases now provide an excellent point of reference for researchers. In combination with comparative genomic analyses, proteomics and *in silico* drug design, these reference genomes can aid the development of new therapies against protozoan infections[2,3].

Leishmania spp. are responsible for different diseases ranging from mild skin or mucosal lesions to fatal visceral leishmaniasis, and there are an estimated 12 million infected individuals worldwide (www.who.int/leishmaniasis/en/). Previous comparative genomic analyses of the *Leishmania major*, *Leishmania braziliensis* and *Leishmania infantum* genomes[4] identified a high degree of gene conservation and synteny between the three species but revealed few genes that could be associated with the diversity of clinical diseases or could be used as drug targets. However, a recent report presents a reference genome for *Leishmania donovani*[5] and includes a novel analysis using 16 related clinical lines of the parasite that exhibit differing susceptibility to antimonial drugs such as sodium stibogluconate (SSG). The reference genome was generated using a single cloned line with 454 and Illumina sequencing technologies. The result was a genome of 32.4 Mb contained within 2,154 fragments, in which 50% of the genome was contained in fragments of 45.5 kb or larger, meeting the quality standards to be used as a reference genome.

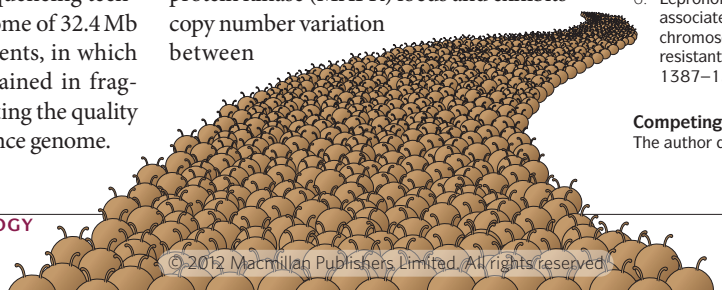
Comparison of the reference genomes of all four *Leishmania* species revealed substantial single nucleotide polymorphism (SNP) variation between the three previously reported genomes and that of *L. donovani*, but interestingly only a low level of variation was found among the 17 *L. donovani* lines. This was unexpected, as high rates of nucleotide changes are often related to differences in drug susceptibility. Nevertheless, those genes that contained SNP differences between SSG-resistant and SSG-susceptible lines did show positive selection and encode products that are probably important during adaptation to drug pressure, including surface proteins, amino acid transporters, proteins involved in differentiation and enzymes involved in antioxidant metabolism.

In addition, the authors found extensive variation in chromosome and gene copy number, with a unique chromosome copy number pattern for each *L. donovani* line. Interestingly, a previous report in *L. infantum*[6] correlates antimonial drug resistance to copy number of aneuploid chromosomes, but the authors of the *L. donovani* study could not identify a clear relationship between SSG susceptibility and aneuploidy in the 17 *L. donovani* lines.

The authors observed other important features contributing to gene dosage differences, including the expansion and contraction of genes in tandem arrays that are expressed through polycistronic transcription. Changes in the arrayed nature and copy number of such genes could regulate the dosage and expression of genes encoding antigenic proteins. In addition, the authors identified an episome that contains a mitogen-activated protein kinase (MAPK) locus and exhibits copy number variation between the different lines. The locus contains an acid phosphatase gene, and a MAPK gene with a probable role in signal transduction during the infection. Episomes have previously been reported to be stable only after *in vitro* drug selection[6].

The combination of SNPs, changes in ploidy and variation in gene copy number provide a strategy for adaptation that has not previously been described for kinetoplastid parasites and is probably important for the degree of SSG susceptibility among the *L. donovani* population. This analysis will therefore provide the basis for a more detailed molecular profile that could be used to improve vaccine and drug design.

Whole-genome sequencing of parasite populations, along with association studies, could be the next step to understanding the evolution and epidemiology of a wide range of important parasitic diseases.



Alejandro Sanchez-Flores is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro2738

1. Chain, P. S. G. *et al.* Genome project standards in a new era of sequencing. *Science* **326**, 236–237 (2009).
2. Zucca, M. & Savoia, D. Current developments in the therapy of protozoan infections. *Open Med. Chem. J.* **5**, 4–10 (2011).
3. Pérez-Victoria, F. J., Sánchez-Cañete, M. P., Castanys, S. & Gamarro, F. Phospholipid translocation and miltefosine potency require both *L. donovani* miltefosine transporter and the new protein LdRos3 in *Leishmania* parasites. *J. Biol. Chem.* **281**, 23766–23775 (2006).
4. Peacock, C. S. *et al.* Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nature Genet.* **39**, 839–847 (2007).
5. Downing, T. *et al.* Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **21**, 2143–2156 (2011).
6. Leprohon, P. *et al.* Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Res.* **37**, 1387–1399 (2009).

Competing interests statement
The author declares no competing financial interests.