J Genet Resour2021;7(2): 211-219 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2021.20849.1238*

J|G|R

Phylogenetic and Population Analysis of *Lettuce mosaic virus* Isolates Based on the Coat Protein Gene

Mahsa Abadkhah, Hossein Hamidi and Davoud Koolivand*

Department of Plant Protection, Faculty of Agriculture, University of Zanjan, Zanjan, Iran

ARTICLE INFO

Article history:
Received 25 March 2021
Accepted 29 May 2021
Available online 05 June 2021

Keywords:
Coat protein gene
Lettuce mosaic virus
Molecular evolution
Negative selection pressure

Corresponding authors:
✉ D. Koolivand
koolivand@znu.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Lettuce mosaic virus (LMV) is one of the most important and destructive members of *Potyviruses* in the family *Potyviridae* that is transmitted by aphid and seed which cause economic damage and serious yield losses to different crops around the world. To explore phylogenetic relationships, population evolution, and the effect of selection forces on the complete coat protein (CP) gene of LMV isolates, 36 complete coat protein gene sequences retrieved from the GenBank database in several different countries in four continents: Central and East Asia (China, Taiwan, South Korea, and Turkey), Europe (France), Africa (Tunisia), and America (Chile and Brazil). The phylogenetic tree of LMV isolates was grouped into independent clades with the significance of F_{ST} values (>0.27). The ratio of dN/dS is calculated less than one and showed that the LMV-CP has been under negative selection. Statistical tests (Tajima's D, Fu and Li's D* and F*) were used to estimate non-significantly negative values for all clades and geographic populations except for the Brazilian population and Clade III in a phylogenetic group. The negatives values revealed that there is less polymorphism estimation. All geographic populations in four phylogroups of LMV seem to be at equilibrium because all neutrality test statistics were non-significant. The findings suggested that the dynamics of LMV molecular evolution may be dependent on mutation, recombination, and negative selection. No recombination events were observed in this part of the LMV genome. Therefore, this study provides the first time evolution and differentiation between populations of LMV isolates from around the world and suggested that they may be occurred by the transmission of the virus among lettuce plants by types of aphid species, migration in different geographical areas by infected seeds and plant material, and broad host range which seems that these events have played an important role in shaping the LMV population structure.

Please cite this paper as: Abadkhah M, Hamidi H, Koolivand D. 2021. Phylogenetic and population analysis of *Lettuce mosaic virus* isolates based on the coat protein gene. *J Genet Resour* 7(2): 211-219. doi: 10.22080/jgr.2021.20849.1238.

Introduction

Potyvirus genus is the biggest genus in the family *Potyviridae* and other genera, and one of the most important viruses caused economic serious damage to a wide range of plants. Lettuce (*Lactuca sativa* L.) is a member of the *Lactuca* genus and the *Asteraceae* family. One of the economically important viruses that affect lettuce, is *Lettuce mosaic virus* which belongs to the genus *Potyvirus* with worldwide distribution (Fakhfakh *et al*., 2001) and it was recorded for the first time in 1920 in Florida, the USA by Jagger (Jagger, 1921) and thereafter has been subsequently reported from all continents (German-Retana *et al*., 2008; Sharma *et al*., 2016). Like other potyviruses, LMV is a seed-borne and aphid-borne virus in lettuce cultivars and causes severe damage and reduction of yield and quality in crops, and it can be transmitted via several aphid species including cotton aphid (*Aphis gossypii*), the potato aphid (*Macrosiphum*

euphorbiae), and the green peach aphid (*Myzus persicae*) in a non-persistent transmission manner (Moreno *et al*., 2007b). Its genome consists of a single-strand positive-sense RNA (+ssRNA) which encodes into a unique polyprotein and is divided into ten proteins as follows: proteinase 1 (Pro 1), helper component-protease (HC-Pro), P3, 6K1, cylindrical inclusion (CI), 6K2, nuclear inclusion A (NIa), a viral protein linked genome (VPg), nuclear inclusion B (NIb), and CP. Also, a short polypeptide (PIPO) is expressed within the P3 cistron by frameshifting (Riechman *et al*., 1992; Lim *et al*., 2014).

Lettuce plants infected with LMV show an array of symptoms and these general symptoms are chlorosis, reduction of plant growth, mosaic, dwarfism, yellowing, necrotic spots, wilting, and mottling. Indeed, the severity of the symptoms caused by LMV relies on many events including the type of cultivar, the environmental conditions, stage of the host infection, and virus genotype (Moreno *et al*., 2007a; German-Retana *et al*., 2008; Soleimani *et al*., 2011).

According to previous studies conducted in the world, LMV isolates based on the molecular characterization have been divided into three main groups including LMV-Yar (Yemen), LMV-Greek (Greece), and LMV-RoW (Rest of the World) which LMV-RoW group containing two large subgroups of seed-borne isolates, LMV-Most and LMV-Common (Krause-Sakate *et al*., 2002; German-Retana *et al*., 2008). In recent years, a new phylogroup has been added for LMV isolates called LMV-Cr (*Catharanthus roseus*) which the new group is not associated with three previously groups (Svanella-Dumas *et al*., 2014). To produce healthy lettuce crops, both genetic control and virus-free seeds were used. For genetic control of LMV, three resistance genes of LMV have been identified in lettuce, two recessive allelic resistance genes $mo1^1$ (formerly named *g*) and $mo1^2$ (formerly named *mo*), which encodes protein elF4E (eIF4E) in lettuce crops (Nicaise *et al*., 2003) and a dominant gene *Mo2* (Krause-Sakate *et al*., 2005). However, LMV isolates in the LMV-Most group are transmitted by seeds and there has a potential to break two recessive resistance genes in hosts including mol^1 and mol^2 (Bos *et al*., 1994; Revers *et al*., 1997; Lim *et al*., 2014).

Recombination and mutation events are common and the most important evolutionary factors in genetic diversity and variation in plant RNA viruses especially in Potyvirus populations (Gibbs *et al*., 2020). Indeed, plant RNA viruses have the abundant potential for adaptation to new conditions (Hajizadeh *et al*., 2019). So far, there has not been any study on the population genetic structure and molecular evolution of LMV; hence, the main aims of the recent work were to investigate the genetic variation and population analysis of LMV isolates, survey the genetic relationships of LMV isolates from different countries that retrieved in the GenBank database.

Materials and Methods

LMV sequences and phylogeny

The complete coat protein sequences of 36 isolates of LMV were selected from the GenBank database from several countries based on the region, collection date, and plant host (Table 1). Table 1 shows plant hosts, regions (country), collection dates, and accession numbers of LMV isolates used in this work. Based on the NCBI information, there are many sequences related to *Lettuce mosaic virus* in different sizes and different genomic regions, but most of them are partial sequences related to a short region of the LMV genome. In the current study, the coat protein gene has been selected for future analysis because the coat protein genes are representative to show the complete genome properties in potyviruses and there are a few (around 28) complete sequences of LMV deposited in the GenBank. In this research, the complete coat protein gene of LMV isolates from different countries, different hosts, and different times was retrieved from the Genbank (Table 1).

For population genetic analysis, multiple sequence alignment of 36 nucleotide sequences of complete CPs without outgroup was performed with the ClustalW algorithm (Larkin *et al*., 2007) implemented in the program MEGA v.7.0.21 (Kumar *et al*., 2016). To evaluate the evolutionary relationship of LMV isolates, an unrooted phylogenetic tree for the complete coat protein genes of 36 LMV isolates from four continents of the world: American continent

(Brazil and Chile), Central and East Asian continent (Turkey, China, South Korea, and Taiwan), European continent (France), and African continent (Tunisia) populations, was designed by the Neighbor-Joining (NJ) method based on Tajima's model implemented in MEGA v.7.0.21, and branch support was computed using the bootstrap method based on 1000 replications and all branches with < 50% bootstrap were collapsed (Kumar *et al*., 2016). Moreover, Sequence Demarcation Tool version 1.2 software (SDT v1.2) was used to obtain a pairwise nucleotide sequence identity matrix (Muhire *et al*., 2014).

Table 1. List of the origins, hosts, and accession numbers of LMV isolates/strains.

Location	Isolates/strains	Host	Accession Number	Collection Date
France	FR25, 13, -	Dimorphotheca sp., Lactuca sativa, -	KJ161186, KJ161173, NC_003605	2008, -, -
China	Beijing, Yuhang, HZ, Shandong	Lettuce, Lettuce, Lactuca sativa, lettuce	EF423619, AJ306288, AJ488153, EF633502	-, -, -, -
South Korea	Muju	lettuce	KF955619	2011
Taiwan	CH, NT, HL	Pisum sativum, lettuce, Pisum sativum	MH844631, MK483102, MH844632	2016, 2015, 2015
Chile	CL208, CL246, CL281, CL394,CL427, CL427-ev, CL464, CL574, CL574-ev, CL117	Lactuca sativa, Lactuca virosa, Lactuca sativa, Lactuca sativa, Lactuca virosa, Lactuca virosa, Lactuca sativa, Lactuca virosa, Lactuca virosa, Dimorphotheca sp.	KJ161176, KJ161177, KJ161178, KJ161179, KJ161180, KJ161181, KJ161182, KJ161183, KJ161184, KJ161175	2005, 2005, 2005, 2006, 2007, 2007, 2007, 2007, 2007, 2003
Turkey	BLK 385, BLK 383, BLK 366, BRS 190, CNK 382, CNK 345, CNK 249	Lettuce, lettuce, lettuce, lettuce, lettuce, lettuce, lettuce	KX378997, KX378996, KX378994, KX378993, KX378987, KX378983, KX378982	2015, 2014, 2013, 2014, 2013, 2013, 2015
Tunisia	Ham6, Tn516-ev, Tn517, Tn517-ev	Dimorphotheca sp., Dimorphotheca sp., Dimorphotheca sp., Dimorphotheca sp.	KJ161187, KJ161191, KJ161192, KJ161193	2002, 2005, 2005, 2005
Brazil	AF199-2013, Br6, AF199-2013-Cr-Var1,-	Lettuce, Lactuca sativa, Lettuce, Hypochaeris chillensis	KF268954, KJ161174, KF268955, MK140596	2013, -, 2013, 2016

Population genetic analysis

Aligned LMV sequences were investigated by using the DnaSP version 6.10.01 program to estimate the genetic differentiation and number of haplotypes (H), haplotype diversity (Hd), the number of polymorphic (segregation) sites (S), the total number of mutations η (Eta), the average number of nucleotide differences (k), average pairwise nucleotide diversity (π = Pi), the total number of synonymous sites (SS), the total number of non-synonymous sites (NS), and the ratio of non-synonymous nucleotide diversity to synonymous nucleotide diversity dN/dS was investigated (Rozas *et al*., 2017). Generally, in nature, the gene is under three types of selection pressures including negative (purifying), neutral, and positive (diversifying) selection when dN/dS (ω) ratio is <1, =1, and >1, respectively (Rozas *et al*., 2017). DnaSP, DNA Sequence Polymorphism, is a software package for the analysis of DNA polymorphisms using data from a single locus (a multiple sequences aligned -MSA data), or from several loci (a Multiple-MSA data, such as formats generated by some assembler RAD-seq software). DnaSP can estimate several measures of DNA sequence variation within and between populations in noncoding, synonymous, or nonsynonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow, and gene conversion parameters. Three statistical tests, Tajima's D (Tajima 1989), Fu and Li's D* & F* (Fu and Li, 1993) implemented in the DnaSP v.6.10.01 program, to assess the neutral selection hypothesis operating on complete CP genes between populations were used. To calculate statistical tests of population differentiation including K_S*, K_{ST}*, Z*, Snn, and F_{ST} (Hudson 2000) between phylogroups and geographical populations of LMV isolates were performed using DnaSP v.6.10.01. The coefficient of F_{ST} (genetic differentiation) was used to estimate inter-population diversity and

the *P-value* of F_{ST} ranges can take from 0, no genetic differentiation, to 1, complete genetic differentiation as fully differentiated populations (Tsompana *et al*., 2005). Values of F_{ST} above 0.25 suggest a large genetic differentiation within the population (Gao *et al*., 2015). DnaSP version 6 (DNA sequence polymorphism) has been used for population analysis because it is an interactive computer program for the analysis of DNA polymorphism from nucleotide sequence data. The program calculates several measures of DNA sequence variation within and between populations. More explanations have been added to the manuscript for DnaSP.

Results

Phylogeny

An unrooted phylogenetic tree was procured based on the complete coat protein sequences by the Neighbor-Joining (NJ) method (Fig. 1).

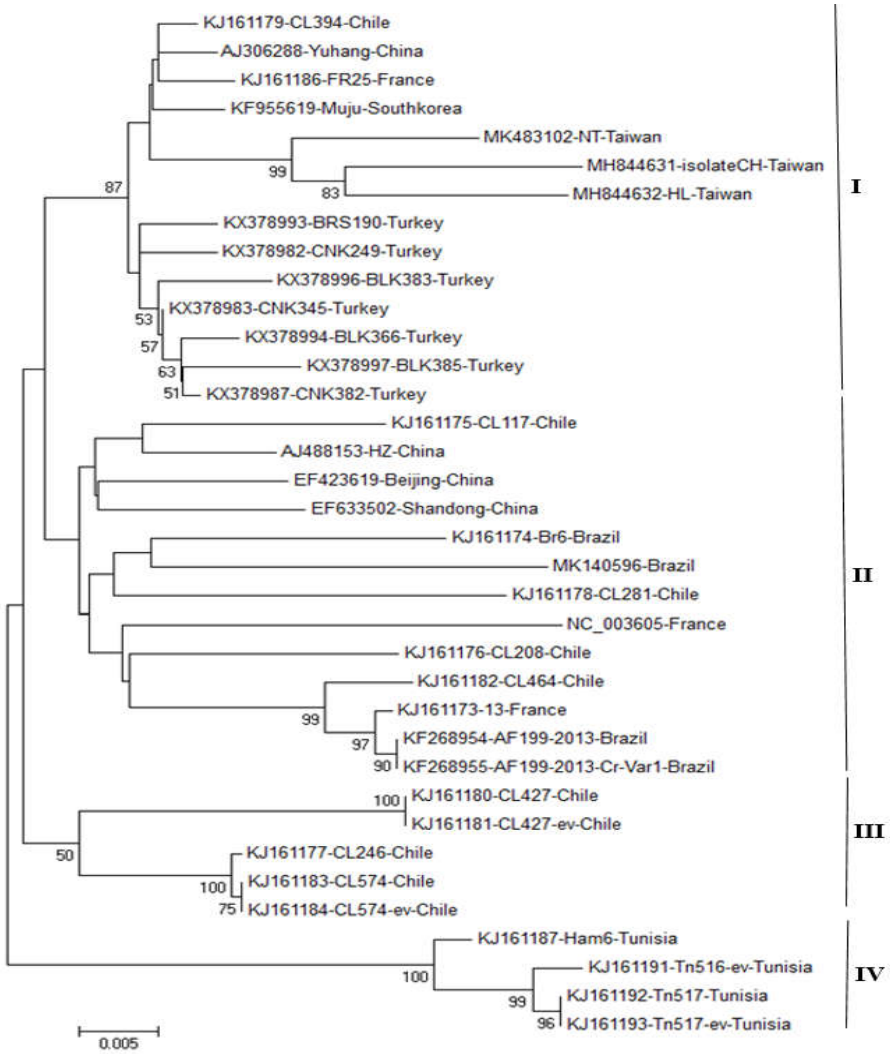

Fig. 1. Phylogenetic relationship of global *Lettuce mosaic virus* isolates constructed based on the complete CP gene sequences using Neighbor-Joining (NJ) method and Tajima model implemented in MEGA v.7.0.21. Branch support was computed using the bootstrap method based on 1000 replicates. Bootstrap values over 50 % are given at the nodes.

Fig. 1 showed that the 36 isolates were divided into four main phylogroups inconsistent with geographic origins and hosts (I, II, III, and IV) that each phylogroup was divided again into several branches (Fig. 1). Phylogroup I includes isolates from Chile, China, France, South Korea,

Taiwan, and Turkey. Phylogroup II consists of isolates from three continents such as America, East Asia, and Europe. Phylogroups III and IV just contain isolates from Chile and Tunisia, respectively (Fig. 1). The alignment of the LMV sequences showed the identity among all isolates was from 93% to 99% (Fig. 2).

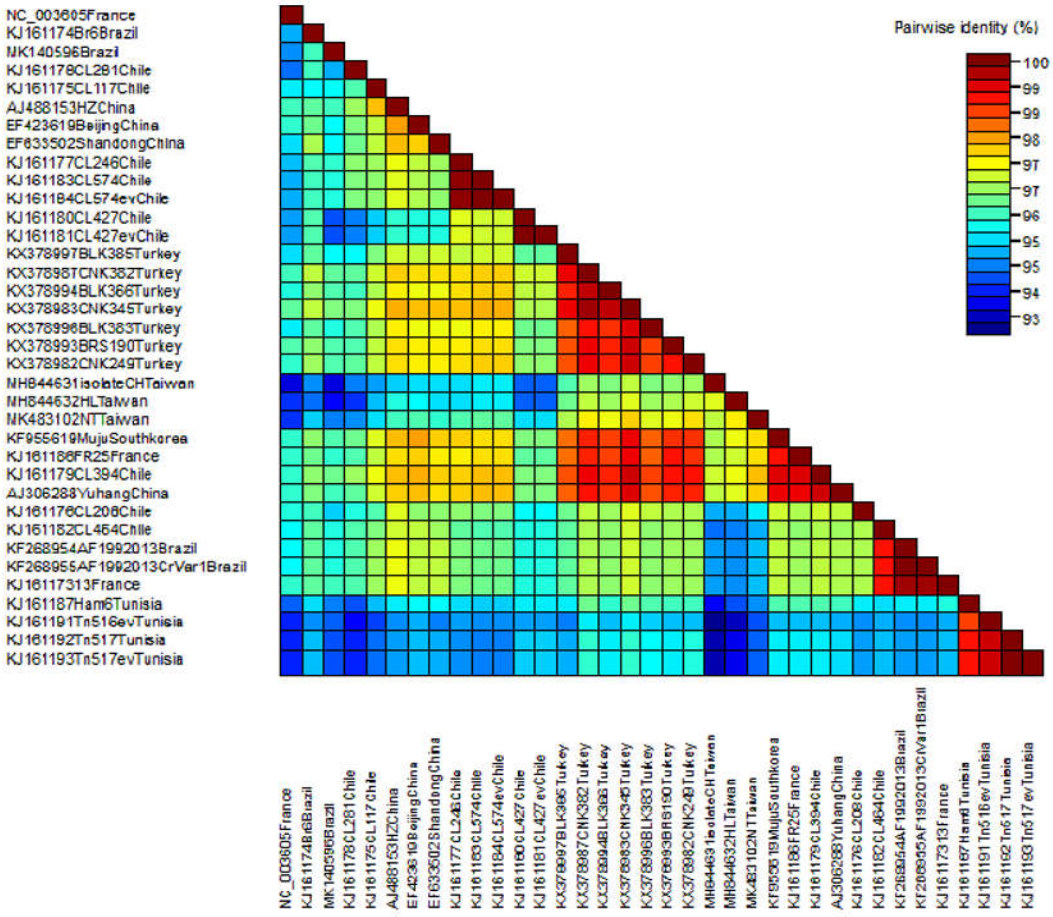

Fig. 2. Two-dimensional pairwise sequence identity color plot of the LMV isolates.

Population evolution and neutrality test parameters

For genetic variation and polymorphism analyses of the LMV populations based on the complete CP gene, all the alignment sequences were calculated by several genetic diversity parameters (Table 2). Population evolution results showed that the largest nucleotide diversity, π value (0.03917), and the greatest overall average number of differences between the sequences (isolates), k (32.667 nt), belonged to the French LMV population. The greatest number of segregation (parsimonious) sites, S (86), and mutation within the segregation sites, η (Eta) (88), were obtained for the Chilean LMV population. Also, the smallest values of π (0.00679), k (5.667nt), S (11), and η (Eta) (11) were estimated for the Tunisian population (Table 2). The highest values of π (0.03493) and k (29.128 nt) were calculated for clade II and the lowest values of π (0.00679) and k (5.667 nt) were obtained for clade IV (Table 2). The ratio of dN/dS (ω) was estimated at < 1 for all LMV populations. The maximum and the minimum ω values were calculated for Tunisian (0.6829) and French (0.0723) populations, respectively (Table 2). These findings showed that the complete CP gene of LMV populations was under negative selection. To obtain the molecular variation patterns from segregated sites of LMV populations based on the complete CP gene sequences, several test statics such as Tajima's D, Fu's, and Li's D* and F* were tested (Table 3).

Table 2. Genetic variation and polymorphism analysis of LMV CP from different populations.

Phylogroup	N	H	Hd	S	η	K	π	SS	NS	P_i(s)	P_i(a)	ω
All	36	32	0.994	187	204	29.990	0.03596	183.89	650.11	0.11940	0.01236	0.1035
Group I	14	14	1.000	78	80	15.209	0.01824	183.61	650.39	0.05567	0.00767	0.1377
Group II	13	12	0.987	**115**	**121**	**29.128**	**0.03493**	183.65	650.35	0.12132	0.01053	0.0867
Group III	5	3	0.800	25	25	15.000	0.01799	184.57	649.43	0.05205	0.00831	0.1596
Group IV	4	3	0.833	**11**	**11**	**5.667**	**0.00679**	184.79	649.21	0.00902	0.00616	**0.6829**
Geographic origin												
China	4	4	1.000	37	37	19.333	0.02318	183.08	650.92	0.07737	0.00794	0.1026
Taiwan	3	3	1.000	36	36	24.000	0.02878	183.67	650.33	0.07986	0.01435	0.1796
France	3	3	1.000	49	49	**32.667**	**0.03917**	183.61	650.39	0.14160	0.01025	**0.0723**
Turkey	7	7	1.000	26	26	7.905	0.00948	183.64	650.36	0.02749	0.00439	0.1596
Tunisia	4	3	0.833	**11**	**11**	**5.667**	**0.00679**	184.79	649.21	0.00902	0.00616	**0.6829**
Chile	10	8	0.956	**86**	**88**	27.578	0.03307	184.10	649.90	0.11450	0.01002	0.0875
Brazil	4	3	0.833	50	52	28.333	0.03397	183.72	650.28	0.09083	0.00717	0.0789

N, sequence number; H, number of haplotypes/isolates; Hd, haplotype diversity; S, number of polymorphic (Segregating) sites; η(Eta), the total number of mutations; k, the average number of nucleotide differences between sequences; π, nucleotide diversity; SS, the total number of synonymous sites analyzed; NS, the total number of non-synonymous sites analyzed; P_i(s), synonymous nucleotide diversity; P_i(a), non-synonymous nucleotide diversity. Maximum and minimum values between populations are in bold.

Table 3. Summary of neutrality test statistic values in LMV populations.

Comparisons	π [a]	Tajima's D	Fu and Li's D*	Fu and Li's F*
All	0.03596	-1.47118 ns	-1.82883 ns	-2.02587 ns
Group I	0.01824	-1.75650 ns	-2.23127 ns	-2.41410 ns
Group II	0.03493	-1.15505 ns	-1.26829 ns	-1.41774 ns
Group III	0.01799	1.86069*	1.86069**	2.00047**
Group IV	0.00679	-0.55827 ns	-0.55827 ns	-0.56160 ns
Geographic origin				
China	0.02318	-0.43601 ns	-0.43601 ns	-0.46025 ns
Taiwan	0.02878	nd	nd	nd
France	0.03917	nd	nd	nd
Turkey	0.00948	-1.44835 ns	-1.54243 ns	-1.67954 ns
Tunisia	0.00679	-0.55827 ns	-0.55827 ns	-0.56160 ns
Chile	0.03307	-0.56103 ns	-0.45097 ns	-0.54077 ns
Brazil	0.03397	-0.01112 ns	0.11124 ns	0.09457 ns

P > 0.10, P < 0.05, 0.10 > P > 0.05; ns: Non-significant; *: P < 0.05; **: P < 0.02; [a] π: Nucleotide diversity per site
nd, Four or more sequences are needed to compute Tajima's and Fu and Li's statistics.

Among geographic regions, findings showed that the non-significantly positive values were obtained only for the Brazilian population for Fu and Li's D* and Fu and Li's F* statistical tests (Table 3). The significantly positive values were estimated in all three statistical tests for clade III (Table 3).

Differentiation of LMV populations

To access the degree of virus gene flow (migration), the F_{ST} statistical test was evaluated using DnaSP v.6.10.01 (Hudson, 2000). Results of the genetic differentiation analysis of LMV populations showed that four phylogroups with significant K_S^*, Z*, and Snn independent tests were completely distinct. It is also confirmed by high F_{ST} (>0.27) (Table 4). Among the LMV populations, the Tunisian LMV population was distinct from the other populations (Snn values were significantly 1.000) and F_{ST} values were higher than (0.25) (Table 4). However, the non-significant Z* values were indicated no significant differentiation between the French population with the Chinese, Taiwanese, Brazilian, and Chilean populations, and also, between the Chilean and Chinese populations. The highest F_{ST} value (0.816) was estimated for Turkish versus Tunisian populations, and the lowest F_{ST} value (0.036) was calculated when comparing the French and Chilean populations (Table 4).

Discussion

Despite the occurrence of viral infections in lettuce crops around the world, there is not any report of population genetic analysis of LMV

isolates based on the coat protein (CP) gene. According to the previous reports, mutation and recombination have significant effects in shaping the population evolution of Potyviruses that are the largest genus in the family of *Potyviridae* (Gibbs and Ohshima, 2010; Abadkhah *et al*., 2020) and the most important step to managing viral infections is prevention and use of resistance genes (German-Retana *et al*., 2008). In this survey, we determined the genetic diversity and molecular evolution of 36 LMV isolates to identify haplotypes and relationships between LMV populations based on the complete coat protein genes.

Table 4. Summary of gene flow and genetic differentiation estimates between populations of LMV based on the CP.

Comparisons	[a]K_S*	[a]K_{ST}*	*Ks*, Kst*p-value*	[a]Z*	*P-value*	Snn	*P-value*	[b]F_{ST}
GroupI (n=14)/ GroupII (n=13)	2.95229	0.07547	0.0000 ***	4.48609	0.0000 ***	0.814	0.0000 ***	0.278
GroupI (n=14)/ GroupIII (n=5)	2.52407	0.12366	0.0000 ***	3.71585	0.0000 ***	1.000	0.0000 ***	0.450
GroupI (n=14)/ GroupIV (n=4)	2.49579	0.16494	0.0000 ***	3.49335	0.0000 ***	1.000	0.0000 ***	0.740
GroupII (n=13)/ GroupIII (n=5)	3.03969	0.08506	0.0000 ***	3.61620	0.0000 ***	1.000	0.0000 ***	0.348
GroupII (n=13)/ GroupIV (n=4)	3.04891	0.10678	0.0000 ***	3.34016	0.0000 ***	1.000	0.0000 ***	0.597
GroupIII (n=5)/ GroupIV (n=4)	1.91880	0.34799	0.0100 *	1.93318	0.0110 *	1.000	0.0110 *	**0.752**
Geographic origin								
China (n=4)/ Taiwan (n=3)	3.07913	0.07580	0.0400 *	1.53005	0.0400 *	0.857	0.0650 ns	0.353
China (n=4)/ France (n=3)	3.17736	0.01090	0.4340 ns	1.94065	0.1170 ns	0.428	0.5700 ns	0.009
China (n=4)/ Turkey (n=7)	2.37722	0.09641	0.0050 **	2.65329	0.0050 **	0.863	0.0250 *	0.276
China (n=4)/ Tunisia (n=4)	2.33943	0.25204	0.0240 *	1.67052	0.0240 *	1.000	0.0240 *	0.689
China (n=4)/ Chile (n=10)	3.12101	0.02403	0.1330 ns	3.43608	0.1180 ns	0.607	0.4330 ns	0.117
China (n=4)/ Brazil (n=4)	2.98601	0.07257	0.0320 *	2.01320	0.0320 *	0.791	0.0520 ns	0.181
Taiwan (n=3)/ France (n=3)	3.36409	0.04907	0.0620 ns	1.60454	0.1970 ns	0.666	0.1860 ns	0.281
Taiwan (n=3)/ Turkey (n=7)	2.30610	0.16015	0.0070 **	2.36367	0.0070 **	0.900	0.0070 **	0.394
Taiwan (n=3)/ Tunisia (n=4)	2.18460	0.31693	0.0150 *	1.42834	0.0240 *	1.000	0.0240 *	0.712
Taiwan (n=3)/ Chile (n=10)	3.15625	0.05888	0.0000 ***	3.04701	0.0020 **	0.897	0.0290 *	0.342
Taiwan (n=3)/ Brazil (n=4)	3.04671	0.12659	0.0330 *	1.42373	0.0330 *	1.000	0.0330 *	0.415
France (n=3)/ Turkey (n=7)	2.35522	0.12188	0.0060 **	2.50152	0.0050 **	0.700	0.4710 ns	0.176
France (n=3)/ Tunisia (n=4)	2.28284	0.26990	0.0310 *	1.45344	0.0200 *	1.000	0.0310 *	0.547
France (n=3)/ Chile (n=10)	3.18899	0.02269	0.1920 ns	3.31197	0.1890 ns	0.666	0.3960 ns	**0.036**
France (n=3)/ Brazil (n=4)	3.14495	0.00206	0.2690 ns	2.14790	0.3970 ns	0.547	0.3000 ns	N/A
Turkey (n=7)/ Tunisia (n=4)	1.99385	0.29861	0.0010 **	2.35881	0.0000 ***	1.000	0.0020 **	**0.816**
Turkey (n=7)/ Chile (n=10)	2.75457	0.08409	0.0010 **	3.63511	0.0000 ***	0.794	0.0060 **	0.272
Turkey (n=7)/ Brazil (n=4)	2.36333	0.17802	0.0040 **	2.57662	0.0030 **	0.818	0.0530 ns	0.390
Tunisia (n=4)/ Chile (n=10)	2.85266	0.14139	0.0000 ***	2.91474	0.0000 ***	1.000	0.0000 ***	0.607
Tunisia (n=4)/ Brazil (n=4)	2.31512	0.26525	0.0170 *	1.67827	0.0060 **	1.000	0.0170 *	0.602
Chile (n=10)/ Brazil (n=4)	3.11129	0.05051	0.0340 *	3.33746	0.0120 *	0.785	0.0800 ns	0.135

*, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001; [a] K*, Kst*, Z*, and Snn are test statistics of genetic differentiation; [b] F_{ST}, coefficient of gene differentiation, which measures inter-population diversity.

As shown, the obtained unrooted phylogenetic tree of complete CP gene sequences in this study consists of four main clades. The clade I was the largest clade including the LMV isolates from Chile, France, South Korea, China, Taiwan, and Turkey; and also Chilean isolates were placed in clades II and III (Fig. 1). The LMV isolates from France, Chile, and China were polyphyletic and speared in more than one phylogenetic group (Fig. 1). Such a phylogenetic relationship between the isolates may be related to virus transmission via propagated materials or (seed) and evolved via genetic drift (Sokhandan-Bashir and Melcher, 2012). As it has been approved in previous studies that the phylogenetic tree of LMV strains/isolates mainly clustered into several phylogroups (Lim *et al*., 2014). Thus, the occurrence of LMV around the world in all continents might be due to the global trade in plant products, insect vectors, seeds, and pollen. No recombination events were detected by RDP4 version beta 80 software (by using all of the algorithms) in this part of LMV CP. However, recombination is considered an important event to make new variants/strains of viruses especially RNA viruses (Rubio *et al*., 2013).

Population genetics studies should be conducted to find the systematic and random forces of evolution factors for virus isolates. In genetic diversity analysis among geographic regions, the lowest nucleotide diversity (π) was observed for the Tunisian LMV population (0.00679). Probably the LMVs from Tunisia were isolated from the same host and region; thus, they may have been under similar host adaptive selection (Garcı´a-Arenal *et al*., 2003; Rubio *et al*., 2013). The haplotype diversity (Hd) size can range from 0, meaning no diversity, to 1.000, which indicates high levels of haplotype diversity (Nei and Tajima, 1981). Our data showed that Hd was 0.994 indicating high levels of diversity for each

population. Evolutionary selection pressure on the studied complete CP gene was examined through the ratio of substitution rates at non-synonymous (dN) and synonymous sites (dS). The estimated ω ratios were less than one (ω<1) for all LMV populations showing the negative selection has occurred for all the LMV populations. Among LMV populations, the French population was under strong negative selection (ω= 0.0723) (Table 2). Negative pressure is one of the main reasons for the evolution and survival of the LMV and other plant RNA viruses such as *Tomato spotted wilt virus* (TSWV) (Abadkhah *et al*., 2018).

Non-significantly negative values were calculated by neutrality statistical tests (Tajima's D, Fu and Li's D* and F*) for all clades and geographic populations except for the Brazilian population in Fu and Li's D* and Fu and Li's F* which were obtained non-significantly positive values and clade III in all three test statics which were obtained significantly positive values. Generally, negative values showed that the estimated polymorphism was less than what was expected. All geographic populations in four phylogroups of LMV seem to be at equilibrium because all neutrality test statistics were non-significant (Table 3).

According to the result of genetic differentiation and gene flow, the Tunisian population was differentiated from the other LMV populations because the Snn values were significantly 1.000 (Table 4). Results from population genetic and phylogenetic analyses are perfectly aligned together. In geographical regions, the maximum F_{ST} value (F_{ST} = 0.816) was found between Turkish and Tunisian populations, and the most important reason for this event could be due to the long distances between these countries and, as shown in Fig. 1, Turkey population placed in cluster I and Tunisia population grouped in cluster IV (Fig. 1).

In conclusion, this study for the first time provides with evolution and differentiation between populations of LMV isolates from around the world and suggested that they may be occurred by the transmission of the virus among lettuce plants by types of aphid species, migration in different geographical areas by infected seeds and plant material, and broad host range which seems these events have been playing an important role in shaping the LMV population structure. Our research represents one of a few attempts to understand patterns and causes of spatial population genetic structure in LMV which is a destructive pathogen in lettuce plants. Negative selection pressure and mutation have powerful potential drivers to the dynamics of the genetic variability of global LMV isolates.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Abadkhah M, Hajizadeh M, Koolivand D. 2020. Global population genetic structure of *Bean common mosaic virus*. *Arch Phytopathol Pflanzenschutz* 53(5-6): 266-281.

Abadkhah M, Koolivand D, Eini O. 2018. A new distinct clade for Iranian *Tomato spotted wilt virus* isolates based on the polymerase, nucleocapsid, and non-structural genes. *Plant Pathol J* 34(6): 514-531.

Bos L, Huijberts N, Cuperus C. 1994. Further observations on variation of *lettuce mosaic virus* to lettuce (*Lactuca sativa*), and a discussion of resistance terminology. *Eur J Plant Pathol* 100: 293-314.

Fakhfakh H, Le Gall O, Candresse T, Marrakchi M. 2001. Analysis of the variability of Tunisian isolates of *lettuce mosaic virus* using biological and molecular properties. *J Plant Pathol* 83(1): 3-11.

Gao F, Lin W, Shen J, Liao F. 2016. Genetic diversity and molecular evolution of *Arabis mosaic virus* based on the CP gene sequence. *Arch Virol* 161:1047-51.

Garcı´a-Arenal F, Fraile A, Malpica JM. 2003. Variation and evolution of plant virus populations. *Int J Microbiol* 6: 225-232.

German-Retana S, Walter J, Gall OL. 2008. *Lettuce mosaic virus*: from pathogen diversity to host interactors. *Mol Plant Pathol* 9(2): 127-136.

Gibbs AJ, Hajizadeh M, Ohshima K, Jones RAC. 2020. The Potyviruses: an evolutionary synthesis is emerging. *Viruses* 12(2): 132.

Gibbs A, Ohshima K. 2010. Potyviruses and the digital revolution. *Annual Rev Phytopathol* 48: 205-223.

Hajizadeh M, Gibbs AJ, Amirnia F, Glasa M. 2019. The global phylogeny of Plum pox virus is emerging. *J Gen Virol* 100:1457-1468.

Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155: 2011-2014.

Jagger IC. 1921. A transmissible mosaic disease of lettuce. *J Agric Res* 20(10): 737-739.

Krause-Sakate R, Redondo E, Richard-Forget F, Salomao Jad ao A, Houvenaghel MC, German-Retana S, Gall OL. 2005. Molecular mapping of the viral determinants of systemic wilting induced by a *Lettuce mosaic virus* (LMV) isolate in some lettuce cultivars. *Virus Res* 109: 175-180.

Krause-Sakate R, Le Gall O, Fakhfakh H, Peypelut M, Marrakchi M, Varveri C, Candresse T. 2002. Molecular and biological characterization of *Lettuce mosaic virus* (LMV) isolates reveals a distinct and widespread type of resistance breaking isolate: LMV-Most. *Virology* 92 (5): 563-572.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7): 1870-1874.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.

Lim S, Zhao F, Yoo RH, Igori D, Lee SH, Lim HS, Moon JS. 2014. Characteristics of a *Lettuce mosaic virus* isolate infecting lettuce in Korea. *Plant Pathol J* 30(2): 183-187.

Moreno A, Bertolini E, Olmos A, Cambra M, Fereres A. 2007a. Estimation of vector propensity for *Lettuce mosaic virus* based on viral detection in single aphids. *Span J Agric Res* 5(3): 376-384.

Moreno A, Nebreda M, Diaz B. M, Garcý ìa M, Salas F, Fereres A. 2007b. Temporal and spatial spread of *Lettuce mosaic virus* in lettuce crops in central Spain: factors involved in *Lettuce mosaic virus* epidemics. *Ann Appl Biol* 150: 351-360.

Muhire BM, Varsani A, Martin DP. 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *Plos One* 9:e108277.

Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.

Nicaise V, German-Retana S, Sanjuan R, Dubrana MP, Mazier M, Maisonneuve B, Candresse T, Caranta C, Le Gall O. 2003. The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus Lettuce mosaic virus. *Plant Physiol* 132: 1272-1282.

Revers F, Lot H, Souche S, Le Gall O, Candresse T, Dunez J. 1997. Biological and molecular variability of *lettuce mosaic virus* isolates. *Phytopathol* 87: 397-403.

Riechman JL, Lain S, Garcia JA. 1992. Highlights and prospects of potyvirus molecular biology. *J Gen Virol* 73: 1-16.

Rubio L, Guerri J, Moreno P. 2013. Genetic variability and evolutionary dynamics of viruses of the family *Closteroviridae*. *Front Microbiol* 4: 1-15.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol Biol Evol* 34(12): 3299-3302.

Sharma P, Sharma S, Singh J, Saha S, Baranwal VK. 2016. Incidence of *Lettuce mosaic virus* in lettuce and its detection by polyclonal antibodies produced against recombinant coat protein expressed in *Escherichia coli*. *J Virol Methods* 230: 53-58.

Sokhandan-Bashir N, Melcher U. 2012. Population genetic analysis of *Grapevine fanleaf virus*. *Arch Virol* 157:1919-1929.

Soleimani P, Mosahebi G, Koohi Habibi M. 2011. Identification of some viruses causing mosaic on lettuce and characterization of *Lettuce mosaic virus* from Tehran province in Iran. *Afr J Agric Res* 6(13): 3029-3035.

Svanella-Dumas L, Verdin E, Faure C, German-Retana S, Gognalons P, Danet JL, Marais A, Candresse T. 2014. Adaptation of *Lettuce mosaic virus* to *Catharanthus roseus* involves mutations in the central domain of the VPg. *Mol Plant Microbe Interact* 27(5): 491-497.

Tsompana M, Abad J, Purugganan M, Moyer JW. 2005. The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. *Mol Ecol* 14: 53-66.