ORIGINAL ARTICLE

Phylogenetic analysis and genetic structure of new isolates of *Tomato mosaic virus* in Iran

Farshad Rakhshandehroo[1]*, Samaneh Sadat Hashemi[1], Nooh Shahraeen[2]

[1] Department of Plant Pathology, College of Agriculture and Natural Resources, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Research Department of Plant Virology, Iranian Research Institute of Plant Protection (IRIPP), Tehran, Iran

Vol. 58, No. 1: 25–35, 2018

DOI: 10.24425/119114

Received: October 9, 2017
Accepted: February 12, 2018

*Corresponding address:
rakhshandehroo_fa@srbiau.ac.ir

Abstract

The present report describes the new occurrence of *Tomato mosaic virus* (ToMV) in cabbage, bean and *Malva neglecta* plants in Iran. In this study, sequence analyses of a partial RNA dependent RNA polymerases (RdRp) and complete movement protein (MP) and the coat protein (CP) nucleotide sequences of three new ToMV isolates collected from major crop fields in Iran revealed low genetic variation of RdRp gene compared to the CP and MP genes. The different topologies of the phylogenetic trees constructed, using available open reading frame (ORF1), ORF2 and ORF3 sequences from ToMV isolates, indicated different evolutionary constraints in these genomic regions. Statistical analysis also revealed that with the exception of CP other tested ToMV genes were under negative selection and the RdRp gene was under the strongest constraints. According to the phylogenetic tree it can be inferred from the nucleotide sequences of the complete CP and MP genes, that isolates from Iran and Egypt formed separate groups, irrespective of host origin. However, isolates clustered into groups with correlation to geographic origin but not the host. Analysis of the K_s^*, Z^* and S_{nn} values also indicated genetic differentiation between ToMV populations. The Tajima's D, Fu and Li's statistical values were significantly negative for the RdRp gene of the Asian population which suggests the sudden expansion of ToMV in Asia. Taken together, the results indicate that negative selection and genetic drift were important evolutionary factors driving the genetic diversification of ToMV.

Key words: genetic differentiation, genetic variation, phylogenetic tree, *Tobamovirus,* ToMV

Introduction

Tobamovirus is a genus of positive single stranded RNA species with rod-shaped virions belonging to the family Virgaviridae and it contains 29 definitive and six unclassified devastating plant virus species such as *Tobacco mosaic virus* (TMV) and *Tomato mosaic virus* (ToMV) (Adams *et al*. 2009). Like other members of the family, tobamoviruses contain the characteristics of the family including: coat proteins of 19–24 K, t-RNA like structure at the 3′ end of the genome, cell to cell movement protein (MP) of the '30K' superfamily and transmission by mechanical inoculation (Adams *et al*. 2009). ToMV is the typical member of the *Tobamovirus* genus and was first considered to be a strain of TMV because of its close biological properties and serological relationship with TMV. In 1971 it was identified as a distinct species (Holling and Huttinga 1995). ToMV has a wide natural distribution and has been reported in red spruce trees, water, and cloud and soil samples (Hu *et al*. 2012). The virus RNA is positive-sense and serves directly as a messenger RNA that is translated using host ribosomes. A cap structure is found at the 5′ terminus of the genome followed by a 60–71 nucleotides (nt) long 5′ nontranslated (NTR) sequence. The 3′ side of the genome has A NTR of 180–415 nt long that can be folded into pseudoknots followed by 3′-terminal sequences that can be folded into

a tRNA-like, amino acid-accepting structure. The subgenomic mRNAs transcribed in infected cells also have a 5′-terminal cap and 3′-tRNA-like structure (Ding *et al.* 2004). At least four protein encoding genomic regions constructed the ToMV genome which is about 6,400 nucleotides. Genome encodes for the 126 kDa and 183 kDa proteins, which are translated from genomic RNA, were considered as two replicase components (Ishikawa *et al.* 1986). The 180 kDa protein is synthesized by a read-through of the amber termination codon of the 126 kDa protein. The 126 kDa protein is also responsible for vial pathogenicity as the suppressor of host's virus-induced gene silencing (Ding *et al.* 2004), whereas the 30 kDa and 17.5 kDa proteins that are translated from the subgenomic mRNAs were considered as viral movement protein (MP) and coat protein (CP), respectively (Takamatsu *et al.* 1983).

Tobamoviruses cause persistent infections in their systemic hosts, and their populations are usually very large and heterogeneous within infected hosts (Gibbs *et al.* 2015). Many ToMV strains have been reported which showed different pathogenesis (Rangel *et al.* 2011; Aghamohammadi *et al.* 2013). Thus, the control of this virus is very difficult in the field. Analysis of the genetic structure of virus populations and the sources of the genomic variation like mutations caused by genetic drift, gene flow and acquisition of extra genomic components may help identify which factors determine virus evolution which is often necessary for control purposes. Effectiveness of control strategies can be compromised by the evolution of the populations of pathogens (Harrison 2002). From 2009 to 2010, epidemics of ToMV severely affected the production of snap horticultural crops in the central, northern and southern regions of Iran. Our previous findings showed that the ToMV populations in Iran consist of several strains that have high nucleotide diversity (Aghamohammadi *et al.* 2013). Effective dissemination of ToMV on non-solanaceous plants such as French radish and nettle-leaf weed (*Chenopodistrum murale*) suggested a new expansion of different genotypes into Iran. There is currently not enough information available about the genetic variability and molecular evolution of the ToMV isolates obtained from various regions of Iran to describe ToMV structure and subgroup distribution in Iran and the world. Analyzing different genomic regions of recently detected ToMV isolates in Iran, we investigated the precise phylogenetic relationships and population genetics of detected isolates. This allowed us to monitor the population dynamics of this virus carefully in order to design some efficient control strategies for its emergence.

Materials and Methods

Surveys and sample collection

Field-grown horticultural crop species including: cabbage (*Brassica oleracea* var. *capitata*), common bean (*Phaseolus vulgaris* L.), cucumber (*Cucumis sativus* L.), eggplant (*Solanum melongena* L.), French radish (*Raphanus sativus* convar. *radicula*), pepper (*Capsicum annum* L.), zucchini (*Cucurbita pepo* L.cv. Zucchini), tomato (*Solanum lycopersicum* L.) and watermelon (*Citrullus lanatus* Schard.) were surveyed from April 2012 to September 2013 in Alborz, Mazandaran and Tehran provinces of Iran (Table 1). A total of 594 samples were collected. Forty-five symptomatic weeds belonging to Chenopodiaceae (*Chenopodium album* L.), Malvaceae (*Malva neglecta* Wallr.), and Asteraceae (*Sonchus oleraceus* L.) were also sampled randomly within fields and the borders of various vegetable fields in the surveyed provinces (Table 2). Overall, 21 fields located in seven districts were surveyed. Samples (20–30 samples per field) were collected during the growing season.

Virus testing

Mature leaf blades were tested in duplicate using double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA), with coating antibodies against ToMV as well as alkaline phosphatase conjugates obtained from Agdia (Agdia, Inc., South Bend, IN, USA) (Converse and Martin 1999).

RT-PCR and sequencing analyses

Total RNA extraction was performed using LiCl buffer according to the previously described protocol (Letschert *et al.* 2002). Total RNAs were isolated from leaves of symptomatic crop plants and weed species. First-strand cDNA synthesis was carried out according to the manufacturer's instructions (MBI, Fermentas, Germany) in a 20 µl reaction containing 10 ng of total RNA and 20 pmol reverse primers (Table 3). Primers Tob-Uni 1 and ToMV-spec were used for the genus-specific detection of ToMV isolates. Primer pairs listed in Table 3 were designed (Letschert *et al.* 2002; Hu *et al.* 2012) to amplify the 3' end of RdRp and complete nucleotide sequences of the MP and CP genes. The primers were synthesized by MWG Biotech (Germany).

To determine the nucleotide sequences of the RdRp, MP and CP genomic segments of three representative Iranian ToMV isolates, PCR was carried out in a thermocycler amplification system (Eppendorf, Hamburg, Germany) using *Pfu* DNA polymerase (Stratagen, USA) and initial denaturation at 95°C for 5 min, followed by 30 cycles of denaturation at 94°C for 60 s, annealing at 55 to 58°C for 45 s, synthesis at

Table 1. Incidence of *Tomato mosaic virus* (ToMV) on horticultural crops collected from the surveyed provinces of cultivation in Iran

			Sampled horticultural crops[a]								
			Solanaceous			Cucurbitaceous			Brassicas		Legumes
Province	Region	Infected samples[b]	tomato	pepper	eggplant	zucchini	cucumber	water-melon	cabbage	radish	bean
Alborz	Karaj	3/32	2/9	ND	1/13	ND	0/6	ND	ND	ND	0/4
	Savejbo-lagh	37/101	2/9	1/9	1/7	0/8	ND	ND	**32/51**	0/4	**1/13**
	Shahriar	13/99	6/50	2/20	**2/6**	ND	**3/9**	0/2	ND	0/4	0/8
		(22.8%)[c]									
Mazan-daran	Behshahr	20/81	6/20	**12/40**	0/5	ND	1/5	ND	1/7	ND	0/4
		(24.7%)									
Tehran	Shahre-Rey	3/45	3/24	0/17	ND	ND	0/4	ND	ND	ND	ND
	Varamin	56/216	**54/136**	1/73	ND	0/2	1/4	ND	0/1	ND	ND
	Pakdashat	4/20	2/10	ND	2/7	0/2	ND	ND	ND	ND	0/1
		(22.4%)									
Total		136/594 (22.9)[d]	75/258 (**29**)[e]	16/159 (10)	6/38 (15.7)	0/12 (0)	5/28 (17.8)	0/2 (0)	33/59 (55.9)	0/8 (0)	1/30 (3.3)

The highest incidence for ToMV infection in each surveyed host species is shown in bold and was highest for tomato; [a]data obtained by DAS-ELISA for ToMV. no. of infected samples over number of samples collected for each crop; [b]no. of infected samples over number of samples collected for each province; [c]percentage of ToMV infection in each province; [d]average percentage of ToMV infection for all provinces; [e]average percentage of ToMV infection in each crop species; ND – not determined

Table 2. Relative incidence of *Tomato mosaic virus* (ToMV) on selected weed species collected from the surveyed vegetable fields

District	Infection ratio	Sampled weed species		
		Sonchus oleraceus	*Chenopodium album*	*Malva neglecta*
Pakdashat	0/4[a]	0/2	ND	0/2
Varamin	0/1	ND	ND	0/1
Behshahr	2/10	1/6	ND	1/4
Shahriar	0/5	0/2	0/3	ND
Shahre-Rey	0/8	0/4	0/2	0/2
Savejbolagh	1/13	0/2	1/5	0/6
Karaj	1/4	1/2	ND	0/2
Total	4/45 (8.9%)[b]	2/18 (11.1%)	1/10 (10%)	1/17 (5.8%)

[a]no. of infected samples/collected samples, determined by RT-PCR; [b]average incidence of ToMV infection on selected weed species; ND – not determined

72°C for 60 s, and a final extension at 72°C for 5 min. The ToMV amplicons were cloned into the *pTZ57R/T* vector (Fermentas, Germany) and then sequenced by the Macrogen (Seoul, South Korea) using the dideoxyl chain reaction termination method. A minimum of three cloned DNAs were sequenced in both directions. The sequences obtained were compared with virus sequences available in the GenBank database using BLAST (Altschul *et al.* 1997).

Sequence and phylogenetic analysis

Three representative ToMV isolates were selected from distinct host and geographical locations with typical

Table 3. Characteristics of primers used to amplify different genomic regions of Iranian *Tomato mosaic virus* (ToMV) isolates

Primer name	Sequence [from 5′ to 3′]	Position in ToMV genome	Amplicon sizes [bp]
L-S55[a]	TGAGGTACAAGGTGAGACTT	3,191–3,210	862
L-S53[a]	GATACACAATCGTTTGCAACG	4,053–4,033	
L-S75[b]	ACGCTGTTGGGGAGGTTCATA	4,801–4,820	1074
L-S73[b]	TGAGGGAAAGGTTTCCACACCT	5,875–5,854	
ToMV-Spec[c]	CGGAAGGCCTAAACCAAAAAG	5,597–5,618	686
Tob-Unit[c]	ATTTAAGTGGASGGAAAAVCACT	6,283–6,260	

[a]primer sequences (Hu *et al*. 2012) used to amplify a portion at the 3′ end of the RNA polymerase genomic region of the selected Iranian ToMV isolates
[b]primer sequences (Hu *et al*. 2012) used to amplify the complete sequence of the movement protein genomic region of the selected Iranian ToMV isolates
[c]primer sequences (Letschert *et al*. 2002) used to amplify the complete sequence of the coat protein genomic region and also detect the Iranian ToMV isolates

symptoms of ToMV infection common in each region. Isolates T194 (collected from tomato plants) and A356 (collected from common sowthistle plants) isolated from different tomato cultivations of Tehran province and isolate To52 (collected from tomato plants) were isolated from Mazandaran province. Phylogenetic analysis of the three, new representative Iranian ToMV isolates was conducted by comparing separately the 480 and 795 nt of the whole CP and MP and 862 nt of the partial RdRp genes with the comparable sequences of other ToMV isolates from GenBank. Only those isolates with their full-length genome sequences available at Gen -Bank were considered for this study. The corresponding homologous sequences were aligned by the CLUSTALX 1.8 (Pearson and Lipman 1988) algorithm. Multiple sequence alignments of nucleotide and deduced amino acid sequences were used for the analysis of variability and the construction of a phylogenetic tree using the neighbor-joining method (Saitou and Nei 1987), the p-distance method (Nei and Kumar 2000) and bootstrap consisting of 1,000 pseudo-replicates. Finally, this was evaluated using the interior branch test method with MEGA 5.05 software (Tamura *et al*. 2011). Branches with <70% bootstrap value were collapsed.

Statistical analysis of genetic distance, genetic differentiation, gene flow and selection pressure

Using MEGA 5.05 (Tamura *et al*. 2011) software and the Tamura-3-parameter nucleotide substitution model, the genetic distance for ToMV sequences derived within and between geographical groups was calculated. Genetic differentiation between populations was examined by three different permutation-based statistical tests K_s^*, Z^* and S_{nn} (Hudson 2000). Only geographical areas with more than two isolates of the virus analysed were taken into account. The S_{nn} (the nearest-neighbor statistic) is a measurement of how often the ‘nearest neighbors’ in the sequences were from the same locality in the geographic space. The level of gene flow between populations was measured by estimating N_m and F_{st} statistics using DnaSP 5.10 (Librado and Rozas 2009). Generally, if $N_m < 1$, genetic drift will result in substantial local differentiation; if $N_m > 1$, gene flow between populations is higher and the extent of genetic differentiation is smaller (Slatkin 1987). The absolute values of Fst from 0 to 1 represent undifferentiated to fully differentiated populations (Weir and Cockerham 1984; Wei et al. 2009). Selection pressure was estimated using the dN/dS ratio, where dN represents the average number of nonsynonymous substitutions per non-synonymous site and dS represents the average number of synonymous substitutions per synonymous site. The values of dN and dS were estimated separately using the Pamilo and Bianchi (1993) method in MEGA 5.05 (Tamura *et al*. 2011).

Population genetics and demography analysis

DnaSP 5.10 was used to estimate Tajima’s D, Fu & Li’s D and F statistical tests, and haplotype diversity. Tajima’s D and Fu & Li’s D and F tests measure the departure from neutrality for all mutations in a genomic region (Tajima 1989; Fu and Li 1993). Haplotype diversity refers to the frequency and number of haplotypes in the population. Nucleotide diversity estimates the average pairwise differences among sequences.

Results

Incidence of ToMV infection in symptomatic horticultural crops and weed samples

The results showed that 15 out of 21 fields surveyed had ToMV infection. Of 594 crop leaf samples showing symptoms that were tested by ELISA, ToMV was found in 136 samples (22.9% incidence). ToMV was unevenly distributed among all surveyed provinces with the highest incidence rate in Mazandaran province (24.7%)

located in the north of Iran (Table 1). The prevalence of ToMV plummeted to about 36.6% in Savejbolagh district in the center of Iran. There were differences in the relative incidence and severity (data not shown) of ToMV among the horticultural crops tested. ToMV was distributed on Brassicaceae crops with an average incidence rate of 49.2%, however it was incidenced on solanaceous plants with an average incidence rate of 21.3% (Table 1). ToMV infection rates in crop samples showing symptoms, in decreasing order, were cabbage (55.9%), tomato (29%), cucumber (17.8%), eggplant (15.7%), pepper (10%) and bean (3.3%) (Table 1). It was not detected in radish, watermelon and zucchini plants. Of 45 weed leaf samples tested by DAS-ELISA, ToMV was found in four samples. ToMV was detected in two samples of *S. oleraceus*, one sample of *Ch. album*, and one sample of *M. neglecta* (Table 2). Using the ToMV-spec and Tob-Uni 1 primers, infection of the different crop and weed samples with ToMV was confirmed by RT-PCR with amplicons of the expected size (686 bp) for the tested isolates.

Sequence and phylogenetic analysis

The 862 nt sequences of a portion of the RdRp gene of the new Iranian ToMV isolates were obtained and deposited in the GenBank as accession numbers KF527462, KF527465 and KJ160239 for new isolates A356, T195 and To52 respectively (Table 4). The 795 nt sequences of the entire MP sequences and 480 nt sequences of the entire CP sequences were also obtained and deposited in the GenBank under accession numbers KF527463, KF527468, KJ000537 and KF527464, KF527466, KC534879 for the isolates A356, T195 and To52, respectively (Table 4). BLAST analysis on the basis of the 862 nt sequences of the RdRp gene disclosed 97% identity at the nucleotide level and 98–100% identity at the amino acid level among the Iranian ToMV isolates, respectively. However, relative to the reference ToMV nucleotide sequence they were 97–100% and 98–100% identical at the nucleotide and amino acid levels, respectively (Table 4). BLAST analysis of the sequences of the MP and CP genes disclosed 96% and 95% identity at the nucleotide level and 92–99% and 98–100% identity at the amino acid level among the Iranian ToMV isolates, respectively (Table 4). Minimum nucleotide and amino acid sequences identities of the RdRp, MP and CP genes were observed between the Iranian isolate A356 and those deposited previously in GenBank (Table 4). In general, CP gene showed the highest overall variability to the reference ToMV isolate (KR537870) and the RdRp gene appeared to

Table 4. Accession numbers, host, isolate name, origin and percentage [%] of nucleotide (nt) and amino acid (aa) identity between the coat protein (CP), movement protein (MP) and RNA dependent RNA polymerase (RdRp) gene sequences of selected Iranian isolates and the world with reference isolate (KR537870)

Accession no.			Origin	Isolate	Host	Nucleotide (nt) and amino acid (aa) sequences identity with reference [%]					
						CP		MP		RdRp	
CP	MP	RdRp				nt	aa	nt	aa	nt	aa
AF332868	AF332868	AF332868	Australia	Queensland	Tomato	96	100	97	100	98	100
AJ417701	AJ417701	AJ417701	China	Camellia	Camellia	95	100	97	100	98	100
FN985165	FN985165	FN985165	China	XJT-1	Tomato	96	100	97	100	97	100
GQ280794	GQ280794	GQ280794	China	N5	Tomato	96	100	97	100	98	100
LN827934	LN827934	–	Egypt	AH5	Tomato	96	99	96	90	–	100
LN827937	LN827937	–	Egypt	AH8	Tomato	96	99	96	99	–	100
KU321698	KU321698	KU321698	Egypt	AH4	Tomato	96	99	96	99	97	100
KF527464	KF527463	KF527462	Iran	A356	Common sow thistle	95	98	96	92	97	98
KF527466	KF527468	KF527465	Iran	T194	Tomato	95	98	96	99	97	100
KC534879	KJ000537	KJ160239	Iran	To52	Tomato	95	98	96	99	97	98
AB083196	AB083196	AB083196	Japan	L11A	Tomato	96	99	97	100	98	100
AB355139	AB355139	AB355139	Japan	L11Y	Tobacco	96	99	97	100	98	100
Z92909	Z92909	Z92909	Kazakhstan	K2	Tomato	95	99	97	100	98	100
KR537870	KR537870	KR537870	USA	99-1	Winter Jasmine	100	100	100	100	100	100
KJ207374	KJ207374	KJ207374	Taiwan	Penghu	Peppino	96	99	97	100	98	100
JX534224[a]	JX534224	JX534224	China	Xiamen	Pepper	–	–	–	–	–	–

[a]*Tobacco mild green mosaic virus* (TMGMV), a member of the genus *Tobamovirus*, was used as an out-group species

be the most highly conserved genomic region among the three studied isolates and the reference ToMV isolate (Table 4). To determine phylogenetic relationships among the Iranian ToMV isolates, phylogenetic trees were constructed using samples that showed homogeneous populations and were composed of a predominant genetic variant obtained in this research and those available in the GenBank database. Phylogenetic analysis based on the nucleotide sequences of the CP gene resulted in the classification of ToMV isolates into three major groups with high bootstrap support in which isolates in group I were divided into three subgroups (Fig. 1A) indicating the existence of divergent isolates of ToMV in Iran. Iranian ToMV isolates clustered in groups II (collected from the Tehran province in the centre of Iran) and III (collected from the Mazandaran province in the north of Iran), with 95% similarity to ToMV-99-1 (the type ToMV species present in group I). Other ToMV isolates (Table 4) were clustered in: the larger subgroup I, sharing 96% nt sequence identities with isolate ToMV-99-1 and comprising isolates from Asia, Australia and America, or subgroup II, composed of isolates only from Egypt or the smaller subgroup III, composed of isolates from Japan and Kazakhstan (Fig. 1A). The phylogenetic tree based on analysis of the CP amino acid sequences showed a different pattern than that of the nucleotide sequences and formed two main groups in which isolates in group I were divided into two subgroups (Fig. 1B). ToMV-tomato infecting isolate To52, was clustered in phylogeny group II together with the tobacco-infecting isolate L11Y from Japan. Other Iranian (tomato and common sow thistle-infecting) isolates clustered in the small subgroup II, and in phylogeny group there were isolates from China and Kazakhstan. No correlation was found between genetic variation and the hosts. However, on the basis of the CP nucleotide sequences, ToMV isolates from Iran in phylogeny groups II and III and isolates from Egypt in subgroup II were distinct from other isolates, indicating a clear correlation between genetic variation and geographical origin of the isolates (Fig. 1A).

The phylogenetic tree constructed with MP nucleotide sequences showed that the ToMV isolates can be clustered into three different subgroups on the basis of geographical origin but irrespective of host origin (Fig. 1C). Subgroup I contained isolates from Australia, China, Japan, Kazakhstan and Taiwan. Subgroup II had only Iranian isolates. In phylogeny subgroup III isolates from Egypt were found beside isolates from China, Japan and USA. The phylogenetic tree based on analysis of the MP amino acid sequences showed a different pattern than that of the nucleotide sequences of the MP gene (Fig. 1D). The sequences of the amino acids of the MP gene formed two groups and two subgroups radiating from one large group I. Isolates from Australia, China, Japan and Kazakhstan were placed in subgroup I. Distribution of Chinese ToMV isolates into two subgroups in the phylogenetic tree (Fig. 1D) agreed with that in the nucleotide-based tree (Fig. 1C and D). The Iranian isolates in phylogeny group II were placed separately from other isolates, sharing 92–99% identity to the reference ToMV-99-1 isolate. Egyptian isolates in subgroup II were placed alongside isolates from China, Taiwan and USA with the highest amino acid identity of 100% found in the reference isolate ToMV-99-1. On the basis of the nucleotide and amino acid sequences the resulting trees for RdRp gene formed two groups (Figs. 1E and F). Isolates were mostly distributed in group I, while Egyptian isolate AH4 (Table 4) formed distinct group II indicating low genetic diversity of the studied ToMV isolates on the basis of the RdRp gene.

Statistical analysis of genetic distance, genetic differentiation, gene flow and selection pressure

To assess the genetic diversity of ToMV populations, we estimated average nt distances for each of the functional genes using the sequences of the three Iranian isolates determined in this study as well as ten other isolates from Asia and Egypt. On the basis of the nucleotide sequences of Asian, Egyptian and Iranian isolates, CP showed greatest mean genetic diversity between (0.027) and within (0.022) studied geographic groups. The mean nucleotide diversities between and within studied geographic groups were 0.004 and 0.015 for the MP gene and 0.015 and 0.014 for the RdRp gene, respectively. On the basis of the CP gene the within-group genetic distances of Asian, Egyptian and Iranian geographic groups were: 0.018 ± 0.001, 0.011 ± 0.001 and 0.037 ± 0.003, respectively. The inter-group genetic distances between Iranian and Asian, Iranian and Egyptian, and Asian and Egyptian geographic groups were: 0.032 ± 0.002, 0.033 ± 0.001 and 0.017 ± 0.001, respectively. On the basis of the RdRp and MP genes the within-group genetic distances were not greater than the intergroup genetic distances of the different geographic groups. Analysis of the K_s*, Z* and S_{nn} p-values for the MP gene indicated significant genetic differentiation between Asian and Egyptian populations. Although the p-value for the S_{nn} statistic calculated for populations from Iran and Egypt was not significant, the K_s* and Z* were significant indicating differentiation between Iranian and Egyptian populations (Table 5). Analysis of p-values of these three statistics for the RdRp gene was significant for ToMV in Iranian and Asian populations suggesting strong genetic differentiation between Iranian and Asian populations. On the basis of the MP and RdRp genes but not the CP

Table 5. Genetic differentiation and gene flow between ToMV populations

Genomic region	N_m			F_{st}			K_s* (p-value)			Z*(p-value)			S_{nn} (p-value)		
	Between Iran and Asia	Between Iran and Egypt	Between Asia and Egypt	Between Iran and Asia	Between Iran and Egypt	Between Asia and Egypt	Between Iran and Asia	Between Iran and Egypt	Between Asia and Egypt	Between Iran and Asia	Between Iran and Egypt	Between Asia and Egypt	Between Iran and Asia	Between Iran and Egypt	Between Asia and Egypt
RdRp	0.36	ND	ND	0.560	ND	ND	0.2404 (0.020)*	ND	ND	2.9160 (0.014)*	ND	ND	0.8051 (0.017)*	ND	ND
MP	0.98	0.61	0.69	0.343	0.390	0.366	2.1493 (0.032)*	2.4886 (0.031)*	1.8491 (0.008)*	2.9404 (0.101)	2.1161 (0.034)*	2.7840 (0.008)*	0.6590 (0.154)	0.7500 (0.082)	0.8863 (0.004)*
CP	1.80	1.62	5.69	0.1223	0.134	0.042	1.8478 (0.081)	2.1169 (0.152)	1.6699 (0.281)	3.1280 (0.093)	2.3084 (0.072)	3.2189 (0.324)	0.5138 (0.296)	0.6875 (0.095)	0.7777 (0.057)

N_m and F_{st} are parameters for gene flow. If the absolute value of N_m is greater than 1 or $F_{st} < 0.33$, the gene flow between two populations is frequent, ns – no significant values; *$0.01 < p < 0.05$. K_s*, Z* and S_{nn} values are parameters for genetic differentiation, $p < 0.05$ was used as the criterion for rejecting the null hypothesis that there is no genetic differentiation between two populations; ND – not determined because of the low corresponding sequences available in GenBank

gene, the N_m values for populations from Asia, Egypt and Iran were <1 (Table 5) and the absolute values of F_{st} for all the populations were >0.33, also indicating infrequent MP and RdRp but frequent CP gene flow between ToMV populations (Table 5). To further analyze the variability of these genes and their selection pressures, we calculated the dN/dS ratio. The dN/dS ratios varied significantly between different genes. It was the smallest for the RdRp gene and the biggest for the CP gene, which implied that the constraint on RdRp was the highest, and that on CP it was the lowest (Table 6). dN/dS ratios for different populations derived from MP and RdRp genes were <1 and bigger than 1 for the CP gene of the Egyptian and Iranian ToMV isolates indicating that different ToMV, ORFs were under different levels of selection (Table 6).

Neutrality tests and population demography

Neutrality tests were used to assess the evidence of selection or demographic forces acting on the ToMV populations. The neutrality values generated from Tajima's D and Fu & Li's F and D tests for the different regions of all the studied populations were negative regardless of host and geographic origin (Table 6). In contrast to the Asian population, values generated from neutrality tests of different studied genes for the Iranian and Egyptian populations were not significantly different ($p > 0.05$), and the result was not conclusive (Table 6). On the basis of the nucleotide sequences for CP and MP genes of ToMV isolates, the Iranian population showed greater genetic variability than other populations and isolates from Egypt had the lowest

Table 6. Neutrality tests, haplotype and nucleotide diversity of *Tomato mosaic virus* (ToMV) populations calculated for different ORFs

Genomic region	Population	Tajima's D	Fu and Li's D	Fu and Li's F	Haplotype diversity	Nucleotide diversity‡	dN/dS
RdRp	Iran	–0.61237	–0.61237	–0.47871	0.500±0.07	0.0172± 0.0091	0.000
	Asia	–1.96119*	–2.29857**	–2.49293**	0.643±0.03	0.01035±0.0021	0.452
	Egypt	ND	ND	ND	ND	ND	ND
MP	Iran	–0.63894	–0.33665	–0.41450	1.000±0.03	0.0282±0.0068	0.959
	Asia	–1.51007	–1.53078	–1.69809	1.000±0.003	0.0078±0.0013	0.345
	Egypt	–0.86356	–0.86356	–0.91038	1.000±0.03	0.0078±0.0016	0.188
CP	Iran	–0.85318	–0.44717	–0.54334	1.000±0.03	0.0381±0.0087	3.452
	Asia	–1.75917*	–1.83370*	–2.02039*	0.972±0.004	0.0137±0.0032	0.571
	Egypt	–0.83379	–0.83379	–0.83370	1.000±0.03	0.0120±0001	2.101

Tajima's D and Fu and Li's F and D are neutrality tests that used as an indication of recent population expansion. Strongly negative and significant values suggest recent population expansion or selection; *significant values $0.01 < p < 0.05$; **significant values $0.001 < p < 0.01$
‡nucleotide diversity was estimated using the average pairwise differences at all sites in the sequences in a given sample, numbers after '±' indicate standard deviations, the dN/dS ratios represent estimated selection pressures, values of dN/dS < 1.0 imply negative selection, values of dN/dS = 1.0 imply neutral selection, and values of dN/dS > 1.0 imply positive selection; ND – not determined because of the low corresponding sequences available in GenBank

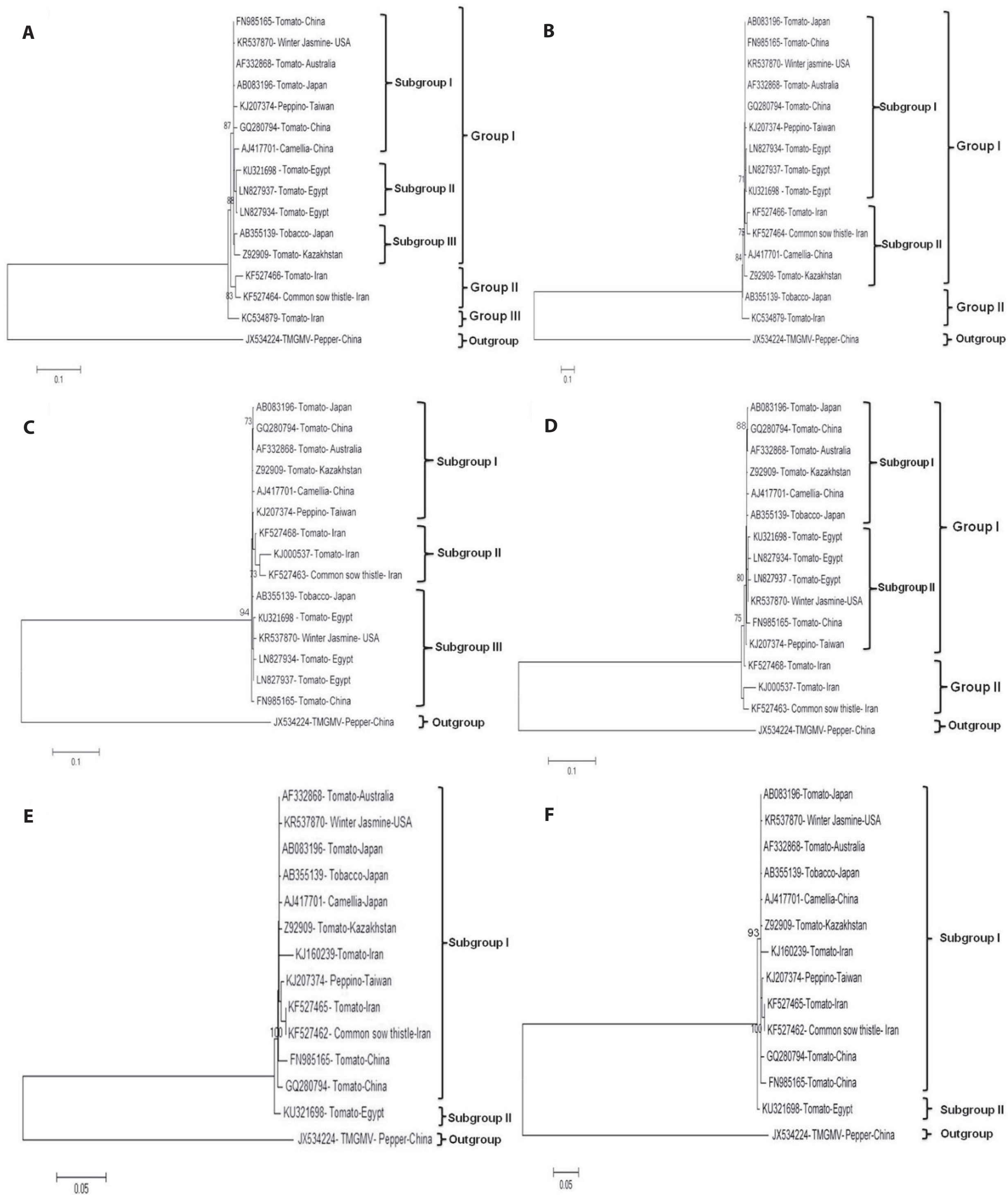

Fig. 1. Phylogenetic trees of ToMV

Phylogram generated from the alignment of nucleotide (nt) and amino acid (aa) sequences of three new selected Iranian ToMV isolates, together with the homologue sequences of those from Asia, Australia, Egypt and USA (Table 4). The phylogenetic trees of ToMV were constructed on the basis of the sequences of CP (a: nt and b: aa), MP (c: nt and d: aa) and RdRp (e: nt and f: aa) genes, respectively. The corresponding genomic sequences of *Tobacco mild green mosaic virus* (TMGMV; accession number: JX534224) was used as outgroup to root the tree. Nodes with less than 70% bootstrap support were collapsed; only bootstrap values >70% were shown

values (Table 6). Haplotype diversity for different genes was calculated based on the frequency and number of haplotypes in the population. The haplotype diversity in Asian, Egyptian and Iranian populations was 1.000 however, it was lower than one for the RdRp gene, confirming the low genetic variation of this gene compared to the others (Table 6).

Discussion

The present report describes the occurrence of an infection by ToMV in new crops and herbaceous weed plants. ToMV was found in 55.9% of tested cabbage and 3.3% of tested bean plants in the researched

provinces. It is noteworthy that ToMV, a virus hitherto unreported from cabbage in Iran, had high rates of incidence in Alborz province. Detection of ToMV in cabbage could be because of the exchange of infected materials between solanaceous and brassicas cultivations in the studied regions in Iran.

Although, ToMV was previously reported from common sow thistle (*S. oleraceus*) in Iran (Hashemi *et al.* 2014), the natural occurrence of this virus on *M. neglecta* had not been previouisly reported. To our knowledge, this is the first report of the occurrence of ToMV in cabbage, bean and *M. neglecta* in Iran. The uneven distribution of this virus within different crops in the researched regions could be attributed to the presence of resistant sources among the crop plant species or different levels of virus inoculum in the fields. A previous study suggested that temperature and relative humidity are the most critical factors in ToMV disease development (Imran *et al.* 2013).

In this study, the CP, MP and RdRp sequences shared 95 to 100%, 96 to 100% and 97 to 100% nucleotide identity with the reference isolate (KR537870), respectively, indicating that the variability of CP is greater than that of MP or RdRp nucleotide sequences. Consistent with this observation, CP showed the greatest mean genetic diversity between (0.027) and within (0.022) studied geographic groups. Analysis of selective pressure on these genes suggested that in contrast to the CP, other genes were under purifying selection. Previously, positive selection was reported worldwide as well as in Iranian ToMV isolates on the basis of the CP gene sequence (Rangel *et al.* 2011; Aghamohammadi *et al.* 2013). The presence of positively selected sites in the CP gene of ToMV isolates might be necessary for adaption to different biological conditions. In nonvector-borne viruses like ToMV, surface structural genes which frequently contain sites undergoing adaptive evolution are generally more subject to positive (diversifying) selection (Woelk and Holmes 2002). It has been elucidated that positive selection in the N-terminal region of CP gene plays a key role in virus–host interactions and several other functions of potyviruses (Ullah *et al.* 2003). The presence of positive selection pressure, could also lead to the development of resistance-breaking strains of Tomato spotted wilt virus (Sundaraj *et al.* 2014).

The dN/dS ratios varied significantly between different genes (Table 6). It was the smallest for the RdRp gene and the largest for the CP gene, which implied that the constraint on RdRp was the highest, and that on CP, the lowest. RdRp plays essential roles in viral life cycles including replication and suppressing RNA silencing of tobamoviruses (Kubota *et al.* 2003). Higher constraints on the RdRp gene will help RdRp protein maintain its structure and fulfill its functions (Garcia-Arenal *et al.* 2001). High genetic stability of the suppressor of RNA silencing has been previously reported for *Tobacco vein banding mosaic virus* (TVBMV), a distinct species of the genus *Potyvirus* (Zhang *et al.* 2011). However, it is in contrast to what has been reported for viral suppressors of RNAi (VSRs) of RNA viruses. It has been hypothesized for plant viruses that VSRs diversify faster than other genes by positive selection (Murray *et al.* 2013). Genetic bottleneck(s) may also have contributed to this observed low genetic diversity and function to minimize the extent of genetic variation of RdRp. Phylogenetic trees on the basis of the CP and MP genes' nucleotide and amino acid sequences indicated the distribution of Asian ToMV isolates into different subgroups (Fig. 1) which suggests an old presence of this virus in Asia. The close genetic relationships between some geographically distant isolates suggest long-distance migration, probably due to the international exchange of infected plant material such as seeds. This has also been observed for other plant viruses (Walia *et al.* 2014). The international trade of infected seed contaminated with tobamoviruses has facilitated the spread of the viruses to different parts of the world (Gibbs *et al.* 2015). ToMV isolates were clustered in three groups according to their CP genes. In contrast to our previous study (Aghamohammadi *et al.* 2013), here, we provided new evidence for further clustering isolates of the first group into three subgroups and Iranian isolates into three distinct groups with some geographical origin specificity.

The trees constructed based on RdRp, MP and CP nucleotide sequences showed different groupings (Figs. 1A, C and E). The different topologies of the phylogenetic trees constructed, using available ORF1, ORF2 and 3 sequences from ToMV isolates, indicated different evolutionary constraints in these genomic regions. In the phylogenetic tree inferred from the nucleotide sequences of the complete CP and MP gene sequences, isolates from Iran and Egypt formed separate groups, irrespective of host origin (Figs. 1A, C and E), however, isolates clustered into groups with correlation to geographic origin but not the host. This finding indicates that there is genetic differentiation between ToMV populations.

In line with this finding, Fraile and co-workers (Fraile *et al.* 1996) in their research on genetic diversity of different global *Tobacco mild green mosaic viruses* (TMGMV), another member of the Tobamovirus genus, showed that different selection mechanisms may operate on the different TMGMV geographical populations. It was suggested that the formation of the distinct TMGMV badnavirus lineages was mostly attributed to mechanisms such as bottlenecks and founder effects in new geographical environments.

Genetic distances between different geographical groups were also assessed and results showed that ToMV could be classified into different genotypes with respect to the geographic origin. In line with these results N_m values for different populations on the basis of RdRp and MP genes were < 1, suggesting gene flow was infrequent between the populations and genetic drift will result in substantial geographical differentiation (Table 5). Absolute values of F_{st} for all the populations were > 0.33, also indicating infrequent MP and RdRp gene flow between ToMV populations. However, CP gene flow between ToMV populations (N_m values > 1 and F_{st} < 0.33) was observed which is similar to the results that have previously been reported for populations of this virus in different geographic regions (Rangel *et al.* 2011).

Natural selection and genetic drift are two main evolutionary mechanisms limiting genetic variation of virus populations (Moya *et al.* 2004). Genetic drift can lead to a decrease in diversity within populations and an increase in diversity between populations (Hall 2006). Thus, Iranian and Egyptian isolates analysed here could be considered to be two distinct populations that have evolved separately. In accordance with these findings, results obtained for K_s*, Z* and S_{nn} statistics showed significant genetic differentiation between the RdRp of the Iranian and Asian populations as well the MP of the Asian and Egyptian populations (Table 5). Natural selection and genetic drift drive the evolution of virus populations within their hosts and therefore strongly influence virus emergences (Fabre *et al.* 2012). On the basis of the CP nucleotide sequence Iranian populations had conspicuously higher nucleotide diversity values than other populations (Table 6). High genetic diversity has often been shown to be positively correlated with indicator values of individual fitness and adaptability to changing environmental conditions (Booy *et al.* 2000). The Tajima's D, Fu and Li's statistics values were negative for all the studied populations, suggesting ToMV populations are increasing in different parts of the world (Table 6). However, a significantly negative value for the RdRp gene of Asian populations suggests the sudden expansion of ToMV population in Asia and a strong negative or purifying selection on the RdRp gene (Table 6). The significantly negative values of neutrality tests obtained for Asia may also suggest that the population of ToMV isolates in Asia is increasing in size. Because the genetic variation and evolutionary factors that shaped ToMV populations have not been studied extensively, the results of this study may provide useful information for understanding the molecular evolution present within the global ToMV population and give insight into viral disease management.

References

Adams M.J., Antoniw J.F., Kreuze J. 2009. Virgaviridae: a new family of rod-shaped plant viruses. Archives of Virology 154 (12): 1967–1972. DOI: https://doi.org/10.1007/s00705-009-0506-6

Aghamohammadi V., Rakhshandehroo F., Shams-Bakhsh M., Palukaitis M. 2013. Distribution and genetic diversity of *Tomato mosaic virus* isolates in Iran. Journal of Plant Pathology 95 (2): 339–347. DOI: http://dx.doi.org/10.4454/JPP.V95I2.030

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25 (17): 3389–3402. DOI: https://doi.org/10.1093/nar/25.17.3389

Booy G., Hendriks R.J.J., Smulders M.J.M., van Groenendael J.M., Vosman B. 2000. Genetic diversity and the survival of populations. Plant Biology 2: 379–395. DOI: 10.1055/s-2000-5958

Converse R.H., Martin R.R. 1999. ELISA methods for plant viruses. p. 179–196. In: "Serological Methods for Detection and Identification of Viral and Bacterial Plant Pathogens, A Laboratory Manual" (R. Hampton, E.M. Ball, S.H. De Boer, eds.). APS Press, St. Paul, MN, 389 pp.

Ding X.S., Liu J., Cheng N.H., Folimonov A., Hou Y.M., Bao Y., Katagi C., Carter S.A., Nelson R.S. 2004. *The tobacco mosaic virus* 126-kDa protein associated with virus replication and movement suppresses RNA silencing. Molecular Plant--Microbe Interactions 17 (6): 583–592. DOI: https://doi.org/10.1094/mpmi.2004.17.6.583

Fabre F., Montarry J., Coville J., Senoussi R., Simon V., Moury B. 2012. Modelling the evolutionary dynamics of viruses within their hosts: a case study using high-throughput sequencing. PLoS Pathogens 8 (4): e1002654. DOI: https://doi.org/10.1371/journal.ppat.1002654

Fraile A., Malpica J.M., Aranda M.A., Guez-Cerezo E.R., Garcia-Arenal F. 1996. Genetic diversity in *Tobacco mild green mosaic tobamovirus* infecting the wild plant *Nicotiana glauca*. Virology 223 (1): 148–155. DOI: https://doi.org/10.1006/viro.1996.0463

Fu Y.X., Li W.H. 1993. Statistical tests of neutrality of mutations. Genetics 133 (3): 693–709.

Garcia-Arenal F., Fraile A., Malpica J.M. 2001. Variability and genetic structure of plant virus populations. Annual Review of Phytopathology 39: 157–186. DOI: https://doi.org/10.1146/annurev.phyto.39.1.157

Gibbs A.J., Wood J., Garcia-Arenal F., Ohshima K., Armstrong J.S. 2015. *Tobamoviruses* have probably co-diverged with their eudicotyledonous hosts for at least 110 million years. Virus Evolution 1 (1): vev019. DOI: https://doi.org/10.1093/ve/vev019

Hall G. 2006. Selective constraint and genetic differentiation in geographically distant barley yellow dwarf virus populations. Journal of General Virology 87 (10): 3067–3075. DOI: https://doi.org/10.1099/vir.0.81834-0

Harrison B.D. 2002. Virus variation in relation to resistance-breaking in plants. Euphytica 124: 181–192.

Hashemi S.S., Rakhshandehroo F., Shahraeen N. 2014. First Report of *Tomato mosaic virus* on common sow thistle in Iran. Plant Disease 98 (8): 1164–1164. DOI: https://doi.org/10.1094/pdis-03-14-0220-pdn

Holling M., Huttinga H. 1995. Tomato mosaic virus. CMI/AAB Descriptions of Plant Viruses No 156. [Available on: http://www.dpvweb.net/dpv/showdpv.php?dpvno=156]

Hudson R.R. 2000. A new statistic for detecting genetic differentiation. Genetics 155 (4): 2011–2014.

Hu Q., Jiang T., Xue C., Zhou X. 2012. Characterization and complete nucleotide sequence of two isolates of Tomato mosaic virus. Journal of Phytopathology 160 (3): 115–119. DOI: https://doi.org/10.1111/j.1439-0434.2011.01866.x

Imran M., Khan M.A., Fiaz M., Azeem M., Mustafa M. 2013. Influence of environmental conditions on *Tomato mosaic virus* disease development under natural condition. Pakistan Journal of Phytopathology 25 (2): 117–122.

Ishikawa M., Meshi T., Motoyoshi T., Takamatsu N., Okada Y. 1986. *In vitro* mutagenesis of the putative replicase genes of *Tobacco mosaic virus*. Nucleic Acids Research 14 (21): 8291–8305. DOI: https://doi.org/10.1093/nar/14.21.8291

Kubota K., Tsuda S., Tamai A., Meshi T. 2003. *Tomato mosaic virus* replication protein suppresses virus-targeted post-transcriptional gene silencing. Journal of Virology 77 (20): 11016–11026. DOI: https://doi.org/10.1128/jvi.77.20.11016-11026.2003

Letschert B., Adam G., Lesemann D.E., Willingmann P., Heinze P. 2002. Detection and differentiation of serologically cross-reacting tobamoviruses of economical importance by RT-PCR and RT-PCR-RFLP. Journal of Virological Methods 106 (1): 1–10. DOI: https://doi.org/10.1016/s0166-0934(02)00135-0

Librado P., Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25 (11): 1451–1452. DOI: https://doi.org/10.1093/bioinformatics/btp187

Moya A., Holmes E.C., Gonzalez-Candelas F. 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nature Reviews Microbiology 2 (4): 279–288. DOI: https://doi.org/10.1038/nrmicro863

Murray G.G.R., Kosakovsky Pond S.L., Obbard D.J. 2013 Suppressors of RNAi from plant viruses are subject to episodic positive selection. Proceedings of The Royal Society B: Biological Science 280 (1756): 20130965–20130965. DOI: https://doi.org/10.1098/rspb.2013.0965

Nei M., Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York, NY, USA, 352 pp.

Pamilo P., Bianchi N.O. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Molecular Biology and Evolution 10 (2): 271–281. DOI: https://doi.org/10.1093/oxfordjournals.molbev.a040003

Pearson W.R., Lipman D.J. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences 85 (8): 2444–2448. DOI: https://doi.org/10.1073/pnas.85.8.2444

Rangel E.A., Alfaro-Fernández A., Font-San-Ambrosio M.I., Luis-Arteaga M., Rubio L. 2011. Genetic variability and evolutionary analyses of the coat protein gene of *Tomato mosaic virus*. Virus Genes 43 (3): 435-438. DOI: https://doi.org/10.1007/s11262-011-0651-3

Saitou N., Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4 (4): 406–425. DOI: https://doi.org/10.1093/oxfordjournals.molbev.a040454

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science 236 (4803): 787–792. DOI: https://doi.org/10.1126/science.3576198

Sundaraj S., Srinivasan R., Culbreath A.K., Riley D.G., Pappu H.R. 2014. Host plant resistance against Tomato spotted wilt virus in peanut (*Arachis hypogaea*) and its impact on susceptibility to the virus, virus population genetics, and vector feeding behavior and survival. Phytopathology 104 (2): 202–210. DOI: https://doi.org/10.1094/phyto-04-13-0107-r

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Takamatsu N., Ohno T., Meshi T., Okada Y. 1983. Molecular cloning and nucleotide sequence of the 30K and the coat protein cistron of TMV (tomato strain) genome. Nucleic Acids Research. 11 (11): 3767–3778. DOI: https://doi.org/10.1093/nar/11.11.3767

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28 (10): 2731–2739. DOI: https://doi.org/10.1093/molbev/msr121

Ullah Z., Chai B., Hammar S., Raccah B., Gal-On A., Grumet R. 2003. Effect of substitution of the amino termini of coat proteins of distinct potyvirus species on viral infectivity and host specificity. Physiological and Molecular Plant Pathology 63 (3): 129–139. DOI: https://doi.org/10.1016/j.pmpp.2003.11.001

Walia J. J., Willemsen A., Elci E., Caglayan K., Falk B. W., Rubio L. 2014. Genetic variation and possible mechanisms driving the evolution of worldwide *Fig mosaic virus* isolates. Phytopathology 104 (1): 108–114. DOI: https://doi.org/10.1094/phyto-05-13-0145-r

Wei T.Y., Yang J.G., Liao F.R., Gao F.L., Lu L.M., Zhang X.T., Li F., Wu Z.J., Lin Q.Y., Xie L.H., Lin H.X. 2009. Genetic diversity and population structure of rice stripe virus in China. Journal of General Virology 90 (6): 1025–1034. DOI https://doi.org/10.1099/vir.0.84058-0

Weir B.S., Cockerham C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38 (6): 1358–1370. DOI: https://doi.org/10.2307/2408641

Woelk C.H., Holmes E.C. 2002. Reduced positive selection in vector-borne RNA viruses. Molecular Biology and Evolution 19 (12): 2333–2336. DOI: https://doi.org/10.1093/oxfordjournals.molbev.a004059

Zhang C.L., Gaob R., Wanga J., Zhang G.M., Li X.D., Liuc H.T. 2011. Molecular variability of *Tobacco vein banding mosaic virus* populations. Virus Research 158 (1–2): 188–198. DOI: https://doi.org/10.1016/j.virusres.2011.03.031