

Assembly, annotation and gene editing of the genome of the PH8 strain of *Leishmania amazonensis* with focus on multigene families encoding virulence factors

Carlos Rodolpho Ferreira Brasil, João Luís Reis Cunha, Anderson Coqueiro-dos-Santos, Viviane Grazielle da Silva, Daniella Bartholomeu, Ana P. S. M. Fernandes, Santuza Maria Ribeiro Teixeira, Wanessa Moreira Goes

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Leishmania (*Leishmania*) *amazonensis* is one of the etiological agents of cutaneous leishmaniasis, a disease that has 21, 000 cases / year in Brazil. Different molecules of the parasite have already been studied as they play a crucial role in the establishment of infection in the mammalian host and contribute to the pathogenesis of leishmaniasis. Among the most studied virulence factors are multigene protein-coding families such as amastins, GP63 metalloproteases and A2 proteins. To deepen the understanding of the role of these virulence factors, it is essential to obtain well-assembled and annotated complete genomes of different isolates of *L. amazonensis* and also to be able to manipulate the genes that encode these factors using the CRISPR-Cas9 technology. Here, we report the sequencing and chromosome level assembly of the PH8 strain of *L. amazonensis* using a strategy based on the combination of long PacBio sequencing reads, short Illumina reads and synteny data with the *Leishmania mexicana* genome. The initial contigs were generated using only the PacBio reads and the Canu assembler. The scaffolding step was performed using SSPACE and gap filling with GapFiller, based on short paired-end reads Illumina. Finally, the assembly was polished using Pilon, and the scaffolds were ordered based on the chromosomes of *L. mexicana* using Abacas. The final assembly, composed of 34 chromosomes and 44 small scaffolds not incorporated in the 34 chromosomes, represents a genome of 32 Mb. The annotation of the *L. amazonensis* genome was transferred from the annotation of 8225 genes present in the genome of *L. mexicana*. Of these, 7072 are protein coding genes, among which 82 encode tRNAs, 13 rRNAs and 270 ncRNAs. In addition, genes belonging to the families of amastins and metalloproteins GP63 were characterized. Amastins and GP63 were identified based on homology with proteins from other *Leishmania* spp and *T. cruzi*, adding up to 33 and 5 genes, respectively in *L. amazonensis*. The alignment of amastin genes, using MUSCLE and phylogeny analyzes, using the neighbor-joining algorithm with 1000 bootstrappings resulted in groupings corresponding to the four sub-classes of amastins known as alpha (α), beta (β), gamma (?) and delta (δ) amastins. Analysis of the sequences encoding GP63 showed the conservation of important domains, such as HExxH and SRYD, which are important for protein structure and binding to macrophage surface receptors, respectively. Finally, we tested a CRISPR-Cas9 protocol to generate knockout cell lines of the Miltefosine Transporter gene (TM) of *L. amazonensis* as a proof of concept. Expression of *Streptococcus pyogenes* Cas9 (SpCas9) in *L. amazonensis* promastigotes was achieved after transfection with pLDCN, an episomal vector. Alternatively, *Staphylococcus aureus* Cas9 (SaCas9) was also expressed in *Escherichia coli*. Promastigotes were

transfected with the recombinant SaCas9 ribonucleoprotein complex and an SgRNA. Using both strategies to express Cas9 we were able to disrupt the TM gene in parasites that were transfected with in vitro transcribed sgRNA and a donor oligonucleotide and obtained knockout parasite cell lines that were approximately 6 fold more resistant to miltefosine than wild parasites.

Funding: CAPES

Link to Video: