

# The metagenomic analysis of Colorado potato beetles public NGS data and the nanopore sequencing of its genetic material

Starchevskaya M.E.<sup>1,2\*</sup>, Kamanova E.P.<sup>1,2</sup>, Vyatkin Y.V.<sup>1,2,3</sup>, Tregubchak T.V.<sup>2</sup>, Bauer T.V.<sup>2</sup>, Rotskaya U.N.<sup>4</sup>, Kosman E.S.<sup>4</sup>, Antonets D.V.<sup>1,2</sup>

<sup>1</sup> State Research Center of Virology and Biotechnology “Vector”, Novosibirsk, Russia

<sup>2</sup> Novel Software Systems LLC, Novosibirsk, Russia

<sup>3</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>4</sup> Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia

\* [starchevskayamaria@mail.ru](mailto:starchevskayamaria@mail.ru)

**Key words:** viral metagenomics, Colorado potato beetle, viruses of insects, nanopore sequencing

**Motivation and Aim:** The Colorado potato beetle (*Leptinotarsa decemlineata* from Chrysomelidae family) is one of the most serious insect pests. High ecological plasticity and ability to adapt to a variety of Solanaceae plants has led to its global expansion worldwide [1]. *L. decemlineata* is also known for its ability to rapidly develop resistance to insecticides [2]. Another promising approach to control insect pests is the use of biological insecticides based on viruses [3], but the information on viruses, which infect leaf feeding beetles, is scarce. Baculoviridae and Iridoviridae viruses were proposed to infect Chrysomelidae beetles [4]. Besides, leaf feeding beetles were shown to transmit different plant viruses [5, 6]. However, despite the undoubted importance of Colorado potato beetle virome for agriculture, it has not been studied yet and, as far as we know, our work is the first attempt to discover the viral material in DNA- and RNA-Seq data of *L. decemlineata*. We also hope that identification of new viruses will help to extend the arsenal of biopesticides.

**Methods and Algorithms:** 119 genomic and transcriptomic samples of *Leptinotarsa decemlineata* were selected from the NCBI SRA database. Genomic data obtained from muscle tissue (PRJNA508767, PRJNA369863) and whole insect tissue (PRJNA171749). The transcriptome datasets were obtained from antennas (PRJNA280017), heads (PRJNA400685), intestines (PRJNA400685, PRJNA336167), larvae (PRJNA275431) and whole insect (PRJNA438159, PRJNA384383, PRJNA297027, PRJNAJ427580, PRJNA297027, PRJNAJ42764380, PRJNA297027, PRJNAJ42764380). The quality control of reads was carried out and low-quality sequences were filtered, and possible contamination was eliminated. The reads that were not aligned to *L. decemlineata* reference genome (assembly GCA\_000500325.2 Ldec\_2.0) were assembled using metaSPAdes [7]. More than 3.5 million contigs were obtained, 7726 contigs with a coating of more than 20 and a length of more than 500 bp were selected. The selected contigs were analyzed with BLAST against NCBI nt database. The sequences of unidentified contigs were translated into 6 reading frames and analyzed with BLAST using NCBI nr and NCBI CDD databases. Possible gene sequences were predicted using MetaGeneMark [8] and the obtained gene sequences were also analyzed with BLAST using NCBI nr/nt and CDD databases. Contigs and non-aligned reads were also analyzed with Kraken2 software [9].

**Results:** 4334 contigs were attributed to different taxonomic groups. Among the identified viral families, the most reads were attributed to Baculoviridae, Iridoviridae,

Polydnaviridae, Iflaviridae, Mimiviridae, Pandoraviridae, Phycodnaviridae, and Poxviridae families. Also, full-length genome sequences of several plant viruses were assembled: Potato virus S, and Potato virus Y, Grapevine virus, and Ligustrum virus. The largest number of viral fragments was found in intestinal tissues. We have also found that some genes, predicted with MetaGeneMark, probably belong to bracoviruses. Bracoviruses belong to Polydnaviridae virus family and are associated with parasitoid wasps. Further examination of Colorado potato beetle reference genome revealed numerous long fragments of *Bracovirus* genomic sequences (over 25,000 bp in total) attributed to *Cotesia sesamiae Mombasa bracovirus* and *Cotesia vestalis bracovirus*. We analyzed 18 more *Coleoptera* genomes and found that Colorado potato beetle genome was the fourth most enriched with bracoviral sequences. As far as we know, this is the first report of a polydnavirus found in Chrysomelidae species. Bracoviruses are known to integrate into genomes of their hosts that is why we decided to examine if these sequences could be found in the genetic material extracted from *L. decemlineata* eggs collected by our colleagues from ISEA SB RAS. 2 contigs 1.5–2 kbp long, attributed to *Cotesia sesamiae Mombasa bracovirus*. A were selected and a number of oligonucleotide flanking the contig sequences were designed, and a set of primers for Sanger sequencing. The presence of bracoviral fragments in *L. decemlineata* genetic material was confirmed with PCR and with Sanger sequencing. The genetic material extracted from Colorado potato beetle eggs was sequenced with MinION sequencer (Oxford Nanopore Technologies). Over 1 million reads with a total length of 4,74 Gb were obtained. The average read length was above 4.6 kb and the maximal length was over 1 Mb. The bracoviral genetic sequences were found to be integrated into the genomic DNA of Colorado potato beetle.

**Conclusion:** Metagenomic studies of insects are extremely important. The identified viruses allow us to expand our understanding of the ecology and evolution of both the viruses themselves and their hosts. Discovery of new viruses will potentially expand the arsenal of biocontrol agents.

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

## References

1. Cingel A. et al. Extraordinary adaptive plasticity of Colorado potato beetle: “Ten-striped Spearman” in the Era of Biotechnological Warfare. *Int J Mol Sci.* 2016;17(9):1538.
2. Alyokhin A. et al. Colorado potato beetle resistance to insecticides. *Am. J. Potato Res.* 2008;85:395-413.
3. Cuartas P.E. et al. Novel biopesticide based on Erinnyis ello betabaculovirus: characterization and preliminary field evaluation to control Erinnyis ello in rubber plantations. *Pest Manag Sci.* 2019;75:1391-1399.
4. Lashomb J. et al. *Edovum puttleri* (Hymenoptera: Eulophidae), an egg parasitoid of Colorado potato beetle (Coleoptera: Chrysomelidae): development and parasitism on eggplant. *J. Econ. Entomol.* 1987;80(1):65-68.
5. Wickizer S.L., Gergerich R.C. First report of Japanese Beetle (*Popillia japonica*) as a vector of Southern bean mosaic virus and Bean pod mottle virus. *Plant Dis.* 2007;91(5):637-637.
6. Kersch-Becker M.F., Thaler J.S. Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores. *Oecologia.* 2014;174:883-892.
7. Bankevich A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477.
8. Zhu W. et al. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 2010;38:e132.
9. Wood D.E., Salzberg S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46.