

Molecular characterization of two sugarcane streak mosaic virus isolates from Iran with emphasis on its population structure

Z. MORADI¹, E. NAZIFI², M. MEHRVAR^{1*}

¹Department of Plant Pathology, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran; ²Department of Biology, Faculty of Basic Sciences, University of Mazandaran, Babolsar, Iran

Received January 18, 2017; revised August 15, 2017; accepted September 29, 2017

Summary. – Sugarcane streak mosaic virus (SCSMV; the genus *Poacevirus*, the family *Potyviridae*) is an economically important causal agent of sugarcane mosaic disease in Asia. In this study, for the first time, we determined the complete genomic sequences of two Iranian SCSMV isolates, IR-Khuz6 and IR-Khuz57 from sugarcane. The sequences of both isolates were 9,782 nucleotides (nt) long, excluding the 3' poly(A) tail. Both of them contained a 5'-untranslated region (UTR) of 199 nt, an open reading frame of 9,393 nt encoding a polyprotein of 3,130 amino acids (aa), and 3'-UTR of 190 nt. SCSMV-IR-Khuz6 and IR-Khuz57 genome nucleotide sequences were in 97.7% identical and shared identities of 81-92.4% with 10 other SCSMV isolates available in the GenBank. The highest identity was shared with the isolate PAK (NC_014037) from Pakistan. When separate genes were compared, most of the genes shared the highest identities with Pakistani isolate. Phylogenetic analysis of the complete genomic nucleotide and polyprotein amino acid sequences reveals that all SCSMV isolates clustered into two main groups. Both IR-Khuz6 and IR-Khuz57 clustered with isolates from Pakistan (PAK) and India (IND671) in group II but formed a separate subgroup. Population genetic analysis revealed greater between-group than within-group evolutionary divergence values, further supporting the results of the phylogenetic analysis. The results indicate that gene flow and selection pressure are important evolutionary factors shaping the genetic structure of SCSMV populations with implications for global exchange of sugarcane germplasm.

Keywords: sugarcane streak mosaic virus; complete genome; phylogenetic analysis; Iran

Introduction

Sugarcane mosaic disease (SMD), one of the most widespread virus diseases of sugarcane, causes significant economic losses to the sugar industry (Grisham, 2000; Li *et al.*, 2011). Sugarcane mosaic virus (SCMV, *Potyvirus*), sorghum mosaic virus (SrMV, *Potyvirus*) and sugarcane streak mosaic virus (SCSMV, *Poacevirus*) are three major RNA viruses which are widely prevalent in most of the Asian countries and are the causative agents of mosaic disease in sugarcane

(Hema *et al.*, 2002; Chatenet *et al.*, 2005; Parameswari *et al.*, 2013; Luo *et al.*, 2016). The typical symptoms of SCSMV are similar to those caused by SCMV and estimated yield losses vary from 30 to 80% (Viswanathan *et al.*, 2008; Bagyalakshmi *et al.*, 2012, Parameswari *et al.*, 2013). SCSMV can be transmitted mechanically, and no insect vector has yet been identified (Hema *et al.*, 2002; Xu *et al.*, 2010; King *et al.*, 2012). SCSMV has flexuous filamentous particles (890×15 nm) which encapsidate a single-stranded, positive-sense RNA genome of about 10,000 nucleotides (Hema *et al.*, 1999). The genomic RNA of SCSMV contains single open reading frame (ORF) which encodes a large polyprotein that is autocatalytically cleaved into at least ten mature proteins (Tatineni *et al.*, 2009; Xu *et al.*, 2010; Li *et al.*, 2011). Similar to other members of the family *Potyviridae*, an additional protein (pretty interesting *Potyviridae* ORF, PIPO) (Chung *et al.*,

*Corresponding author. E-mail: mehrvar@um.ac.ir; phone: +98-51-38805831.

Abbreviations: PIPO = pretty interesting *Potyviridae* ORF; SCSMV = sugarcane streak mosaic virus; SCMV = sugarcane mosaic virus; UTR = untranslated region

2008) was also predicted to be expressed as a fusion protein with the N-terminal part of P3 (P3N-PIPO) via ribosomal frameshifting or transcriptional slippage (Viswanathan *et al.*, 2015). SCSMV was first identified by Hall *et al.* (1998) from quarantined sugarcane germplasm imported from Pakistan into the USA and subsequently detected in Asia (Chatenet *et al.*, 2005; He *et al.*, 2016). So far it has been reported from sugarcane in India, Pakistan, China, Japan, Thailand, Sri Lanka, Bangladesh, Myanmar, Vietnam, Indonesia and most recently in Iran (Hall *et al.*, 1998; Chatenet *et al.*, 2005; Damayanti and Putra, 2011; Li *et al.*, 2011; Parameswari *et al.*, 2013; Moradi *et al.*, 2015). Sugarcane is one of the most economically important crops in north and southwest of Iran. The crop is reported to be affected by several viral diseases, of which mosaic is most widely spread (Moradi *et al.*, 2016, 2017b). Knowledge of the evolutionary history of these viruses will be helpful to understand effects of variation caused by mutation, recombination, selection pressure and adaptation in viral populations which leads to understanding of such important features of their biology as changes in virulence, population structure, and their 'emergence' as new epidemics (Gibbs and Ohshima, 2010; He *et al.*, 2014). Complete sequence data are critical for understanding of different aspects of the virus biology, diversity and evolutionary history, and designing new control strategies. In addition, sequence analysis at the complete genome level, provides new insights into the relationship among isolates from different areas, and possibly sheds light on the possible origins of the virus. To date, complete genome sequences of several SCSMV isolates from different sugarcane-producing Asian countries have been reported (Xu *et al.*, 2010; Li *et al.*, 2011; Parameswari *et al.*, 2013; Liang *et al.*, 2016). Sequence data for SCSMV isolates from Iran are only available from the partial of NIB and CP, and the complete genome sequences of SCSMV Iranian isolates have not yet been determined. Therefore, to gain more insight into the genetic properties of Iranian isolates, the complete nucleotide sequences of two SCSMV isolates from sugarcane were determined and compared with those of the SCSMV isolates available up to now. In addition, the genetic variation and phylogenetic relationship of the two new isolates relative to 10 Asian SCSMV isolates were described.

Materials and Methods

Virus isolates, RNA isolation and RT-PCR. Sugarcane leaves ($n = 93$) showing mosaic and streaking symptoms were collected from sugarcane growing areas in south-western (Khuzestan province, $n = 65$) and northern (Mazandaran province, $n = 28$) Iran during August to October 2014–2015. The collected samples were from common sugarcane cultivars: CP48-103, CP57-614, CP69-1062, CP63-588, CP78-1628. Total RNA was extracted from the

symptomatic samples using RNeasy mini kit (Qiagen, Germany) and used as template for reverse transcription test. Primary detection was performed by RT-PCR using specific primers described by Moradi *et al.* (2015), amplifying 570 bp from the NIB and CP gene sequences. Two positive samples namely IR-Khuz-6 and IR-Khuz-57 were further processed for complete genome characterization. These samples were stored at -20°C and maintained by mechanical inoculation in a sorghum (*Sorghum bicolor* cv. Rio). In order to determine complete genome sequences of both Iranian SCSMV isolates, RT-PCR analysis was performed by use of specific primer pairs designed by Xu *et al.* (2010) and Parameswari *et al.* (2013). Synthesis of cDNA was performed using antisense primers from each set of primers and the MMuLV reverse transcriptase (Thermo Scientific, USA) according to the manufacturer's instructions. PCR was carried out with method described previously (Xu *et al.*, 2010; Parameswari *et al.*, 2013) with minor modifications. The 5' terminal sequence was obtained by two 5' RACE reactions with two virus-specific primers, IR-R (5'-ACACTAATTCCTCACAGCAATCT-3') and PAK5endY (5'-CATAAGCTTCCATATCTGCCATCCA-3') (Xu *et al.*, 2010), and the GeneRacer 5' primers using the GeneRacer™ kit (Invitrogen, USA) following manufacturer's instructions. The PCR amplification products were separated by electrophoresis in 1% agarose gel and then purified using Qiaquick gel extraction kit (Qiagen).

Cloning, sequencing, and phylogenetic analysis. The purified PCR products were ligated into pTZ57R/T vector (Thermo Scientific) according to manufacturer's instructions. After the recombinant vectors were transformed into competent cells of *E. coli* DH5 α , the positive clones were screened by colony-PCR with M13-forward and reverse primers. Plasmid DNA from recombinant clones was purified using a Plasmid miniprep kit (Qiagen) and two independent clones for each amplicon were sequenced in both directions (Macrogen Inc., South Korea). The full-length genome sequences of two SCSMV isolates were assembled from nine overlapping fragments using the contigExpress program in vector NTI 11 (Invitrogen, USA), and a 1268 nt 5'-end was obtained from two 5'-RACE clones. The sequences were submitted under accession numbers KX430771 for IR-Khuz6 and KX430772 for IR-Khuz57. The complete genomic nucleotide sequence and predicted amino acid sequence of the polyprotein of two SCSMV isolates from this study and ten isolates from GenBank database were aligned using DNAMAN7 software (Lynnon Biosoft, Canada). Phylogenetic trees were constructed using neighbor-joining (NJ) and maximum likelihood (ML) methods in the MEGA6 software (Tamura *et al.*, 2013).

Analysis of genetic distance and selection pressure. The genetic distances of SCSMV isolates within and between groups were calculated by the maximum composite likelihood method in MEGA6 (Tamura *et al.*, 2013). Selection pressure on each SCSMV coding region was estimated by calculating the ratio between the number of nonsynonymous substitutions (d_N) and the number of synonymous substitutions (d_S) (d_N/d_S). d_N and d_S values were estimated separately using DnaSP v.5.10.01 software (Librado and Rozas, 2009).

Table 1. Genomic structure of SCSMV-IR-Khuz6 and -IR- Khuz57 and putative cleavage sites

| Region | Start-end position | Size nt/aa | Cleavage sites (C-terminus) |
|---------|--------------------|------------|-----------------------------|
| 5'-UTR | 1-199 | 199/- | - |
| P1 | 200-1273 | 1074/358 | EELVIFY/TD |
| HC-Pro | 1274-2683 | 1410/470 | MKYRIG/GV |
| P3 | 2684-3673 | 990/330 | LIHHAQ/GK |
| 6K1 | 3674-3820 | 147/49 | TTSSPE/SP |
| CI | 3821-5788 | 1968/656 | IYHGGQ/EA |
| 6K2 | 5789-5932 | 144/48 | TLIMHA/GK |
| NIa-VPg | 5933-6526 | 594/198 | QALASE/TI |
| NIa-Pro | 6527-7240 | 714/238 | HGAIEI/QHG |
| NIb | 7241-8746 | 1506/502 | AAVGEQ/GT |
| CP | 8747-9592 | 846/281 | - |
| 3'-UTR | 9593-9782 | 190/- | - |

Neutrality tests. Neutrality tests were performed using Fu & Li's D* and F* test (Fu and Li, 1993), Tajima's D test (Tajima, 1989) and McDonald and Kreitman test (McDonald and Kreitman, 1991) implemented in the DnaSP v.5.10.01 software (Librado and Rozas, 2009). The three tests were used to test the hypothesis that all mutations were selectively neutral (Kimura, 1983).

Genetic differentiation and gene flow. Genetic differentiation between populations was examined using several statistics: Ks*, Z, Z*, Kst, Kst* and Snn based on permutation statistical tests with 1000 replicates. Ks* and Z are the sequence-based statistics considered by Hudson (2000). The Z* statistic is a logarithmic variant of the Z statistic. The nearest-neighbour statistic (Snn) is a measure of how often the "nearest neighbours" of sequences are from the same locality in geographic space. Snn is expected to be near one when the populations at the two localities are highly differentiated and near one-half when the populations at the two localities are part of the same pan-mictic population (Hudson, 2000). The null hypothesis (no genetic differentiation) was rejected, if the test statistics (Ks*, Z, Z*, Kst, Kst*, and Snn) were strongly supported by P-values <0.05. The extent of genetic differentiation or the level of gene flow between SCSMV populations was calculated by estimating the absolute value of the standardized variance in allele frequencies across populations (F_{ST}) (Wright, 1951). The absolute value of F_{ST} ranges from 0 to 1 for undifferentiated to fully differentiated populations, respectively. Generally, an absolute value of F_{ST} >0.33 suggests infrequent gene flow, while absolute value of F_{ST} <0.33 suggests frequent gene flow (Rozas *et al.*, 2003; Librado and Rozas, 2009). These analyses were performed using DnaSP v.5.10.01 (Librado and Rozas, 2009).

Results

Seventeen samples (26.1%, from Khuzestan province) were found to be infected with SCSMV by RT-PCR ampli-

fication, while 28 samples from Mazandaran province were negative.

Genome structure

The complete genome sequences of IR-Khuz6 (KX430771) and IR-Khuz57 (KX430772) were 9,782 nt long, excluding the 3'-terminal poly(A) tail. The genome contained a large single ORF (9,393 nt) which is translated into a large polyprotein with 3,130 aa. In addition, the small ORF PIPO (~141 aa), which is conserved throughout the members of the family *Potyviridae* (Chung *et al.*, 2008) was detected in the P3 cistron of both isolates with a conserved G₂A₇ leading motif (nt 3,085–3,093). Nine putative protease cleavage sites were identified based on recognition motifs for the viral proteases (Adams *et al.*, 2005) (Table 1), and the polyprotein was also predicted to code ten proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP). All of nine cleavage sites are completely conserved among the two isolates (Table 1). Several potyviral conserved motifs were identified in the polyprotein sequences of both isolates. For example, in the N-terminal of P1 protein a conserved motif L₆₇FKA₇₀ and a cysteine-rich zinc finger motif C₈₆-X₂-C₈₉-X₂-C₉₂-X₁₀-H₁₀₃-X₂-C₁₀₆ were present, implicating the role of SCSMV P1 in suppression of gene-silencing similar to that of members of the genus *Tritimovirus* (family *Potyviridae*) and triticum mosaic virus (Valli *et al.*, 2008; Young *et al.*, 2012; Tatineni *et al.*, 2012). We have also found, in the carboxylic end of P1 protein, the serine-type protease domain H₂₆₃-X₈-D₂₇₂-X₃₃-G₃₀₆-X-S₃₀₈-G₃₀₉ and proteolytic domain F₃₃₂VVLG₃₃₆ (Dougherty and Semler, 1993). These domains represent the active sites involved in the autoproteolytic activity at the N-terminal region (Adams *et al.*, 2005). The SCSMV HC-Pro protein lacks the conserved motifs associated with aphid transmission, genome amplification, systemic movement and suppression of RNA silencing at its N-terminal and central regions (Urququi-Inchima *et al.*, 2001; Xu *et al.*, 2010). However, in the N-terminal region of the HC-Pro protein contained a conserved motif C₄₁₀-X₂-C₄₁₃ in which two cysteine residues were determined to be essential for transmission of wheat streak mosaic virus by eriophyid mites (Young *et al.*, 2007). Also in the C-terminal of HC-Pro, the conserved residues of the putative proteinase active site C₇₁₅-X₇₁-H₇₈₇ in which C was located in the cysteine proteinase motif G₇₁₃YCY₇₁₆ were found. In 6K1 protein, a motif A₁₁₇₀-X₂₂-K₁₁₉₃ found in all potyviruses might play an important role in the function of this protein (Xu *et al.*, 2010). In RNA helicase CI, the nucleotide-binding motif G₁₂₉₉GVGTGKSTX₂P₁₃₁₀ and the characteristic RNA helicase domains P₁₃₁₈-X₅-PSR-X₄-N₁₃₃₁, D₁₃₈₇ECH₁₃₉₀, T₁₄₁₆ATPP₁₄₂₀, P₁₄₄₁-S₁₄₄₂, T₁₅₂₆-T-X₂-Q-X-G₁₅₃₂ and N₁₅₆₆-X₅-QR-X-G-R-X-G-R₁₅₇₉ (Adams *et al.*, 2005; Kadare and Haenni, 1997) were also present. In NIa-VPg protein, a conserved tyrosine

(Y₁₉₈₉) required for linking VPg to the 5'-end of the genomic RNA (Murphy *et al.*, 1996), was found in the context of L₁₉₈₈YDLD₁₉₉₂. The cysteine proteinase catalytic motif H₂₁₅₂-X₃₄-D-X₆₆-G-X-C-G₂₂₅₇ (Dougherty and Semler, 1993) was found in NIa-Pro. In NIb protein, the RNA-dependent RNA polymerase motifs S₂₅₁₁LKAEL₂₅₁₆, I₂₅₈₆DADGSRDSS₂₅₉₆ and S₂₆₅₁G-X₃-T₂₆₅₆-X₃-N₂₆₆₀T-X₂₆-G₂₆₈₈-D-D₂₆₉₀ (Riechmann *et al.*, 1992; Urcuqui-Inchima *et al.*, 2001) were present. The conserved residues Y₂₉₁₅-X-P-X₁₇-W₂₉₃₅, A₃₀₁₂-X-P-X₂-R-X₂-M-X₆-A₃₀₂₇ and D₃₀₆₂-F₃₀₆₃ found in monopartite potyviruses (Xu *et al.*, 2010), were also found in CP of both Iranian isolates. The DAG motif necessary for aphid transmission was not found in any isolates, whereas the conserved residues R₃₀₁₇ and D₃₀₆₂, which are presumably involved in virion assembly and cell-to-cell movement, were present (Dolja *et al.*, 1994, 1995; Urcuqui-Inchima *et al.*, 2001).

Identities with other SCSMV isolates

The identities between the complete genomes of isolates IR-Khuz6 and IR-Khuz57 were 97.7% at nucleotide (nt) level and 98.6% at amino acid (aa) level. The entire genomes of IR-Khuz6 and IR-Khuz57 shared 81–92.4% nt identity with 10 other fully sequenced SCSMV isolates, with the highest identity with the isolate PAK (NC_014037) from Pakistan, and the lowest with isolate TPT (GQ246187) from India. The complete ORF and putative polyprotein of IR-Khuz6 and IR-Khuz57 had the highest identities of 92.1 and 91.1% at nt level and 98.3 and 97.8% at aa level, with isolate PAK. IR-Khuz6 shared nt identities of 73.6–98% with other SCSMV isolates for individual genes in the polyprotein, while IR-Khuz57 shared 75–96.5% nt identities. When separate genes were compared, most genes of IR-Khuz6 and IR-Khuz57 shared the highest identities with a Pakistani isolate followed by an Indian isolate IND671 (JN941985). The 6K1 and NIa-Vpg of IR-Khuz6 were more identical to the IND671 isolate from India; the P3 and NIa-Vpg of IR-Khuz57 were more identical to the IND671, while the remaining eight proteins and 5', 3'-UTRs showed higher identities to the Pakistani isolate (Table 2). IR-Khuz6 and IR-Khuz57 shared identities of 85.2–98% (nt) and 94.7–98.9% (aa) with other SCSMV isolates in the CP gene (Table 2).

Phylogenetic analysis of SCSMV isolates

The complete genomic sequences of 12 SCSMV isolates, including those determined in this study, were subjected to phylogenetic analyses, with one triticum mosaic virus (TriMV) isolate (NC_012799) as an outgroup. Because both ML and NJ trees showed almost identical topologies, only the NJ tree is shown in Fig. 1. The results showed that SCSMV isolates were clustered to two distinct groups and members of each group were divided into two subgroups: A, B (Fig. 1).

Group I consisted of eight SCSMV isolates from sugarcane in China, Japan, India, Indonesia, Thailand and Myanmar. The TPT isolate from India, formed a separate subgroup (IB) in this group. The isolates in group I shared within-group identities of 95.3–99.5 and 98.4–99.6% at nt and aa level, respectively. Group II included four isolates from Iran, Pakistan and India among which Iranian isolates formed separate subgroup (IIA). These isolates shared with within-group identities of 90.5–97.5 and 96.5–98.6% at nt and aa level, respectively. The isolates in group II shared identities of 80.6–82.3 and 94–95.3% at nt and aa levels, respectively, with those in group I, which were obviously lower than those within two groups. The genetic distance within group II was high (0.08 ± 0.016), indicating that the isolates in this group were the most genetically divergent. On the contrary, the genetic distance within group I was low (0.026 ± 0.004). The between-group genetic distances of the two groups (0.22 ± 0.019) were significantly higher than those in the within-group, suggesting that the phylogenetic clustering results of these isolates were reasonable. A similar phylogenetic tree was obtained from an alignment of the polyprotein aa sequences of the same isolates. Such results were consistent with similarity plot analysis (data not shown).

Selection pressure on SCSMV

In order to determine the direction of mutation and degree of constraints on SCSMV isolates, the d_N/d_S ratio in each coding region (Table 3) was calculated. The results showed that the predominant selective pressure acting on the SCSMV genome was negative because its d_N/d_S value was <1. d_N values were smaller than d_S values and varied considerably among the SCSMV genomic regions (Table 3), indicating that the SCSMV cistrons are under different evolutionary constraints. The d_N/d_S ratio of HC-Pro coding region was the highest of all SCSMV coding regions, indicating that HC-Pro is under tighter functional constraints; in turn, NIa-Pro- and NIa-Vpg-coding regions have the lowest d_N/d_S ratios, indicating that these proteins are under the strongest negative or purifying selection. In addition, PIPO region of SCSMV isolates was also under negative or purifying selection pressure ($d_N/d_S = 0.6514$).

Nucleotide diversity and neutrality tests

The within- and between-group estimates of evolutionary divergence (d) were calculated for the complete polyprotein sequences of the 12 Asian SCSMV isolates (Table 4; two from this study and ten from GenBank) based on the identified SCSMV phylogroups, using MEGA6 software. The overall mean value of nucleotide sequence diversity was 0.127 ± 0.020. Analysis of group-specific sequences revealed that members of group II are more diverged from

Table 2. Percent nucleotide/amino acid sequence identity of the complete genome, polyprotein and individual coding and non-coding regions of SCSMV-IR-Khuz6 and -IR-Khuz57, and other SCSMV isolates

| Isolate | 5'-UTR | P1 | HC-Pro | PIPO | P3 | 6k1 | CI | 6K2 | NIa-VPg | NIa-Pro | Nib | CP | 3'-UTR | Polyprotein | Whole genome |
|------------------------------|------------------------------|--|--|--|--|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------|--|--|--|------------------------------|--|------------------------------|
| IR-Khuz57 | 98.5 (-) ^a | 96.0/97.8 (-) | 97.5/97.2 (-) | 98.6/98.6 (-) | 98.6/99.1 (-) | 98.0/98.0 (-) | 97.9/98.8 (-) | 98.6/100 (-) | 98.3/99.5 (-) | 98.7/99.6 (-) | 97.1/99.2 (-) | 97.3/98.6 (-) | 98.9 (-) | 97.6/98.6 (-) | 97.7 (-) |
| PAK / NC_014037 | 98.5 (98.5) | 89.4/97.8 (88.6/96.6) | 90.5/98.9 (88.9/97.0) | 96.0/92.2 (96.0/92.2) | 90.4/97.0 (90.3/97.3) | 93.8/100 (93.8/100) | 89.0/98.2 (88/97.6) | 93.8/100 (93.8/100) | 95.1/99.0 (94.4/98.5) | 96.3/99.6 (95.9/99.2) | 95.4/98.4 (93.4/98.0) | 98.0/98.2 (96.5/98.9) | 99.5 (99.5) | 92.1/98.3 (91.1/97.8) | 92.4 (91.4) |
| IND671/ JN941985 | 96.5 (97.5) | 87.5/90.8 (88.0/89.9) | 89.0/97.2 (88.2/95.3) | 96.0/92.2 (96.0/92.2) | 90.1/97.3 (90.4/97.6) | 93.1/100 (93.1/100) | 86.4/98.5 (85.9/97.6) | 93.1/100 (93.1/100) | 96.3/100 (95.6/99.5) | 93.4/97.9 (92.8/97.5) | 95.0/98.2 (93.3/97.8) | 96.3/95.7 (94.8/97.2) | 99.5 (98.4) | 90.8/97.0 (90.2/96.5) | 91.1 (90.5) |
| TPT / GQ246187 | - (-) | 83.0/94.7 (83.1/93.6) | 76.3/89.4 (76.2/87.2) | 93.9/87.9 (94.4/87.9) | 85.5/93.9 (85.9/94.2) | 74.3/100 (75.7/100) | 80.3/96.2 (80.7/95.3) | 74.3/100 (75.7/100) | 80.1/96.5 (81.1/96.0) | 82.2/95.8 (82.2/95.3) | 80.9/95.2 (82.5/94.8) | 86.2/96.8 (87.1/97.5) | 94.7 (94.2) | 81.1/94.6 (81.7/94.0) | 81.0 (81.6) |
| MYA- Formosa/ KJ187049 | 94.0 (95.0) | 83.6/95.5 (83.3/94.4) | 76.8/90.4 (76.7/88.3) | 94.4/88.7 (94.8/88.7) | 84.9/94.5 (85.4/94.8) | 75.5/93.9 (77.6/95.9) | 80.1/96.5 (80.5/95.6) | 74.3/100 (75.7/100) | 80.6/97.5 (81.6/97.0) | 82.6/97.5 (82.6/97.0) | 80.1/94.8 (81.3/94.4) | 85.9/94.7 (87.1/96.1) | 95.2 (95.8) | 81.1/95.0 (81.6/94.4) | 81.7 (82.1) |
| THA-NP3/ JN163911 | 93.5 (94.5) | 84.0/95.3 (83.5/94.1) | 76.4/90.4 (76.5/88.3) | 94.1/89.4 (94.6/89.4) | 85.8/94.5 (86.4/94.8) | 74.8/91.8 (76.9/93.9) | 79.8/96.3 (80.4/95.4) | 74.3/100 (75.7/100) | 80.5/98.0 (81.1/97.5) | 83.1/97.9 (83.1/97.5) | 80.5/94.6 (81.9/94.2) | 86.2/95.0 (87.0/96.4) | 94.7 (95.2) | 81.2/94.9 (81.8/94.4) | 81.7 (82.3) |
| ID / JF488066 | 95.0 (96.0) | 83.7/95.5 (83.8/94.4) | 76.7/89.6 (76.7/87.4) | 94.1/89.4 (94.6/89.4) | 86.2/94.5 (86.6/94.8) | 74.8/91.8 (76.9/93.9) | 80.1/96.5 (80.6/95.6) | 73.6/97.9 (75.0/97.9) | 79.8/98.0 (80.8/97.5) | 82.9/97.9 (83.2/97.5) | 80.1/94.0 (81.7/93.6) | 85.7/95.0 (86.8/96.4) | 94.7 (94.2) | 81.2/94.8 (81.8/94.2) | 81.7 (82.3) |
| JP1/ JF488064 | 94.5 (95.5) | 83.9/95.0 (83.2/93.9) | 76.5/89.8 (76.6/88.1) | 94.4/89.4 (94.4/89.4) | 85.8/94.2 (85.8/94.5) | 74.1/91.8 (76.2/93.9) | 80.2/96.5 (80.7/95.6) | 75.0/100 (76.4/100) | 80.1/98.0 (80.8/97.5) | 83.1/97.9 (83.3/97.5) | 80.3/94.2 (81.7/93.8) | 85.6/95.0 (86.6/96.4) | 94.2 (93.7) | 81.2/94.8 (81.7/94.2) | 81.7 (82.2) |
| JP2/ JF488065 | 94.0 (95.0) | 83.8/95.0 (83.1/93.9) | 76.6/89.8 (76.7/88.1) | 94.1/89.4 (94.6/89.4) | 85.6/94.2 (86.2/94.5) | 74.1/91.8 (76.2/93.9) | 80.1/96.3 (80.7/95.4) | 75.0/100 (76.4/100) | 80.0/98.0 (80.6/97.5) | 83.3/97.9 (83.6/97.5) | 80.3/94.2 (81.7/93.8) | 85.5/94.7 (86.5/96.1) | 95.2 (94.7) | 81.2/94.7 (81.7/94.2) | 81.7 (82.2) |
| YN-YZ211/ KJ187047 | 95.0 (96.0) | 83.5/95.0 (83.4/93.9) | 76.7/89.8 (76.7/87.7) | 94.1/89.4 (94.6/89.4) | 85.7/93.9 (86.1/94.2) | 77.6/91.8 (79.6/93.9) | 80.0/96.2 (80.5/95.3) | 75.7/100 (75.7/100) | 79.6/97.5 (80.3/97.0) | 83.1/97.5 (83.3/97.0) | 80.3/94.0 (81.8/93.6) | 85.8/95.0 (86.9/96.4) | 94.2 (93.7) | 81.2/94.6 (81.8/94.0) | 81.7 (82.3) |
| HN-YZ49/ KJ187048 | 94.5 (95.5) | 83.8/94.7 (83.7/93.6) | 76.4/89.8 (76.4/87.7) | 94.1/89.4 (94.6/89.4) | 85.8/94.2 (86.2/94.5) | 76.9/91.8 (78.9/93.9) | 80.1/96.5 (80.6/95.6) | 75.7/100 (75.7/100) | 80.0/98.0 (80.6/97.5) | 82.9/97.9 (83.2/97.5) | 80.5/94.0 (82.1/93.6) | 85.2/95.0 (86.6/96.4) | 94.2 (93.7) | 81.2/94.7 (81.8/94.1) | 81.7 (82.3) |

The isolates are indicated in the first column by isolate name/accession number. The highest values are indicated in bold. ^aNumber outside the parenthesis shows the identity of SCSMV-IR-Khuz6 to other SCSMV isolates, number inside the parenthesis shows the identity between -IR-Khuz57 and other SCSMV isolates.

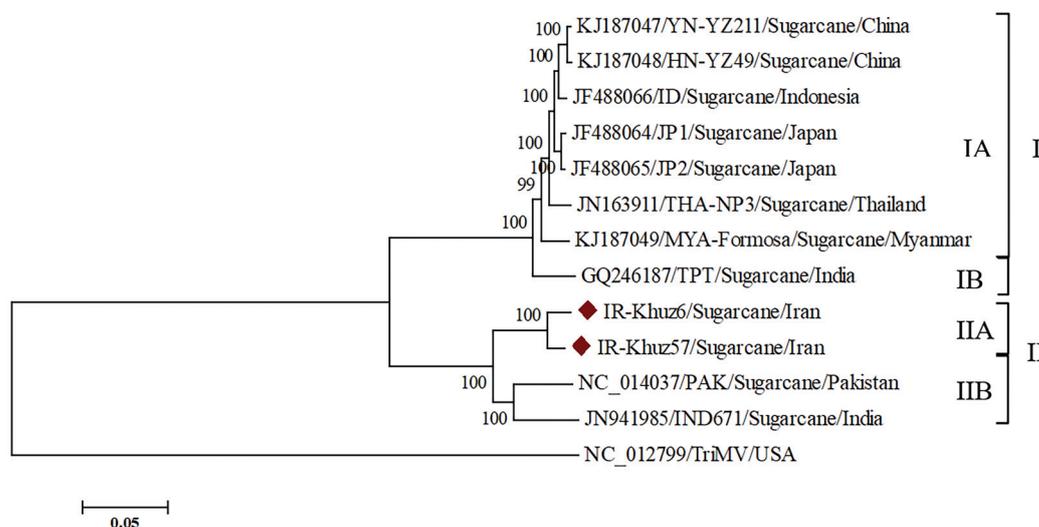


Fig. 1

Neighbor-joining tree calculated from the complete genomic sequences of IR-Khuz6 and IR-Khuz57 with 10 other SCSMV isolates

Isolates are indicated in the tree by accession number/isolate name/host/origin. Numbers at each node indicate bootstrap percentages based on 1000 replications. The nucleotide sequence of a triticum mosaic virus (TriMV) isolate was used as outgroup. The marked isolates were obtained in this research. The scale bar shows a genetic distance of 0.05.

Table 3. Nonsynonymous/synonymous (d_N/d_S) mutation ratio for different SCSMV genomic regions

| Genomic region | d_N | d_S | d_N/d_S |
|----------------|--------|--------|-----------|
| P1 | 0.0268 | 0.8046 | 0.0333 |
| HC-Pro | 0.0348 | 0.4657 | 0.0747 |
| P3 | 0.0184 | 0.5087 | 0.0361 |
| 6K1 | 0.0261 | 0.4300 | 0.0606 |
| CI | 0.0159 | 0.4415 | 0.0360 |
| 6K2 | 0.0145 | 0.4941 | 0.0293 |
| NIa-VPg | 0.0124 | 0.4485 | 0.0276 |
| NIa-Pro | 0.0211 | 0.7859 | 0.0268 |
| NIb | 0.0206 | 0.4065 | 0.0506 |
| CP | 0.0138 | 0.4909 | 0.0281 |

each other relative to isolates belonging to group I (Table 4). Nucleotide polymorphism of the SCSMV populations were estimated using Tajima's D, Fu and Li's D^* and F^* statistical tests (Table 4). Population demographical analyses showed that the values of Tajima's D, Fu and Li's D^* , and F^* in group I were negative whereas in group II and all other populations were positive. Negative values of Tajima's D, Fu and Li's D^* and F^* indicate an excess of low frequency alleles and can result from population expansions or positive selection. Positive values of Tajima's D, Fu and Li's D^* and F^* arise from an excess of intermediate frequency alleles and can result from population bottlenecks, struc-

ture and/or balancing selection (Biswas and Akey, 2006). However, the P -values for Tajima's D, Fu and Li's D and F-test were not statistically significant ($P > 0.10$) in all cases (Table 4), indicating that the results were less convincing. In addition it is plausible that purifying selection may have impact on each of the SCSMV phylogroups. As expected, analysis of between-group evolutionary divergence yielded higher value (0.22 ± 0.019) than those obtained for the within-group comparisons. These findings provide another demonstration for the existence of the two phylogroups identified in Fig. 1. Result of the McDonald and Kreitman test showed significant difference (with high statistical support based on the P -value of the G statistic), indicating that natural selection contributed to the evolution of group-specific SCSMV isolates based on analysis of the polyprotein sequences (Table 4).

Genetic differentiation and gene flow

Measurement of genetic differentiation based value of F_{ST} (0.733) between the two SCSMV phylogroups provided more support for a significant genetic differentiation between group I (East and Southeast Asia) and group II (South and Southwest Asia) isolates in the polyprotein sequence comparison. The F_{ST} value was greater than 0.33 ($F_{ST} = 0.733$), indicating that the gene flow between two different geographical SCSMV populations was not frequent. Further support for this result was found in significantly high values of Snn

Table 4. The within and between group estimates of evolutionary divergence (d) from pairwise comparison of sugarcane streak mosaic virus (SCSMV) polyprotein sequences based on phylogroups identified in Fig. 1 and summary of neutrality statistical tests

| A: Genetic distance and tests of neutrality within groups | | | | | | | | | | | |
|---|----------------------|----------------------|----------------------|-------------|-----------------------|-----------------------|-----------------------|-------------|----------------------|----------------------|----------------------|
| d | Overall (12) | | | d | Group I (8) | | | d | Group II (4) | | |
| | Tajima's D | Fu and Li's D* | Fu and Li's F* | | Tajima's D | Fu and Li's D* | Fu and Li's F* | | Tajima's D | Fu and Li's D* | Fu and Li's F* |
| 0.127±0.020 | 0.2706 ^{ns} | 0.4950 ^{ns} | 0.4973 ^{ns} | 0.026±0.004 | -1.0829 ^{ns} | -1.0850 ^{ns} | -1.2122 ^{ns} | 0.080±0.016 | 0.2183 ^{ns} | 0.3072 ^{ns} | 0.3145 ^{ns} |
| B: Genetic distance and test of natural selection between group I and II | | | | | | | | | | | |
| d | dMKT: G (P-value) | | | | | | | | | | |
| 0.22±0.019 | 11.31 (0.00077***) | | | | | | | | | | |

Note: data based on pairwise analysis of polyprotein sequences of 12 SCSMV isolates. Standard error estimates were obtained by a bootstrap procedure (1000 replicates). The evolutionary divergence analyses were conducted in MEGA6, and the codon positions included were first, second, third, and non-coding. All positions containing gaps and missing data were eliminated from the dataset. Values for neutrality tests were not significant (ns) in all cases ($P > 0.10$). ***, $P < 0.001$. All statistic tests were conducted in DnaSP v.5.10.01 (Librado and Rozas, 2009).

(1.000, P -value = 0.001**), K_s^* (5.615, P -value = 0.001**), Z (17.526, P -value = 0.001**), K_{st} (0.615, P -value = 0.0010**), K_{st}^* (0.132, P -value = 0.001**), **; $P < 0.01$) and Z^* (2.669, P -value = 0.000***, ***; $P < 0.001$) statistics obtained for the same comparison, indicating that infrequent genetic exchange may have occurred between SCSMV populations from South and Southwest Asia (group II) and those from East and Southeast Asia (group I). These results are in agreement with findings of Liang *et al.* (2016) who concluded that global SCSMV populations segregated into two groups of sequence variants that are genetically differentiated from one another and show some geographical structuring.

Discussion

This is the first report of whole genomic sequence of two SCSMV isolates obtained from the sugarcane in Iran. Pairwise comparisons of the complete genome sequences of two isolates revealed that there are no significant differences between the two isolates. The ten putative mature proteins of SCSMV-IR-Khuz6 and -IR-Khuz57 contained the conserved motifs with known functions for other potyviruses. In the present study, all of 12 complete genome sequences of SCSMV isolates were divided into two groups in which group I was consistent with the previous reports (Parameswari *et al.*, 2013; Liang *et al.*, 2016), while in group II were newly proposed isolates including Iranian SCSMV isolates alongside with isolates from Pakistan and India. Iranian SCSMV isolates were clustered together with isolates from Pakistan and India in the group II, but interestingly formed a separate subgroup. The genome of Iranian SCSMV shared the highest degree of sequence identity with SCSMV-PAK (Pakistan) and -IND671 (India) respectively. Phylogenetic trees (NJ and ML) based on the polyprotein and ten individual proteins, grouped Iranian SCSMV isolates in the same cluster separated from PAK and IND671

isolates (data not shown). Analysis of the genetic diversity of SCSMV populations from different geographical regions will be helpful to understand its emergence, epidemiology, movement, and gene flow, which can be important for designing appropriate control strategies to limit virus spreading (Liang *et al.*, 2016). Recently, the genetic diversity and population structure of SCSMV isolates based on analysis of partial virus genome sequences (He *et al.*, 2014, 2016) and whole genome sequences were determined (Liang *et al.*, 2016). In the present study, genetic differentiation measurement between populations from pairwise comparison of 12 SCSMV polyprotein sequences (two from this study and ten from GenBank) based on phylogroups, demonstrated greater genetic differentiation between the populations. No significant departure from neutrality was found for any geographical groups of SCSMV populations when the whole polyprotein sequences were tested for Tajima's D, Fu and Li's D* and F* values; however, the values were negative for the group I and positive for group II and all other populations. According to genetic differentiation and phylogenetic analysis, with the exception of TPT isolate from India, group I consisted of members that originated from East Asia (YN-YZ211, HN-YZ49, JP1 and JP2) and Southeast Asia (MAY-Formosa, ID and THA), while group II consisted of members that originated from Southwest Asia (IR-Khuz6, IR-Khuz57) and South Asia (IND671 and PAK). This suggests that the majority of SCSMV isolates are divided into two groups corresponding with their geographical origins, which is consistent with previous findings (Liang *et al.*, 2016). SCSMV showed a spatial distribution pattern with a greater genetic deviation in the Southwest (Iran) or South (India and Pakistan) Asian lineage than in the subpopulation composed of isolates from East (China and Japan) or Southeast (Thailand, Indonesia and Myanmar) Asia (He *et al.*, 2016; Liang *et al.*, 2016). The higher genetic diversity observed between groups than within groups of SCSMV populations suggests that there is more frequent gene flow within groups than between groups.

Virus transmission characteristics and the presence of physical and quarantine barriers between countries may explain the SCSMV population structuring (Liang *et al.*, 2016). As SCSMV can be transmitted through vegetative set, it seems that Iranian isolates have been introduced into the country from sugarcane-producing Asian countries such as India and Pakistan via contaminated vegetative cuttings. In RNA viruses, viral genome evolutions are directed by recombination and mutation events, which are two important sources of genetic variation (Roossinck, 1997; Posada *et al.*, 2002; Gell *et al.*, 2015). Recombination occurs frequently among members of the family *Potyviridae* (Revers *et al.*, 1996; Chare and Holmes, 2006; Valli *et al.*, 2007; Padhi and Ramu, 2011; Sztuba-Solin'ska *et al.*, 2011; Gell *et al.*, 2015), however, no recombination event was found in Iranian SCSMV isolates and also other members of group II, suggesting that probably this evolutionary force has not shaped the emergence of SCSMV group II variants (in contrast to group I (Liang *et al.*, 2016)). In addition, no recombination events were found between SCSMV group I and group II subpopulations, indicating significant genetic differentiation and limited gene flow between virus isolates belonging to these phylogroups, perhaps due to the presence of physical and quarantine barriers between them (Liang *et al.*, 2016). According to previous studies, SCSMV has recombination events at a low frequency, possibly due to narrow host range of this virus which limits to sugarcane and sorghum in natural conditions (Viswanathan *et al.*, 2008; Srinivas *et al.*, 2010; Xu *et al.*, 2010; He *et al.*, 2016). Natural selection and genetic drift are two main evolutionary mechanisms limiting genetic variation of the SCSMV population (Liang *et al.*, 2016). In this study, adaptive molecular evolution in the polyprotein nucleotide sequences was detected by the G statistic of MKT between two distinct variant groups, which showed significant difference. Similarly, the number of dN was generally smaller than dS and varied between the SCSMV genomic regions, which is suggesting a negative selection against most aa changes for shaping the stability of the genome (Viswanathan *et al.*, 2015; He *et al.*, 2016; Liang *et al.*, 2016). Different selection pressure on the coding regions supports the hypothesis that potyviral cistrons evolved separately though assembled into a single ORF and are expressed as polyprotein (Seo *et al.*, 2009; Bejerman *et al.*, 2013). Among all SCSMV-encoded genes, PIPO and HC-Pro, showed much higher dN/dS ratios than other regions, suggesting that these regions may have experienced different evolutionary pressures. These results are consistent with previous studies (Bagyalakshmi *et al.*, 2012; Viswanathan *et al.*, 2015). Taken together with the previous researches (He *et al.*, 2016; Liang *et al.*, 2016), this study also demonstrates that gene flow and strong purifying (negative) selection might have contributed to the evolution of isolates belonging to the two identified SCSMV groups, with infrequent and frequent genetic exchanges between and within individual geographical subpopulations, respectively.

According to previous researches, mixed infection of three viruses, SCMV, SCSMV, and SrMV in the family *Potyviridae* is common in sugarcane (Viswanathan *et al.*, 2007, 2008; Xu *et al.*, 2008; Xie *et al.*, 2009; Luo *et al.*, 2016). On the other hand, mixed infection constitutes a pre-requisite for the generation of new viral strains/variants by genome exchanges (recombination) (Lai, 1992) which may result in more severe and complex diseases than infection with either virus alone. The intra- and interspecific recombinations are common in potyviruses (Revers *et al.*, 1996; Bousalem *et al.*, 2000; Ohshima *et al.*, 2002; Larsen *et al.*, 2005; Seo *et al.*, 2009; Gell *et al.*, 2015), often leading to emergence of new variants and resistance-breaking virus strains. Hence, generation of resistance-breaking virus strains by mixed infection and recombination is a threat for the sugarcane plants. In Iran, mosaic disease in sugarcane is caused either by SCMV, SCSMV, IJMV (Iranian johnsongrass mosaic virus) alone (Moradi *et al.*, 2016, 2017a) or all three in combinations, but SCMV is the predominant causal agent (Moradi *et al.*, 2017b). According to results of this study, SCSMV is not widely spread in Khuzestan province and was not detected in any of 28 samples collected from Mazandaran province, probably due to lack of vector transmission and agricultural conditions. Looking for the occurrence of these viruses infecting sugarcane crops, there is a need for immediate attention towards studying economic losses caused by mixed infection of either virus in sugarcane. This information provides an insight into the SCSMV population structure and might be helpful in breeding of resistant varieties and developing effective control strategies. However, further genomic sequences of SCSMV isolates from Iran are required to fully assess the molecular diversity and evolutionary history of the virus in the country.

References

- Adams MJ, Antoniw JF, Beudoin F (2005): Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol. Plant Pathol.* 6, 471–487. <https://doi.org/10.1111/j.1364-3703.2005.00296.x>
- Bagyalakshmi K, Parameswari B, Chinnaraja C, Karuppaiah R, Ganesh KV, Viswanathan R (2012): Genetic variability and potential recombination events in the HC-Pro gene of sugarcane streak mosaic virus. *Arch. Virol.* 157, 1371–1375. <https://doi.org/10.1007/s00705-012-1297-8>
- Bejerman N, Giolitti F, de Breuil S, Lenardon S (2013): Sequencing of two sunflower chlorotic mottle virus isolates obtained from different natural hosts shed light on its evolutionary history. *Virus Genes* 46, 105–110. <https://doi.org/10.1007/s11262-012-0817-7>
- Biswas S, Akey JM (2006): Genomic insights into positive selection. *Trends Genet.* 22, 437–446. <https://doi.org/10.1016/j.tig.2006.06.005>
- Bousalem M, Douzery E, Fargette D (2000): High genetic diversity, distant phylogenetic relationships and intraspecies

- recombination events among natural populations of yam mosaic virus: a contribution to understanding potyvirus evolution. *J. Gen. Virol.* 81, 243–255. <https://doi.org/10.1099/0022-1317-81-1-243>
- Chare ER, Holmes EC (2006): A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* 151, 933–946. <https://doi.org/10.1007/s00705-005-0675-x>
- Chatenet M, Mazarin C, Girard JC, Fernandez E, Gargani D, Rao GP, Royer M, Lockhart B, Rott P (2005): Detection of sugarcane streak mosaic virus in sugarcane from several Asian countries. *Proc. Int. Soc. Sugarcane Technol.* 25, 656–662.
- Chung BY-W, Miller WA, Atkins JF, Firth AE (2008): An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. USA* 105, 5897–5902. <https://doi.org/10.1073/pnas.0800468105>
- Damayanti TA, Putra LK (2011): First occurrence of sugarcane streak mosaic virus infecting sugarcane in Indonesia. *J. Gen. Plant Pathol.* 77, 72–74. <https://doi.org/10.1007/s10327-010-0285-7>
- Dolja VV, Haldeman R, Robertson NL, Dougherty WG, Carrington JC (1994): Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants. *EMBO J.* 13, 1482–1491.
- Dolja VV, Halderman-Cahill R, Montgomery AE, Vandenbosch KA, Carrington JC (1995): Capsid protein determinants in cell-to-cell and long distance movement of tobacco etch potyvirus. *Virology* 206, 1007–1016. <https://doi.org/10.1006/viro.1995.1023>
- Dougherty WG, Semler BL (1993): Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* 57, 781–822.
- Fu YX, Li WH (1993): Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- Gell G, Sebestyén E, Balázs E (2015): Recombination analysis of maize dwarf mosaic virus (MDMV) in the sugarcane mosaic virus (SCMV) subgroup of potyviruses. *Virus Genes* 50, 79–86. <https://doi.org/10.1007/s11262-014-1142-0>
- Gibbs AJ, Ohshima K (2010): Potyviruses and the digital revolution. *Annu. Rev. Phytopathol.* 48, 205–223. <https://doi.org/10.1146/annurev-phyto-073009-114404>
- Grisham MP (2000): Mosaic. In Rott P, Bailey R., Comstock J., Croft BJ, Saumtally AS (Eds): *A guide to Sugarcane Diseases*. CIRAD-ISSCT, CIRAD publication Services, Montpellier, France, pp. 249–254.
- Hall JS, Adams B, Parsons TJ, French R, Lane LC, Jensen SG (1998): Molecular cloning, sequencing and phylogenetic relationships of a new potyvirus: sugarcane streak mosaic virus, and a reevaluation of the classification of the Potyviridae. *Mol. Phylogenet. Evol.* 10, 323–332. <https://doi.org/10.1006/mpev.1998.0535>
- He Z, Li W, Yasaka R, Huang Y, Zhang Z, Ohshima K, Li S (2014): Molecular variability of sugarcane streak mosaic virus in China based on an analysis of the P1 and CP protein coding regions. *Arch. Virol.* 159, 1149–1154. <https://doi.org/10.1007/s00705-013-1854-9>
- He Z, Yasaka R, Li W, Li S, Ohshima K (2016): Genetic structure of populations of sugarcane streak mosaic virus in China: Comparison with the populations in India. *Virus Res.* 211, 103–116. <https://doi.org/10.1016/j.virusres.2015.09.020>
- Hema M, Joseph J, Gopinath K, Sreenivasulu P, Savithri HS (1999): Molecular characterization and intervirial relationships of a flexuous filamentous virus causing mosaic disease of sugarcane (*Saccharum officinarum* L.) in India. *Arch. Virol.* 144, 479–490. <https://doi.org/10.1007/s007050050519>
- Hema M, Sreenivasulu P, Savithri HS (2002): Taxonomic position of sugarcane streak mosaic virus in the family Potyviridae. *Arch. Virol.* 147, 1997–2007. <https://doi.org/10.1007/s00705-002-0851-1>
- Hudson RR (2000): A new statistic for detecting genetic differentiation. *Genetics* 155, 2011–2014.
- Kadare G, Haenni A-L (1997): Virus-encoded RNA helicases. *J. Virol.* 71, 2583–2590.
- Kimura M (1983): *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511623486>
- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (Ed.) (2012): *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, 1327 p.
- Lai MMC (1992): RNA recombination in animal and plant viruses. *Microbiol. Rev.* 56, 61–79.
- Larsen RC, Miklas PN, Druffel KL, Wyatt SD (2005): NL-3K strain is a stable and naturally occurring interspecific recombinant derived from bean common mosaic necrosis virus and bean common mosaic virus. *Phytopathology* 95, 1037–1042. <https://doi.org/10.1094/PHYTO-95-1037>
- Li WF, He Z, Li SF, Huang YK, Zhang ZX, Jiang D, Wang X, Luo Z (2011): Molecular characterization of a new strain of sugarcane streak mosaic virus (SCSMV). *Arch. Virol.* 156, 2101–2104. <https://doi.org/10.1007/s00705-011-1090-0>
- Liang SS, Alabi OJ, Damaj MB, Fu WL, Sun SR, Fu HY, Chen RK, Mirkov TE, Gao SJ (2016): Genomic variability and molecular evolution of Asian isolates of sugarcane streak mosaic virus. *Arch. Virol.* 161, 1493–1503. <https://doi.org/10.1007/s00705-016-2810-2>
- Librado P, Rozas J (2009): DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Luo Q, Ahmad K, Fu HY, Wang JD, Chen RK, Gao SJ (2016): Genetic diversity and population structure of sorghum mosaic virus infecting *Saccharum* spp. hybrids. *Ann. Appl. Biol.* Volume, pages?? Moradi N, Rajabi-Memari H, Mehrabi-Koushki M, Taherkhani K, Moazzen-Reza-Mahalle H, Sheikhi F, Nasirpour N, Sanjabifard Z (2015): First report of sugarcane streak mosaic virus in Iran. *New Dis. Rep.* 32, 2.
- McDonald JH, Kreitman M (1991): Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654. <https://doi.org/10.1038/351652a0>
- Moradi Z, Mehrvar M, Nazifi E, Zakiaghl M (2016): The complete genome sequences of two naturally occurring recombinant isolates of sugarcane mosaic virus from Iran. *Virus Genes* 52, 270–280. <https://doi.org/10.1007/s11262-016-1302-5>
- Moradi Z, Mehrvar M, Nazifi E, Zakiaghl M (2017a): Iranian johnsongrass mosaic virus: the complete genome sequence, molecular and biological characterization, and com-

- parison of coat protein gene sequences. *Virus Genes* 53, 77–88. <https://doi.org/10.1007/s11262-016-1389-8>
- Moradi Z, Nazifi E, Mehrvar M (2017b): Occurrence and evolutionary analysis of coat protein gene sequences of Iranian isolates of sugarcane mosaic virus. *Plant Pathol. J.* 33, 296–306.
- Murphy JF, Klein PG, Hunt AG, Shaw JG (1996): Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. *Virology* 220, 535–538. <https://doi.org/10.1006/viro.1996.0344>
- Ohshima K, Yamaguchi Y, Hirota R, Hamamoto T, Tomimura K, et al. (2002): Molecular evolution of turnip mosaic virus; evidence of host adaptation, genetic recombination and geographical spread. *J. Gen. Virol.* 83, 1511–1521. <https://doi.org/10.1099/0022-1317-83-6-1511>
- Padhi A, Ramu K (2011): Genomic evidence of intraspecific recombination in sugarcane mosaic virus. *Virus Genes* 42, 282–285. <https://doi.org/10.1007/s11262-010-0564-6>
- Parameswari B, Bagyalakshmi K, Viswanathan R, Chinnaraja C (2013): Molecular characterization of Indian sugarcane streak mosaic virus isolate. *Virus Genes* 46, 186–189. <https://doi.org/10.1007/s11262-012-0827-5>
- Posada D, Crandall KA, Holmes EC (2002) Recombination in evolutionary genomics. *Annu. Rev. Genet.* 36, 75–97. <https://doi.org/10.1146/annurev.genet.36.040202.111115>
- Revers F, Le Gall O, Candresse T, Le Romancer M, Dunez J (1996): Frequent occurrence of recombinant potyvirus isolates. *J. Gen. Virol.* 77, 1953–1965. <https://doi.org/10.1099/0022-1317-77-8-1953>
- Riechmann JL, Lain S, Garcia JA (1992): Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol.* 73, 1–16. <https://doi.org/10.1099/0022-1317-73-1-1>
- Roossinck MJ (1997): Mechanisms of plant virus evolution. *Annu. Rev. Phytopathol.* 35, 191–209. <https://doi.org/10.1146/annurev.phyto.35.1.191>
- Rozas J, Sanchez-DeI Barrio JC, Messeguer X, Rozas R (2003): DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497. <https://doi.org/10.1093/bioinformatics/btg359>
- Seo JK, Ohshima K, Lee HG, Son M, Choi HS, Lee SH, Sohn SH, Kim KH (2009): Molecular variability and genetic structure of the population of soybean mosaic virus based on the analysis of complete genome sequences. *Virology* 393, 91–103. <https://doi.org/10.1016/j.virol.2009.07.007>
- Srinivas KP, Subba Reddy CV, Ramesh B, Lava Kumar P, Sreenivasulu P (2010): Identification of a virus naturally infecting sorghum in India as sugarcane streak mosaic virus. *Eur. J. Plant Pathol.* 127, 13–19. <https://doi.org/10.1007/s10658-010-9580-6>
- Sztuba-Solin' ska J, Urbanowicz A, Figlerowicz M, Bujarski JJ (2011): RNA-RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* 49, 415–443. <https://doi.org/10.1146/annurev-phyto-072910-095351>
- Tajima F (1989): Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tamura K, Stecher G, Peterson D, Filipiski A, Kumar S (2013): MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tatineni S, Ziems AD, Wegulo SN, French R (2009): Triticum mosaic virus: a distinct member of the family Potyviridae with an unusually long leader sequence. *Phytopathology* 99, 943–950. <https://doi.org/10.1094/PHYTO-99-8-0943>
- Tatineni S, Qu F, Li R, Morris TJ, French R (2012): Triticum mosaic poacevirus enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. *Virology* 433, 104–115. <https://doi.org/10.1016/j.virol.2012.07.016>
- Urcuqui-Inchima S, Haenni A, Bernardi F (2001): Potyvirus proteins: a wealth of functions. *Virus Res.* 74, 157–175. [https://doi.org/10.1016/S0168-1702\(01\)00220-9](https://doi.org/10.1016/S0168-1702(01)00220-9)
- Valli A, López-Moya JJ, García JA (2007): Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* 88, 1016–1028. <https://doi.org/10.1099/vir.0.82402-0>
- Valli A, Dujovny G, García JA (2008): Protease activity, self interaction, and small interfering RNA binding of the silencing suppressor P1b from cucumber vein yellowing ipomovirus. *J. Virol.* 82, 974–986. <https://doi.org/10.1128/JVI.01664-07>
- Viswanathan R, Balamuralikrishnan M, Karuppaiah R (2007): Sugarcane mosaic in India: a cause of combined infection of sugarcane mosaic virus and sugarcane streak mosaic virus. *Int. Sugar J.* 25, 6–14.
- Viswanathan R, Balamuralikrishnan M, Karuppaiah R (2008): Characterization and genetic diversity of sugarcane streak mosaic virus causing mosaic in sugarcane. *Virus Genes* 36, 553–564. <https://doi.org/10.1007/s11262-008-0228-y>
- Viswanathan C, Prabu G (2015): Molecular diversity analysis of pretty interesting Potyviridae ORF (PIPO) coding region in Indian isolates of sugarcane streak mosaic virus. *Sugar Tech.* 18, 214–221.
- Wright S (1951): The genetical structure of populations. *Ann. Eugen.* 15, 323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- Xie Y, Wang M, Xu D, Li R, Zhou G (2009): Simultaneous detection and identification of four sugarcane viruses by one-step RT-PCR. *J. Virol. Methods* 162, 64–68. <https://doi.org/10.1016/j.jviromet.2009.07.015>
- Xu DL, Park JW, Mirkov TE, Zhou GH (2008): Viruses causing mosaic disease in sugarcane and their genetic diversity in southern China. *Arch. Virol.* 153, 1031–1039. <https://doi.org/10.1007/s00705-008-0072-3>
- Xu DL, Zhou GH, Xie YJ, Mock R, Li R (2010): Complete nucleotide sequence and taxonomy of sugarcane streak mosaic virus, member of a novel genus in the family Potyviridae. *Virus Genes* 40, 432–439. <https://doi.org/10.1007/s11262-010-0457-8>
- Young BA, Hein GL, French R, Stenger DC (2007): Substitution of conserved cysteine residues in wheat streak mosaic virus HC-Pro abolishes virus transmission by the wheat curl mite. *Arch. Virol.* 152, 2107–2111. <https://doi.org/10.1007/s00705-007-1034-x>
- Young BA, Stenger DC, Qu F, Morris TJ, Tatineni S, French R (2012): Tritimovirus P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Res.* 163, 672–677. <https://doi.org/10.1016/j.virusres.2011.12.019>