

Molecular detection, and partial nucleotide sequencing of *Cassava common mosaic virus* on pistachio in Iran

Ahmad Hosseini (PhD)^{1*}, Musa Mohammadi (MSc)¹

¹ Department of Plant Protection, College of Agriculture, Vali-e-Asr University of Rafsanjan

Information	Abstract
Article Type: Original Article	Background and Objectives: The pistachio is a semitropical dioecious tree cultivated primarily for its nuts. Despite its agricultural significance, comprehensive screening for plant virus diseases affecting pistachios has been lacking. This study undertakes a bioinformatic analysis of published transcriptomic datasets from the Sequence Read Archive (SRA) within the NCBI database, specifically focusing on samples from Kerman province, to identify potential viral agents that may infect pistachio trees. Materials and Methods: Transcriptomic data were aligned with the Pistachio vera reference genome to filter out plant RNAs. The resulting sequences were assembled into a contiguous format, refined, and subsequently compared with NCBI GenBank using a BLASTn search. To explore the phylogenetic relationships of the virus, multiple sequence alignments were conducted based on the complete genomic sequences, as well as the replicate and coat protein open reading frames (ORFs). Results: Bioinformatic analysis identified a member of the potexviruses in a pistachio sample. The genomic RNA sequence of an isolate (Pis), consisting of 4635 nucleotides, was determined. The virus genome is structured with five open reading frames and includes the complete 3' untranslated region (3' UTR), the coat protein gene, a triple gene block (TGB), and a partial RNA-dependent RNA polymerase (RdRp), reflecting the typical arrangement found in members of the genus Potexvirus. Sequence alignment and nucleotide BLAST analysis indicated that the Pis isolate shares 88% to 94% identity with the complete genome sequence of Cassava common mosaic virus (CsCMV) isolates. Conclusion: According to the sequence and phylogenetic analysis, the pistachio isolate of CsCMV was closely related to other CsCMV isolates. This is the first report of CsCMV in <i>P. vera</i> .
Article History: Received: 11.09.2023 Accepted: 23.12.2023	
Doi: 10.22123/PHJ.2025.463489.1171	
Keywords: Pistachio Cassava common mosaic virus High throughput sequencing Bioinformatics	
Corresponding Author: Ahmad Hosseini Email: hosseini@vru.ac.ir Tel: +98-9171120427	

► Please cite this article as follows:

Hosseini A, Mohammadi M. Molecular detection, and partial nucleotide sequencing of Cassava common mosaic virus on pistachio in Iran. *Pistachio and Health Journal*. 2023;6(3-4):55-68.

1. Introduction

Pistachio (*Pistacia vera*) belongs to the family Anacardiaceae and is native to Central Asia. Pistachio is regarded as the most economically significant plant due to its production of edible nuts, often referred to as the "King of Nuts" [1]. The global annual value of pistachio nuts is estimated at approximately USD ten billion, with Iran being one of the world's largest producers, contributing a wholesale value of around USD four billion (FAO, 2022). There are more than 500,000 hectares of pistachio orchards in Iran and Rafsanjan (Kerman province), with 65,000 hectares is considered as the most important pistachio production pole in the country (FAO, 2018).

The main method of commercial pistachio propagation is grafting; this method favors the disseminating of various pests and diseases, including viruses and viroids. Although pistachio is known as an important nut species worldwide, this crop has not been undergone widespread screening for plant virus diseases. Few viruses have been reported in pistachio orchards across different regions of the world. The first documented viral disease affecting pistachios is Pistachio rosette virus, initially identified in Turkmenistan on non-domesticated trees. Subsequently, this virus was detected in central Asia including Iran [2]. Symptoms associated with Pistachio rosette virus include tree stunting and the formation of rosettes, characterized by deformed twigs with large nodes and shortened internodes. Additionally, a virus known as Pistachio ampelovirus A (PAVA) and a viroid called Citrus bark cracking viroid were identified in pistachio trees using high-throughput sequencing (HTS) [2]. Similarly, in a separate study conducted in Turkey, HTS analysis of pooled RNA from six symptomatic

Pistacia trees revealed contigs with similarities to emaravirus RNAs [3]. However, the observed symptoms could not be definitively linked to the detected virus.

There is little information about the status of viral diseases of pistachio in Iran and reports are limited to very few cases: Maddahian et al. [4] have reported *Hop stunt viroid* in some regions of Kerman province. Phylogenetic studies have demonstrated that the presence of HSVd in pistachio trees has a long-standing history in Iran. Furthermore, the Tunisian HSVd variants found in pistachio likely trace their origins to Iranian sources [4]. Mohammadi et al. [5] identified two novel viruses associated with the family *Kitaviridae* through the analysis of RNA-Seq data obtained from the Sequence Read Archive (SRA) database of NCBI. These viruses were provisionally named Pistachio virus X and Pistachio virus Y. *Cassava common mosaic virus* (CsCMV), a member of the genus *Potexvirus* within the family *Alphaflexiviridae*, is characterized by its +ssRNA genome. Initially reported in southern Brazil, CsCMV is widespread across Latin America and has recently been detected in China [6]. Similar to other potexviruses, CsCMV possesses a monopartite, positive-sense, single-stranded RNA genome that is 6,395 nucleotides (nt) in length, excluding the 3'-poly-A tail. The genome contains five open reading frames (ORFs) that encode an RNA-dependent RNA polymerase (RdRp), three triple gene block (TGB) proteins, and a coat protein (CP) [7].

CsCMV causes mosaic leaf symptoms and leaf blade distortion, especially in young leaves; These symptoms result in a progressive decline in chloroplast functionality and a disruption of

carbon flux in source leaves [8]. Although Similar to other members of the *Potexvirus* genus, *CsCMV* lacks a known biological vector, but it infects host tissues easily and can persist in the propagation material and in agricultural tools, which contributes to its dispersal [9].

In recent decades, advances in plant virology and next generation sequencing technologies have facilitated the identification of novel viruses and the assembly of complete viral genomes. Given the growing interest in identification of viral disease of pistachio in Iran, this study was carried out to ascertain the virus status of this tree. We conducted a bioinformatic analysis of published transcriptomic datasets for pistachio collected from Kerman province to detect the possible presence of viral agents. This idea arose from the fact that publicly available transcriptome data, often contain virus-derived sequences. On the other hand, the majority of viruses that are known to infect plants, possess RNA genomes, which can be detected by metagenomic approaches [10].

In this study, the partial nucleotide genome sequence of an isolate of *CsCMV* which presumably infects *Pistacia vera* assembled, described, and characterized. In addition, the sequence of this isolate and other known genomes, including those of its closely related species compared and analyzed to determine their phylogenetic relationships. The results of the present study can provide good information about the phylogenetic relationship of the isolate.

2. Material and methods

In the current investigation, a bioinformatic analysis of existing transcriptomic datasets related to *Pistacia vera* from Kerman Province was performed to identify potential viral agents associated with pistachio infection.

Transcriptomic datasets from the Sequence Read Archive (SRA) database in the NCBI was retrieved. The data can be accessed using the accession numbers SRR8772755 to SRR8772780 or the BioProject identifier PRJNA526975. As reported by Zeng et al. [11], these datasets were generated through transcriptome sequencing of the Ohadi cultivar in Rafsanjan, conducted between 2015 and 2017. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA) following the manufacturer's recommendations and then sequenced on the Illumina HiSeq X Ten platform to generate 150 bp paired-end reads [11]. Among all these datasets, we focused on data belonging to the transcriptome of 26 (accession number SRR8772759) retrieved from BioSample SAMN11242392. The transcriptomic data of this BioSample was generated from an individual pistachio tree labeled Ohadi2 (Fig. 1), which is located in a plot at the collections of germplasm within the Pistachio Research Institute (PRI) in Rafsanjan, Iran. This tree although did not show any specific viral leaf symptoms such as mosaic or mottling but the symptoms including dieback, gum exudation, and decline were very apparent.



Fig.1. Image of pistachio tree labeled as Ohadi2 which transcriptome of 26 was retrieved from its total RNA (BioSample [SAMN11242392](https://www.ncbi.nlm.nih.gov/biosample/SAMN11242392), accession number SRR8772759). This tree is located in the germplasm collections of the Iranian Pistachio Research Institute (Rafsanjan, Iran) and shows symptoms including dieback, yellowing, and decline.

Fasterq-dump (SRA Toolkit, version 2.10.7) was utilized to convert downloaded SRA files to fastQ format [12]. The raw reads were filtered to remove the failed or low-quality reads using the Fastqc (Version 0.11.9) and Trimmomatic (version 0.36) programs [13]. Qualified reads were mapped on the *P. vera* reference genome (ID: 55403), downloaded from NCBI using Hisat2 (version 2.2.1) [14], to filter out plant RNA sequences. The remaining sequences after removing the reads of plant RNAs, were assembled de novo into contigs using MEGAHIT (version 1.2.9) [15]. Contigs were

annotated by BLASTx comparisons to Viruses NR and Viroids databases downloaded from NCBI GenBank databases. Multiple contigs derived from partial genomes and ORF contigs corresponding to CsCMV were obtained.

To determine the phylogenetic relationships among the isolates, complete genome sequences of viruses closely related to our isolate, including *CsCMV* isolates and other closely related potexviruses, were retrieved from the NCBI database (www.ncbi.nlm.nih.gov) (Table1).

Table 1. List of potexvirus species used for phylogenetic relationships. NCBI acc. numbers, virus names abbreviations, origins and plant hosts are shown.

Virus name	Abbreviation	Acc. No.	host
<i>Bamboo mosaic virus</i>	BaMV	NC_001642	<i>Bambusa oldhamii</i>
<i>Cassava common mosaic virus</i>	CsCMV	NC_001658	<i>Manihot esculenta</i>
<i>Alstroemeria virus X</i>	AlsVX	NC_007408	Alstroemeria
<i>Cactus virus X</i>	CVX	NC_002815	<i>Hylocereus undatus</i>
<i>Clover yellow mosaic virus</i>	CIYMV	NC_001753	<i>Trifolium repens</i>
<i>Cymbidium mosaic virus</i>	CymMV	NC_001812	<i>Cymbidium aloifolium</i>
<i>Foxtail mosaic virus</i>	FoMV	NC_001483	<i>Hordeum murinum</i>

<i>Hydrangea ringspot virus</i>	HdRSV	NC_006943	<i>Hydrangea</i> sp.
<i>Lily virus X</i>	LVX	NC_007192	<i>Lilium candidum</i>
<i>Mint virus X</i>	MVX	NC_006948	<i>Mentha spicata</i>
<i>Narcissus mosaic virus</i>	NMV	NC_001441	<i>Narcissus poeticus</i>
<i>Opuntia virus X</i>	OVX	NC_006060	<i>Opuntia</i> sp.
<i>Papaya mosaic virus</i>	PapMV	NC_001748	<i>Carica papaya</i>
<i>Pepino mosaic virus</i>	PepMV	NC_004067	<i>Solanum lycopersicum</i>
<i>Plantago asiatica mosaic virus</i>	PIAMV	NC_003849	<i>Plantago</i> sp.
<i>Potato aucuba mosaic virus</i>	PAMV	NC_003632	<i>Solanum tuberosum</i>
<i>Potato virus X</i>	PVX	NC_001455	<i>Solanum tuberosum</i>
<i>Scallion virus X</i>	ScaVX	NC_003400	<i>Allium chinense</i>
<i>Tulip virus X</i>	TVX	NC_004322	<i>Tulipa</i> sp.
<i>White clover mosaic virus</i>	WCIMV	NC_003820	<i>Trifolium repens</i>
<i>Zygocactus virus X</i>	ZVX	NC_006059	<i>Schlumberger a</i> sp.
<i>Cassava common mosaic virus</i>	CsCMV	KT002435	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	MN428639	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	MN243731	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	KT002439	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	MT279196	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	MT038420	<i>Manihot esculenta</i>
<i>Cassava common mosaic virus</i>	CsCMV	MW175326	<i>Manihot esculenta</i>

Table 2. Percent of nucleotide identity between Pis isolate and some selected members of the genus *Potexvirus* based on complete nucleotide sequence.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
1	KT002435		98	94	91	91	89	89	89	89	89	54	53	54	48	45	48	47	48	48	44	44	45	44	40	41	41	40	41	42	43	
2	NC_001658	98		93	90	91	88	88	88	88	89	53	52	53	47	44	48	47	47	47	43	44	45	44	40	40	40	39	40	41	42	
3	MT334616	94	93		91	91	88	88	88	88	88	56	55	55	50	49	51	50	51	49	45	45	47	46	42	43	43	42	42	43	44	
4	MN243731	91	90	91		99	89	89	89	89	89	54	54	54	48	45	48	48	47	48	44	44	45	44	40	41	41	40	40	42	43	
5	MN428639	91	91	91	99		89	89	89	89	89	54	53	54	48	45	48	48	47	48	44	44	45	44	40	41	41	39	40	42	43	
6	MW175326	89	88	88	89	89		98	94	94	90	54	53	54	48	45	48	48	47	48	44	44	45	45	40	40	41	39	40	41	42	
7	OM927720	89	88	88	89	89	98		95	94	90	54	53	54	48	45	48	48	47	48	44	44	45	45	40	40	41	39	40	42	42	
8	MT279196	89	88	88	89	89	94	95		93	90	54	53	54	48	45	48	48	48	48	43	44	45	45	40	41	41	40	40	42	42	
9	MT038420	89	88	88	89	89	94	94	93		90	54	53	53	48	45	48	48	47	47	44	44	44	45	40	40	41	39	40	42	42	
10	KT002439	89	89	88	89	89	90	90	90	90		54	53	54	48	45	48	48	47	48	43	44	45	44	41	41	41	39	40	41	42	
11	NC_003849	54	53	56	54	54	54	54	54	54	54		69	58	46	45	47	48	46	46	45	45	47	46	40	42	42	38	42	40	43	
12	NC_004322	53	52	55	54	53	53	53	53	53	53	69		57	45	44	47	47	46	45	44	44	47	45	39	41	41	39	42	41	44	
13	NC_006943	54	53	55	54	54	54	54	54	53	54	58	57		48	46	48	48	48	47	45	46	47	46	40	41	42	38	41	41	43	
14	NC_001748	48	47	50	48	48	48	48	48	48	48	46	45	48		53	50	50	50	45	40	41	41	41	39	40	39	40	39	39	39	
15	NC_001753	45	44	49	45	45	45	45	45	45	45	44	46	53		50	50	49	43	38	40	40	39	41	41	42	39	38	39	39		
16	NC_002815	48	48	51	48	48	48	48	48	48	48	47	47	48	50	50		72	63	45	41	41	42	42	39	40	41	41	39	39	42	
17	NC_006059	47	47	50	48	48	48	48	48	48	48	48	47	48	50	50	72		64	45	42	42	43	42	40	41	41	40	39	39	42	
18	NC_006060	48	47	51	47	47	47	47	48	47	47	46	46	48	50	49	63	64		45	42	42	42	42	40	40	42	40	39	40	41	
19	NC_001455	48	47	49	48	48	48	48	48	47	48	46	45	47	45	43	45	45	45		41	43	44	43	40	40	40	40	40	42	43	
20	NC_001483	44	43	45	44	44	44	44	43	44	43	45	44	45	40	38	41	42	42	41		56	44	42	37	38	38	36	40	39	40	
21	NC_001642	44	44	45	44	44	44	44	44	44	44	45	44	46	41	40	41	42	42	43	56		43	41	38	38	38	38	39	39	40	
22	NC_006948	45	45	47	45	45	45	45	45	44	45	47	47	47	41	40	42	43	42	44	44	43		59	39	40	41	38	42	41	45	
23	NC_007192	44	44	46	44	44	45	45	45	45	44	46	45	46	41	39	42	42	42	43	42	41	59		39	39	40	38	42	41	45	
24	NC_001441	40	40	42	40	40	40	40	40	40	41	40	39	40	39	41	39	40	40	40	37	38	39	39		64	58	50	48	49	43	
25	NC_003400	41	40	43	41	41	40	40	41	40	41	42	41	41	40	41	40	41	40	40	38	38	40	39	64		59	50	48	49	44	
26	NC_007408	41	40	43	41	41	41	41	41	41	41	42	41	42	39	42	41	41	42	40	38	38	41	40	58	59		49	48	50	44	
27	NC_003632	40	39	42	40	39	39	39	40	39	39	38	39	38	40	39	41	40	40	40	36	38	38	38	38	50	50	49		47	48	42
28	NC_001812	41	40	42	40	40	40	40	40	40	40	42	42	41	39	38	39	39	39	40	40	39	42	42	48	48	48	47		55	48	
29	NC_004067	42	41	43	42	42	41	42	42	42	41	40	41	41	39	39	39	39	40	42	39	39	41	41	49	49	50	48	55		47	
30	NC_003820	43	42	44	43	43	42	42	42	42	42	43	44	43	39	39	42	42	41	43	40	40	45	45	43	44	44	42	48	47		

Multiple alignment of genome sequences was performed using the Geneious Prime 2019.1.3 software, with the sequence identity matrix calculated using the default setting. Phylogenetic trees for all identified viruses were constructed utilizing both the neighbor-joining and Maximum Likelihood methods in MEGA 7.0, with confidence levels for the resulting phylogroups assessed through 1000 bootstrap replicates [16].

3. Result

Sequence analysis and genomic organization

The analysis of HTS (High-Throughput Sequencing) data from BioSample SAMN11242392 revealed several contigs. The resulting contigs from the assembly were examined using BLASTx to identify potential viral sequences. In total, six datasets of paired-end reads pertaining to Iranian pistachio plants, a genome with polyA tail, corresponding to potexviruses sequences was identified.

A partial sequence of 4635 nts with a GC content of 47.5% was determined for the

pistachio isolate (Pis) with five ORFs analogous to those found in other potexviruses. This sequence has been deposited in database of the GenBank under the accession number MT334616.

Utilizing the nearly complete viral genome sequence, closely related sequences were identified in GenBank through BLASTn and BLASTx searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The obtained sequence had about 99% coverage and shared 94.24% nt sequence identity with the Colombian isolate of CsCMV genome isolated from *Manihot esculenta* (Acc. no. KT002435).

Sequence comparison revealed a high identity between the Pis isolate and other isolates of CsCMV, so the sequence of CsCMV refseq (Acc. no. NC_001658) was considered a reference for our isolate.

The organization of Pis isolate genome was found to be similar to that of CsCMV and other members of Potexvirus (fig 2).



Fig. 2. Genomic organization of Pis isolate of CsCMV. Replicase is represented by the blue box, triple gene block proteins 1, 2, and 3 by the pink, red, and dark red boxes, respectively, and coat protein by the green box. HEL: helicase; RdRP: RNA-dependent RNA polymerase; CP: coat protein. The size and overlap of proteins are proportional to the actual size.

Partial ORF1 (positions 1 to 2720) which contains some of the RNA polymerase motifs, presumably encodes part of the replication-associated protein, ORF2 (positions 2723 to 3718) encodes a polypeptide of 232 aa, ORF3 (positions 3384 to 3716) overlaps ORF2 by 35 nucleotides and encodes a polypeptide of 111 aa, ORF4 (positions 3571 to 3858) overlaps ORF3

by 146 nucleotides and encodes a polypeptide of 96 aa. ORF2, ORF3, and ORF4, organized in an overlapping manner, encode for the triple gene block (TGB) proteins 1, 2, and 3, respectively which involved in viral movement. ORF5 (positions 3861 to 4550) encodes a polypeptide of 230 aa. This last ORF encodes the putative coat protein.

The genome of the Pis isolate features two intercistronic regions. The first intercistronic region, located between ORF1 and ORF2, consists of two nucleotides (CC), differing from the two nucleotides (CU) found in some other isolates of CsCMV. The second intercistronic region, situated between ORF4 (TGB3) and ORF5 (the coat protein), also consists of two nucleotides (CG). The encoded proteins of ORF2, ORF3, and ORF4 of Pis isolate exhibit similarities to the products of the triple gene block found in CsCMV isolates as well as other potexviruses, and highlighting the presence of this conserved gene module across different viral species. This conservation suggests important functional roles in viral movement and replication. The analysis of the aa sequence of the ORF1 of the several potexviruses, including CsCMV, revealed that there are two regions, the N-terminal region (the first 400 aa) and the C-terminal region (about 850 aa). These two domains are separated by a region of 50 to 359 aa with low homology [17]. Conserved sequence motifs associated with RNA-dependent RNA polymerase (RdRp or POL, characterized by the GDD motif), helicase (HEL, featuring the GKS motif), and methyltransferase (MTR) are commonly present in ORF1 (replicase) of numerous positive-sense RNA viruses [18]. The mentioned domains are typically well conserved and observed within replicase protein of main members of Potexvirus genus. The RdRp of Pis isolate consisted of 905 amino acids. This demonstrated that the 545 amino acids corresponding to N-terminal region is missing when compared to ORF1-encoded RdRp of CsCMV refseq. For this reason, we expected that the methyltransferase motifs was not found in the N-terminal region of the ORF1. The conserved motifs of methyltransferase of the Sindbis-like virus are observed in the amino-end

domain of most potexviruses. The NTP-binding motif of helicase (GXXGKGKS/T) and the motif of RNA polymerase consensus (FLGTLXIMRLXGEXXTFDANTEXXIA) were identified as highly conserved among potexviruses. NTP-binding helicase motif, located in amino acids 190-197 was observed as GAGGSGKS and RNA polymerase consensus motif was found in amino acids 738-763 of Pis isolate RdRp.

ORFs 2, 3, and 4 were assumed to code for proteins of 232 amino acids (aa), 111 aa, and 96 aa, corresponding to molecular weights of approximately 25 kDa (25K protein), 14 kDa (14K protein), and 9 kDa (9K protein), respectively. All three proteins (encoded by ORFs 2, 3, and 4) share significant amino acid similarities with the corresponding proteins from other strains of CsCMV. The genomic arrangement of the Pis isolate is comparable to these strains, featuring overlapping ORFs that are consistent with the typical configuration of triple gene block proteins involved in cell-to-cell movement. This similarity highlights the conserved mechanisms of viral dissemination within host plants across different strains.

ORF2 (nts 2723 to 3418) encodes the first TGB protein, TGB1 (25kDa). The protein encoded by ORF2 contains a typical NTPase/helicase domain, which has been demonstrated to have activity in Potato virus X (PVX). This domain is characterized by seven conserved motifs, two of which are likely involved in nucleotide triphosphate (NTP) binding. TGB1 falls within the superfamily I of RNA helicases, highlighting its functional importance in viral replication and movement. Additionally, this protein may also play a role in the inhibition of RNA silencing, as suggested by Cote et al. [19], aiding the virus in evading host defenses. ORF3 (nts 3384 to 3716) overlaps 35 nucleotides with the 3'-end of ORF2 and extends 146 nts past the start codon of

ORF4. This ORF codes for TGB2 protein which features a consensus motif specific to potexviruses:

PxxGDxxHxL/FPxGGxYxDGTKxxxY. This motif is located in amino acids 39-63 of the Pis isolate and observed as PHVGDNIHALPHGGYYRDGTKTVYY.

ORF4 (nucleotides 3571 to 3858) encodes the third triple gene block (TGB) protein, TGB3. This small protein, with a molecular weight of 9 kDa, exhibits the highest degree of variability among potexviruses. It features a conserved CxV/IxxxG motif, which is a consensus sequence found in TGB3 proteins across the *Potexvirus* genus. This motif was found as CLIIDG (aa 57-63) of Pis isolate TGB3. Our analysis of the nucleotide sequence of ORF4 in CsCMV isolates having full genome sequence revealed that most of these isolates including Pis genome have six-base deletion as GGGCCC in ORF4 (TGB3) when compared to the CsCMV refseq (Acc. no. NC_001658) genome.

The effect of TGB3 in viral symptom induction and cell-to-cell movement has been documented through studies using infectious clones of the potexvirus Bamboo mosaic virus (BMV). Research by Wu et al. demonstrated that various mutations in TGB3 significantly impacted the virus's ability to move between cells during mechanical transmission to *Nicotiana benthamiana* [20]. In a separate study, the absence of TGB3 resulted in ineffective or limited systemic and cell-to-cell propagation of Potato virus X, confining the movement primarily to the initially infected leaves. Additionally, a decrease in the intensity of local symptoms was observed in these leaves [21]. In Potato virus X (PVX), the TGB2 and TGB3 proteins are linked to cell walls and membranes, primarily serving to regulate the activity of TGB1 [19]. Amino acid sequence Alignment of

the ORF5 and CP protein of ten CsCMV isolates shows their high homology. The coat protein (CP) of potexviruses plays a dual role in genome protection and facilitating viral movement within the host [22]. Motif AFAGFDFFDGVMNVAA which is well-known among potexvirus and carlaviruses was observed in the coat protein of Pis isolate genome (aa 150-165). This motif includes a conserved sequence with amphipathic core, which is believed to facilitate the binding of RNA to the coat protein through hydrophobic interactions in potexviruses [23]. An alignment of the amino acid sequences from the coat proteins of filamentous plant viruses indicated that these segments may play a role in forming salt bridges [22]. Specifically, the conserved residues with positive charge (arginine) and negatively charged residues (aspartic acid) are thought to form a salt bridge that is essential for maintaining the structural integrity of the protein [18]. The coat proteins of Pis isolate contained conserved residues (RQ-37X-D) between amino acids 116 and 155 that have been associated with the putative salt bridge and corresponding hydrophobic core.

The 85 nt long 3'-UTR (nts 4551-4635) which followed by the poly(A) tail, is predicted to adopt a tRNA-like secondary structure, analogous to those found in other potexviruses [19]. The 3' non-coding region of RNA contains some motifs conserved among potexviruses. The first motif comprises a conserved region featuring the pentanucleotide sequence ACTTA, which has been demonstrated to play a critical role in the replication of RNA in certain potexviruses [24]. The motif ACUUA is another sequence which is also observed in the 3' non-coding region of all members of potexvirus genus. This motif is recognized as a cis-acting element that plays a role in the

synthesis of RNA viruses with positive and negative senses. Notably, the AATAAA consensus sequence typically associated with polyadenylation sites was absent in the 3'-UTR region of the RNA in the Pis isolate.

Phylogenetic analysis

To explore the phylogenetic relationship between Pis isolate and some other potexviruses (including all CsCMV isolates with full genome sequence), multiple sequence alignments were conducted utilizing the ClustalW algorithm implemented in the Geneious Prime 2019.1.3 program based on the CP, replicase ORFs and full genomic sequences.

Multiple alignments using the partial nucleotide sequence of the Pis isolate and other isolates of CsCMV showed 94% similarity with the Colombian isolate of MCol22 (KT002435), 93% with Brazilian isolate (NC_001658 or U23414), 91% with two Chinese isolates (MN428639, MN243731) and 88% similarity with the remaining isolates of CsCMV.

Comparison of sequence between the Pis isolate genome and its nearest isolates of CsCMV revealed nucleotide similarities from 88-95% for the coat protein and 88-94% for polymerase. Although the coat protein sequences show some degrees of divergence, which are within the range of intra-species variations (more than 72% nt identity) prescribed for potexviruses [25].

Coat protein sequence identities between Pis isolate and the other potexviruses (except CsCMV) revealed nucleotide identities of 36% for *Bamboo mosaic virus* (BaMV) to 57% for *Plantago asiatica mosaic virus* (PIAMV). Pairwise sequence comparisons based on replicase ORF between Pis isolate and these viruses showed nucleotide identities ranging from 47% for *Potato aucuba mosaic virus* (PAMV) and *Narcissus mosaic virus* (NMV) to 59% for *Tulip virus X* (TVX) and PIAMV.

The determination of near complete sequence of Pis isolate revealed that it shares over 90% nucleotide sequence homology with most CsCMV isolates, allowing it to be classified as a member of the CsCMV species. The phylogenetic analysis between Pis isolate and other viruses on the basis of the full genomic sequences, replicase, and coat protein ORFs were performed using character-based (Maximum Likelihood) and distance-based (Neighbor Joining algorithms in MEGA.7.0 software. The trees performed from two algorithms were essentially identical. In the resulting trees, Pis isolate was located in a group with other isolates of CsCMV which was supported by robust bootstrap values and was found to be most closely related to a clad including MCol22 (KT002435) and Brazilian isolate (NC_001658), consistent with the pairwise-comparison results (Fig. 3).

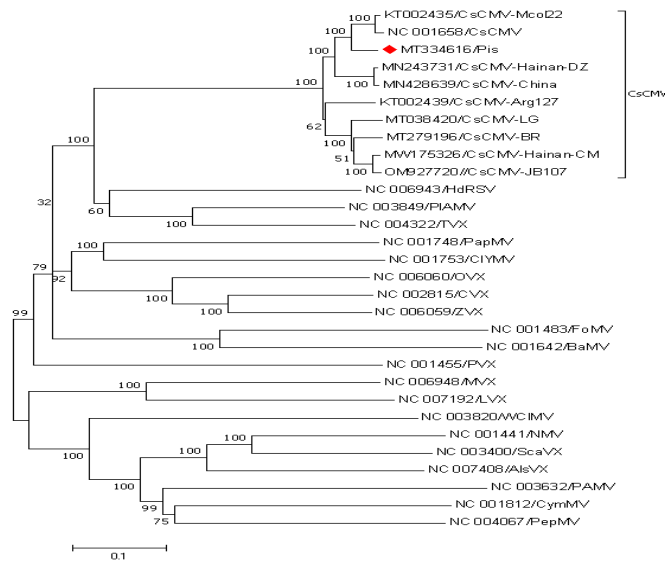


Fig. 3. Phylogenetic analysis of the complete nucleotide sequence of selected members of the genus *Potexvirus*. The phylogenetic tree was generated using the Neighbor Joining algorithms in the MEGA.7.0 analysis package. Branch significances were evaluated by constructing 1,000 trees in bootstrap analysis and the bootstrap values are shown above the horizontal line at each node. Pis isolate is marked with a red diamond. *Bamboosa mosaic virus* (BaMV), *Cassava common mosaic virus* (CsCMV), *Alstroemeria virus X* (AlsVX), *Cactus virus X* (CVX), *Clover yellow mosaic virus* (CIYMV), *Cymbidium mosaic virus* (CymMV), *Foxtail mosaic virus* (FoMV), *Hydrangea ringspot virus* (HdRSV), *Lily virus X* (LVX), *Mint virus X* (MVX), *Narcissus mosaic virus* (NMV), *Opuntia virus X* (OVX), *Papaya mosaic virus* (PapMV), *Pepino mosaic virus* (PepMV), *Plantago asiatica mosaic virus* (PIAMV), *Potato aucuba mosaic virus* (PAMV), *Potato virus X* (PVX), *Scallion virus X* (ScaVX), *Tulip virus X* (TVX), *White clover mosaic virus* (WCIMV) and *Zygocactus virus X* (ZVX).

In the phylogenetic trees constructed for coat protein and the protein associated to replication, most CsCMV isolates and other potexviruses are found in the same topography (data not shown). In constructed trees, CsCMV isolates, including the Pis isolate, consistently grouped together, further reinforcing the close relationship among them.

Based on the phylogenetic grouping and percent of sequence identity with former isolates of CsCMV, the Pis isolate is classified as a potexvirus of CsCMV species [25].

4. Discussion

High-Throughput Sequencing (HTS) represents a contemporary and advanced methodology for virus detection and screening,

enabling the identification of numerous viruses, viral strains, and viral pathogens associated with diseases of previously unknown origins [26]. This technology exhibits significant sensitivity, facilitating not only the detection of known pathogens but also the *de novo* discovery of novel systemic pathogens. As such, HTS has proven to be a valuable tool in certification programs for vegetatively propagated crops, enhancing the ability to ensure plant health and pathogen-free propagation [27].

In this paper, the incidence of CsCMV in pistachio trees in Iran using a bioinformatics approach was described. A near complete sequence was determined for the pistachio isolate which includes part of the gene for the replication-associated protein, the coat protein

cistron, a triple gene block (TGB), and the entire 3' UTR. The genomic organization of this isolate was found to be consistent with that of other members of the *Potexvirus* genus, demonstrating a high degree of similarity.

Although CsCMV has been detected infecting species of at least six families of dicotyledonous plants (Euphorbiaceae, Amaranthaceae, Compositae, Chenopodiaceae, Solanaceae, and Malvaceae) by artificial inoculation, this virus has been found naturally only in cassava and Chaya (*Cnidocolus aconitifolius*) in South America and recently in China [8]. The discovery of this isolate of CsCMV in local pistachio germplasm (so far, the first time identified and reported in Iran) indicates a potentially widespread distribution of the virus. On the other hand, since Ohadi is native to Iranian pistachio cultivars, it can be inferred that the infection of Iranian pistachio with *CsCMV* is not a recent occurrence but likely dates back several decades [4].

Although this study showed the existence of the *Cassava common mosaic virus* in pistachio, but more investigations are necessary for confirmation. To date, Koch's postulates have not been fully satisfied for this virus, so we cannot yet conclude that CsCMV is a viral agent to causes disease in Pistachio. As noted before, a few symptoms including yellowing and decline were observed in the single pistachio sample (Ohadi2) which proved to be a probable host for CsCMV. Whether the symptoms observed in this sample are due to the presence of this virus, or to non-viral agents remains to be studied. Moreover, the potential presence of additional viruses or virus-like agents cannot be ruled out, necessitating more extensive field studies and comprehensive testing to establish a definitive association between the virus and the observed disease.

Our results suggest that next generation sequencing (NGS) is an effective tool for broad virus diagnosis, capable of identifying both known and unknown viruses across various hosts. This technology has been successfully utilized for discovery of new viruses in most of the plants, fueling the current push to incorporate NGS routinely into diagnostics of plant health.

CsCMV, like other members of genus *Potexvirus*, has no known vector. Therefore, its transmission is facilitated through other methods. The spread of the virus has been linked to mechanical transmission through contaminated cutting tools utilized during the cultivation and preparation of planting material. Perennial fruit plants, which are predominantly propagated vegetatively, serve as reservoirs for a wide range of viruses and viroids. In the case of pistachio, propagation is typically achieved through grafting, a common practice involving the attachment of a scion to a rootstock for the clonal reproduction of specific cultivars. However, the inadvertent selection of infected plant material for propagation facilitates the accumulation of viruses within pistachio trees and promotes the dissemination of these pathogens across individual trees, orchards, and even regions.

5. Conclusion

In this study, the incidence of CsCMV in pistachio trees in Iran and the world using a bioinformatics approach was reported. The genomic RNA of Pis isolate was 4635 nts in length (accession number MT334616). This study may help to promote actions aimed at controlling viruses in the pistachio orchards of Kerman province. To comprehensively evaluate the sanitary status of pistachio cultivation, further research is essential to identify and characterize other potential viruses that may infect pistachio. Additionally, the

implementation of a systematic program for the sanitary and clonal selection of local pistachio varieties is crucial. Such a program would ensure the identification and propagation of healthy mother plants, which could then be multiplied and distributed to producers. The findings of this study underscore the need for increased focus on enhancing the phytosanitary conditions of pistachio trees in Iran, emphasizing the importance of mitigating viral infections to safeguard the health and productivity of pistachio orchards.

Acknowledgments

We are grateful to the Vali-e-Asr University of Rafsanjan, Iran, for financial support.

Conflict of Interest

The authors of the present research declare that there is no conflict of interest.

Code of Ethics

In this research, no living thing has been used, and all research stages have been conducted in a laboratory.

References

- 1- Mandalari G, Barreca D, Gervasi T, Roussel MA, Klein B, Feeney MJ, Carughi A. Pistachio nuts (*Pistacia vera* L.): Production, nutrients, bioactives and novel health effects. *Plants*. **2021**;11:18.
- 2- Al Rwahnih M, Rowhani A, Westrick N, Stevens K, Diaz-Lara A, Trouillas FP, Preece J, Kallsen C, Farrar K, Golino D. Discovery of viruses and virus-like pathogens in pistachio using high-throughput sequencing. *Plant Dis*. **2018**;102:1419-25.
- 3- Buzkan N, Chiumenti M, Massart S, Sarpkaya K, Karadağ S, Minafra A. A new emaravirus discovered in *Pistacia* from Turkey. *Virus Res*. **2019**;263:159-63.
- 4- Maddahian M, Massumi H, Heydarnejad J, Hosseini A, Khezri A, Sano T. Biological and molecular characterization of hop stunt viroid variants from pistachio trees in Iran. *J Phytopathol*. **2019**;167:163-73.
- 5- Mohammadi M, Hosseini A, Nasrollanejad S. In silico identification of two novel viruses on Iranian pistachio. *Iran J Plant Pathol*. **2021**;57:81-85.
- 6- Tuo DC, Zhao GY, Yan P, Li RM, Chen X, Wang WQ. First report of *Cassava common mosaic virus* infecting cassava in Mainland China. *Plant Dis*. **2020**;104:997.
- 7- Tuo DC, Zhao GY, Yan P. A *Cassava common mosaic virus* vector for virus-induced gene silencing in cassava. *Plant Methods*. **2021**;17:74.
- 8- Zanini AA, Di Feo L, Luna DF, Paccioretti P, Collavino A, Rodriguez MS. Cassava common mosaic virus infection causes alterations in chloroplast ultrastructure, function, and carbohydrate metabolism of cassava plants. *Plant Pathol*. **2021**;70:195-205.
- 9- Collavino A, Medina R, Nome C, Zanini A, Di Feo L. Evidence of cassava common mosaic virus in cassava shoot apical meristems. *Plant Pathol*. **2023**;72:993-97.
- 10- Gilbert KB, Holcomb EE, Allscheid RL, Carrington JC. Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PLoS One*. **2019**;14:e0219207.
- 11- Zeng L, Tu XL, Dai H, Han FM, Lu BS, Wang MS, Nanaei HA, Tajabadipour A, Mansouri M, Li XL, Ji LL. Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol*. **2019**;20:1-3.
- 12- Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res*. **2010**;39(suppl_1):D19-21.

- 13- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **2014**;30:2114-20.
- 14- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. **2019**;37:907-15.
- 15- Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. **2015**;31:1674-6.
- 16- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. **2018**;35:1547-9.
- 17- Do Kim J, Koo YB, Chang MU. Genome characterization of a Korean isolate of Cymbidium mosaic virus. *Molecules and cells*. **1998**;8:181-8.
- 18- Howitt RL, Beever RE, Pearson MN, Forster RL. Genome characterization of Botrytis virus F, a flexuous rod-shaped mycovirus resembling plant 'potex-like' viruses. *J Gen Virol*. **2001**;82:67-78.
- 19- Côté F, Paré C, Majeau N, Bolduc M, Leblanc E, Bergeron MG, Bernardy MG, Leclerc D. Nucleotide sequence and phylogenetic analysis of a new potexvirus: Malva Mosaic Virus. *Infect Genet Evol*. **2008**;8:83-93.
- 20- Wu CH, Lee SC, Wang CW. Viral protein targeting to the cortical endoplasmic reticulum is required for cell-cell spreading in plants. *J Cell Biol*. **2011**;193:521-35.
- 21- Lozano I, Leiva AM, Jimenez J, Fernandez E, Carvajal-Yepes M, Cuervo M, Cuellar WJ. Resolution of cassava-infecting alphaflexiviruses: Molecular and biological characterization of a novel group of potexviruses lacking the TGB3 gene. *Virus Res*. **2017**;241:53-61.
- 22- Koenig R, Pleij CW, Loss S, Burgermeister W, Aust H, Schiemann J. Molecular characterisation of potexviruses isolated from three different genera in the family Cactaceae. *Arch virol*. **2004**;149:903-14.
- 23- Thompson JR, Jelkmann W. Strain diversity and conserved genome elements in Strawberry mild yellow edge virus. *Arch virol*. **2004**;149:1897-909.
- 24- Solovyev G, Novikov VK, Merits A, Savenkov EI, Zelenina DA, Tyulkina LG, Morozov SY. Genome characterization and taxonomy of Plantago asiatica mosaic Potexvirus. *J Gen Virol*. **1994**;75:259-67.
- 25- Kreuze JF, Vaira AM, Menzel W, Candresse T, Zavrjev SK, Hammond J, Hyun Ryu K, Report Consortium IC. ICTV virus taxonomy profile: Alphaflexiviridae. *J Gen Virol*. **2020**;101:699-700.
- 26- Xu Y, Li S, Na C, Yang L, Lu M. Analyses of virus/viroid communities in nectarine trees by next-generation sequencing and insight into viral synergisms implication in host disease symptoms. *Sci Rep*. **2019**;9:12261.
- 27- Maliogka VI, Minafra A, Saldarelli P, Ruiz-García AB, Glasa M, Katis N, Olmos A. Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses*. **2018**;10:436.