

Complex life cycle: a set of phenotypes on the single genome

M. Nesterenko¹, V. Starunov^{1,2}, S. Shchenkov^{1,3*}, A. Dobrovolskij¹, K. Khalturin^{1,4}

¹ Saint-Petersburg State University, St. Petersburg, Russia

² Zoological Institute of RAS, St. Petersburg, Russia

³ Kovalevsky Institute of Marine Biological Research of RAS, Russia

⁴ Okinawa Institute of science and technology, Okinawa, Japan

* e-mail: sergei.shchenkov@gmail.com

Key words: complex life cycle, RNA-seq, comparative transcriptomics, trematodes

Motivation and Aim: Complex life cycles are characteristic to different groups of invertebrates, free-living as well as parasitic. During the single life cycle implementation several generations contrasting in morphology, physiology and behavior alternate successively and regularly. In the life cycles with heterogony different generations and phases are formed on the basis of single genome. In the trematodes life cycle (which are endoparasites with heterogony) amphimictic generation alternates parthenogenetic ones, and both of them include free-living and parasitic phases. The aim of our research is to reveal the molecular basics of different phenotype formation during trematodes life cycle. As an object of study, we chose two trematodes species from Psilostomatidae family, *Sphaeridiotrema pseudoglobulus* and *Psilotrema simillimum*.

Methods and Algorithms: Transcriptomes of rediae, cercariae and maritae of both investigated species were sequenced using Illumina Hiseq 2500 instrument and assembled *de novo* by Trinity, SOAPdenovo-Trans and TransABYSS. Contigs were clustered (CD-HIT-EST) and only best ones were selected by TransRate. Web-resources FunctionAnnotator and KEGG were used for annotation. The close sequences identification between transcriptomes were performed by reciprocal best hit search and custom Python scripts. To determine the domain architecture of predicted amino acids sequences, hidden Markov models (hmmscan versus Pfam-A database) and custom Python parser were used. Comparison of sets of highly expressed genes both among life cycle phases, and between two species were performed by TROM package for R.

Results: Selected assembling strategy allowed us to obtain high-quality transcriptome for each life cycle phase of both species. Approximately 50 % of the assembled sequences had analogs in NCBI NR protein database, however domain architecture of protein was predicted for contigs with annotation as well as for many unannotated sequences. Among homologous and orthologues sequences at least 1/5 part demonstrate variation in the domain architecture. The corresponding phases of the two trematode species show more significant overlap in highly expressed gene sets, than the phases within a single life cycle.

Conclusion: In current biology genome-phenotype relationship is still weakly studied, however it becomes clear, that not only gene expression or its absence determines the traits. Present results suggest that during the implementation of a complex life cycle both the change in the gene expression level and creation of transcripts isoforms are important.

Acknowledgements: Scientific research were performed at the Research park of St. Petersburg State University “Biobank”, “Center for Molecular and Cell Technologies”, “Computing Center”, and “Center for Culture Collection of Microorganisms”. The research was founded by grant MK-2105.2017.4.