Iranian Journal of Virology 2022;16 (2):74-85.
©2022, Iranian Society of Virology

Original Article

The Major Sources of Genetic Variation Among Grapevine Leafroll-Associated Virus 3 Isolates And Their Relation to Evolution

Pourrahim R*, Farzadfar Sh

Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran.

Abstract

Background and Aims: The genetic variability and population structure of Grapevine leafroll-associated virus 3 (GLRaV-3) have not already been studied in mid-Eurasia Iran. The investigations will contribute to developing efficient and durable control strategies of latent and graft-transmissible virus.

Materials and Methods: During the winter of 2018-2019, a total of 229 dormant cutting samples were collected from 15 vineyards in West Azarbaijan province. All the samples were analyzed by serological and RT-PCR assays. The phylogenetic tree and two-dimensional nucleotide diversity plot were constructed using CP sequences by the SplitsTree4 v.4.12.6, and SDTv1.2 soft wares, respectively. Dnasp v.6.10.04 was used for genetic diversity and demographic analysis. We provide analyses of the codon usage and composition of GLRaV-3 based on 517 nucleotide sequences of the CP gene including 11 full CP sequences from Iran.

Results: Using serological assay 31.87% GLRaV-3 infection was detected. Neighbor-Net method analysis of the virus complete CP gene showed that the Iranian isolates fell into phylogroup I (GI) "as it is dominant in the rest of the world". High haplotype diversities indicate the recent expansion of GLRaV-3. No clustering was found according to where the GLRaV-3 was isolated. Using dN/dS values it was found that the different populations of this virus is under negative (purifying) selection. The highest gene flow was determined between Europe and East Asia. Moderate or low genetic differentiation, and frequent gene flow (FST < 0.33 and Nm > 1) also confirmed with Ks*, Kst* Z, Z* and Snn statistics values. The frequency of amino acid coded by A/G ended optimal codon indicates the overlapping influences of natural selection and mutational pressure on the codon preferences in the CP gene. Codon bias of the CP gene was strongly affected by natural selection rather than mutation according to the effective number of codons-ENC vs. GC3s plot. Principal component plot analysis (PCA), illustrated the possible origin of GLRaV-3 in the Old World.

Conclusions: This analysis is the first demonstration of the population structuring of GLRaV-3 in mid-Eurasian Iran. Indeed, these consequences explain selectively driven codon bias in GLRaV-3 species; and reveal the potential importance of expression-mediated selection in shaping the genome evolution of this virus.

Keywords: GLRaV-3, Iran, Nucleotide diversity, Codon usage bias

Introduction

Grapevine leafroll disease (GLD) includes multiple viruses with symptoms of downward leaf rolling and various types of interveinal reddening in the foliage of red-berried cultivars of grapevine (1). GLD has economic significances disorders related to the grape wine industry due to a potential reduction in grape quality and yield that results in low-quality of different products. Grapevine leafroll-associated virus 3 (GLRaV-3) (genus Ampelovirus; family Closteroviridae) is one of the main virus pathogens of grapevine worldwide and the important etiological factor of GLD (1).

GLRaV-3 can be transmitted by grafting, mealybugs (Homoptera: Pseudococcidae), and/or soft-scale insects (Homoptera: Coccidae)

***Corresponding author:**
Reza Pourrahim, Email: pourrahim@yahoo.com
Tel: +98 21 22403012-16; Fax: +98 21 22403692

(2). However, the long-distance dispersal of GLRaV-3 is largely mediated by the transmission of infected bud wood or through vegetative propagation by cuttings.

It is presumed that viral coat proteins (CP) evolve more rapidly than proteins involved in the replication and expression of virus genomes (3), thus providing a strong incentive to study the diversity of viruses based on the CP gene. The genetic variability of GLRaV-3 has been studied widely and various types of research showed that GLRaV-3 is a highly diverse virus species (4-6). The CP gene analysis showed that the GLRaV-3 isolates were clustered into four supergroups (A, B, C, and D) with nine monophyletic subgroups (I-VII, IX, and X). Five subgroups I, II, III, IV, and, V fell into supergroup A, meanwhile supergroup B consisted of two subgroups VI, and X. Supergroup C consisted of subpopulation VII, and GLRaV-3 isolates belonging to subgroup IX fell into supergroup D (6).

Iran is the twelfth largest producer of grapes in the world, accounting for about 3.3% of total grape production (7). Almost 155000 hectares were used for the cultivation of grapevines, with an annual production of around 2 million tons of grapes in 2019. One of Iran's main grapevine production areas is located in West-Azarbaijan province, North West Iran.

GLRaV-3 has previously been reported in Iran (8). However, the genetic variability and population structure of GLRaV-3 in Iran using complete coat protein gene sequences have not already been studied. Here we provide the occurrence of GLRaV-3 in West Azarbaijan Province of Iran. In addition, a genotyping profile of 11 Iranian GLRaV-3 based on the comparison of the nucleotide sequence of the CP gene with representative isolates available in GenBank using different bioinformatics approaches was investigated. Identification of codon usage patterns provides important information about the host-pathogen co-evolution, phylogenetic relations among specified species, and molecular evolution of genes (9-12).

Therefore, we provide analyses of the codon usage and composition of GLRaV- 3 based on 157 nucleotide sequences of the CP gene.

Material and Methods

Serological Assay

During the winter of 2018-2019, totally 229 dormant cutting samples were collected from 15 vineyards in West Azarbaijan.

Dormant cutting samples were planted in pots in the greenhouse until germination. All the samples were analyzed by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) (13) using specific antibodies to GLRaV-3 (Bioreba AG, Reinach BL 1, Switzerland). Optical densities (O.D.) were measured at 405 nm, using a BioTek microplate reader 2 hr after the addition of the p-nitro-phenyl phosphate substrate. Samples with OD405 values equal to or more than three the mean of the negative controls were considered positive. ELISA-positive plants were considered as those with OD405 values equal to or greater than twice the mean of the negative controls.

RT-PCR and Nucleotide Distance Analysis

Total RNA was extracted from GLRaV-3-infected petiole and midrib of leaf samples using the CTAB method. First-strand cDNA synthesis was performed using M-MuLV reverse transcriptase (Parstous, Iran), according to the manufacturer's instructions.

Subsequent PCR reactions were performed the amplification of GLRaV-3 CP gene using previously reported specific primers (4) and high fidelity Platinum™ Pfx DNA polymerase (Invitrogen, Carlsbad, CA, USA).

PCR products were evaluated by electrophoresis on a 1% agarose gel in TBE buffer including ethidium bromide (final concentration 1 µg/ml). The PCR products were isolated from agarose gel, and purified using Wizard PCR DNA purification Kit (Promega, USA) according to the manufacturer's recommendations. PCR products were sequenced in both directions and in three replications using commercial service (Gene Fanavaran Co., Iran).

Sequenced data were assembled using BIOEDIT version 5.0.9 (14). The GLRaV-3 CP sequences were aligned using CLUSTALX2 (15). Neighbor-Net method in SplitsTree4 v.4.13 was used for the phylogenetic tree (16). Nucleotide diversity was estimated using

Kimura two parameters implemented in MEGAX (15). The pairwise nucleotide diversity are shown using color plot.

Genetic Differentiation and Genetic Flow

Using the CP gene, pairwise fixation (Fst) and genetic flow indices (Nm) were estimated with DnaSP version 6.0 (17). Pairwise average genetic distances between different populations were calculated by Kimura-2-parameter model with MEGAX (15). The genetic differentiation considered as low for Fst < 0.05, moderate for 0.05 < Fst < 0.15, high for 0.15 < Fst < 0.25, and very high for Fst > 0.25 (18).

With reference to the criterion for gene flow values, we defined genetic flow as low for Nm < 1, high for 1 < Nm < 4, and very high for Nm > 4 (18,19).

Gene Flow and Genetic Differentiation among Populations

The Ks *, Kst*, Z*, Snn and Fst values (19) which determine the genetic differentiation on the CP nt sequences among phylogroups were calculated using DnaSP (17). Kst* will be near zero if there is no genetic differentiation (null hypothesis) (20). A smaller Z* means a smaller genetic differentiation among the population (19). The value of Snn describes a range of the exact same population (value of 0.5) (null hypothesis) to distinctly differentiate the population (value of 1) (19).

The null hypothesis in Ks *, Kst*, Z*, and Snn is rejected by a significant P value (19. 20). Fst ranges between the exact same population (value of 0) to fully distinct populations (value of 1) (20). Fst > 0.25, in most cases, indicates a large gene flow and a big genetic differentiation in the tested populations (21).

Demographic History and Neutrality Test

With the DnaSP version 6.0 software (17), the Tajima's D and Fu's Fs values were estimated to test demographic expansion (22), and we assessed their significance with 1,000 permutations at levels of each geographical populations and overall population. The demographic history of whether GLRaV-3 experienced a range expansion was also investigated by mismatch distributions of pairwise differences among CP sequences calculated with parametric bootstrapping (1,000 replications). The expected values were calculated assuming that the sudden population growth model through coalescent simulations.

According to simulations, demographically stable or admixed populations must present a multimodal distribution, whereas populations that have undergone a recent expansion generally exhibit a unimodal distribution (23).

Codon Usage Bias and Relative Synonymous Codon Usage of the CP Gene

The CodonW 1.4.2 package (http://codonw.sourceforge.net) was used for assessing the nucleotide mixtures at the 3rd codon position (A3, C3, T3, and G3%). In this analysis 517 full CP genes (n= 506 CP sequences from GenBank and n=11 Iranian CP sequences) was used. After deleting five non-bias codons including AUG (start codon), UGG (encoding Trp), and three termination codons UAA, UGA, and UAG, the component parameters of the CP sequences were calculated. The total percent of nucleotide composition and the overall GC and AU contents were estimated by MEGAX (15).

The relative synonymous codon usage (RSCU) values for the CP coding sequences were determined to find out the aspects of synonymous codon usage without the complicated effect of amino acid composition and coding sequence size of CP gene samples according to the previously described method (11). The RSCU values >1.6 and <0.6 were indicated as “overrepresented” and “underrepresented”, respectively.

Effective Number of Codons (ENC) and ENC-plot Analysis

The ENC value was determined using CodonW v1.4.2. ENC analysis to quantify the absolute codon usage bias by evaluating the degree of codon usage bias exhibited by the GLRaV-3 CP gene, regardless of gene length and the number of amino acids. ENC value ranged from 20 (a high codon usage bias using only one of the possible synonymous codons for the matching amino acid) to 61 (no bias using all available synonymous codons equally for the matching amino acid). Generally, the lower ENC value reflects a highly expressed gene with a great codon preference, whereas, those genes with low expression have more rare codons and show higher ENC values (24).

Using the ENC values versus GC3s values plots (ENC-plot analysis), the effect of mutational pressure on codon usage bias is analyzed. When the points are located on the standard curve it shows that mutation pressure is the lonely factor for driving the codon usage bias. Otherwise, the effect of other factors like natural selection must be considered.

Principal Component Analysis (PCA)

The significant tendency in codon usage variation of the GLRaV-3 CP gene was examined by PCA analysis (24). In this respect, a plot of the 1st axis and the 2nd axis of the isolated strains according to each phylogroup and geographical isolation were drawn.

Results

Serological Assay

According to the ELISA cut-off value, 31.87% of samples were infected by GLRaV-3. Vein banding, leaf deformation, and yellowing were observed with positive ELISA samples (Fig. 1).

RT-PCR and Nucleotide Distance Analysis

The full-length CP sequences of 11 GLRaV-3 Iranian isolates were deposited in GenBank under Accession Numbers OK274299-OK274 309. The complete CP gene of Iranian isolates was 942 nt long with an open reading frame (ORF) of 313 amino acids.

Phylogenetic analyses using Neighbor-Net method clustered the GLRaV-3 isolates into 10 phylogroups in which Iranian isolates with Hungarian (Kis252), American (Cha138b, Cha103, 9-1010, 9-25), Brazilian (TRAJ-BR), and Chinese (Bj-1) isolates fell into the phylogroup I (GI) (Fig. 2).

The diversity plot also confirmed 10 phylogroups. The lowest nucleotide diversities (0.0 to 11.2%) was detected among phylogroups I, II, III, IV and V (Fig. 3), whereas the nucleotide diversity among phylogroups VI, IX, and X were 11.2% to 28.0%. Two phylogroups VII and VIII showed the highest nucleotide diversity (28.0 to 32.5%) (Fig. 3).

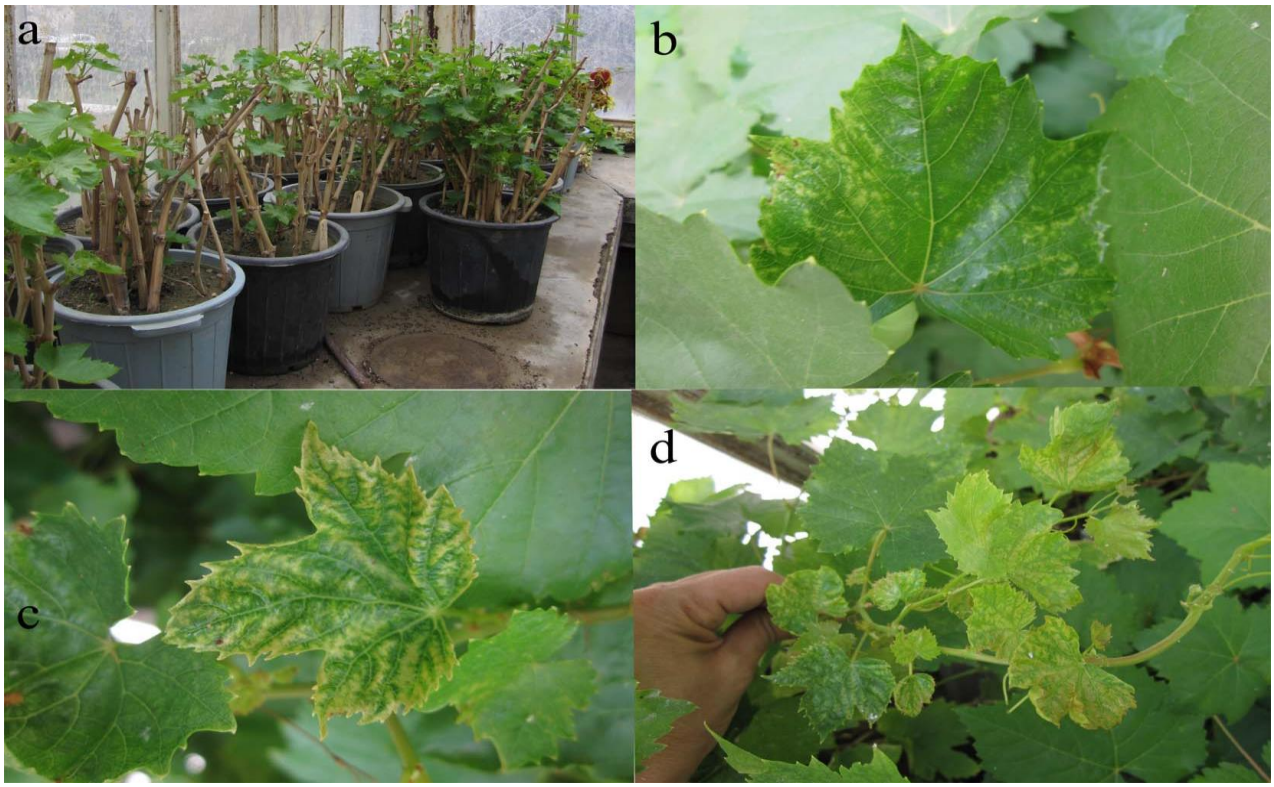


Fig. 1. (a) Rooted and germinated grapevine cuttings in the greenhouse. (b, c, and d) chlorosis and leaf deformation symptoms of samples infected with GLRaV-3

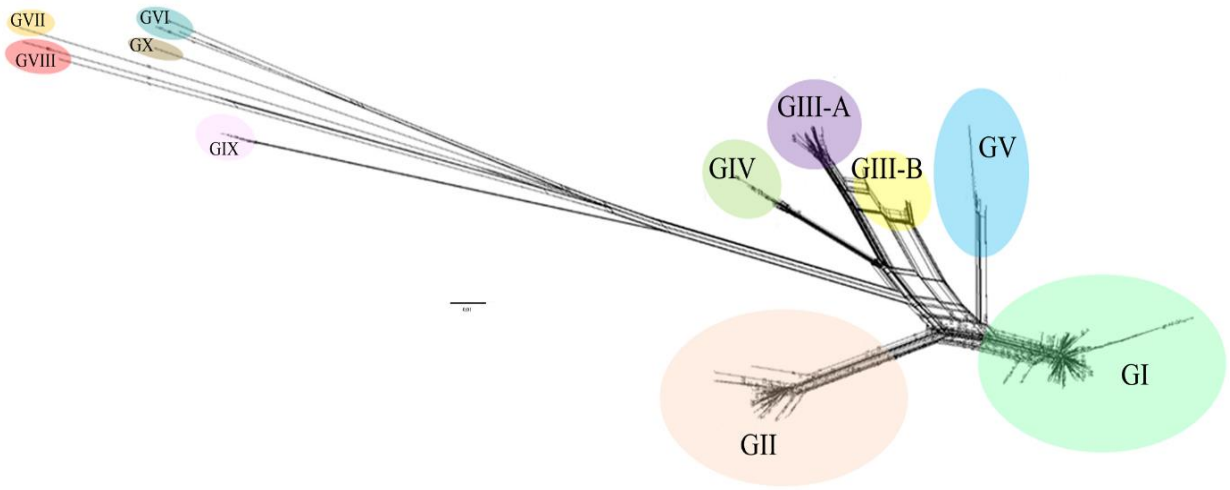
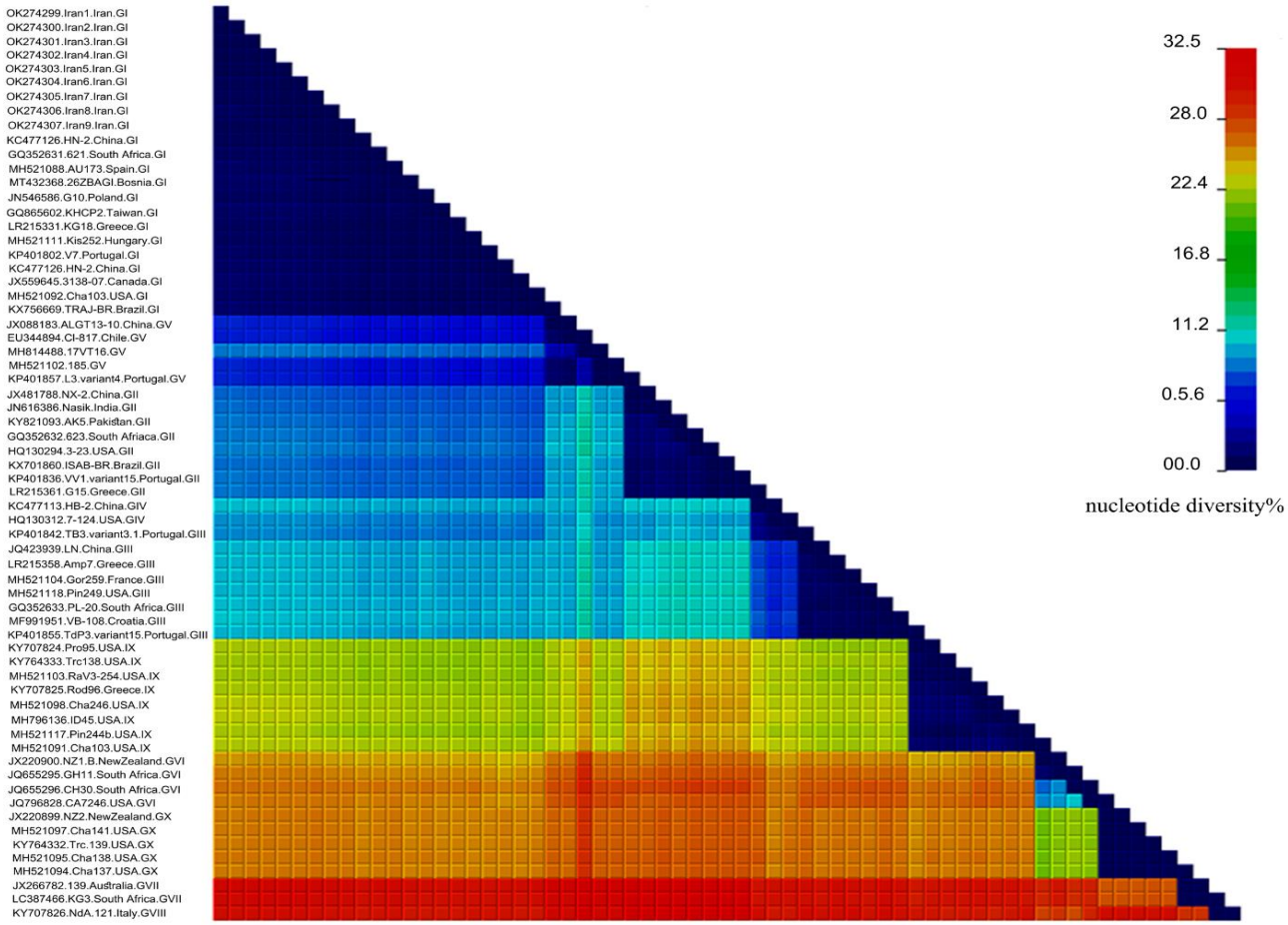


Fig. 2. Neighbor-Net method tree clustered the GLRaV3 isolates into 10 phylogroups. Iranian isolates fell into phylogroup I (GI).

Fig. 3. two-dimensional nucleotide diversity plots showing the relationship among 56 representatives of each phylogroup of *Grapevine leafroll-associated virus -3* isolates and nine Iranian isolates sequenced in this study. The accession number, name of each isolate, and country of its origin are shown.



Genetic Diversity and Polymorphism Analysis

In the *CP* gene analysis, 404 distinct haplotypes were detected among the 495 tested isolates. Haplotype diversity (Hd) for all 495 sequences was calculated to be 0.997. The average number of nucleotide differences (k) was shown to be 55.675, and nucleotide diversity (Pi) was 0.05929. (Table 1). The highest haplotype diversity (Hd), and nucleotide diversity (Pi) were detected in Oceania population. However, Nucleotide diversity (Pi) and haplotype diversity (Hd) are lower in West Asia (Table 1).

Nucleotide differences within haplotypes (k) are highest in Oceania followed by Africa, America, Europe, East Asian, and West Asian populations. Using dN/dS values, it was found that the different populations of GLRaV-3 are under negative (purifying) selection with the ω ratio less than 1 (Table 1).

Gene Low and Genetic Differentiation among Populations

Kst* will be near zero if there is no genetic differentiation (null hypothesis) (30). A smaller Z* means a smaller genetic differentiation among the population (31).

The value of Snn describes a range of the exact same population (value of 0.5) (null hypothesis) to distinctly differentiate the population (value of 1).

The Ks*, Kst*, Z*, and Snn values obtained from each CP gene comparison among their respective populations were all different, which means these populations are distinct from each other. Fst value, which is a measure of population differentiation, was calculated for all the geographical populations. Fst in all the populations varies from 0.04578 (between Europe and East Asia populations) to 0.43746 (between East Asia and Oceania populations) (Table 2).

Pairwise FST values didn't show a strong genetic differentiation between GLRaV-3 populations (Table 2). In addition gene flow (Nm) in all the populations varied from 5.21 (between Europe and East Asia populations) to 0.32 (between East Asia and Oceania populations) (Table 2). The highest gene flow was determined between Europe and East Asia (Table 2). Moderate or low genetic differentiation, and frequent gene flow (FST < 0.33 and Nm > 1) also confirmed with Ks*, Kst* Z, Z* and Snn statistics values (Table 2).

Table 1. Demography test statistics between the *CP* gene of GLRAV-3 populations

population	N	h	Hd	K	Pi	dS	dN	ω	Neutrality test and significance test	
									Tajima's D	Fu's Lis F
All	495	404	0.997	55.675	0.05929	0.237±0.017	0.016±0.002	0.0675	-1.78963*	0.86515*
AF	7	7	1.00000	160.76190	0.17121	0.957±0.133	0.041±0.007	0.0428	-0.09025ns	0.86515ns
AM	93	54	0.97639	81.38663	0.08667	0.397±0.037	0.021±0.003	0.0528	-0.46540ns	-0.36889ns
EA	209	180	0.99706	37.33244	0.03976	0.136±0.013	0.011±0.002	0.0801	-1.72157*	-5.36524**
EU	169	150	0.99810	50.49683	0.05378	0.184±0.016	0.015±0.003	0.0815	-1.72157ns	-4.43291**
OC	3	3	1.00000	199.66667	0.21264	1.677±0.354	0.041±0.011	0.0244	ND	ND
WA	14	10	0.95604	20.14286	0.02145	0.060±0.010	0.010±0.000	0.166	-0.64720ns	0.86515ns

N – number of isolates; h – number of haplotypes; Hd – haplotype diversity; k – average number of nucleotide differences between sequences; Pi– nucleotide diversity (per site); dS – synonymous nucleotide diversity; dN – non-synonymous nucleotide diversity; ω – dN/dS; * – P < 0.05; ns – not significant; ND – not determined.
AF=Africa, AM=America, EA=East Asia, EU=Europe, OC= Oceania, WA=West Asia

Table 2. Genetic differentiation measurement between populations from pairwise comparison of GLRaV-3 sequences based on geographical isolation

	Ks*	Kst*	*Z**	*Snn*	*P* value	F_{ST}	*Nm*
EA and AF	3.10137	0.01389	9.03670	0.97917	0.0651 ns	0.17995	1.14
EA and AM	3.24984	0.02053	9.64542	0.87097	0.0096 **	0.06536	3.57
EA and EU	3.22670	0.01692	10.10368	0.88054	0.0752 ns	0.04578	5.21
EA and OC	3.07096	0.01717	8.99685	0.99057	0.0633 ns	0.43746	0.32
EA and WA	3.00138	0.01259	9.02713	0.99103	0.0458 *	0.08925	2.55

Probability (p value) obtained by the permutation test (PM test) with 1000 replicates. *, 0.01 < P < 0.05; **, 0.001 < P < 0.01; ***, P < 0.001; ns, not significant. The PM test was performed using DnaSP v. 6.10.04. FST > 0.33 indicates infrequent gene flow; FST < 0.33 suggests frequent gene flow. Nm is the migration fraction per generation.

codon usage is more affected by natural selection rather than the mutation pressure (Fig. 5).

Relative Synonymous Codon Usage and Codon Usage Bias of the CP Gene

High frequency of A and G nucleotides were detected in the GLRaV-3 CP sequences, with average compositions of 31.41 ± 0.25% and 26.20 ± 0.29% respectively, in comparison with T (U) (22.39 ± 0.41%) and C (19.97 ± 0.41%). In contrast, the nucleotide composition was different from the nucleotide compositions at the 3rd position of synonymous codons. The most frequent nucleotide was G3s (33.28% ± 0.83), followed by A3s (28.48% ± 0.54), T3s (19.43% ± 0.97) and C3s (18.79% ± 0.96). The compositions of AU and GC in the CP coding sequence were 53.80% ± 0.35 and 50.34% ± 0.41, respectively, informing that there is an AU-biased composition in GLRaV-3 *CP* gene.

RSCU value of each codon in the *CP* gene analysis showed that the GLRaV-3 CP coding sequences was AU-rich, whereas G-ending codons were favored in the *CP* gene (Table 3). Six out of 18 frequently used codons were G-ending, while 5, 4, and 3 were ended to the A, T(U), and C (Table 1, Fig. 4). The RSCU value > 1.6 was detected for four of the optimal synonymous codons (UUG, UCG, CCA, and AGA), with the highest preferred value for the AGA codon (2.882) (Fig. 4).

Effective Number of Codons (ENC) and ENC-plot Analysis

Low codon usage bias in all CP coding sequences of GLRaV-3 was found with ENC average value (53.79 ± 2.118) representing an approximately constant and conserved genomic composition. By GC3s values against ENC plot the data points clustered together under the normal ENC curve, which indicated that the

Principal Component Analysis (PCA)

Principal component analysis was used to analyze the major trends in codon usage patterns among the GLRaV-3 CP gene.

This plot demonstrated that the GLRaV-3 belonging to phylogroups I and V (Fig. 6) are clustered together and near to the origin of plots, however, those isolates from other phylogroups are dispersal (Fig. 6). Indeed, the PCA analysis based on the geographical Isolation of each population demonstrated that the Chines (Fig. 7b), and European (Fig. 7c) Isolates are closer to the origin of the plots. Among them, the GLRaV-3 isolates belonging to phylogenetic group V (including four Portugalian and two Chinese isolates) were the nearest to the origin.

Discussion

Studying the genetic diversity of viruses provides critical information for understanding virus evolution, geographical origin, virulence variations, and the occurrence of emerging new epidemics. Phylogenetic analyses clustered the GLRaV-3 isolates into 10 phylogroups (5, 6), where Iranian isolates fell into the major phylogroup I “as it is dominant in the rest of the world” (Fig. 2). The higher fitness or transmission efficiency by vectors may be related to the wide distribution and prevalence of GLRaV-3 subgroup I variants than the others (25).

As previously reported, no correlation was detected between geographical distributions with phylogenetic groups (6).

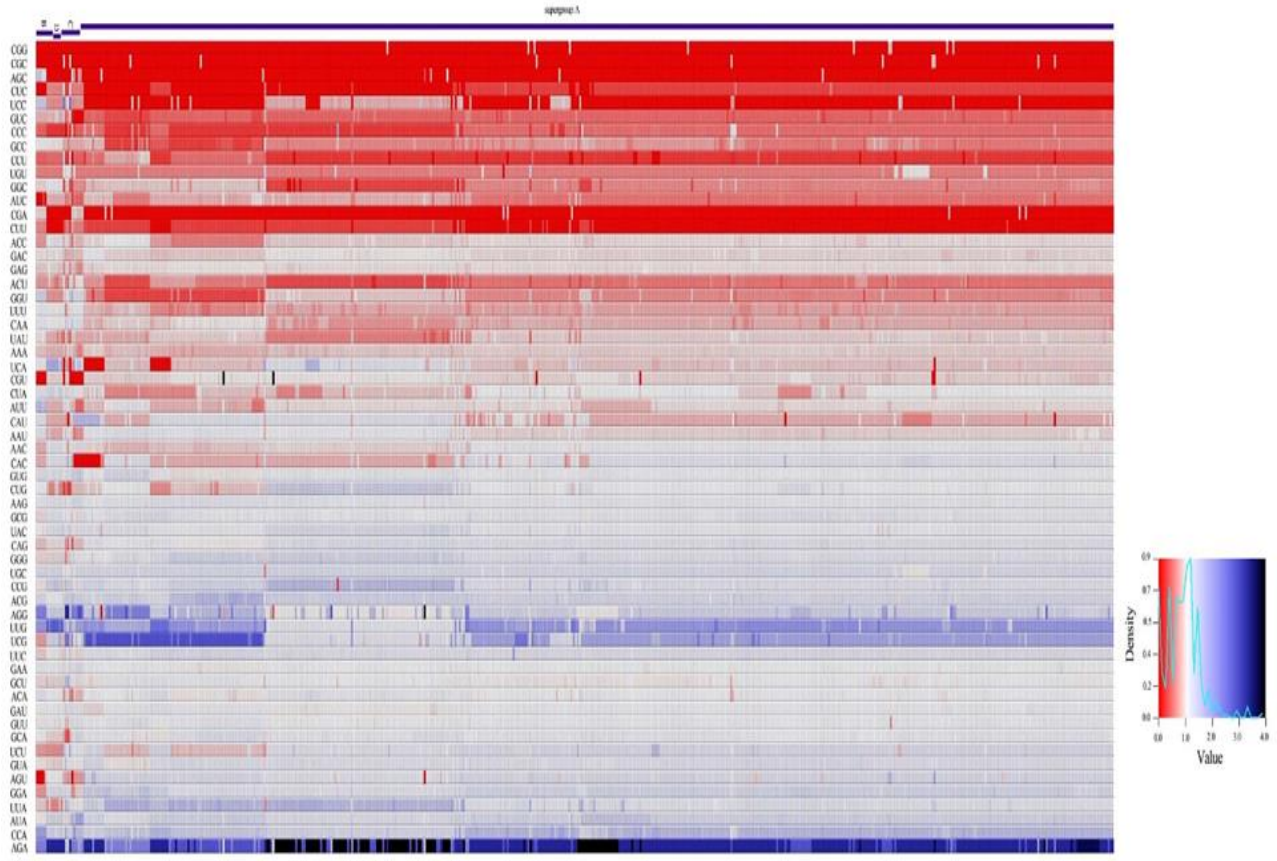


Fig. 4. RSCU value of each codon in the *CP* gene of each *Grapevine leafroll-associated virus -3* isolates is shown. Rows indicated the 59 nondegenerate, non-stop codons. Over-represented codons (RSCU > 1) are shown by blue cells and underrepresented codons (RSCU < 1) by red cells.

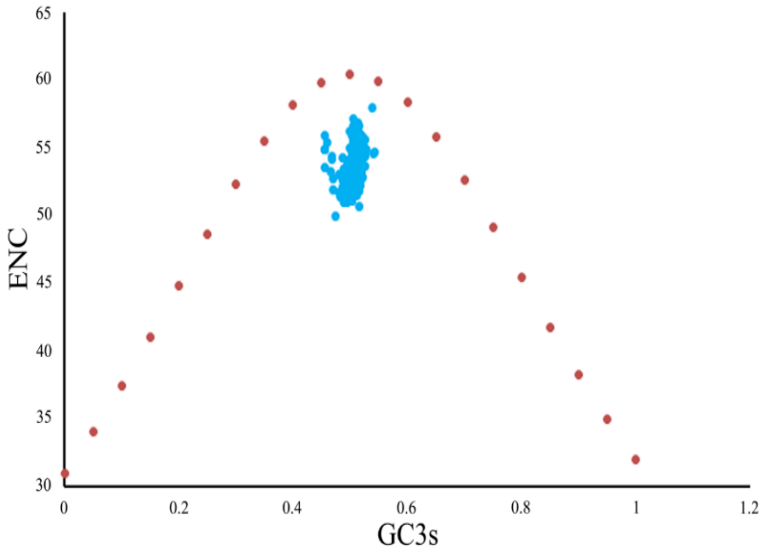


Fig. 5. ENC-plot analysis of the GLRaV-3 *CP* gene with ENC curve drawn against GC3s. The standard curve was plotted while using the codon usage bias (calculated by the GC3s composition only) indicated by blue points.

Table 3. The RSCU value of 59 codons encoding 18 amino acids according to hosts of grapevine leafroll-associated virus 3 coat protein gene.

Codon	aa	A	B	C	D	All
UUU	F	0.63	1.07	0.95	**1.09**	0.935
UUC	**F**	***1.37**	0.93	**1.05**	0.91	**1.065**
UUA	L	1.58	1.05	1.59	0.66	1.220
UUG	**L**	**2.12**	**2.04**	**2.02**	**2.63**	**2.202**
CUU	L	0.1	0.56	0.37	0.2	0.307
CUC	L	0.17	0.49	0.61	0.32	0.397
CUA	L	0.87	1.11	1.1	1.4	1.120
CUG	L	1.17	0.74	0.31	0.78	0.750
AUU	I	0.88	0.55	1.29	1.15	0.967
AUC	I	0.54	0.38	0.59	0.46	0.492
AUA	**I**	**1.57**	**2.07**	**1.12**	**1.39**	**1.537**
GUU	V	1.11	1.26	0.9	1.04	1.077
GUC	V	0.4	0.04	0.66	0.72	0.455
GUA	V	1.36	1.09	1.24	0.89	1.145
GUG	**V**	**1.13**	**1.61**	**1.21**	**1.35**	**1.325**
UCU	S	1.36	1.17	1.58	0.57	1.170
UCC	S	0.17	0.59	1.74	1	0.875
UCA	S	0.85	1.46	0.32	1.78	1.102
UCG	**S**	**2.16**	**1.61**	**1.74**	**1.3**	**1.702**
AGU	S	1.45	0.44	0.63	0.74	0.815
AGC	S	0.01	0.73	0	0.61	0.337
CCU	P	0.26	0.29	0.85	0.4	0.450
CCC	P	0.38	0.44	0.21	0.28	0.327
CCA	**P**	**1.86**	**1.95**	**1.65**	**1.9**	**1.840**
CCG	P	1.5	1.32	1.28	1.42	1.380
ACU	T	0.43	0.98	0.88	0.59	0.720
ACC	T	0.86	0.68	0.81	0.99	0.835
ACA	T	1.11	0.68	0.71	1.18	0.920
ACG	**T**	**1.6**	**1.66**	**1.59**	**1.24**	**1.522**
GCU	**A**	**1.06**	0.98	**1.78**	0.85	**1.167**
GCC	A	0.5	0.68	0.63	1.04	0.712
GCA	A	1.2	1.38	0.49	0.95	1.005
GCG	A	1.24	0.95	1.09	1.16	1.110
UAU	Y	0.77	0.75	0.62	0.81	0.737
UAC	**Y**	**1.23**	**1.25**	**1.38**	**1.19**	**1.262**
CAU	**H**	0.92	**1.83**	0.5	0.92	**1.042**
CAC	H	1.08	0.17	1.5	1.08	0.957
CAA	**Q**	0.69	**1.36**	**1.29**	**1.16**	**1.125**
CAG	Q	1.31	0.64	0.71	0.84	0.875
AAU	N	0.97	0.61	1.07	0.93	0.895
AAC	**N**	**1.03**	**1.39**	**0.93**	**1.07**	**1.105**
AAA	K	0.8	0.68	0.78	0.72	0.745
AAG	**K**	**1.2**	**1.32**	**1.22**	**1.28**	**1.255**
GAU	**D**	**1.1**	**1.26**	**1.29**	**1.05**	**1.175**
GAC	D	0.9	0.74	0.71	0.95	0.825
GAA	**E**	**1.05**	**1.31**	**1.31**	**1.16**	**1.207**
GAG	E	0.95	0.69	0.69	0.84	0.792
UGU	**C**	0.52	**1.38**	**1**	**1.5**	**1.100**
UGC	C	1.48	0.63	1	0.5	0.902
CGU	R	0.89	0	0.43	0.53	0.462
CGC	R	0.01	0	0.43	0	0.110
CGA	R	0.02	0.83	0	0.33	0.295
CGG	R	0.02	0	0	0	0.005
AGA	**R**	**3.38**	**2.48**	**2.57**	**3.1**	**2.882**
AGG	**R**	**1.68**	**2.69**	**2.57**	**2.04**	**2.245**
GGU	G	0.55	1.57	1	1.01	1.032
GGC	G	0.52	0.5	0.53	1.03	0.645
GGA	G	1.53	0.71	1.33	1.1	1.167
GGG	G	1.4	1.21	1.13	0.86	1.150

* The most frequently used codons are shown in bold.

The highest similarities; and the lowest nucleotide diversity were detected for phylogroup I (GI), which may indicate strong selection pressures (e.g. due to fitness) or evolutionary bottlenecks/founder effects (26).

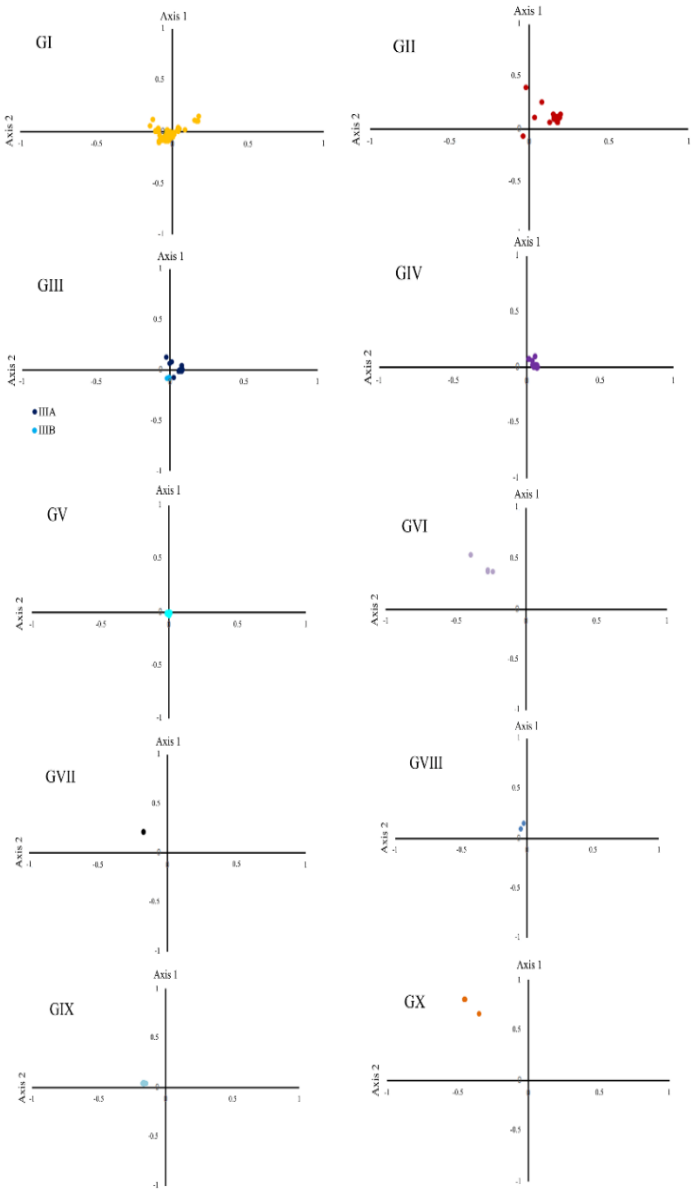


Fig. 6. Principal component analysis, 1st axis plotted against 2nd axis according to the *Grapevine leafroll-associated virus-3* phylogroups.

The role of negative selection pressure in the genetic variation of this virus can be inferred from dN/dS values (Table 1). Generally, high haplotype diversities, frequent gene flow and low genetic differentiation among populations indicate the recent distribution of GLRaV-3 (Table 2). Meanwhile, Fu's Fs test was not significant for Africa, America, and West Asia geographic region ($P > 0.05$), but the result was significant for East Asia and Europe populations as also indicates for Tajima's D test. The negative values of Tajima's D suggest

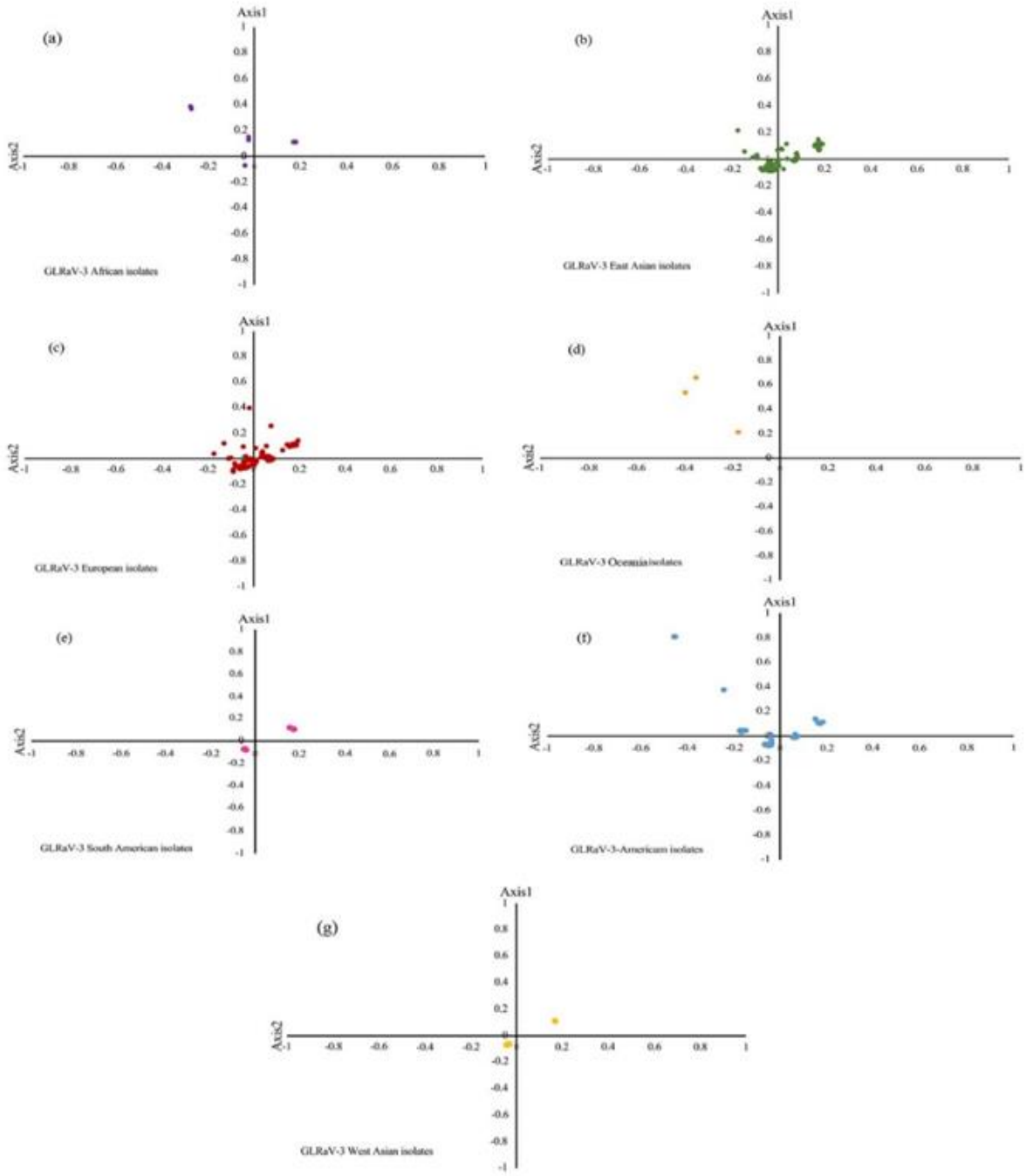


Fig. 7. Principal component analysis, 1st axis plotted against 2nd axis according to the geographical isolation of *Grapevine leafroll-associated virus-3*.

a possible recent population expansion or from genetic hitchhiking. On the other hand, positive Tajima D values associated with multimodal mismatch distributions point to a possible population bottleneck and/or population subdivision. As shown in Table 2, significant genetic differentiation and frequent gene flow were identified among GLRaV-3 populations except between East Asia and Oceania (Table 1), which indicates geographical isolation may not have played a role in GLRaV-3 population structure.

Nucleotide composition analysis indicated that the GLRaV-3 CP gene was AU-rich. However, codons with G or A in the third position are preferred in the CP gene (Table 3), which indicates codon usage bias. This unequal use of nucleotides indicates the overlapping influences of mutational pressure and natural selection on the codon preferences in the present CP gene sequences as previously reported for the

citrus tristeza virus (CTV) (27). This study showed that overall codon usage patterns within the CP gene of GLRaV-3 is slightly biased.

The low ENC values represent an approximately constant and conserved CP composition. The lower codon usage bias has been previously reported for some plant viruses including begomoviruses (9), papaya ringspot virus-PRSV (28), citrus tristeza virus-CTV (27), potato virus M-PVM (11), and apple mosaic virus-ApMV (29). A low CUB of RNA viruses has an advantage for efficient replication in the host cells by reducing the competetion between the virus and host in using the synthesis machinery (30).

PCA was used to analyze the major trends in codon usage patterns among the GLRaV-3 CP gene. This plot demonstrated that the majority of GLRaV-3 isolates belonging to phylogroups I, and V are clustered near the origin of plots, which may suggest that the main codon usage trend is somewhat identical. However, the isolates from other phylogroups were dispersal, which may reflect independently evolved due to biological variation and/or dispensation diversities. GLRaV-3 isolates belonging to subgroup V (including four Portugalian and two Chinese isolates) were the nearest to the origin of the plot (Fig. 6), which is in line with the theory about the possible origin of the virus in the Old World (5). It assumes that the origin of the graph coincides with the ancestral (original) virus (31, 32).

Conclusion

The investigation on the distribution of GLRaV-3 and its molecular characteristics can give us more data about the epidemiology and control of graft-transmissible viruses.

According to vegetative propagation of grapevines, it is not easy to eliminate virus from infected trees, therefore virus control is reduced to propagation through healthy scions and virus-resistant rootstocks. This analysis is the first demonstration of the population structuring of GLRaV-3 in mid-Eurasian Iran. Indeed, these consequences explain selectively driven codon bias in GLRaV-3 species; and reveal the potential importance of expression-mediated selection in shaping the genome evolution of this virus.

Acknowledgment

This work was carried out in the laboratory of the Plant Virus Research Department, Iranian Research Institute of Plant Protection (IRIPP), Agricultural Research, Education and Extension Organization (AREEO). Financial support from the IRIPP project (Projects No. 961463 and 83063) is appreciatively acknowledged.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

None

Ethics Approval and Consent to Participate

Not applicable.

References

1. Burger JT, Maree HJ, Gouveia P, Naidu RA. Grapevine leafroll-associated virus3. Grapevine Viruses: Molecular Biology, Diagnostics, and Management. Springer, Cham, 2017. pp. 167-195.

2. Le Maguet J, Beuve M, Herrbach E, Lemaire O. Transmission of six ampeloviruses and two vitiviruses to grapevine by *Phenacoccus aceris.* Phytopathology 2012;102(7):717-23.

3. Callaway A, Giesman-Cookmeyer D, Gillock ET, Sit TL, Lommel SA, The multifunctional capsid proteins of plant RNA viruses. Ann Rev Phytopathol. 2001;419-60.

4. Liu MH, Li MJ, Qi HH, Guo R, Liu WM, Wang Q, Cheng DY-Q. Occurrence of Grapevine leafroll-associated viruses in China. Plant Dis. 2013;97(10): 1339-45.

5. Maree HJ, Pirie MD, Oosthuizen K, Bester R, Rees DJ, Burger JT. Phylogenomic analysis reveals deep divergence and recombination in an economically

important grapevine virus. PLoS One 2015;10: e0126819.
6. Crnogorac A, Panno S, Mandić A, Gasˇpar M, Caruso AG, Noris E, Davino S, MatićI S. Survey of five major grapevine viruses infecting Blatina and ˇZilavka cultivars in Bosnia and Herzegovina. PLoS One 2021;16: e0245959.
7. FAO, Food and agriculture organization of the United Nations; (cited 2021 January 3). http:// www.fao.org/ faostat/en/#home. 2019.
8. Pourrahim R, Ahoonmanesh A, Farzadfar Sh, Rakhshandehro F, Golnaraghi AR. Occurrence of *Arabis mosaic virus* and *Grapevine leaf roll associated virus-3* on grapevines in Iran. Plant Dis. 2004;88(4):424.
9. Xu X-Zh, Liu Qi-po, Fan L-ji, Cul Xi-fe, Zhou Xu-pi. Analysis of synonymous codon usage and evolution of begomoviruses. J Zhejiang Uni Sci. 2008;9:667-74.
10. Liu XS, Zhang YG, Fang YZ, Wang Y-Lu. Patterns and influencing factor of synonymous codon usage in porcine circovirus. Virol J. 2012;9:68.
11. He Z, Gan H, Liang X. Analysis of synonymous codon usage bias in *Potato virus M* and its adaption to hosts. Viruses 2019;11(8):752.
12. Gu H, Chu DKW, Peiris M, Poon LLM. Multivariate analyses of codon usage of SARS-CoV-2 and other betacoronaviruses. Virus Evol. 2020: veaa032.
13. Clark MF, Adams AN. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J Gen Virol. 1977:34(3):475-83.
14. Hall TA, BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Sym Ser. 1999;41: 95-8.
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGAX: Molecular evolutionary genetics analysis across Computing Platforms Sudhir. Mol Biol Evol. 2018;35(6):1547-49.
16. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254-67.
17. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, GuiraoRico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol Biol Evol. 2017;34: 3299–302.
18. Govindaraju RD. Variation in gene flow levels among predominantly self pollinated plants. J Evol Biol. 1989;2:173-81.
19. Hudson RR. A new statistic for detecting genetic differentiation. Genetics. 2000;155:2011-14.
20. Tsompana M, Abad J, Purugganan M, Moyer J. The molecular population genetics of the Tomato spotted wilt virus (TSWV) genome. Mol Ecol. 2005;14(1):53-66.
21. Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. Mol Biol Evol. 1992;9(1):138-51.
22. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 1989;123(3):585–95.
23. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. Genetic traces of ancient demography. Proc Natl Acad Sci U S A. 1998;95(4): 1961–67.
24. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14 (13):5125-43.
25. Gouveia P, Santos MT, Eiras-Dias JE, Nolasco G. Five phylogenetic groups identified in the coat protein gene of grapevine leafroll-associated virus 3 obtained from Portuguese grapevine varieties. Arch Virol. 2011: 156(3):413-20.
26. Farooq ABU, Ma YXi, Wang Z, Zhuo N, Wenxing X, Wang GP, Hong N. Genetic diversity analyses reveal novel recombination events in Grapevine leafroll-associated virus 3 in China. Virus Res. 2013;171(1):15-21.
27. Biswas K, Palchoudhury S, Chakraborty P, Bhattacharyya U, Ghosh D, Debnath P, et. al. Codon usage bias analysis of *Citrus tristeza virus*: Higher codon adaptation to citrus reticulata host. Viruses. 2019;11(4):331.
28. Chakraborty P, Das S, Saha B, Sarkar P, Karmakar A, Saha A, Saha H, Saha A. Phylogeny and synonymous codon usage pattern of *Papaya ringspot virus* coat protein gene in the sub-Himalayan region of north-east India. Can J Microbiol. 2015;61(8):555-64.
29. Pourrahim R, Farzadfar Sh. Insights into factors affecting synonymous codon usage in apple mosaic virus and its host adaptability. J Plant Biotechnol. 2022;49(1):46-60.
30. Zhang Z, Dai W, Wang Y, Lu C, Fan H. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). Arch Virol. 2013;158(1):145-54.
31. Zhang W, Zhang L, He W, Zhang X, Wen B, Wang C, et al. Genetic evolution and molecular selection of the HE gene of Influenza C virus. Viruses 2019;11(2):167.
32. Konishi T. Principal component analysis of coronaviruses reveals their diversity and seasonal and pandemic potential. PLoS One 2020;15:e0242954.