J. Agr. Sci. Tech. (2023) Vol. 25(1): 213-223

Metagenomic Analysis of *Malva sylvestris* from Iran Displays a Malva Vein Clearing Virus Genome

Z. Moradi[1], and M. Mehrvar[2]*

ABSTRACT

To date, only the complete genome of one malva vein clearing virus (MVCV) has been sequenced worldwide. Here, for the first time, the genomic sequence of an isolate of MVCV affecting *Malva sylvestris* in Iran was determined, using RNA deep sequencing confirmed by reverse-transcription polymerase chain reaction and Sanger sequencing. The sequence of IR1 was 11,055 nucleotides in length and contained a single open reading frame of 10,527 nucleotides encoding a large polyprotein of 3,508 amino acids with predicted molecular weight of 395,08 KDa. The sequence contained nine putative proteolytic cleavage sites and motifs conserved in homologous proteins of other potyviruses. The P1 of the IR1 was 1236 nucleotides longer than that of the only recently reported sequence from China (MVCV-SX, MN116683). The complete genome sequence obtained from the study showed 81.24% and 91.93% identities to its Chinese counterpart at the nucleotide (nt) and deduced amino acid (aa) levels, respectively. The low nt and aa sequence identity with known MVCV isolate seems to indicate that IR1 is a novel strain. Phylogenetic analysis of the coat protein gene also showed that the Iranian isolate was most closely related to the Dutch isolate NAKT-NL (FJ539084), with identities of 94.06 (at nt level) and 96.04% (at aa level). The results of this study will be useful for understanding the global molecular epidemiology of MVCV.

Keywords: Complete genome, High-throughput sequencing, MVCV.

INTRODUCTION

Malva spp. are annual, biennial, or herbaceous perennial plants that comprise a widespread group of tropical and temperate plants belonging to the *Malvaceae* family. The origin of *Malva* species is uncertain, however, the Mediterranean and South Western Asia are suggested as the plausible centers of diversity. Commonly, these are used as ornamental plants in landscape designs, although they can be used for food and medicinal purposes (Sharifi-Rad *et al*., 2019). As a medicinal plant, the aerial parts of *Malva* species such as flowers and leaves, are most widely used for a variety of therapeutic purposes, including digestive, respiratory, genitourinary, skeletal, and skin disorders. *Malva* species have diuretic, lenitive, spasmolytic, and laxative effects and are used as expectorant, antitussive, antidiarrheal, and profoundly applied for skin care as antiseptic and demulcent (Gasparetto *et al*., 2012; Al-Rubaye *et al*., 2017; Benzie and Wachtel-Galor, 2011). However, the risk of viral infection is becoming more serious due to the commercial cultivation of the *Malva* spp., leading to an undesirable effect on its pharmacodynamic quality and ornamental values.

Malva spp. are reported to be affected by several viral agents including malva vein clearing virus (MVCV) (Lunello *et al*., 2009), watermelon mosaic virus (Kim *et al*., 2019), cherry rasp leaf virus (Villamor and

[1] Department of Plant Pathology, Faculty of Crop Sciences, Sari Agricultural Sciences and Natural Resources University, Sari, Islamic Republic of Iran.
[2] Department of Plant Pathology, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Islamic Republic of Iran.
*Corresponding author; e-mail: mehrvar@um.ac.ir

[DOI: 10.52547/jast.25.1.213]

Eastwell, 2016), cucumber vein yellowing virus (Janssen *et al*., 2002), and tomato leaf curl Kerala virus associated with yellow vein net disease (Snehi *et al*., 2018). These viruses have a significant negative impact on malva plants provoking a wide range of symptoms, reducing both ornamental and economic values. MVCV (genus *Potyvirus*, family *Potyviridae*) has flexuous filamentous particles about 700-900 nm in length, is transmitted by aphids in a non-persistent manner, and naturally infects only plant species of the *Malvaceae* family (Lunello *et al*., 2009). Molecular analysis of the viral genome is needed for characterization and taxonomic classification of viruses. In the case of the potyviruses, comparisons of the whole genome and *coat protein* (cp) gene are the most commonly used criteria (Adams *et al*., 2005; Moradi *et al*., 2017). To devise new control methods and evaluating its risk, complete sequence data are essential for the understanding the genome structure, diversity, phylogenetic relationships, evolutionary biology, and epidemiology of a virus (Moradi *et al*., 2017). Currently, the complete genome sequence of only one MVCV isolate (SX, MN116683) from hollyhock (*Alcea rosea*) in China is available in the GenBank. MVCV has been recently reported in Iran based on CP sequence (Valouzi *et al*., 2017); however, to date, no studies have been performed on the complete genomic properties of this virus in the country. To characterize MVCV more accurately, in this study, we sequenced the whole genome of one MVCV isolate from Iran (IR1) and compared its sequence with the sole full MVCV sequence recorded in the GenBank. High throughput sequencing techniques have significantly broadened possibilities for rapid diagnosis of new viruses as well as virus strains and their genomes (Roossinck *et al*., 2015; Wylie *et al*., 2011; Al-Jaberi *et al*., 2021). Thus, RNA deep-sequencing technique was applied to identify the virus. Additionally, the *cp* gene sequence of isolate IR1 and 19 other publicly available sequences were analyzed to determine their phylogenetic relationships and to know more about the genetic diversity and molecular evolution of MVCV population.

MATERIALS AND METHODS

RNA Isolation, Library Construction, RNA-Sequencing, and Bioinformatics Analysis

In the spring of 2020, *Malva sylvestris* plants in desert areas near Azadshahr, Golestan Province, in the north of Iran, displayed a significant occurrence of virus-like symptoms consisting of yellow vein mosaic, green mosaic, and leaf rugosity (Figure 1). In order to look for the causal agent(s), total RNA was isolated using Promega SV Total RNA Isolation Kit (USA) on a leaf sample collected from one diseased malva plant. Libraries (one from a symptomatic plant and one from a healthy-appearing plant) were generated by Illumina TruSeq Stranded Total RNA sample preparation kit with Ribo-Zero Plant and sequenced using an Illumina NovaSeq 6000 (Macrogen, Seoul, South Korea) with 2×151 bp paired-end reads. Raw Illumina reads were trimmed for quality and cleaned by removing adaptors. The cleaned reads of one symptomatic plant were mapped to reads from one healthy-looking plant by CLC Genomics Workbench version 20 (CLC Bio, Qiagen). Unmapped reads were used for viral genome *de novo* assembly using CLC Genomics Workbench. The assembled contigs were subjected to BlastN and BlastX analyses against GenBank nucleotide and protein databases.

Sequence Analysis, Database Search, and Phylogenetic Analysis

Open Reading Frames (ORF), mature peptides, and domains encoded by them were predicted within Geneious Prime v. 2020.1.2 (Biomatters Ltd., Auckland, New Zealand). Clustalw multiple sequence

[DOI: 10.52547/jast.25.1.213]

Figure 1. Malva leaves showing yellow vein mosaic, green mosaic, and rugosity collected from Golestan Province, Iran.

alignments, nucleotide (nt), and amino acid (aa) sequence identities between the isolates were carried out using the Geneious Prime (Biomatters). The *cp* gene sequences of 20 MVCV isolates (one from this study and 19 retrieved from the GenBank) were used to determine phylogenetic correlation and genetic diversity. Phylogenetic tree was constructed using the Neighbor-Joining (NJ) method in MEGA11 (Tamura *et al*., 2021), after codon-based alignment with ClustalW. Two-dimensional nucleotide diversity matrix was implemented by the MUSCLE-based pairwise alignment option using SDT v1.2 software (Muhire *et al*., 2014). The haplotype and nucleotide diversity values, Nonsynonymous substitution rate (dN) and Synonymous substitution rate (dS) were calculated using the DnaSP6 (Rozas *et al*., 2017). Possible codons subjected to positive selection were identified using Single-Likelihood Ancestor Counting (SLAC), Fast Unbiased Bayesian Approximation (FUBAR), Fixed Effects Likelihood (FEL), Internal Fixed Effects Likelihood (IFEL), and mixed effects model of evolution (MEME) in Datamonkey (http://www.datamonkey.org/) (Kosakovsky Pond and Frost, 2005). Only selections determined to be significant by at least two methods were considered as positive selections.

RESULTS

Genomic Characterization of Iranian Isolate of MVCV

Around 67,005,840 reads from malva RNA-seq library were produced by Illumina sequencing. *De novo* assembly generated three MVCV contigs and 611,517 reads mapped to the contig of interest. One long contig was identified from the assembled non-plant reads that mapped to MVCV sequence in the GenBank. Its sequence was confirmed by Sanger sequencing using reverse-transcription polymerase chain reaction with potyvirus degenerate primers designed to amplify the Helper Component Protease (HC-Pro) and Cylindrical Inclusion protein (CI) coding regions (Ha *et al*., 2008) and MVCV-specific primers (Valouzi *et al*., 2017). There was no evidence of other viral agents in the NGS contigs. The verified complete genome for the Iranian isolate MVCV-IR1 comprised 11,055 nucleotides, excluding the poly(A) tail (GenBank accession no. MZ555807). The 5′ and 3′ Untranslated regions (UTRs) were 240 nt and 288 nt in length, respectively. The genome contained a single large ORF that started with the AUG codon at nt position 241-243 and ended with a termination codon (UGA) at nt position 10,765-10,767,

[DOI: 10.52547/jast.25.1.213]

comprising 10,527 nt and putatively encoding a polyprotein of 3,508 amino acid residues with an estimated molecular mass of 395,08 kDa. A small putative protein, pretty interesting *Potyviridae* ORF (PIPO) (Chung *et al*., 2008) was also identified within the P3 cistron, which was likely expressed via ribosomal frameshifting or transcriptional slippage of viral RNA-dependent RNA polymerase. Its predicted start was a highly conserved motif, identified as a G1A7 in MVCV-IR1, at position 4,174-4,181 and a UAG stop codon at position 4,371–4,373, giving it a total length of 65 aa. Two highly conserved potyboxes 'a' (UCAACACAACAU) and 'b' (CAAGCA), are located in the 5'-UTR of many potyviruses (Turpen, 1989). The nucleotide sequence $\underline{A}_{126}$CAACACAACA$\underline{C}_{137}$ (underlined base is different) for potybox 'a', and the sequence C_{152}AA$\underline{CACA}$GCA$_{161}$ (underlined bases inserted), which is slightly different from potybox 'b', were found in the 5'-UTR of MVCV-IR1. The 5′-UTR had a high content of AU (72.92%) and a low content of GC (27.08%), like values reported for other potyviruses (Moradi *et al*., 2017). The 3′-UTR, whose secondary structure might be involved in genome replication, was AU-rich (59.72%) and contained TATA box-like sequences. The large polyprotein precursor of MVCV-IR1 was predicted to be proteolytically processed into ten smaller functional peptides: Protein 1 (P1), Helper Component-Protease (HC-Pro), Protein 3 (P3), 6 kilodalton (kDa) peptide 1 (6K1), Cylindrical Inclusion protein (CI), 6 kDa peptide 2 (6K2), Viral Protein genome-linked (VPg), Nuclear Inclusion-a Protease (NIa-Pro), Nuclear Inclusion-b protein (NIb), and Capsid Protein (CP). Its cleavage sites were in consensus to those of other known potyviruses, whose dipeptides are Y/S, G/G, Q/G, Q/S, Q/S, Q/G, E/S, Q/S, and Q/A (Figure 2). The genome positions of ORFs and UTRs, and the expected cleavage sites are presented in Figure 2. Conserved motifs of potyviruses were also identified, including the serine-type protease domain H_{606}-8X-D-31X-G-X-S_{649}-$G_{650,}$ and proteolytic domain F_{670}VVRGR$_{675}$ in P1; the F_{880}RNK$_{883}$, C_{991}CCVT$_{995}$, P_{1009}TK$_{1011}$, V_{758}GN$_{760}$, G_{1041}YCY$_{1044}$ motifs and the putative proteinase active site C_{1043}-72X-H_{1116} in HC-Pro; the G_{1649}AVGSGKSTG$_{1658,}$ V_{1669}LILEPTRPL$_{1678}$, D_{1738}ECH$_{1741}$, and V_{1867}ATNIIENGVTL$_{1878}$ motifs in CI; the protease motif H_{2492}-34X-D-67X-G-X-CG-14X-H_{2613} in NIa-Pro; the S_{2859}LKAEL$_{2864}$, C_{2934}DADGS$_{2939}$, and G_{3040}DD$_{3042}$ motifs in NIb; and the M_{3356}VWCIENGTSP$_{3366}$, R_{3397}-43X-D_{3441}, and Q_{3459}MKAAA$_{3464}$, motifs in CP (Riechmann *et al*., 1992; Dougherty and Semler, 1993; Kadare and Haenni 1997; Revers *et al*., 1999; Urcuqui-Inchima *et al*., 2001; Adams *et al*., 2005; Moradi *et al*., 2017). Also, the conserved motif DAG, which is an important factor of virus transmission by aphids (Atreya *et al*., 1995),

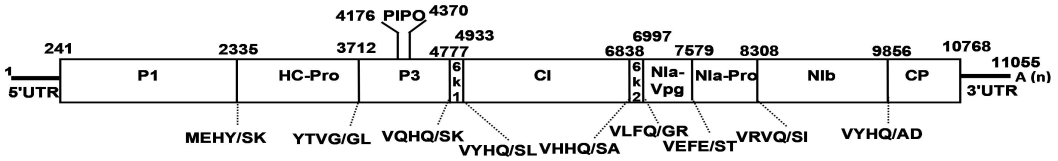

Figure 2. Schematic representation of the genomic organization of Iranian isolate of malva vein clearing virus (MVCV-IR1). Open reading frames are represented by boxes. The numbers above the genome indicate the start nucleotide position of each protein gene. The putative proteinase cleavage sites in the polyprotein are indicated below the genome.

[DOI: 10.52547/jast.25.1.213]

was changed to $N_{3211}AG_{3213}$, in concordance with previous reports (Lunello *et al*., 2009; Wylie *et al*., 2002; Nigam *et al*., 2019). NAG is one of the most frequent deviations of DAG in potyviruses, which has been proven to be efficient in aphid transmission as the canonical DAG motif (Wylie *et al*., 2002; Atreya *et al*., 1991; Nigam *et al*., 2019).

Sequence Comparison and Phylogenetic Analysis

The complete genomic sequence of MVCV-IR1 shared identity of 81.24% with the lone MVCV complete sequence (SX) available in the GenBank. Aligned complete polyprotein sequence of MVCV-IR1 shared 80.84 and 91.93% nt and aa sequence identity, respectively, with the corresponding sequence of MVCV-SX from China (Table 1). High sequence variability (from 76.34 to 88.67% at nt level and 80.77 to 99.18% at aa level) was observed over most of the genome of two MVCV isolates, suggesting that IR1 and SX were different strains. Among putative gene products of MVCV, the P1 was the most variable protein (80.77%), while NIa-Pro was the most conserved (99.18%) (Table 1). The differences between two MVCV isolates may be due to different mutation occurring in a common ancestor virus or mutations produced by different selection pressures associated with different environmental pressures in two different agro-ecological conditions. The P1 of the MVCV-IR1 isolate (2094 nucleotides) was 1236 nucleotides longer than that of the SX isolate (858 nt), and their identity was 81.12%. Interestingly, P1 gene as long as 2,228 nucleotides is in some potyviruses like sweet potato mild mottle virus. The *cp* gene of 20 MVCV isolates was used to investigate the genetic diversity and molecular mechanisms underlying the evolution of this virus. Using RDP4 (Martin *et al*., 2015) implemented algorithms, no significant recombination event was detected in the MVCV *cp* gene sequences. Comparison of the complete *cp* gene sequences revealed that IR1 had 87.46-94.06% nt identity (90.10-96.04% aa identity) with 19 other MVCV isolates, of which only partial sequences are available in the GenBank (Figure 4). MVCV-IR1 shared the lowest nt identity (87.46%) (92.08% aa identity) with Italian isolate Napoli (LN651191) and the highest nt

Table 1. The size of nucleotide and amino acid sequence in different genomic regions of Iranian isolate of malva vein clearing virus (MVCV-IR1) and percentage nucleotide and amino acid sequence identity to MVCV-SX.

Genomic region	Size in nucleotide	Size in amino acid	Percentage of nucleotide identity	Percentage of amino acid identity
			MVCV-SX	
Polyprotein	10,527	3,508	80.84%	91.93%
5′-UTR	240	-	51.47%	-
P1	2,094	698	81.12%	80.77%
HC-Pro	1,377	459	80.25%	94.34%
P3	1,065	355	76.34%	81.97%
6K1	156	52	80.77%	96.15%
CI	1,905	635	77.53%	92.44%
6K2	159	53	79.87%	92.45%
NIa-VPg	582	194	83.51%	96.91%
NIa-Pro	729	243	82.30%	99.18%
NIb	1,548	516	81.91%	95.35%
CP	912	303	88.67%	93.73%
3′-UTR	288	-	94.14%	-

[DOI: 10.52547/jast.25.1.213]

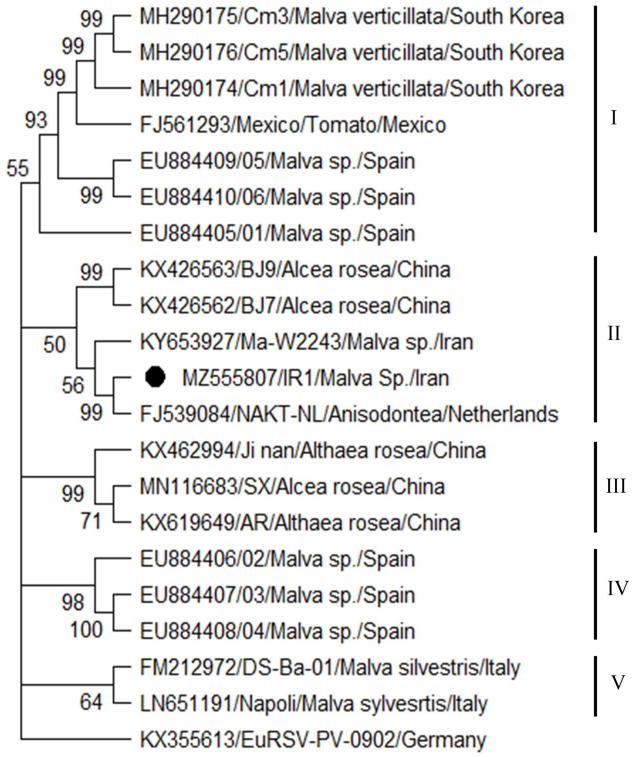

Figure 3. A neighbor-joining phylogenetic tree of the complete coat protein sequences of 20 malva vein clearing virus isolates. The tree was rooted with the corresponding sequence of euphorbia ringspot virus (EuRSV; accession No. KX355613). The isolate IR1 (marked) was obtained in this study. The significance of each branch was evaluated by constructing 1,000 trees in bootstrap analysis, and the bootstrap values (%) higher than 50 are shown. Isolates are indicated in the tree by accession number/isolate name/host origin/ geographical origin of collection.

sequence identity (94.06%) (96.04% aa identity) with a Dutch isolate NAKT-NL (FJ539084). In the phylogenetic analyses based on the *cp* gene nucleotide sequences, 20 isolates were clustered into five phylogroups I, II, III, IV and V (Figure 3). Phylogroup I contained seven MVCV isolates from South Korea, Spain, and Mexico. The Iranian isolates (IR and Ma-W2243) were clustered together with isolates from Netherlands, and China into phylogroup II, and categorized into a distinct sub-lineage with NAKT-NL isolate from Netherlands. Phylogroups III and IV consisted of, respectively, three isolates from China and Spain. Two Italian isolates were grouped in the phylogroup V (Figure 3).

Pairwise comparisons showed that members of phylogroups I, II, III, IV, and V shared, respectively, within-group identities of 89.44–100 (average 94.69%), 88.01-99.67 (average 90.86%), 99.23-99.45 (average 99.33%), 94.50-99.01 (average 96.10%), and 90.76%. At the deduced amino acid level, the isolates in phylogroups I, II, III, IV, and V shared identities of 93.40-100 (average 96.84%), 90.10-99.01 (average 93.15%), 99.01-99.67 (average 99.33%), 97.03-99.67 (average 98.01%), and 93.40%, respectively. This suggested that isolates in phylogroups II and V had a greater level of genetic variation than those belonging to the other phylogroups. The overall mean value of genetic distance (d) among 20 isolates was 0.113±0.007, indicating a high genetic diversity. As expected, the between-group genetic divergence (d> 0.1) was obviously higher than the within-group ones

[DOI: 10.52547/jast.25.1.213]

(Phylogroup I: d= 0.06±0.01; II: d= 0.10±0.01; III: d= 0.01±0.00; IV: d= 0.04±0.01; and V: d=0.10±0.01) (Table 2), confirming the phylogenetic grouping. As depicted in Figure 3, the geographical location did not seem to play a role in formation of phylogroups I and II. However, the isolates in the other phylogroups tended to cluster according to their geographical locations. This result suggested that the MVCV CP diversification could be explained in part by geography-driven adaptation. A similar phylogenetic tree was obtained from an alignment of the aa sequences of *cp* gene of the same isolates (data not shown). Haplotype diversity (Hd) and nucleotide diversity (π) for all MVCV isolates were 0.984 and 0.101, respectively, which was higher than 0.500 and 0.005, indicating a high genetic diversity in MVCV population. The global selection pressure (dN/dS= ω) in CP of all MVCV isolates was 0.0857, suggesting that purifying selection was restricting variability in the population. The ω values for phylogroup I and phylogroup II were, respectively, 0.070 and 0.103, indicating that phylogroup I was subjected to more intense purifying selection than phylogroup II. Using HYPHY, there were three codon positions (10, 35, and 76) for MVCV isolates that were shown to be under positive selection (Table 3).

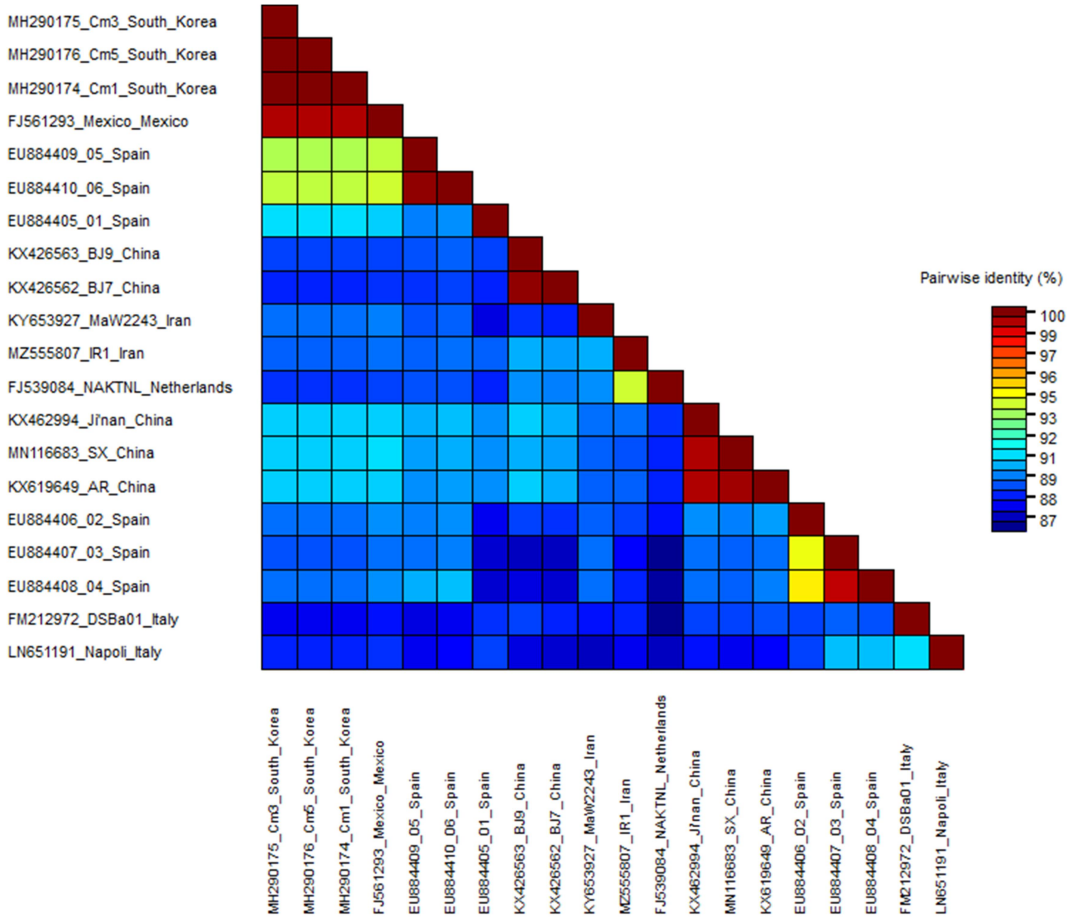

Figure 4. Pairwise nucleotide sequence identity matrix of 20 *Malva vein clearing virus* isolates generated using SDT software with the MUSCLE alignment.

[DOI: 10.52547/jast.25.1.213]

Table 3. Codon positions of coat protein coding region of 20 *Malva vein clearing virus* isolates significantly affected by positive selection by different codon-based maximum-likelihood algorithms. [a]

Model				
SLAC	FEL	IFEL	FUBAR	MEME
-	**10, 35**, **76**	**10**, 120	**10**, **35**	**10,** 12, **35**, 43, 57, **76**

[a] Note: In bold are the sites identified as being under positive selection by more than two methods.

DISCUSSION

In this paper, we obtained the full-length sequence of MVCV from Iran. Nucleotide and amino acid sequence identities between the complete polyprotein-coding genomic sequence and most of the individual genes displayed higher levels of identity for the amino acid sequences (Table 1), indicating that many of the nucleotide substitutions in IR1 were silent. On the contrary, in P1, the value of the amino acid sequence identity was lower than that estimated from the nucleotide sequence, suggesting a prevalence of nonsynonymous mutations in this ORF. Phylogenetic comparison of the MVCV isolate from our study with those identified by others and deposited in the GenBank revealed five main clades. For phylogroups I and II, the phylogenetic analysis of the CP sequences did not show a clear clustering of MVCV isolates by the geographical position. This is probably due to the occurrence of genetic exchange or the limited number of geographically different isolates. These results are consistent with the analysis of Parrella *et al*. (2015), who also concluded that there was no statistically significant grouping among nine different MVCV isolates (from Italy, Spain, and Mexico). According to the complete CP sequence identity analysis, MVCV-IR1 most closely resembled the Dutch isolate NAKT-NL. The movement and exchange of infected propagation materials can be one of the possible explanations for such high sequence identities observed between the geographically separated MVCV isolates. In the selection pressure analysis at individual codon positions of CP, most of the codons were found to be under negative selection or neutral evolution, while the codons at positions 10, 35, and 76 were under positive selection. Notably, among positively selected codons, only the amino acid position 10 (in the N-terminal region of CP) was confirmed by four methods (Table 3). The variable N-terminus of the potyvirus CP has been shown to be involved in aphid transmission, and systemic infection, as well as in the adaptation of the virus to its host (Ullah *et al*., 2003; Desbiez *et al*., 2014; Garcia-Ruiz, 2018). Molecular adaptation is often identified with positive selection and is normally characterized by high dN/dS ratios (Parto and Lartillot, 2018; Moradi and Mehrvar, 2019). The presence of diversifying selection at a number of codons in MVCV CP reflects the adaptive evolution by fixing beneficial genotypes and enhancing population diversification. Molecular adaptation due to positive selection has been reported for some potyviruses, for example: mutations in the N-terminus of tobacco etch virus CP correlated with its ability to break resistance (Decroocq *et al*., 2009); amino acid substitutions at specific positions of plum pox virus CP N-terminus determined host-dependent pathogenicity (Carbonell *et al*., 2013); and codons under positive selection at the N-terminal region of the potato virus Y CP affected fitness in tobacco and potato (Moury and Simon, 2011). These confirm the necessity of genetic flexibility in the CP and in the aphid-transmission motif of potyviruses to sustain functionality in genetically various hosts and vectors, as pointed out by Nigam *et al*. (2019). According to our finding, an SNP C→A (30% variant frequency) and C→T (25% variant frequency) occurring in codon

[DOI: 10.52547/jast.25.1.213]

position 10 (at nucleotide 28) represented the polymorphism that led to a change in the encoded amino acid from proline (P, codon CCA) to threonine (T, codon ACA) or serine (S, codon TCA). Replacement of nonpolar amino acid proline to polar amino acids threonine or serine with different physicochemical properties may modify protein formation due to hydrophobic to hydrophilic molecule changes. However, the specified effect of the aforementioned mutations in this stretch should be further studied. All data obtained in this work could assist future studies on protein localization and host protein interactions, and provide further valuable information on pathogenicity and the processes involved in virus evolution and epidemiology to improve our understanding of disease evolution and management strategies. More genomic sequences of MVCV isolates from different geographical areas and hosts are needed to offer precise information on the virus's evolutionary characteristics. In addition, the potential infection of *Malvaceae* crop plants by this virus has to be explored.

REFERENCES

1. Adams, M. J., Antoniw, J. F. and Beaudoin, F. 2005. Overview and Analysis of the Polyprotein Cleavage Sites in the Family Potyviridae. *Mol. Plant Pathol.,* **6**: 471–487.
2. Al-Jaberi, M. S., Moradi, Z., Mehrvar, M., Al-Inizi, H. R. and Zakiaghl, M. 2021. Whole Genome Characterization of Wisteria Vein Mosaic Virus from Iran and Its Relationship to Other Members of Bean Common Mosaic Virus Group. *3. Biotech.,* **11**: 407.
3. Al-Rubaye, A. F., Kaizal, A. F. and Hameed, I. H. 2017. Phytochemical Screening of Methanolic Leaves Extract of *Malva sylvestris*. *Int. J. Pharmacogn. Phytochem. Res.,* **9**: 537–552.
4. Atreya, P. L., Atreya, C. D. and Pirone, T. P. 1991. Amino Acid Substitutions in the Coat Protein Result in Loss of Insect Transmissibility of a Plant Virus. *Proc. Natl. Acad. Sci. USA,* **88**: 7887±91.
5. Atreya, P. L., Lopez-Moya, J. J., Chu, M., Atreya, C. D. and Pirone, T. P. 1995. Mutational Analysis of the Coat Protein N-Terminal Amino Acids Involved in Potyvirus Transmission by Aphids. *J. Gen. Virol.,* **76**: 265-270.
6. Benzie, I. F. and Wachtel-Galor, S. 2011. *Herbal Medicine: Biomolecular and Clinical Aspects, Oxidative Stress and Disease.* 2nd Edition, CRC Press, Florida, 499 PP.
7. Carbonell, A., Maliogka, V. I., Pérez Jde, J., Salvador, B., León D. S., García J. A. and Simón-Mateo, C. 2013. Diverse Amino Acid Changes at Specific Positions in the N-Terminal Region of the Coat Protein Allow Plum Pox Virus to Adapt to New Hosts. *Mol. Plant-Microbe Interact.*, **26**: 1211–1224.
8. Chung, B. Y-W., Miller, W. A., Atkins, J. F. and Firth, A. E. 2008. An Overlapping Essential Gene in the *Potyviridae*. *Proc. Natl. Acad. Sci. USA,* **105**: 5897–5902.
9. Decroocq, V., Salvador, B., Sicard, O., Glasa, M., Cosson, P., Svanella-Dumas, L., Revers, F., García, J.A. and Candresse, T. 2009. The Determinant of Potyvirus Ability to Overcome the RTM Resistance of *Arabidopsis thaliana* Maps to the N-Terminal Region of the Coat Protein. *Mol. Plant Microbe Interact.,* **22**: 1302–1311.
10. Desbiez, C., Chandeysson, C. and Lecoq, H. 2014. A Short Motif in the N-Terminal Part of the Coat Protein is a Host-Specific Determinant of Systemic Infectivity for Two Potyviruses. *Mol. Plant Pathol.,* **15**: 217-21.
11. Dougherty, W. G. and Semler, B. L. 1993. Expression of Virus-Encoded Proteinases: Functional and Structural Similarities with Cellular Enzymes. *Microbiol, Rev.,* **57**: 781-822.
12. Garcia-Ruiz, H. 2018. Susceptibility Genes to Plant Viruses. *Viruses*, **10**: 484.
13. Gasparetto, J. C., Martins, C. A., Hayashi, S. S., Otuky, M. F. and Pontarolo, R. 2012. Ethnobotanical and Scientific Aspects of *Malva sylvestris* L: A Millennial Herbal Medicine. *J. Pharm. Pharmacol.,* **64**: 172–189.
14. Ha, C., Coombs, S., Revill, P. A., Harding, R. M., Vu, M. and Dale, J. L. 2008. Design and Application of Two Novel Degenerate

[DOI: 10.52547/jast.25.1.213]

Primer Pairs for the Detection and Complete Genomic Characterization of Potyviruses. *Arch. Virol.,* **153**: 25-36.
15. Janssen, D., Ruiz, L., Velasco, L., Segundo, E. and Cuadrado, I. M. 2002. Non-Cucurbitaceous Weed Species Shown to Be Natural Hosts of *Cucumber Vein Yellowing Virus* in South-Eastern Spain. *Plant Pathol.,* **51**: 797.
16. Kadare´, G. and Haenni, A-L. 1997. Virus-Encoded RNA Helicases. *J. Virol.,* **71:** 2583–2590.
17. Kim, J. E., Kwak, H. R., Choi, H. S., Kim, M., Jung, W. K., Seo, J. K., Kim, J. S. and Cha, B. 2019. First Report of *Watermelon Mosaic Virus* on *Malva verticillata* in Korea. *Plant Dis.,* **103**: 380-381.
18. Kosakovsky Pond, S. L. and Frost, S. D. W. 2005. Datamonkey: Rapid Detection of Selective Pressure on Individual Sites of Codon Alignments. *Bioinformatics,* **21**: 2531–2533.
19. Lunello, P., Tourino, A., Nunez, Y., Ponz, F. and Sanchez, F. 2009. Genomic Heterogeneity and Host Recovery of Isolates of *Malva Vein Clearing Virus*. *Virus Res.,* **140**: 91-97.
20. Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. 2015. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.,* **1**: vev003.
21. Moradi, Z., Mehrvar, M., Nazifi, E. and Zakiaghl, M. 2017. *Iranian Johnsongrass* Mosaic *Virus*: The Complete Genome Sequence, Molecular and Biological Characterization, and Comparison of *Coat Protein* Gene Sequences. *Virus Genes,* **53**: 77-88.
22. Moradi, Z. and Mehrvar, M. 2019. Genetic Variability and Molecular Evolution of *Bean Common Mosaic Virus* Populations in Iran: Comparison with the Populations in the World. *Eur. J. Plant Pathol.,* **154**: 673–690.
23. Moury, B. and Simon, V. 2011. dN/dS-Based Methods Detect Positive Selection Linked to Trade-Offs between Different Fitness Traits in the Coat Protein of *Potato Virus Y*. *Mol. Biol. Evol.,* **28:** 2707–2717.
24. Muhire, B. M., Varsani, A. and Martin, D. P. 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS One,* **9**: e108277.
25. Nigam, D., LaTourrette, K., Souza, P. F. N. and Garcia-Ruiz, H. 2019. Genome-Wide Variation in Potyviruses. *Front. Plant Sci.,* **10**: 1439.
26. Parrella, G., Nappo, A. G. and Delecolle, B. 2015. Cytopathology, Biology and Molecular Characterization of Two Italian Isolates of *Malva Vein Clearing Virus. Plant Sci. Today,* **2**: 69-73.
27. Parto, S. and Lartillot, N. 2018. Molecular Adaptation in Rubisco: Discriminating between Convergent Evolution and Positive Selection Using Mechanistic and Classical Codon Models. *PLoS One,* **13**: e0192697.
28. Revers, F., Le Gall, O., Candresse, T. and Maule, A. J. 1999. New Advances in Understanding the Molecular Biology of Plant: Potyvirus Interactions. *Mol. Plant Microbe Interact.,* **12**: 367-376.
29. Riechmann, J. L., Lain, S. and Garcia, J. A. 1992. Highlights and Prospects of Potyvirus Molecular Biology. *J. Gen. Virol.,* **73**: 1–16.
30. Roossinck, M. J., Martin, D. P. and Roumagnac, P. 2015. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology,* **105**: 716–727.
31. Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sanchez-Gracia, A. 2017. DnaSP6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.,* **34:** 3299-3302.
32. Sharifi-Rad, J., Melgar-Lalanne, G., Hernández-Alvarez, A.J., Taheri, Y., Shaheen, S., Kregiel, D. 2019. Malva Species: Insights on Its Chemical Composition towards Pharmacological Applications. *Phytother. Res.,* **34**: 546–567.
33. Snehi, S. K., Parihar, S. S., Gupta, G., Purvia, A. S. and Singh, V. 2018. Molecular Identification of a Begomovirus Associated with Yellow Vein Net Disease on *Malva parviflora L*. from India. *Microbiol. Curr. Res.,* **2**: 24-29.
34. Tamura, K., Stecher, G. and Kumar, S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.,* **38**: 3022–3027.
35. Turpen, T. J. 1989. Molecular Cloning of a *Potato Virus Y* Genome: Nucleotide Sequence Homology in Non-Coding Regions of Potyviruses. *Gen. Virol.,* **70**: 1951-1960.

[DOI: 10.52547/jast.25.1.213]

36. Ullah, Z., Chai, B., Hammar, S., Raccah, B., Gal-On, A. and Grumat, R. 2003. Effect of Substitution of the Amino Termini of the Coat Proteins of Distinct Potyvirus Species on Viral Infectivity and Host Specificity. *Physiol. Mol. Plant Pathol.,* **63**: 129–139.
37. Urcuqui-Inchima, S., Haenni, A. and Bernardi, F. 2001. Potyvirus Proteins: A Wealth of Functions. *Virus Res.,* **74**: 157-175.
38. Valouzi, H., Golnaraghi, A. and Rakhshandehroo, F. 2017. Natural Occurrence of *Malva Vein Clearing Virus* in Malva in Iran. *New Dis. Rep.,* **35**: 15.
39. Villamor, D. E. V. and Eastwell, K. C. 2016. *Sambucus nigra* subsp. Caerulea and *Malva* spp.: Newly Identified Hosts of *Cherry Rasp Leaf Virus. Plant Dis.,* **100**: 867–867.
40. Wylie, S. J., Kueh, J., Welsh, B., Smith, L. J., Jones, M. G. K. and Jones, R. A. C. 2002. A Nonaphid-Transmissible Isolate of Bean Yellow Mosaic Potyvirus Has an Altered NAG Motif in Its Coat Protein. *Arch. Virol.,* **147**: 1813–1820.
41. Wylie, S. J., Luo, H., Li, H. and Jones, M. G. K. 2012. Multiple Polyadenylated RNA Viruses Detected in Pooled Cultivated and Wild Plant Samples. *Arch. Virol.,* **157**: 271–284.

تجزیه و تحلیل متاژنومیک گیاه پنیرک از ایران آلودگی به Malva Vein Clearing Virus را نشان می‌دهد

ز. مرادی، و م. مهرور

چکیده

تا به امروز تنها ژنوم کامل یک جدایه از ویروس رگبرگ روشنی پنیرک (MVCV) در سراسر جهان توالی یابی شده است. در این بررسی، برای اولین بار توالی ژنومی MVCV از گیاه پنیرک (*Malva sylvestris*) در ایران با استفاده از توالی‌یابی عمیق RNA و سپس RT-PCR و توالی یابی سنگر تعیین گردید. براساس نتایج به دست آمده ترادف ژنوم جدایه ایرانی (IR1) شامل ۱۱۰۵۵ نوکلئوتید بدون دنباله (A) Poly بود. ژنوم حاوی یک ORF منفرد به اندازه ۱۰۵۲۷ نوکلئوتید بوده که کدکننده پلی‌پروتئینی حاوی ۳۵۰۸ آمینواسید و وزن مولکولی ۳۹۵/۰۸ می‌باشد. این ژنوم دارای نه سایت برش پروتئولیتیک و موتیف‌های محافظت شده در پروتئین‌های همولوگ موجود در سایر پوتی ویروس‌ها بود. ژن P1 جدایه IR1 به اندازه ۱۲۳٦ نوکلئوتید بلندتر از P1 تنها توالی گزارش شده این ویروس از چین (MVCV-SX, MN116683) می‌باشد. ژنوم کامل جدایه به دست آمده در این بررسی ۸۱/۲۴% تشابه را در سطح نوکلئوتیدی و ۹۱/۹۳% تشابه را در سطح آمینواسیدی با همتای چینی خود نشان داد. تشابه کم در سطح نوکلئوتیدی و آمینواسیدی با تنها جدایه MVCV شناخته شده بیانگر استرینی جدید از این ویروس می‌باشد. آنالیز فیلوژنتیکی ژن پروتئین پوششی نیز نشان داد که جدایه ایرانی بیشترین تشابه (۹٤/۰٦% در سطح نوکلئوتیدی و ۹٦/۰٤% در سطح آمینواسیدی) را با جدایه هلندی این ویروس (NAKT-NL) (FJ539084) دارد. نتایج این پژوهش برای درک بهتر اپیدمیولوژی MVCV در سطح جهانی مفید خواهد بود.

[DOI: 10.52547/jast.25.1.213]