doi 10.18699/bgrs2024-1.1-23

Genome analysis of novel strains of *Cydia pomonella* granulovirus from Russia

Lakhova T.[1, 2, 3]*, Klimenko A.[1, 2, 3], Tsygichko A.[4], Ismailov V.[4], Vasiliev G.[1, 2], Asaturova A.[4], Lashin S.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Federal State Budgetary Scientific Institution "Federal Research Center of Biological Plant Protection", Krasnodar, Russia*

** tlakhova@bionet.nsc.ru*

Key words: *Cydia pomonella* granulovirus; codling moth; genome sequencing; comparative genomic analysis

Motivation and Aim: *Cydia pomonella* is a pest of fruit trees such as apple and pear, as it damages their fruits and thus causes damage to agriculture. *Cydia pomonella* granulovirus (CpGV) is a natural pathogen for *C. pomonella* that is used as a biocontrol agent of the insect population. One of the earliest isolates of CpGV isolated and widely used to control *C. pomonella* populations is the Mexican isolate. However, it has recently been observed that some natural moth populations may exhibit resistance to formulations based on well-known CpGV isolates. Thus, the study of novel CpGV isolates and comparison with known isolates is necessary for identifying promising isolates that could form the basis for highly effective bioinsecticides in the future. In our study, we investigated 16 new strains obtained from infested *C. pomonella* caterpillars collected in Russia (Krasnodar Krai and Rostov Oblast) and 2 strains from Kazakhstan.

Methods and Algorithms: Quality control of raw data was carried out by FastQC and fastp. We used the NC_002816.1 genome from the NCBI RefSeq database as a reference virus genome. BWA is used for mapping our data to the reference genome. Draft assemblies are performed by two assemblers: Spades for *de novo* assembly and MinYS for reference-guided assembly. To reduce possible assembly errors, the Pilon tool is applied to each assembly. To produce a more complete assembly, genome polishing was performed with the Gfinisher tool based on the draft assemblies. Quast is applied after each assembly step to control the quality of the assembly. Genome annotation is performed using the Prokka tool. We compared our genomes to CpGV genomes from NCBI Virus that have complete assemblies using average nucleotide identity (ANI) with FastANI tool. We also used progressiveMauve to examine multiple alignments of our genomes relative to NC_002816.1 to identify gene rearrangement or loss, as well as regions of low similarity. Comparison of gene content with two reference genomes: NC_002816.1 and KM217575 was performed using GeneAPI. In addition, we used CNVpytor to confirm the absence of genes listed in GenAPI that may have resulted from deletions.

Results: We obtained 18 assemblies consisting of four contigs on average. FastANI demonstrated high values (ANI>99 %) against 14 genomes from NCBI Virus. Multiple alignments of the genomes showed large regions with low sequence similarity in the following genomes: BZR GV 2, BZR GV 5, BZR GV 12, BZR GV L-2, BZR GV L-5

and BZR GV L-7. We consider the region to be large if its size is higher than 300 bp in length. GeneAPI showed that some genes are missing. The orf6 and IFEMGEHL_00128 genes were absent in the most isolates. Gene absence was defined if the level of gene sequence similarity was <90 % and coverage <50 %. CNV analysis confirmed deletions for the genes identified as absent in GenAPI and missing genes in progressiveMauve for BZR GV L-4 and BZR GV L-6 as true deletions. We also detected SNPs when we analyzed genes and their products involved in the virosis infectious process.

Conclusion: Comparative genomic analysis of the 18 genomes of the new strains showed high similarity to the reference isolates, but they also have several differences that probably affect the virulence of the strains or the mechanism of interaction with the codling moth. These differences require further studies to understand their impact.

Funding: The study is supported by Russian Science Foundation (No. 23-16-00260).