

Prediction of operon-like gene clusters in the *Intoshia linei* genome

Bondarenko N.I.

Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, St. Petersburg,
Russia

Natalya.lannik@gmail.com

Key words: Orthonectida; genome reduction; operon; adaptations

Motivation and Aim: Orthonectida is a small, rare, and in many aspects enigmatic group of organisms with a unique life cycle and a highly simplified adult free-living stage parasitizing various marine invertebrates [1, 2]. Several studies have shown that parasitism can lead to a dramatic reduction of the body plan, morphological structures and also affects organisms at the genomic level [3, 4]. In previous studies were shown that Orthonectida has extremely compact genomes due to the significant reduction of gene number, intergenic regions, intron length, and repetitive elements [5, 6]. As a result, gene density in these species much higher than in other metazoans. For example, in *Intoshia linei* genes cover 53.8 % of the genome, an average density of 202 genes per Mb and median intergenic distance consists of 1549 bp [6]. The close proximity of Orthonectida genes suggested that they might be possess operon-like gene organization and co-transcribed from a single upstream promoter. In this study we sequenced *I. linei* transcriptome and compared the experimentally determined polycistronic transcripts with computational analysis of intergenic distances in the *I. linei* genome.

Methods and Algorithms: To identify potential orthonectid operon-like gene clusters we employed our pipeline with the prediction method based on the intergenic distance between reading frames as a primary predictor of the number of operons in the genome. All predicted operon-like gene clusters then were validated using the RNA-seq data. Total RNA was isolated using Quick-RNA Miniprep (Zymo Research). The RNA-seq library was prepared using NEBNext RNA library preparation protocol (New England BioLabs). The library was sequenced on a HiSeq 2000 sequencing system, producing 52.2 M 110-bp paired-end reads. The reads were adapter trimmed with Trimmomatic v0.30 and assembled into transcripts using the rnaSPAdes. To find polycistronic transcripts we used TransDecoder and our own scripts for excluding false positive prediction.

Results: Computational analysis of the genome assembly for *I. linei* predicted that 4497 genes (51.5 %) have intergenic distance ≤ 3500 bp and might be organized into 1847 operons. The average number of genes per operon are 2.4, minimum two genes per operon, and maximum 9 genes per operon. The median intergenic distance within the set of all these predicted operons in *I. linei* was observed to be 1285 nt, while 934 operons (~50.5 % of the total 1847 operons) have genes with intergenic distances < 500 nt. According to this, most of the predicted *I. linei* operons are characterized by close apposition of the cotranscribed genes. To characterize of *I. linei* transcriptomes and investigate the presence of operon-derived mRNAs total RNA-sequencing was used. Sequencing revealed 403 dicistronic and 55 polycistronic transcripts. More than 4 % of the transcriptome of *I. linei* is organized as operons. Thereby we validated 982 genes

organized into operons out of 4497 potential candidates. Most of the genes in *I. linei* operons are separated from one another by approximately 1289 bp.

Conclusion: In this study we have predicted and validated that *I. linei* genome has operon-like nonoverlapping gene clusters which are transcribed as polycistronic transcripts.

Funding: The study is supported with Russian Science Foundation 24-24-00449.

References

1. Kozloff E.N. The genera of the phylum Orthonectida. *Cah Biol Mar.* 1992;33:377-406
2. Slyusarev G.S. Phylum Orthonectida: morphology, biology, and relationships to other multicellular animals. *Zh Obshch Biol.* 2008;69:403-427
3. Burke M. et al. The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. *Nematology.* 2017;(6):621-637. doi 10.1163/15685411-00002901
4. Wang J., Gao S., Mostovoy Y., Kang Y. et al. Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* 2017;27(12):2001-2014. doi 10.1101/gr.225730.117
5. Mikhailov K.V., Slyusarev G.S., Nikitin M. et al. The genome of *Intoshia linei* affirms orthonectids as highly simplified spiralian. *Curr Biol.* 2016;26:1768-1774
6. Slyusarev G., Starunov V., Bondarenko A., Zorina N., Bondarenko N. Extreme genome and nervous system streamlining in the invertebrate Parasite *Intoshia variabilis*. *Curr Biol.* 2020;30(7):1292-1298.e3. doi 10.1016/j.cub.2020.01.061