J. Agr. Sci. Tech. (2022) Vol. 24(2): 477-490

The Complete Genome Sequences of Two Recombinant Isolates of *Squash Mosaic Virus* from Iran

M. Gerami Nooghabi[1], M. Mehrvar[1*], and M. Zakiaghl[1]

ABSTRACT

Squash Mosaic Virus **(SqMV) is a *Comovirus* that infects many cucurbit crops worldwide. In this study, the first two complete genome sequences of *SqMV* (BSQ and TSQ) from Iran were determined. The RNA genomes of isolates BSQ and TSQ were, respectively, 5,754 and 5,755 (RNA1) and 3290 and 3271 (RNA2) nucleotides (nt) in length, excluding the 3'-terminal poly (A) tail. RNA1 of both isolates encodes a single polyprotein of 1858 amino acids (aa). The identity between the two Iranian isolates (BSQ and TSQ) was 94.24% nt and 94.82% aa for RNA1 and 88.80% nt and 89.50% aa for RNA2. In comparison to other *SqMV* isolates, BSQ and TSQ shared the highest nucleotide sequence identities of 95.12 % to 93.56 % (RNA1), and 87.59 % to 87.19 % (RNA2), respectively, with the Spanish isolate (RZ-SqMV). Phylogenetic analysis based on complete genome sequences reveals that SqMV isolates cluster into three distinct groups. BSQ was clustered alongside a Spanish isolate in one group and TSQ was separately clustered with a Chinese and US isolates in another group. Recombination analysis revealed that BSQ (RNA1, 2) and TSQ (RNA2) were putative recombinants. BSQ had 6 recombination sites within 5'-UTR, helicase, protease, RdRP (in RNA1), SCP and 3'-UTR (in RNA2) regions, whereas TSQ had 4 recombination sites within 5'-UTR, MP (two breaking points) and LCP region.**

Keywords: *Comovirus*, Cucurbits, Phylogenetic analysis, Recombination, *SqMV*.

INTRODUCTION

Secoviridae is a new plant virus family that was created by combining two former families, *Comoviridae* and *Sequviridae*, and currently contains eight genera, including *Comovirus* (Sanfacon, *et al*., 2009). *Squash Mosaic Virus* (SqMV) is an important virus of cucurbit crops (*Comovirus*, *Secoiviridae*) (Sanfacon, *et al*., 2009), with worldwide distribution (Haudenshield and Palukaitis, 1998; Hu *et al*., 2009). It is a seed-borne virus mainly transmitted by beetles (*Chrysomelidae*) (Nelson and Knuhtsen, 1973; Sanfacon, 2015).

SqMV has a bipartite single-stranded RNA genome encapsidated in separate isometric particles (28 nm). Both RNA molecules contain a genome-linked Viral Protein (VPg) at the 5' end and a poly (A) tail at the 3' end (Haudensheild and Palukaitis, 1998). RNA-1-encodes a large polyprotein that yields the viral polymerase (RdRp), a helicase (hel), a protease, a protease Cofactor (Co-pro) and the genome-linked Viral Protein (VPg), while RNA-2-encodes a polyprotein that yields the Movement Protein (MP) and Large (LCP) and Small Capsid Proteins (SCP) (Yoshida *et al*., 1980).

SqMV was first reported in the USA (California) in 1934 (Kendrick, 1934). Since then, it has frequently been reported in North and South America, Australia, Israel, China, Spain and Japan (Haudensheild and Palukaitis, 1998; Li *et al*., 2015; Nelson and Knuhtsen, 1973; Yoshida *et al*., 1980).

[1] Department of Plant Protection, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Islamic Republic of Iran.

[*] Corresponding author; e-mail: mehrvar@um.ac.ir

In Iran, *SqMV* was first identified in southern province of Bushehr in melon fields (Izadpanah, 1987), and then the occurrence of the virus was confirmed in 18 other provinces by serological and molecular tests (Bananej and Vahdat, 2008; Gholamalizade, *et al*., 2008).

To date, complete genome sequences of several *SqMV* isolates have been sequenced from around the world (Han *et al*., 2002; Hu *et al*., 2009; Li *et al*., 2015; Maina *et al*., 2017). In this study, we aimed to identify the distribution of *SqMV* in Iran and elucidate the full genome sequence of two *SqMV* isolates, for better diagnostics and management of this virus in cucurbits in Iran.

MATERIALS AND METHODS

RNA Extraction and RT-PCR

Severely infected melon (*Cucumis melo* var. inodorus (Persian melons)) leaves (n= 62) with dark green mosaic, blistering, vein clearing, and yellowing symptoms were collected from east Iran (Southern Khorasan Province) between April to June 2017. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Germany) and used as a template for RT-PCR, which was performed using specific primers designed against the Japanese *SqMV* isolates (NC-003799 and NC-003800) (Supplementary Table 1). Cycling conditions consisted of an initial denaturation step at 94°C for 3 minutes, followed by 35 cycles at 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 2.5 minutes, and a final elongation step at 72°C for 7 minutes. Out of 62 samples, 37 *SqMV* positive isolates were identified, two of which were selected for further investigation (BSQ from Boshroye City, and TSQ from Tabas City).

Rapid amplification of cDNA end technology (5' RACE) was used to elucidate the 5’ end of the RNA1 and RNA2 genome of the virus. Following cDNA synthesis with specifically designed RACE primers (Supplementary Table 1), the enzyme Terminal deoxynucleotidyl Transferase (TdT) was used to add a string of identical nucleotides, known as a homopolymeric tail. PCR was then carried out, using reverse specific primers (Supplementary Table 1) and a forward Universal Primer (UP) that binds to the homopolymeric tail (Bertioli and Burrows, 1995; Frohman *et al*., 1988).

Cloning, Sequencing, Phylogenetic Analysis and Recombination Detection

The gel extracted cDNA fragments were cloned into pTG19-T Vector (Vivantis, Malaysia) following the manufacturer’s instructions, and then confirmed by colony-PCR with M13 and specific primers (Supplementary Table 1). Random clones were selected to extract from plasmid DNA using a Qiagen Plasmid Miniprep Kit (Qiagen, Germany). Clones were sequenced (Macrogen Inc., South Korea) and analyzed using ContigExpress program in Vector NTI 11 (Invitrogen, USA).

Phylogenic analysis was carried out using MEGA7 (using Maximum Likelihood (ML) and Neighbor-Joining (NJ) methods), SDT v.1.2 software (by color-coded pairwise identity matrix) (Kumar *et al*. 2016; Muhire *et al*. 2014) [Figure 1 (c, d)], DNAMAM 7 software (with MUSCLE algorithm) (Lynnon ;Biosoft, Quebec, Canada).

The RDP4 program was used to analyze the *SqMV* isolates for putative recombination sites using different algorithms conformed in this software (recombination detection program ver.4; Martin *et al*., 2010). Evidence of recombination was further assessed using the Neighbor-Net method in SplitsTree4 v.4.15.1 (Huson and Bryant, 2006).

RESULTS

Genome Structure

The RNA1 molecule of BSQ (MT724705) and TSQ (MT709101) were, respectively, 5,754 and 5,755 nt in length, excluding the 3'-terminal poly (A) tail. The putative 5'-

UTRs of *SqMV* RNA1 for BSQ and TSQ are 132 and 128 nt in length, respectively, and begins with *CCGGCTCTGCA*, which is a unique 5' sequence for both Iranian isolates compared with other *SqMV* isolates and *Comoviruses* (Chen and Burening, 1992; Di *et al*., 1999) which have a consensus sequence of UAUUAAA. In addition, the 5' UTR of *SqMV*-BSQ and TSQ are shorter by 105 nt compared with other *Comoviruses*. The 3'-UTRs of RNA1 for BSQ and TSQ are 52 and 51 nt in length, respectively. Moreover, the 5' ends of RNA2 of Iranian isolates are 137 (BSQ) and 120 nt (TSQ) in length and the 3' UTRs are 126 (BSQ) and 124 nt (TSQ). Comparable to other *Comoviruses*, the 3' and 5'-UTRs, of both Iranian SqMV isolates, are rich in U residues (Krengiel *et al*., 1993; Yijun *et al*., 2000).

Analysis of BSQ and TSQ RNA1 nucleotide sequences revealed the existence of a single long Open Reading Frame (ORF) that begins at nt position 133 (AUG 133-135) (TSQ) and 129 (AUG 129-131) (BSQ). The termination codon UAA was positioned at nt 5,703-5,705 (BSQ) and 5707-5709 (TSQ). Therefore, a single long ORF consisted of 5574 nt, in which a polyprotein of 1,858 aa's would be produced, for both isolates, with a predicted molecular mass of 204.38 KDa. The base composition of BSQ and TSQ RNA1 were similar to the reported RNA1 sequences from Chinese and Japanese SqMV isolates (T/U 30.4%, C 19%, G 24.2%, and A 26.4%) (Han *et al*., 2002, Hu *et al*., 2009).

The BSQ (MT709103) and TSQ (MT709102) RNA2 were 3,290 and 3,271 nt in length, respectively, excluding the 3'-terminal poly (A) tail. Both sequences encoded a single ORF, which started at AUG_{121} (TSQ) and AUG_{138} (BSQ) and terminated at UAG_{3165} (BSQ) and UAG_{3148} (TSQ). Thus, both isolates encode a polypeptide containing 1,009 aa with a predicted 110.99 KDa molecular mass.

Like other members of *Comoviuses*, *SqMV* RNA1 polyprotein contains domains with homology to the Co-pro, hel, VPg, the protease, and the RdRp. The exact cleavage sites of *Comoviruses* have been identified only in *Cowpea Mosaic Virus* (CPMV) (Wellink *et al*., 1986). However, it has been suggested that amino acid pairs for cleavage sites at hel/VPg (Q/S) and VPg/protease (Q/M) are conserved among *CPMV*, *BPMV* (Bean Pod Mottle Virus), *CPSMV* (Cowpea Sever Mosaic Virus), and *RCMV* (Red Clover Mottle Virus), while other cleavage sites are variable (Petrzik *et al*., 2005). Similarly, the cleavage sites at hel/VPg (Q_{912}/S_{913} in TSQ and Q_{910}/S_{911} in BSQ) and VPg/protease (Q_{940}/M_{941} in TSQ and Q_{938}/M_{939} in BSQ) were found. In contrast, the putative cleavage sites at Co-Pro/hel (Q/D) (Q_{304}/D_{305} in TSQ and Q_{310}/D_{311} in BSQ) and protease/RdRp (Q/C) (Q_{1147}/C_{1148} in TSQ and Q_{1145}/C_{1146} in BSQ) cleavage were unique to *SqMV* (Han *et al*., 2002).

SqMV RNA2 polyprotein has two cleavage sites at MP/LCP (Q/N) and LCP/SCP (Q/S) (Haudensheild and Palukaitis, 1998; Hu *et al*., 1993). In both Iranian isolates, these cleavage sites are conserved and located at Q_{450}/N_{451} and Q_{824}/S_{825} positions for MP/LCP and LCP/SCP, respectively. However, these cleavage sites can be variable (Shanks *et al*., 1986; Van *et al*., 1983) and many potential proteolytic cleavage sites could be presented in the deduced amino acid sequences (Hu *et al*., 1993). Nevertheless, all known motifs were conserved in the Iranian *SqMV* isolates, with only BSQ VPg motif being different from the others (Table 1).

Sequence Comparisons and Phylogenetic Analysis of SqMV Isolates

The complete RNA1 sequences of Iranian isolates BSQ and TSQ shared, respectively, 94.24% nt and 94.82% aa identities. For RNA2, their identities were 88.80% nt and 89.50% aa. Compared with other *SqMV* isolates, BSQ and TSQ shared the highest nt sequence identity with the Spanish isolate (RZ-SqMV, KP223323) with 95.12 to 93.56% (RNA1), and 87.59 to 87.19% (RNA2), respectively. In the VPg gene the

Table 1. All known *Comoviruses* motifs are conserved in Iran *SqMV* isolates, except for the BSQ VPg (genome-linked Viral Protein) motif.

Gene Name	Motifs	Function	Conservation of motifs in TSQ and BSQ
RdRP (RNA depended RNA Polymerase)	GDD CDYS/KXFDG TDGXDK FLKRXF	Catalyzes the polymerase reaction (Rajakaruna *et al*., 2007)	Conserved
Protease	H-Xn-E/D-Xn-C-G-Xn-G-Xn-H-Xn-G	Potential catalytic triad enzyme (Rott *et al*., 1995)	Conserved
MP (Movement Protein)	$LxDx_{14}VAx_4GR$	Virus movement within the host (Holness *et al*., 1989)	Conserved
Polyprotein1	$F\text{-}X_{27}\text{-}W\text{-}X_{11}\text{-}L\text{-}X_{21}\text{-}LX_1E$	A co-factor viral protease (Peter *et al*., 1992)	Conserved
VPg	$E/DX_{1\text{-}3}YX_3NX_{4\text{-}5}R$	Virus replication (Mayo and Ftsch, 1994)	In TSQ ($DX_3YX_3NX_4R$) and in BSQ ($EX_3YX_3NX_4R$)

BSQ isolate shared only 67.86% aa sequence identity with all other *SqMV* isolates, including TSQ (Supplementary Table 2).

The RNA1 polyprotein (Haudensheild and Palukaitis, 1998) and large CP amino acid (Kendrick, 1934) sequences were selected for phylogenetic analysis, with *Bean Pod Mottle Virus* (BPMV, KY G-7 isolate) (NC_003496 and NC-003495) as an outgroup. The topologies of NJ and ML trees were identical and showed that the large CP of *SqMV* isolates formed three distinct clusters. The *SqMV* BSQ isolate clustered with the Spanish isolate (KP223324) while the TSQ isolate clustered with the US (AF059532) and Chinese isolates (EU42060) (Figure 1a). For the RdRp based phylogenetic tree, both Iranian *SqMV* isolates clustered with the Spanish isolate (KP223323) (Figure 1-b).

RNA Recombination Analysis

To check for potential recombination events and the break points within Iran *SqMV* RNA1 and RNA2 genomes, aligned sequences of Japan (NC-003799, AB054688, NC-003800, AB054689), Australia (MF166754, MF166755, MF166756, MF166757), Spain (KP223323, KP223324), China (EU421059, EU421060) and the USA (AF059532, AF059533) were examined using the RDP4 recombination program. Two putative recombination events were supported for RNA1 of BSQ and RNA2 of TSQ and a single recombination event in BSQ RNA2 (Table 2). Based on RDP4 results, TSQ RNA2 was an intraspecific recombinant of two isolates, BSQ and MF166754 (AUS). The pattern of these recombination results is shown in Figure 2. All recombination events were strongly supported using the RDP, GENECONV, CHIMERA, SISCAN, BOOTSCAN, MAXCHI and 3SEQ methods with P-values less than 5×10^{-2} (Figure 1, supplementary Table 3, supplementary Figure 1).

Since reassortments and recombination events can lead to contradictory tree topologies in phylogenetic studies, a NeighborNet phylogenetic network was created and showed that most of the SqMV isolates formed a reticulate network structure, indicating putative recombination events occurred among them [Supplementary Figure 2 (a, b)]. Divergence patterns show a star-like and a

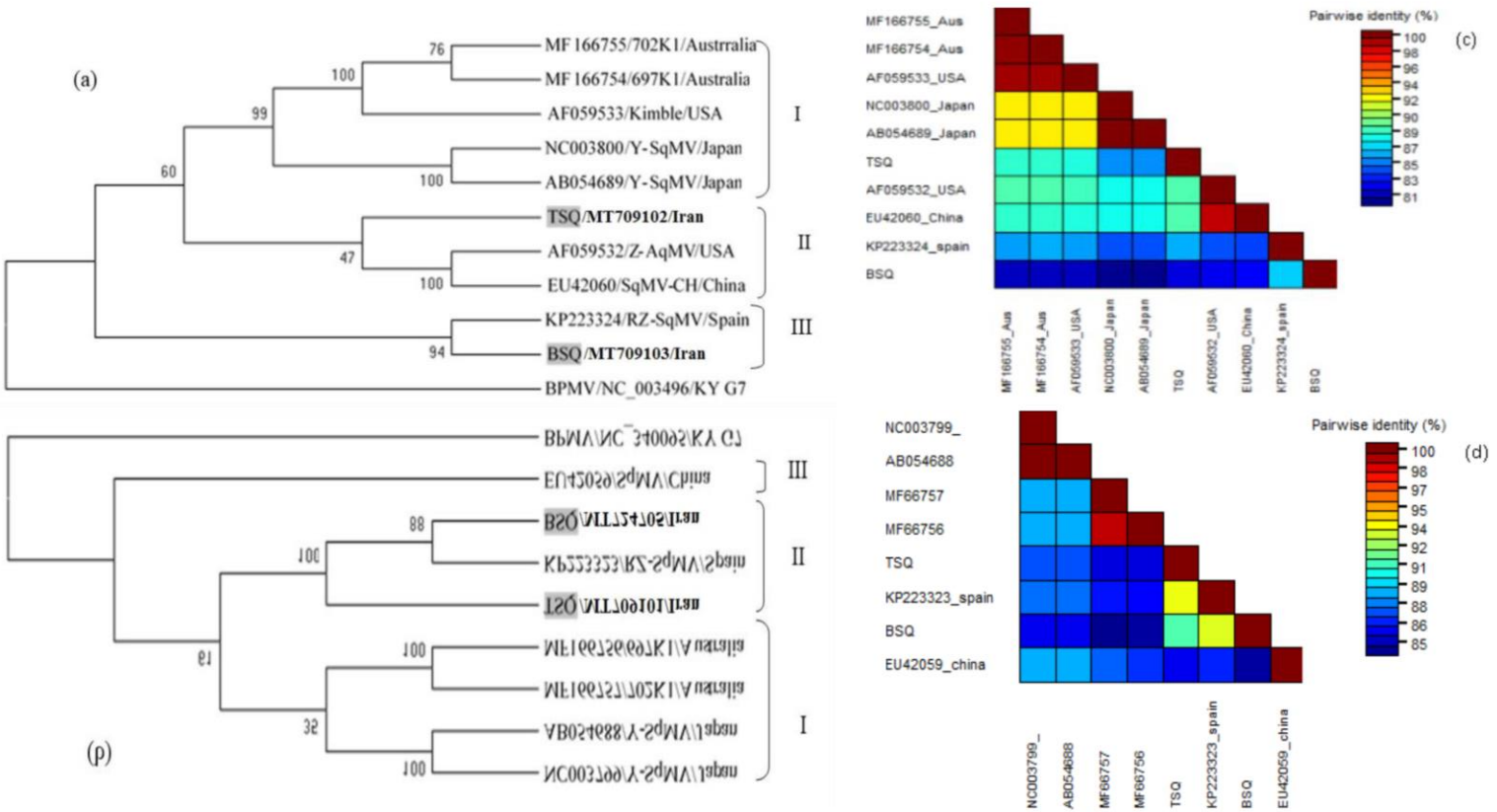


Figure 1. Maximum likelihood tree calculated from: (a) Large Coat Protein (LCP) amino acids sequences and (b) RNA depended RNA Polymerase (RdRP) nucleotides sequence of TSQ and BSQ with 8 other *squash mosaic virus* isolates. Isolates are indicated in the tree by accession number/ isolate name and geographical origin of collection. Multiple sequence alignments were generated by ClustalW assembled in MEGA 7. Numbers at each node indicate bootstrap percentages based on 1000 replications. The nucleotide sequence of a *BPMV* isolate was used as an outgroup. The shaded isolates are presented in this research. Two-dimensional nucleotide diversity plots (c) and (d) based on SDT Muscle alignment.

radiation pattern among the recombinant isolates in group II [Figure 1 (a, b)].

DISCUSSION

In this study, we present, for the first time, the complete genome sequences of two Iranian *SqMV* isolates, *SqMV*-BSQ and *SqMV*-TSQ. The Iranian isolates were most closely related to *SqMV* isolate from Spain (RZ-SqMV). The phylogenetic analysis of *SqMV* showed that the Iranian isolates were in different clades to each other.

Consistent with previous results, high genetic diversity was found in the VPg and MP genes of Iranian isolates (Kwak *et al*., 2013). VPg is linked to virus RNA by the B-OH group of Serine (Jaegle *et al*., 1987) and mutation in this amino acid inhibits virus replication. Although previous mutational analysis of VPg's have shown that aa changes have no impact on virus production in infected cells (Quentin *et al*., 1990), serine is conserved in all *SqMV* VPg proteins, including the BSQ isolate.

The phylogenetic analysis based on LCP amino acid sequences showed that all of the *SqMV* isolates clustered in three main groups, in agreement with previous studies (Han *et al*., 2002., Hu *et al*., 2009., Knuhtsen and Nelson, 1968). BSQ nucleotide sequences of RNA1 and RNA2 shared 94.82 and 87.34% identity with its Spanish counterpart (RZ-SqMV) and only 85.23-86.28 and 81.26-82.58% identity with other isolates of *SqMV*, respectively. This high level of identity with Spanish isolate and low level of identity with other isolates could be explained by recombination. Recombination is an important source of evolution of RNA viruses (Gracia-Arenal *et al*., 2003; Kwak *et al*., 2013), that may occur in various regions of the genome (Boulila, 2007). Recombination has previously been reported for *Comoviruses* (Zhang *et al*., 2007; Bradshow *et al*., 2011), yet this is the first report of possible recombination in *SqMV*. Both Iranian isolates showed putative recombination events, confirmed by the RDP4 program and phylogenetic and nucleotide analysis. Previous studies suggest that recombination hot spots have an A-U rich region (Zhang *et al*., 2007). It is suggested that the A-U rich region acts as a signal to pause, due to the weak A-U pairing during replication, causing the two strands to separate, creating the possibility of changing the template RNA strand (Zhang *et al*., 2007). Interestingly, helicase genes can include hot spots that influence the severity of the symptoms (Gu and Ghabrial, 2005). For the Iranian *SqMV* isolates, 2 recombination events (BSQ RNA1 and TSQ RNA2) occurred in the 5' A-T rich region and 1 event (BSQ RNA2) in the 3'-UTR A-U rich region (Figure 2) (Krengiel *et al*., 1993; Yijun *et al*., 2000) (as *Secoviridae* members) (Walker *et al*., 2014). However, novel recombination patterns were also found in both TSQ and BSQ isolates, in the helicase and protease protein (BSQ RNA1) and the MP (TSQ RNA2).

Split networks (by Splittree4 program) showed reticulate topologies representing several potential recombination events among the different *SqMV* isolates. This suggests that several recombination events might have occurred among other *SqMV* isolates (Supplementary Figure 2).

In *Comoviruses*, the LCP is used as one of the criteria for species demarcation (King *et al*., 2012). Comparable with nucleotide analysis results, phylogenetic analysis of LCP, placed BSQ in phylogenic group III, alongside the Spanish isolate, and TSQ in group II. Previous studies have established that in all of the plant-infecting RNA viruses, such as *Secoviridae*, the coat protein domain evolves, so, viruses can adapt to new environments, host and vector variants (Thompson *et al*., 2014). The changes in the LCP may occur over time due to specific molecular interactions between virus and vector and plant hosts (Chare and Holmes, 2004; Woelk and Holmes, 2002). This may explain the variation observed in the LCP gene of the Iranian *SqMV* TSQ isolate.

Genomic differences between the Iranian *SqMV* isolates, as shown in the phylogenetic

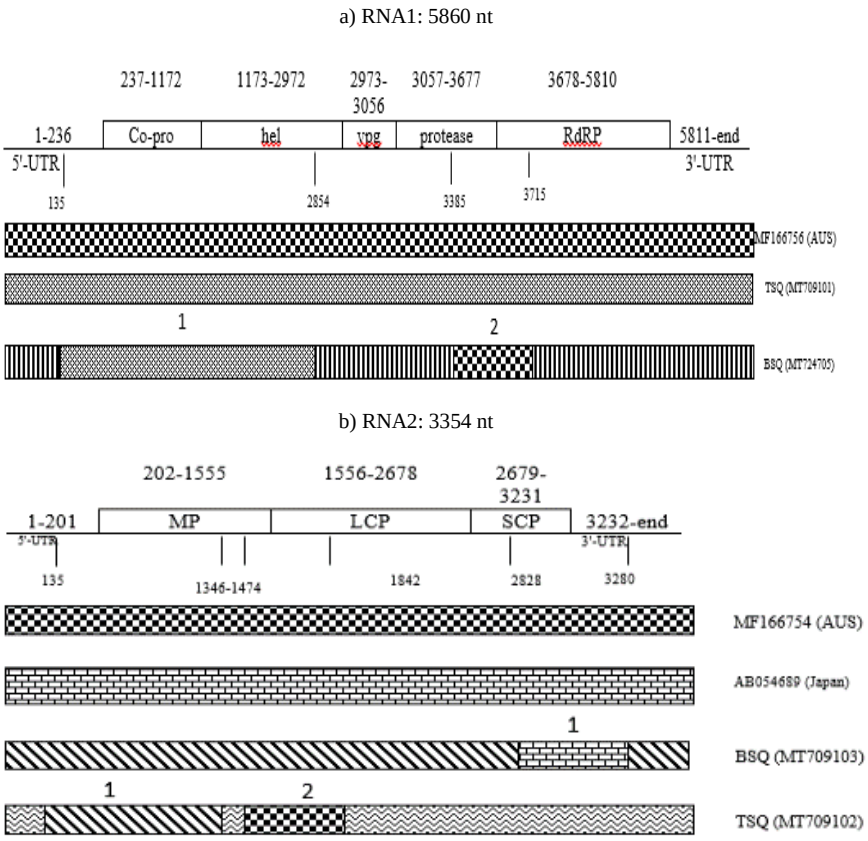
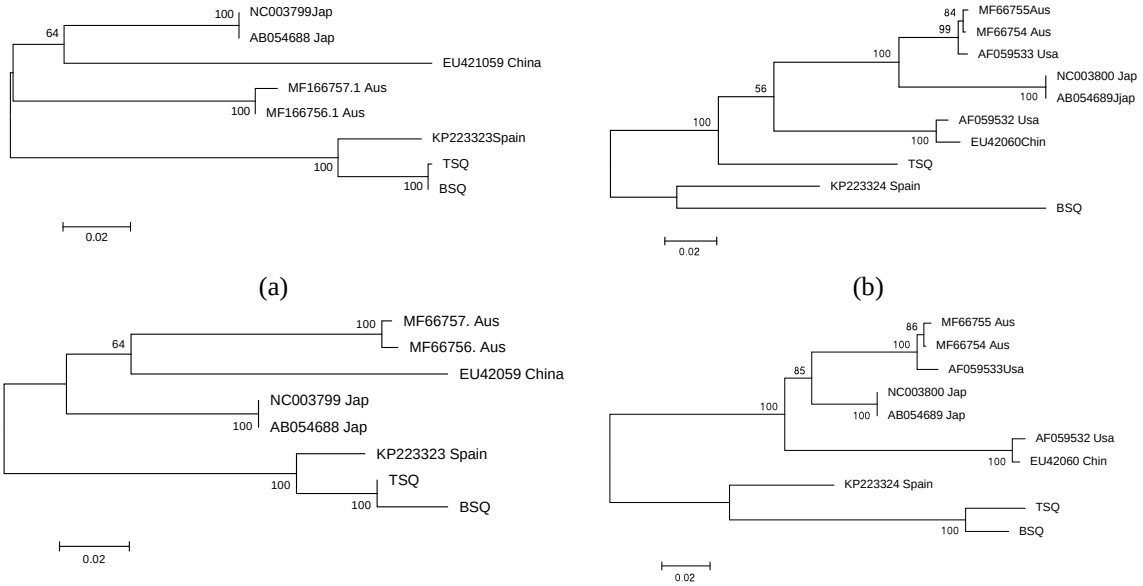


Figure 2. Recombination pattern: (a) Recombination events of RNA1 and (b) Recombination events of RNA2 of TSQ and BSQ genomes detected by RDP4 software.

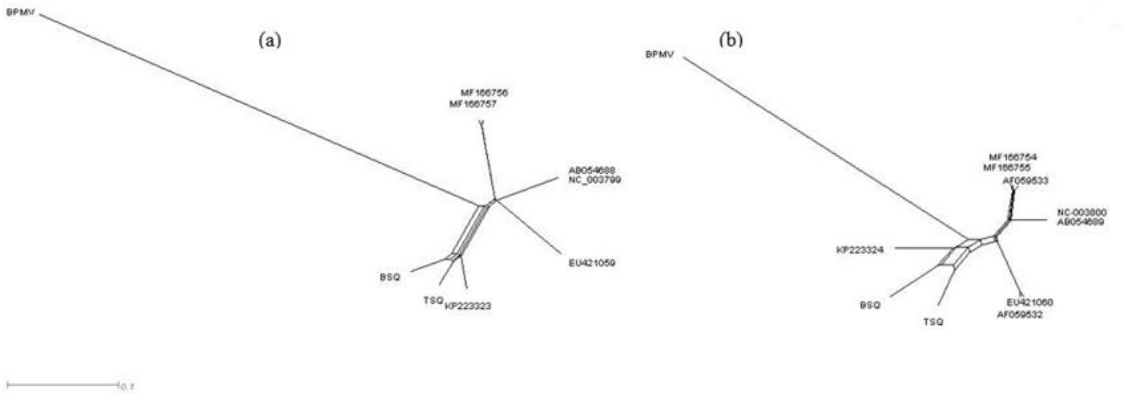
Table 2. Recombination events observed in full-length nucleotide sequences of the *Squash Mosaic Virus* (SqMV) using seven different methods in RDP4.

Recombination (Site)	TSQ (RNA2) (MT709101) Recombination event number		BSQ (RNA1) (MT724705) Recombination event number		BSQ (RNA2) (MT709103) Recombination event
	1	2	1	2	1
Major parent	AF059532 (USA)	EU421060 (China)	KP223323 (Spain)	KP223323 (Spain)	AF059532 (USA)
Minor parent	BSQ (Iran)	MF166754 (AUS)	TSQ (Iran)	MF166756 (AUS)	AB054689 (Japan)
	P value determined by seven different program[a]				
RDP	3.792×10^{-42}	3.736×10^{-13}	3.047×10^{-42}	1.307×10^{-16}	1.141×10^{-11}
GENECONV	1.794×10^{-42}	4.593×10^{-05}	1.005×10^{-37}	7.188×10^{-12}	1.941×10^{-05}
BOOTSCAN	7.806×10^{-30}	2.706×10^{-09}	6.568×10^{-41}	3.357×10^{-12}	1.618×10^{-07}
MAXCHI	1.196×10^{-26}	1.404×10^{-14}	7.787×10^{-23}	1.846×10^{-05}	4.253×10^{-10}
CHIMERA	6.483×10^{-29}	2.567×10^{-12}	7.339×10^{-21}	5.601×10^{-07}	1.733×10^{-10}
SISCAN	1.098×10^{-34}	4.306×10^{-06}	3.441×10^{-21}	9.623×10^{-06}	3.028×10^{-08}
3SEQ	2.984×10^{-13}	1.177×10^{-13}	8.427×10^{-23}	2.863×10^{-12}	1.176×10^{-10}
Beginning break point (nt)	135	1474	135	3385	2828
Ending break point (nt)	1346	1842	2854	3715	3280

[a] P values following each detection method for the respective recombination event are mentioned.



Supplementary Figure 1. Recombination events were confirmed by trees using the segments corresponding to each recombination event and breakpoints. Phylogenetic tree based on nucleotide sequences of Co-factor Protein (a), Large Coat Protein (b), helicase (c) and Movement Protein (d) region of Squash mosaic virus constructed using MEGA7 by the maximum likelihood method, with 1000 bootstrap replications and with a 50% bootstrap threshold score. Recombination event Predict that BSQ Co-facor and helicase regions clustered with TSQ and KP223323 (Spain) as a minor and major parent. Also, TSQ Movement Protein, clustered with BSQ isolates in one group, and based on Large Coat Protein region, TSQ isolate clustered with MF166754 (Australia) in another group which confirmed recombination event and breakpoints.



Supplementary Figure 2. Neighbor-Net phylogenetic networks were inferred from the polyprotein sequences of (a) polyprotein 1 (b) polyprotein2 that were created using the SplitsTree4 v.4.15.1 software program.

Supplementary Table 1 - specific primer pairs designed based on Japanese isolates (NC_003799, NC_003800) to make the complete genome sequence of Squash mosaic virus (a) and specific reverse primers for 5' RACE of it.

Primer name	Accession number	Location in Genome	Length of primer	sequence	Length of fragment
Specific primer pairs to make the complete genome sequence of Squash mosaic virus					
F1-2	NC_003799	105-127	22	5'-CCGGCTCTGCACTTTACGGCCC-3'	1617
R1	NC_003800	1702-1723	22	5'-GTTTTTCCAACGCGAGAAGTTC-3'	1722
F2	NC_003799	1702-1723	22	5'-GAACTTCTCGCGTTGGAAAAAC-3'	1660
R2	NC_003800	3341-3362	21	5'-AACATCCCAAGTAGTGTGTGG-3'	1660
F3	NC_003799	3341-3362	21	5'-CCACACACTACTTGGGATGTT-3'	1337
R3	NC_003800	4658-4678	20	5'-AGGAGATTCATTCTCATGGT-3'	1337
F4	NC_003799	4658-4678	20	5'-ACCATGAGAATGAATCTCCT-3'	1226
R4	EU421059	5764-5784	22	5'-AAAAGAAAAAGCAACATAGTAA-3'	1226
F5-2	NC_003799	121-143	22	5'-GCACTTTACGGCTTCGGTAGAT-3'	1589
R5	NC_003800	1689-1711	23	5'-CGAAGGGAACTAGTATCATCCAA-3'	1710
F6	NC_003799	1364-1384	24	5'-TGTGTACAAGATTGGTGGAGATGC-3'	1977
R6	NC_003800	3359-3380	21	5'-AGGCTTCTAAAGCGAACTGGG-3'	1977
R7	NC_003800	3328-3354	26	5'-AAAAAGAAAAGAGAATGCATCAAAGA-3'	2129
specific primer for 5' RACE					
R150-1	NC_003800	1-150	22	5'-AGGAGGGAGAAAAGCAATTGCA-3'	150
R250-1	NC_003799	1-250	21	5'-AAATTCATTTTGAATTAATTTA-3'	250
R150-2	NC_003800	1-150	21	5'-TTAGAATGGTTGAGGAAGAAA-3'	150
R250-2	NC_003800	1-250	22	5'-TTATGGTAACCCTCAAAACACT-3'	250

Supplementary Table 2- Percent amino acids (Top) and nucleotides (bottom) sequence identity of Squash mosaic virus TSQ (Iran isolate) with other Squash Mosaic Virus isolates.

Isolate	country	Accession No.	Poly protein 1	Poly protein 2	RdRP	Helicase	Co-factor	VPg	Protease	MP	LCP	SCP	Total CP
Y-SqMV	Japan	NC_003799 (RNA1)	96.23	91.77	94.66	97.37	96.15	100	98.07	91.13	89.57	97.83	92.29
		NC_003800 (RNA2)	87.93	85.25	87.34	87.89	88.46	88.10	89.21	82.56	85.20	91.85	87.40
702K1	Australian	MF166757(RNA1)	95.91	92.07	94.37	97.00	95.19	100	98.55	90.91	90.37	98.37	93.01
		MF166755(RNA2)	86.48	86.63	85.79	87.00	86.97	84.52	86.80	81.97	88.59	94.02	90.38
697K1	Australian	MF166756 (RNA1)	95.80	92.07	94.23	96.83	95.19	100	98.55	90.91	90.37	98.37	93.01
		MF166754 (RNA2)	86.57	86.67	85.61	87.17	87.50	84.52	86.96	81.89	88.50	94.57	90.50
RZ-SqMV	Spain	KP223323 (RNA1)	97.79	92.77	95.78	99.17	99.04	100	98.55	93.57	89.3	97.83	92.11
		KP223324 (RNA2)	95.12	87.19	93.95	96.17	95.41	96.43	95.49	87.66	86.10	88.22	86.80
Y-SqMV	Japan	AB054688 (RNA1)	96.23	91.77	94.66	97.33	96.15	100	98.07	91.13	89.57	97.83	92.29
		AB054689 (RNA2)	87.93	85.25	87.34	87.89	88.46	88.10	89.21	82.56	85.20	91.85	87.40
SqMV-CH	China	EU421059(RNA1)	94.67	91.18	93.53	95.50	93.27	100	97.58	89.36	90.37	97.28	92.65
		EU421060 (RNA2)	86.43	86.07	85.98	86.22	86.75	85.71	88.08	80.56	89.22	93.12	90.50
Z-SqMV	USA	AF059532 (RNA2)	-	90.88	-	-		-	-	89.14	89.84	97.28	92.29
				86.17						80.71	89.04	93.66	90.56
Kimble	USA	AF059533 (RNA2)	-	92.07	-	-		-	-	91.13	90.11	98.37	92.83
				86.14						81.37	88.24	93.48	89.96
BSQ	Iran		94.24	88.80	91.70	96.83	100	67.86	90.34	97.12	77.01	92.39	82.08
			94.82	89.50	91.42	98.11	99.89	86.90	90.34	96.08	81.55	89.49	84.17

Supplementary Table 3- Percent amino acids (Top) and nucleotides (bottom) sequence identity of Squash mosaic virus BSQ (Iran isolate) with other SqMV isolates.

Isolate	country	Accession No.	Poly protein 1	Poly protein 2	RdRP	Helicase	Co-factor	VPg	Protease	MP	LCP	SCP	Total CP
Y-SqMV	Japan	NC_003799 (RNA1)	93.54	88.40	93.39	94.33	96.15	67.86	91.30	90.02	83.96	93.48	87.10
		NC_003800 (RNA2)	86.59	83.60	86.03	86.50	88.57	77.38	86.96	82.78	79.95	92.93	84.23
702K1	Australian	MF166757(RNA1)	93.16	88.31	92.97	94.00	95.19	67.86	91.79	89.80	83.69	94.02	87.10
		MF166755(RNA2)	85.62	83.37	84.48	85.56	87.07	73.81	89.05	82.63	80.93	90.04	83.93
697K1	Australian	MF166756 (RNA1)	93.16	88.31	93.11	93.83	95.19	67.86	91.79	89.80	83.69	94.02	87.10
		MF166754 (RNA2)	86.86	83.23	84.62	85.72	87.61	73.81	89.86	82.14	81.02	89.67	83.87
RZ	Spain	KP223323 (RNA1)	95.05	89.69	94.37	96.17	99.04	67.86	91.79	93.13	83.69	93.48	86.92
		KP223324 (RNA2)	93.56	87.59	93.35	94.39	95.51	84.52	89.53	88.47	86.01	88.59	86.86
Y-SqMV	Japan	AB054688 (RNA1)	93.54	88.4	93.39	94.33	96.15	67.86	91.30	90.02	83.96	93.48	87.10
		AB054689 (RNA2)	86.59	83.60	86.03	86.50	88.57	77.38	86.96	82.78	79.95	92.93	84.23
SqMV-CH	China	EU421059(RNA1)	92.03	87.22	92.41	92.50	93.27	67.86	90.82	88.47	83.96	90.76	86.20
		EU421060 (RNA2)	85.23	82.38	84.76	84.78	86.86	75.00	86.96	80.78	82.00	86.96	83.63
Z	USA	AF059532 (RNA2)	-	86.82	-	-		-	-	88.25	83.16	90.76	85.66
				82.28						80.93	81.73	86.59	83.33
Kimble	USA	AF059533 (RNA2)	-	88.4	-	-		-	-	90.02	83.69	94.02	87.10
				83.04						81.89	80.84	90.22	83.93
TSQ	Iran		94.24	88.80	91.70	96.83	100	67.86	90.34	97.12	77.01	92.39	82.08
			94.82	89.50	91.42	98.11	99.89	86.90	90.34	96.08	81.55	89.49	84.17

trees, may be due to climatic differences. The *SqMV* BSQ isolate was collected from a temperate zone in Iran, compared with *SqMV* TSQ, which was collected from a tropical region. Another contributing factor could be the host's composition. Both Iranian *SqMV* isolates were isolated from melon, but in Boshroye City (BSQ), melons and pumpkins were planted together in the same field, therefore, the genetic variation observed in the Iranian *SqMV* isolates could be due to exposure to different environmental conditions, hosts, and vectors, as has been shown for other viruses (Gracia Arenal *et al*., 2003).

SqMV is usually transmitted via infected seed and mechanically by beetles during agricultural practices (Alvarez and Campbell, 1978; Luis-Arteaga, 1994; Luis-Arteaga and Alvarez, 1998). Movement of the virus, via infected seeds or beetles, may explain the high sequence similarities observed between *SqMV* BSQ and TSQ and other intercontinental *SqMV* isolates, and is an important factor in the epidemiology of the melon disease caused by this virus (Han *et al*., 2002). The origin of melon (*Cucumis melo* var. Inodorus) as a major crop worldwide (Mirtalebi and Banihashemi, 2019) is not clear (Naroui Rad *et al*., 2010), however, historical reports show that melon was previously cultivated in Persia (Iran), and was exported from Iran to Europe (Hasandokht *et al*., 2013; Walters, 1989). Therefore, the genetic similarities observed between the *SqMV* isolates from distant geographical isolates (Asian and European isolates) may be due to economic exchange (importation and exportation) of plant material (seeds and fresh cucurbits), leading to the long-distance movement of the virus (Sanfacon, 2015).

ACKNOWLEDGEMENTS

The author Mohadese Gerami Nooghabi received a research grant from the Ministry of Science, Research and Technology of Iran.

REFERENCES

1. Alvarez, M. and Campbell, R. N. 1978. Transmission and Distribution of Squash Mosaic Virus in Seeds of Cantaloupe. *Phytopathology*, **68(3):** 257-263.
2. Bananej, K. and Vahdat, A. 2008. Identification, Distribution and Incidence of Viruses in Field Grown Cucurbit Crops of Iran. *Phytopathol. Mediterr*., **47**: 247-257.
3. Bertioli, D. J. and Burrows, P. R. 1995. A Simple RACE Method Based on CTAB Precipitation. *Method. Mol. Cell. Biol*., **5:** 118–121.
4. Boulila, M. 2007. Phylogeny and Genetic Recombination of Grapevine Fanleaf Virus Isolated from Naturally Infected Vineyards in Tunisia. *Phytopathol. Mediterr*., **46:** 285–294.
5. Bradshow, J. D., Zhang, C., Hill, J. H. and Rice, M. E. 2011. Landscape Epidemiology of *Bean Pod Mottle Comovirus*: Molecular Evidence of Heterogenous Sources. *Arch. Virol*., **156:** 1615-1619.
6. Chen, X. and Bruening, G. 1992. Nucleotide Sequence and Genetic Map of Cowpea Severe Mosaic Virus RNA 2 and Comparisons with RNA 2 of Other *Comoviruses*. *Virology*, **187(2):** 682–692.
7. Di, R., Hu, C. C. and Ghabrial, S. A. 1999. Complete Nucleotide Sequence of Bean Pod Mottle Virus RNA1 Sequence Comparisons and Evolutionary Relationships to Other *Comoviruses*. *Virus Genes*, **18(2):** 129–137.
8. Chare, E. R. and Holmes, E. C. 2004. Selection Pressures in the Capsid Genes of Plant RNA Viruses Reflect Mode of Transmission. *J. Gen. Virol*., **85(10):** 3149–3157.
9. Frohman, M. A., Dush, M. K. and Martin, G. R. 1988. Rapid Production of Full-Length cDNAs from Rare Transcripts. *PNAS*, **85(23):** 8998–9002.
10. Gibbs, A. J. and Ohshima, K. 2010. *Potyviruses* and the Digital Revolution. *Annu. Rev. Phytopathol.*, **48:** 205-223.
11. Gholamalizadeh, R., Vahdat, A., Keshavarz, T., Elahinia, S. A., Sharaeen, N. and Bananej, K. 2008. Occurrence of *Ourmia Melon Virus* in the Guilan Province, Iran. *Acta Virol.*, **5:** 113-118.
12. Gracia Arenal, F., Fraile, A. and Malpica, J. M. 2003. Variation and Evolution of Plant

Virus Populations. *Int. Microbiol*., **6(4):** 225-232.
13. Gu, H. and Ghabrial, S. A. 2005. The *Bean Pod Mottle Virus* Proteinase Cofactor and Putative Helicase Are Symptom Severity Determinants. *Virology*, **333:** 271-283.
14. Han, S. S., Yoshida, K., Karasev, A. V. and Iwanami, T. 2002. Nucleotide Sequence of a Japanese Isolate of *Squash Mosaic Virus*. *Arch. Virol*., **147(2):** 437–443.
15. Hasandokht, M. R., Zamani, Z., Moghadam, M. R. F. and Kashi, A. 2013. Genetic Diversity among Melon Accessions from Iran and Their Relationships with Melon Germplasm of Diverse Origins Using Microsatellite Markers. *Plant Syst. Evol*., **300(1):** 139-151.
16. Haudenshield, J. S. and Palukaitis, P. 1998. Diversity among Isolates of *Squash Mosaic Virus*. *J. Gen. Virol.,* **79(10):** 2331–2341.
17. Holness, C. L., Lomonossoff, G. P., Evans, D. and Maule, A. J. 1989. Identification of the Initiation Codons for Translation of *Cowpea Mosaic Virus* Middle Component RNA Using Site-Directed Mutagenesis of an Infectious cDNA Clone. *Virology*, **172:** 311–320
18. Hu, J. S., Pang, S. Z., Nagpala, P. G., Siemieniak, D. R., Slightom, J. L. and Gonsalves, D. 1993. The Coat Protein Genes of *Squash Mosaic Virus*, Cloning, Sequence Analysis, and Expression in Tobacco Protoplasts. *Arch. Virol*., **130(1-2):** 17-31.
19. Hu, J., Zhou, T., Liu, L., Peng, B., Li, H., Fan, Z. and Gu, Q. 2009. The Genomic Sequences of a Chinese Isolate of Squash Mosaic Virus with Novel 5' Conserved Ends. *Virus Genes,* **38:** 475–477.
20. Huson, D. H. and Bryant, D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.,* **23:** 254–267.
21. Izadpanah, K. 1987. Squash Mosaic Virus as Cause of Melon Vien-Banding Mosaic in Iran. *J. Phytopathol.*, **120:** 276-282.
22. Jaegle, M., Wellink, J. and Goldbach, R. 1987. The Genome-Linked Protein of *Cowpea Mosaic Virus* Is Bound to the 5_ Terminus of Virus RNA by a Phosphodiester Linkage to Serine. *J. Gen. Virol.,* **68:** 627–632.
23. Kehoe, M. A., Coutts, B. A., Buirchell, B. and Jones, R. A. C. 2014b. Split Personality of a *Potyvirus*: To Specialize or Not to Specialize. *PLoS One,* **9:** e105770.
24. Kendrick, J. B. 1934. Cucurbit Mosaic Transmitted by Muskmelon Seed. *Phytopathology,* **24:** 820-823.
25. King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. 2012. Genus *Comovirus*. In: "*Virus Taxonomy: The Classification and Nomenclature of Viruses*". The Ninth Report of the International Committee on Taxonomy of Viruses, Academic Press, San Diego, USA, PP. 887-889.
26. Knuhtsen, H. K. and Nelson, M. R. 1968. Identification of Two Serotypes in *Squash Mosaic Virus. Phytopathology*, **58**: 345–347.
27. Krengiel, R., Vincente, A. C. P., Weyne, M., Shindo, N., Brioso, P. S. T., Felix, D. B. and Villaroel, R., Oliveira, D.E. and Timmerman, B. 1993. Molecular Cloning and Sequence Analysis of a Segment from *Andean Potato Mottle Virus* B RNA Encoding the Putative RNA Polymerase. *J. Gen .Virol*., **74(2):** 315–318.
28. Kumar, S., Stecher, G. and Tamura, K. 2016. Mega7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33(7):** 1870-1874.
29. Kwak, H. R., Kim, M. K., Lee, Y. J., Seo, J. K., Kim, J. S., Kim, K. H., Cha, B. N. and Choi, H. S. 2013. Molecular Characterization and Variation of *Broad Bean Wilt Virus* 2 Isolates Based on Analysis of Complete Genome Sequences. *Plant Pathol. J*., **29(4):** 397-409.
30. Li, R., Gao, A. S., Berendsen, B. S., Fei, C. Z. b. and Linga, b. K. 2015. Complete Genome Sequence of a Novel Genotype of *Squash Mosaic Virus* Infecting Squash in Spain. *Genome Announc*., **3(1):** 13-14.
31. Luis-Arteaga, M. 1994. Enfermedades Producidas por Virus. In: *"Enfermedades de las Cucurbitáceas en España"*. Agropubli Edition, Monografías de la Sociedad Española de Fitopatología, Valencia Phytoma-España, PP. 73-91.
32. Luis-Arteaga, M. and Alvarez, J. M. 1998. Occurrence, Distribution, and Relative Incidence of Mosaic Viruses Infecting Field-Grown Melon in Spain. *Plant Dis*., **82(9):** 979-982.
33. Maina. S., Owain, R. E. and Roger, A. C. 2017. Two Complete Genome Sequences of

Squash Mosaic Virus from 20-Year-Old Cucurbit Leaf Samples from Australia. *Genome Announc*., **5(32):** 00778-17.
34. Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. And lefeuvre, P. 2010. RDP3: A Flexible and Fast Computer Program for Analysis Recombination. *Bioinformatics*, **26(19):** 2462-2463.
35. Mayo, M. A. and Ftsch, C. 1994. A Possible Consensus Sequence for VPg of Viruses in the Family *Comoviridae*. *FEBS J.*, **354(2):** 129–130.
36. Mirtalebi, M. and Banihashemi, Z. 2019. Effect of Salinity on Root Rot of *Cucumis melo* L. Caused by *Phytophthora melonis*. *J. Agr. Sci. Tech.*, **21:** 209-220.
37. 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(9):** e108277.
38. Naroui Rad, M. R., Allahdoo, M. and Fanaei, H. R. 2010. Study of Some Yield Traits Relationship in Melon (*Cucumis melo* L) Germplasm Gene Bank of Iran by Correlation and Factor Analysis. *Trakia J. Sci*., **8(1):** 27-32.
39. Nelson, M. R. and Knuhtsen, H. K. 1973. Squash Mosaic Virus Variability Review and Serological Comparisons of Six Biotypes. *Phytopathology*, **63(7):** 920-926.
40. Peter, S. A., Voorhorst, W. G. B., Wery, J., Wellink, J. and Van Kammen, A. 1992. A Regulatory Role for the 32K Protein in Proteolytic Processing of *Cowpea Mosaic Virus* Polyproteins. *Virology*, **191**: 81-89.
41. Petrzik, K., Hola, M. and Spak, J. 2005. Complete Nucleotide Sequence of *Radish Mosaic Virus* RNA Polymerase Gene and Phylogenetic Relationships in the Genus *Comovirus*. *Acta Virol.*, **49(4):** 271–275.
42. Quentin, R., Richard, J. K. and Eckard, W. 1990. Characterization of *Poliovirus* Clones Containing Lethal and Nonlethal Mutations in the Genome-Linked Protein Vpg. *J. Virol.*, **64(6):** 2967-2975.
43. Rajakaruna, P., Khandekar, S., Meulia, T. and Leisner, S. M. 2007. Identification and Host Relations of *Turnip Ringspot Virus*, a Novel *Comovirus* from Ohio. *Plant Dis.*, **91(10):** 1212-1220.
44. Rott, M. E., Gilchrist, A., Lee, L. and Rochon, D. A. 1995. Nucleotide Sequence of Tomato Ring Spot Virus RNA1. *J. Gen. Virol.*, **76:** 465-473.
45. Sanfacon, H., Wellink, J., Le Gall, O. Kaeasev, A., van der Vlugt, R. and Wetzel, T. 2009. *Secoviridae*: A Proposed Family of Plant Viruses within the Order Picornavirales that Combines the Families *Sequiviridae* and *Comoviridae*, the Unassigned Genera *Cheravirus* and *Sadwavirus*, and the Proposed Genus *Torradovirus*. *Arch. Virol.*, **154(5):** 899–907.
46. Sanfaçon, H. 2015. *Secoviridae: A Family of Plant Picorna-Like Viruses with Monopartite or Bipartite Genomes.* Encyclopedia of Life Sciences, John Wiley and Sons, Ltd., Chichester, PP. 1-14.
47. Shanks, M., Stanley, J. and Lomonossoff, G. P. 1986. The Primary Structure of *Red Clover Mottle Virus* Middle Component RNA. *Virology*, **155(2):** 697-706.
48. Thompson, J. R., Kamath, N. and Perry, K. L. 2014. An Evolutionary Analysis of the *Secoviridae* Family of Viruses. *PloS One*, **9(9):** e106305.
49. Valli, A., Lopez-Moya, J. J. and Gracia, J. A. 2007. Recombination and Gene Duplication in the Evolutionary Diversification of P1 Proteins in the Family *Potyviridae*. *J. Gen. Virol*., **88**: 1016-1028.
50. Van Wezenbeek, P., Verver, J., Harmsen, J., Vos, P. and Van Kammen, A. 1983. Primary Structure and Gen Organization of the Middle Component RNA of *Cowpea Mosaic Virus. EMBO J.*, **2(6):** 941-946.
51. Walker, M., Chisholm, J., Wei, T., Ghoshal, B., Saeed, H., Rott, M. and Sanfacon, H. 2014. Complete Genome Sequence of Three *Tomato Ringspot Virus* Isolates: Evidence for Reassortment and Recombination. *Arch. Virol*. **160(2):** 543-457.
52. Walters, T. W. 1989. Historical Overview on Domesticated Plants in China with Special Emphasis on the *Cucurbitaceae*. *Econ. Bot.*, **43(3):** 297–313.
53. Wellink, J., Rezelman, G, Goldbach, R. and Beyreuther, K. 1986. Determination of the Proteolytic Processing Sites in the Polyprotein Encoded by the Bottom-Component RNA of *Cowpea Mosaic Virus*. *J. Virol.*, **59(1):** 50–58.
54. Woelk, C. H. and Holmes, E. C. 2002. Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol.* 19(12), 2333–2336.

55. Yijun, Q., Xueping, Z. and Debao, L. 2000. Complete Nucleotide Sequence and Infectious cDNA Clone of the RNA 1 of a Chinese Isolate of *Broad Bean Wilt Virus* 2. *Virus Genes*, **20(3):** 201-207.
56. Yoshida, K., Goto, T., Nemoto, M. and Tsuchizaki, T. 1980. Squash mosaic virus isolated from melon (*Cucumis melo* L.) in Hokkaido. *Annu. Phytopathol. Soc. Japan*, **46(3):** 349–356.
57. Zhang, C., Gu, H. and Ghabrial, S. A. 2007. Molecular Characterization of Naturally Occurring RNA1 Recombinants of the *Comovirus Bean Pod Mottle Virus*. *Phytopathology*, **97:** 1255-1262.

توالی یابی کامل ژنوم دو جدایه نوترکیب ویروس موزاییک کدو (Squash Mosaic Virus, SqMV) از ایران

م. گرامی نوقابی، م. مهرور، و م. زکی عقل

چکیده

ویروس موزاییک کدو (SqMV) یک کوموویروس است که کدوییان را در سرتاسر دنیا آلوده می کند. در این پژوهش، برای اولین بار دو جدایه ایرانی ویروس SqMV (BSQ و TSQ) به طور کامل توالی یابی شدند. RNA ژنومی جدایه های BSQ وTSQ بدون دنباله Poly A به ترتیب 5754 و 5755 (برای RNA1) و 3290 و 3271 (برای RNA2) نوکلئوتید طول داشتند. RNA1 هر دو جدایه ایرانی پلی پروتئین واحدی به طول 1858 آمینواسید را کد می نماید. میزان شباهت دو جدایه ایرانی (BSQ و TSQ) 94.24٪ در سطح نوکلئوتیدی و 94.82٪ در سطح آمینواسیدی، برای RNA1 و 88.80٪ در سطح نوکلئوتیدی و 89.50٪ در سطح آمینواسیدی برای RNA2 می باشد. در مقایسه با سایر جدایه های این ویروس، BSQ و TSQ به ترتیب، بیشترین درصد شباهت یعنی 93.56٪ و 95.12٪ (RNA1)، و 87.19٪ و 87.59٪ (RNA2) در سطح نوکلئوتیدی را با جدایه اسپانیائی(-RZ SqMV) این ویروس دارند. بررسی های فیلوژنتیکی بر اساس توالی کامل جدایه ها نشان داد که جدایه های ویروس SqMV در سه گروه متفاوت قرار می گیرند. BSQ در کنار جدایه اسپانیایی در یک گروه و TSQ به همراه جدایه های چینی و آمریکایی در گروهی مجزا قرار می گیرند. بررسی نوترکیبی نشان داد که BSQ (RNA1 و RNA2) و TSQ (RNA2) نوترکیب هستند. BSQ دارای نواحی نوترکیبی در 5'-UTR، هلیکاز، پروتئاز، RdRP (در RNA1)، پوشش پروتئینی کوچک (SCP) و 3'-UTR (در RNA2) می باشد، در حالی که TSQ دارای 4 ناحیه نوترکیبی در 5'-UTR، پروتئین حرکتی (دو ناحیه نوترکیب) و پوشش پروتئینی بزرگ (LCP) است.