

Viral metagenomic analysis of publicly available genomic and transcriptomic samples from Simuliidae and Ceratopogonidae insects

Kamanova E.P.^{1,2*}, Starchevskaya M.E.^{1,2}, Antonets D.V.^{1,2}

¹ State Research Center of Virology and Biotechnology “Vector”, Novosibirsk, Russia

² Novel Software Systems LLC, Novosibirsk, Russia

* kamanovae@gmail.com

Key words: metagenomics, *Nematocera*, viruses of insects

Motivation and Aim: Insects make the largest group of multicellular organisms. They also often act as carriers of numerous infection diseases of plants, animals, and humans. Blood-sucking insects from the suborder Nematocera deserve special attention, as they can transmit dangerous human viral infections such as Zika and Dengue fever, yellow fever and many more. Viral infections of livestock animals, transmitted by blood-sucking insects, can cause serious economic losses. To date, the Culicidae family is the most well-studied family of *Nematocera* insects. And there are numerous studies throughout the world dedicated to mosquito-borne viral infections. However, other important groups of *Nematocera* insects such as Simuliidae and Ceratopogonidae families as well as viruses associated with them remain largely unknown. Nevertheless, it has been shown that the members of Simuliidae family can transmit vesicular stomatitis virus to domestic pigs [1]. They were also found to play a role in transmission of Venezuelan equine encephalomyelitis virus and myxomatosis virus [2].

Methods and Algorithms: We decided to investigate the available datasets to identify the sequences belonging to known and new viruses. From the NCBI SRA database we collected 8 genomic and 3 transcriptomic samples of Simuliidae family and 8 genomic and 144 transcriptomic samples of Ceratopogonidae family. The analysis of Simuliidae samples was started. The sequencing reads were classified against the NCBI Nucleotide database (nr) with Kraken2 software [3]. More than 20 % of reads were remained unclassified. The unclassified reads and those attributed to viruses were extracted and assembled with metaSPAdes [4]. The resulting contigs with a coverage of no less than 20 and with a length of 500 bp or more were used for functional analysis. The open reading frames were predicted with Prodigal [5].

Results: The obtained amino acid sequences were analyzed for conserved domains and several genes were proposed to encode proteins specific for nucleopolyhedroviruses belonging to Baculoviridae family known to infect different groups of insects. Many sequences were attributed to bacteriophages of Myoviridae family and unclassified dsDNA viruses and a large number of contigs was still not attributed to any taxon.

Conclusion: Metagenomic studies are very important as they can provide us with better understanding microorganisms diversity, ecology and evolution and also help us to identify new pathogens and their transmission routes. We believe that metagenomic studies of Simuliidae and Ceratopogonidae species are of critical importance. However, the available genomic and transcriptomic data for these insects is scarce, and especially, for Simuliidae family.

Acknowledgements: This work was supported by the Ministry of Science and Higher Education of Russian Federation (agreement No. 075-15-2019-1665).

References

1. Mead D.G. et al. Biological transmission of vesicular stomatitis virus (New Jersey serotype) by *Simulium vittatum* (Diptera: Simuliidae) to domestic swine (*Sus scrofa*). *J Med Entomol.* 2004;41:78-82.
2. Manguin S. et al. Main Topics in Entomology: Insects as Disease Vectors. In: Lopez O., Fernandez-Bolanos J. (Ed.). *Green Trends in Insect Control*. RSC Green Chemistry, 2011:1-52.
3. Wood D.E. et al. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):257.
4. Nurk S. et al. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27(5):824-834.
5. Hyatt D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119.