Long non-coding RNAs potentially involved with Schistosoma japonicum sexual maturation

Lucas Ferreira Maciel, David Abraham Morales Vicente, Sergio Verjovski-Almeida

University of São Paulo

Abstract

Schistosoma japonicum is a flatworm which causes schistosomiasis, a neglected tropical disease. There is only one efficient drug for treatment, which may lead to resistance emergence. Due to the importance of sexual maturation for the parasite lifecycle and host immunopathogenesis, Wang et al. (Nature Communications 8: 14693, 2017) performed RNA-seq analyses of females and males obtained from 14 up to 28 days post-infection (dpi) in mouse in order to better understand the molecular mechanisms of sexual maturation. They identified protein-coding (PC) genes and specific pathways whose expression levels are related to sexual development; however, this work did not include an analysis of long non-coding RNAs (lncRNAs), transcripts that in mammals were shown to be key regulators of vital processes. There is one paper in the literature reporting the presence of 3, 000 lncRNAs in S. japonicum, but the annotation was performed with an old version of the genome, and only one male and one female RNA-Seq library were used. Our group has recently shown that lncRNAs expression is stage-specific in S. mansoni. Therefore, the aim of the present work is to identify and annotate a more complete set of lncRNAs that complements the most updated PC transcriptome annotation by re-analyzing all RNA-seq datasets in the public domain, including those generated by Wang et al. (2017), to identify stage-specific lncRNAs related to sexual maturation. For this purpose, 66 RNA-seq libraries from five different life-cycle stages were downloaded from the SRA-NCBI. Reads quality control was performed using fastp and aligned against the genome ASM636876v1 using STAR. Uniquely mapped reads were then used for transcripts reconstruction with Scallop, followed by TACO meta-assembly. Coding potential calculation was performed with FEELnc and CPC2. Transcripts classified as lncRNAs were then submitted to annotation with eggnog-mapper to remove possible remaining mRNAs. Synteny analysis was performed between S. japonicum and S. mansoni genomes. The lncRNAs found were included in the transcriptome dataset and expression quantified with RSEM. Weighted gene co-expression network analyses (WGCNA) were then performed in order to identify modules related to sexual maturation. Our pipeline was able to identify 12, 291 lncRNAs in S. japonicum genome. Synteny analysis identified that 80% of all intergenic lncRNAs were contained inside syntenic blocks of at least 5 pairs of orthologous PC genes. WGCNA analysis identified 7 different modules that demonstrate that lncRNAs have a dynamic expression throughout sexual maturation.

Funding: This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant number 2018/23693-5 to SV-A. LM received FAPESP fellowships grant number 2018/19591-2 and DMV received a fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). SV-A laboratory was also supported by institutional funds from Fundação Butantan and received an established investigator fellowship award from CNPq, Brasil.