بیماری‌های گیاهی / جلد ۵۷ / شماره ۱ / سال ۱۴۰۰: ۸۱-۸۵ DOI: 10.22034/IJPP.2021.527978.351

مقاله علمی کوتاه

شناسایی بیوانفورماتیک دو ویروس جدید از پسته ایرانی

موسی محمدی[1*]، احمد حسینی[2] و سعید نصراله‌نژاد[1]

(تاریخ دریافت: ۱۴۰۰/۱/۲۲؛ تاریخ پذیرش: ۱۴۰۰/۱۰/۷)

پسته یکی از گیاهان پر اهمیت اقتصادی بوده و به خانواده Anacardiaceae تعلق دارد. گزارش‌های مرتبط با عوامل ویروسی در این گیاه به ویروئید کوتولگی رازک، آمپلوویروس آ پسته و ویروس ب پسته محدود است. در مطالعه حاضر به منظور بررسی حضور احتمالی ویروس‌ها در پسته ایرانی به واکاوی داده‌های ترنسکریپتومیک عمومی پرداخته شد. دوازده داده با خوانش جفت شده از پایگاه داده SRA قابل دسترس در وب‌سایت NCBI اخذ گردید. آرانِاهای گیاهی پس از همردیف سازی داده‌ها بر روی ژنوم مرجع پسته به وسیله نرم‌افزار Hisat2، حذف گردید. توالی‌های غیر گیاهی به وسیله نرم افزارهای MEGAHIT و Trinity مونتاژ و نتیجهٔ حاصله به منظور شناسایی توالی‌های ویروسی توسط نرم‌افزار BLASTxدر توالی‌های مرجع ویروسی مورد جستجو قرار گرفت. دو ویروس ناشناس متعلق به خانواده Kitaviridae شناسایی و موقتاً ویروس ایکس پسته (MT334618-20) و ویروس وای پسته (MT362605-06) نام گذاری گردیدند. بر اساس ساختار ژنوم، تشابه ژنتیکی و روابط فیلوژنتیک پیشنهاد می‌شود این دو ویروس ناشناخته به ترتیب به عنوان عضوی از جنس‌های Higrevirus و Cilevirus در نظر گرفته شوند. تعداد بالای توالی‌های ویروسی در نمونه‌های ریشه و برگ، احتمال آلودگی سیستمیک ویروسی ایکس پسته را مطرح می‌نماید. براساس بررسی‌های صورت گرفته، این اولین گزارش از توالی ژنومی عامل ویروسی احتمالی بر روی پسته ایرانی می‌باشد.

واژگان کلیدی: RNA-Seq، Kitavirdae، پسته، ویروس، RNA

* مسئول مکاتبات، پست الکترونیکی: m2.musa90@gmail.com

۱. دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گروه گیاه پزشکی، گرگان، ایران.

۲. دانشگاه ولی عصر رفسنجان، گروه گیاه پزشکی، رفسنجان، ایران.

In silico identification of two novel viruses on Iranian pistachio

M. Mohammadi[1], A. Hosseini[2], and S. Nasrollanejad[1]

(Received: 11.4.2021; Accepted: 28.12.2021)

The pistachio is a commercial nut crop that belongs to the family Anacardiaceae. Reports for viral agents of this plant are limited to hop stunt viroid, pistachio ampelovirus A, and pistacia virus B. We investigated of the presence of viruses in Iranian *Pistacia vera* by analyzing publicly available RNA-Seq data. Twelve paired-end datasets were retrieved from the Sequence Read Archive (SRA) database of NCBI. Transcriptomic data were mapped to *P. vera* reference genome using Hisat2 to eliminate plant RNAs. The remaining data were de novo assembled by MEGAHIT and Trinity programs and the resultant contigs subsequently were queried versus viral reference sequences using BLASTx. Two novel viruses related to the family *Kitaviridae* were identified and tentatively named "pistachio virus x" (MT334618-20) and "pistachio virus y" (MT362605-06). Based on the genome structure, sequences similarity, and phylogenetic relationships we propose these novel viruses as possible members of genera *Higrevirus* and *Cilevirus*. High reads number on different plant tissues, root and leaf, of pistachio virus x led us to make a conclusion on possible systemic infection of this virus on *Pistacia vera*. To the best of our knowledge, it is the first report of possible viral agent from Iranian pistachios.
Keywords: RNA-Seq, Kitavirdae, Pistachio, Virus, RNA.

The *Pistacia vera* (pistachio), a commercial nut crop, with the production of more than 1 million tons around the world, is one of the most economically important members of the family Anacardiaceae (Zeng et al. 2019). Despite its great economic value, the viral agents of pistachios are incompletely understood due to the cryptic nature of some viral infections. Reports are limited to very few cases: hop stunt viroid in Tunisia and Iran (Elleuch et al. 2013, Maddahian et al. 2019), pistachio ampelovirus A in the USA (Al Rwahnih et al. 2018), and pistacia virus B (*Emaravirus*) in Turkey (Buzkan et al. 2019). Kerman province is one of the largest pistachio cultivation centers in Iran, the second largest producer in the world after the USA. Previous studies have shown that the flora of this province is the host for unknown plant viruses (Nury et al. 2020) and detecting viruses that infect pistachios is not unexpected. Today, with the assist of high-throughput sequencing methods such as RNA-Seq, the complexity of identifying novel viruses has been greatly reduced. Various studies have demonstrated the ability of these methods in detecting unknown viruses (Buzkan et al. 2019; Ciuffo et al. 2020; Read et al. 2020). As the majority of viruses that are known to infect plants possess RNA genomes, publicly available transcriptome data, often contain virus derived sequences, which can be detected by metagenomic approaches (Gilbert et al. 2019). In this regard, we undertook a bioinformatic analysis of published transcriptomic datasets for *P. Vera* originated from Kerman province to discover possible viral agents.

The transcriptomic datasets were downloaded from the Sequence Read Archive (SRA) database in the NCBI. Data can be accessed through SRR8772755 to SRR8772780 accession numbers or BioProject number of PRJNA526975. According to the metadata and the related publication (Zeng et al. 2019), these data were produced by sequencing transcriptome of Ohadi cultivar under salinity stress in Rafsanjan during 2015–2017. Downloaded SRA files were converted to fastq format using the fasterq-dump (version 2.10.7, SRA Toolkit) (Leinonen et al. 2010). The Fastqc (Version 0.11.9) and Trimmomatic (version 0.36) (Bolger et al.

* Corresponding author's email: m2.musa90@gmail.com
1. Gorgan University of Agricultural Sciences and Natural Resources, Department of Plant Protection, Gorgan, Iran.
2. Vali-e-Asr University of Rafsanjan, Department of Plant Protection, Rafsanjan, Iran.

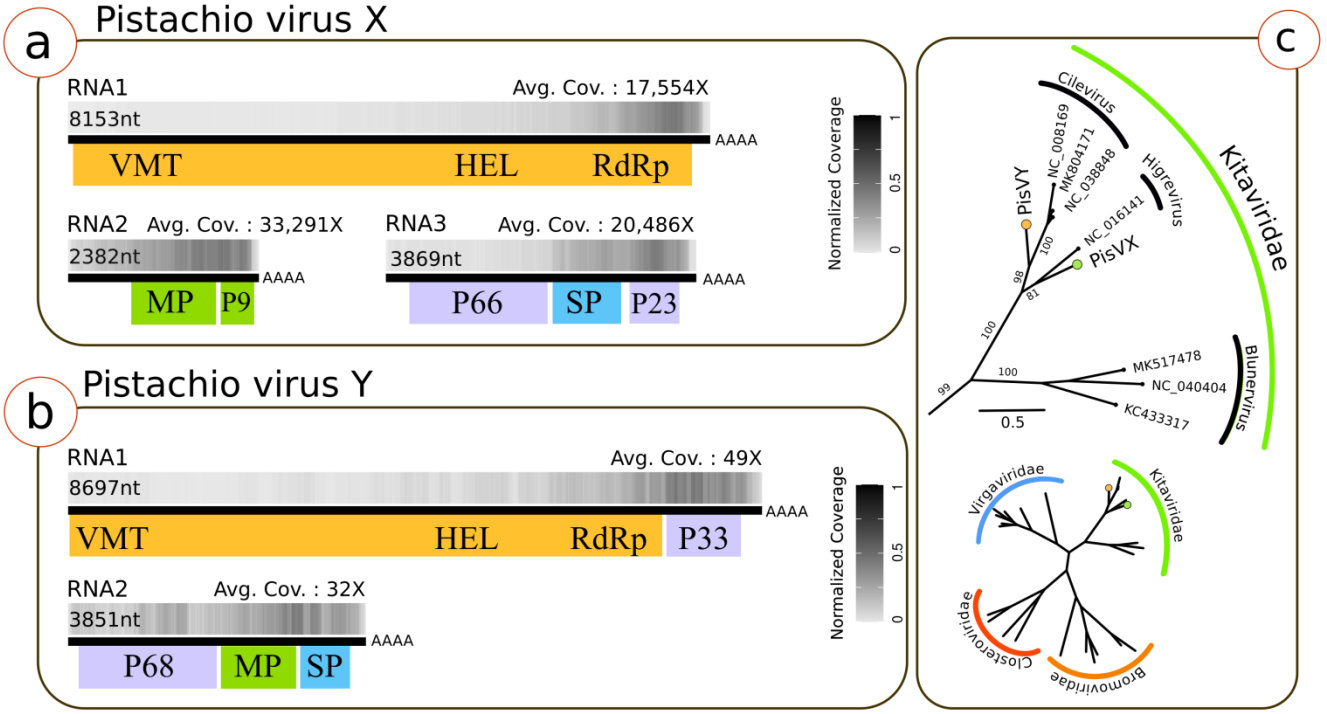

Fig. 1 Schematic representation for the possible genomic structure of a. pistachio virus x (PisVX) b. pistachio virus y (PisVY). c. Unrooted maximum-likelihood tree including PisVX, PisVY and viruses form Kitaviridae, Virgaviridae, Closteroviridae, and Bromoviridae constructed with RaxML-NG with the LG+I+G4+F substitution model and 1000 bootstrap replicates.

2014) programs were used to quality control and filter out low-quality reads. To eliminate plant RNA sequences, paired-end reads were mapped on the *P. vera* reference genome (ID: 55403), downloaded from NCBI by Hisat2 (version 2.2.1) (Kim et al. 2019), and the unmapped reads were captured as none plant reads by Samtools (version: 1.10) (Li et al. 2009). After depletion of plant RNAs, the remaining sequences were de novo assembled by MEGAHIT (version 1.2.9) (Li et al 2015). To detect viral sequences, resultant contigs were queried against the viral reference sequences using BLASTx (version 2.10.1). Additionally, to check the possibility of getting longer contigs, candidate sequences were used as a reference to map back the reads with BWA-mem (version 0.7.17) (Li and Durbin 2009), then output files were used by Trinity (version 2.9.1) (Grabherr et al. 2011) to perform reference-based de novo assembly. Finally, the resulted contigs from the assembly were checked again by BLASTx to detect possible viral sequences.

Generally, in six paired-end datasets related to Iranian pistachios, we identified polyadenylated genome for two putative novel members of the family *Kitaviridae*, tentatively named them “pistachio virus x” (PisVX) (Fig 1 a) and “pistachio virus y” (PisVY) (Fig 1 b). Mapping back the trimmed datasets to PisVX and PisVY indicated that in the best cases a total of 2,029,507 (SRR8772752) and 3,104 (SRR8772753) reads were used to assemble genome segments of PisVX and PisVY, respectively. PisVY only exists on leaf samples with the average per-base coverage of 41X and 26X for RNA1 and RNA2, while PisVX were detected on several samples for root and leaf with an average coverage of 17,554X, 33,291X, and 20,486X for RNA1, RNA2, and RNA3, respectively (Table 1). Excluding the poly-A of 3′-end, PisVX RNA1 (GenBank accession number MT334620) is 8153 nucleotides (nt) in size, analysis with NCBI ORFfinder and Conserved Domain Database (CDD) provided us the evidence for a large ORF (2632 aa) containing conserved domains of Viral methyltransferase (VMT), RNA helicase (HEL) and RNA dependent RNA polymerase (RdRP). To detect possible homologous sequences, the ORF was evaluated versus the NCBI

Table 1. Number of reads mapped to pistachio virus x (PisVX) and pistachio virus y (PisVY) genome using BWA.

		PisVX			PisVY	
Dataset	Tissue	RNA1	RNA2	RNA3	RNA1	RNA2
SRR8772752	leaf	962,714	533,389	533,404	461	161
SRR8772753	leaf	111,569	359,099	151,804	2,428	676
SRR8772754	root	489	1,162	558	-	-
SRR8772755	root	204	200	109	-	-
SRR8772759	leaf	13,588	13,914	10,253	-	-
SRR8772779	leaf	35,942	105,480	48,516	-	-

non-redundant (nr) protein database using BLASTp. As result indicated a protein (YP_004928118.1) related to hibiscus green spot virus 2 with 40.24% identity and 99% query cover is the most identical record to the ORF. RNA2 (MT334619) codes for two ORFs with lengths of 357 and 88 aa. BLASTp showed that they are most related to movement proteins, p39 (40.45%) and p9 (40.74%), of hibiscus green spot virus 2, respectively. The ORFfinder showed that the RNA3 (MT334618) contains three putative ORFs, two (p66 and p23) without significant similarity to any known protein, and one (p31) codes for conserved virion membrane domain (SP24) homologous to structural protein (YP_004928121) of hibiscus green spot virus 2 with 35.83% identity. PisVY RNA1 (MT362606) encodes two ORFs with lengths of 2484 aa and 311 aa. The ORF1 with 38.21% identity to polymerase (QFU28430) of passion fruit green spot virus, codes for conserved domains of VMT, HEL, and RdRP. Also, the ORF2 is homologous to citrus leprosis virus C p29 protein (AKJ79133.1) with 31.37% identity. PisVY RNA2 (MT362605) contains three putative ORFs, the first of which is 597 aa long and with unknown function, second ORF (325 aa) with 44.83% identity is homologous to citrus leprosis virus C (ABG33782) and codes for 3A conserved RNA movement domain. The ORF3 is 203 aa long with SP24 conserved domain and as BLASp indicates, it is close to structural protein (QIH54362) of passion fruit green spot virus with 56.91% identity. To investigate for taxonomic relationships, we performed a phylogenetic analysis with the RdRP domain of PisVX and PisVY alongside some reference sequences of *Kitaviridae*, *Bromoviridae*, *Virgaviridae*, and *Closteroviridae*. For reconstructions of the unrooted maximum likelihood tree, all amino acid sequences were aligned using MAFFT (version 7.453) and subjected to RaxML-NG (version 1.0.1) (Kozlov et al. 2019) with the LG+I+G4+F substitution model and 1000 bootstrap replicates. The substitution model was predicted with ModelTest-NG (Darriba et al. 2020). According to the ML tree, both viruses are related to *Kitaviridae*, PisVX groups with a higrevirus (HQ852052) and PisVY share a branch with cileviruses (Fig 1 C).

Structurally speaking, RNA1 of PisVX and PisVY are quite similar to higreviruses and cileviruses, respectively. However, analyzing the genome organization of PisVX indicated some difference with the higreviruses, the RNA2 lacks triple gene block which serves as the MPs of the virus (Melzer et al. 2012), nonetheless, a recent study has proposed that proteins p39 and p9 of higreviruses, are sufficient for virus movement and could be as binary movement block (Quito-Avila et al. 2020). In the case of RNA3, the organization of ORFs is completely different from higreviruses, but both share the presence of structural protein. Also, some differences could be found in RNA2 of PisVY when comparing to cileviruses, however, both conserved domains of movement and structural protein are present as other cileviruses. The family *Kitaviridae* was recently accepted by ICTV as a monophyletic group with three genera *Blunervirus*, *Cilevirus*, and *Higrevirus*. As proposed by the ICTV Kitaviridae study group, to distinguish a new species the aa identity of polymerase (RNA1) should be less than 75%, which in both cases, PisVX and PisVY, are well satisfied with 40.24% and 38.21% identities, respectively. Also, in both cases difference in structural of genome organization could be detected in comparison with cileviruses, and higreviruses. Based on this information we propose that PisVX and PisVY could be considered as possible novel members of the genera *Higrevirus* and *Cilevirus*, respectively. To the best of our knowledge, it is the first genomic sequence report for possible viral agent from Iranian pistachio. Great sequencing coverages on different samples for PisVX led us to make a conclusion on possible systemic infection of this virus on *Pistacia vera*, which could not be considered in the case of PisVY. We should note that this report concludes on presence of two novel viruses on Iranian pistachio. Thus, to achieve complete viral genomic sequences application of the RACE technique

will be necessary. Also, more research is needed to study plant symptoms caused by these isolates and identification of possible vectors of two viruses

منابع

Al Rwahnih M. Rowhani A. Westrick N. Stevens K. Diaz-Lara A. Trouillas FP et al. 2018. Discovery of viruses and virus-like pathogens in pistachio using high-throughput sequencing. Plant Disease 102(7):1419–25.

Bolger A. M. Lohse M. Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics btu170.

Buzkan N. Chiumenti M. Massart S. Sarpkaya K. Karadağ S. Minafra A. 2019. A new emaravirus discovered in Pistacia from Turkey. Virus Research 263:159–63.

Elleuch A. Hamdi I. Ellouze O. Ghrab M. Fkahfakh H. Drira N. 2013. Pistachio (Pistacia vera L.) is a new natural host of Hop stunt viroid. Virus Genes 47(2):330–7.

Ciuffo M. Kinoti W. M. Tiberini A. Forgia M. Tomassoli L. Constable F. E. et al. 2020. A new blunervirus infects tomato crops in Italy and Australia. Archives of Virology 165(10):2379–84.

Darriba D. Posada D. Kozlov A. M. Stamatakis A. Morel B. Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37(1):291–4.

Gilbert K. B. Holcomb E. E. Allscheid R. L. Carrington J. C. .2019. Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. PLoS ONE 14:1–51.

Grabherr M. G. Haas B. J. Yassour M. Levin J. Z. Thompson D. A. Amit I. et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7):644–52.

Kim D. Paggi JM. Park C. Bennett C. Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology 37(8):907-15.

Kozlov A. M. Darriba D. Flouri T. Morel B. Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35(21):4453–5.

Leinonen R. Sugawara H. Shumway M. 2010. International Nucleotide Sequence Database Collaboration. The sequence read archive. Nucleic acids research 39:19-21.

Li D. Liu C. M. Luo R. Sadakane K. Lam T. W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–6.

Li H. Handsaker B. Wysoker A. Fennell T. Ruan J. Homer N. et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–9.

Li H. Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–60.

Maddahian M. Massumi H. Heydarnejad J. Hosseinipour A. Khezri A. Sano T. 2019. Biological and molecular characterization of hop stunt viroid variants from pistachio trees in Iran. Journal of Phytopathology. 167: 163–173.

Melzer M. J. Sether D. M. Borth W. B. Hu J. S. 2012. Characterization of a virus infecting Citrus volkameriana with citrus leprosis-like symptoms. Phytopathology, 102(1):122–7.

Nury S, Hosseini A, Gibbs AJ, Mohammadi M (2020) Poison hemlock virus Y (PHVY), a novel potyvirus from Iranian Conium maculatum (Apiaceae). Australasian Plant Pathology 49(2):119–26.

Read D. A. Muoma J. Thompson G. D. 2020. Metaviromic analysis reveals coinfection of papaya in western Kenya with a unique strain of Moroccan watermelon mosaic virus and a novel member of the family Alphaflexiviridae. Archives of Virology 165(5):1231–4.

Quito-Avila D. F. Freitas-Astúa J. Melzer M. J. 2020. Bluner-, Cile-, and Higreviruses (Kitaviridae). Reference Module in Life Sciences 1–5.

Zeng L. Tu XL. Dai H. Han FM. Lu BS. Wang M. S. et al. 2019. Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. Genome Biology 20(1):1–13.