Improvement of Angiostrongylus costaricensis genome annotation using RNA-Seq data

Karina Mastropasqua Rebello, Makedonka Mitreva, James McKerrow, Ana Gisele da Costa Neves Ferreira, Fabio Passetti, ESDRAS MATHEUS GOMES DA SILVA

IOC/Fiocruz

Abstract

Angiostrongylus costaricensis is a roundworm species that causes an intestinal inflammatory disease, known as abdominal angiostrongyliasis. The rodents are typically its definitive hosts, where they are usually found in the mesenteric arteries. Humans are accidental hosts, being contaminated by the ingestion of infective third stage larvae present on contaminated water and food. Currently, there is no drug available that acts directly on this parasite, mostly due to the sparce understanding of its molecular characteristics. Thus, aiming to provide a better understanding of its molecular aspects we present here the improved annotation of A. costaricensis protein-coding genes and transcripts using RNA-Seq data. First, the transcripts of both male and female adult worms were sequenced using RNA-Seq Illumina technology, generating short-paired reads. These RNA-Seq reads were aligned to the genome draft (version WBPS15) and used as extrinsic evidence for predicting protein-coding genes and transcripts, using the software BRAKER2. These predicted genes and transcripts were used to increment the WBPS15 genome annotation. The functional annotation of the complete ORFs of the WBPS15 improved was achieved using the software blast2GO from the OmicsBox package. The different gene expression (DGE) analysis between male and female worm genes was performed using the DESeq2 R package. The WBPS15 improved genome annotation comprises 14, 588 genes, 27, 788 mRNAs and 21, 584 complete ORFs. Overall, 72% of complete ORFs sequences were completely annotated by Blast2GO. It was identified 2, 573 genes more expressed in male and 747 genes more expressed in female worms, with adjusted P value (FDR) = 0.01 and Log2 fold change = 2 thresholds. Among the overrepresented terms of male are: non-membrane spanning protein tyrosine kinase activity (GO:0004715), protein kinase activity (GO:0004672) and phosphoprotein phosphatase activity (GO:0004721) and of female (among 7 GO terms over-expressed) are: protein tyrosine phosphatase activity (GO:0004725), mRNA binding (GO:0003729) and protein phosphatase 1 binding (GO:0008157). We believe that this improved genome annotation of protein-coding genes and transcripts contributes to better understand the molecular diversity of A. costaricensis, being this is a key step for the selection of therapeutic proteins.

Funding: CAPES

Link to Video: