https://doi.org/10.17221/149/2020-PPS

Phylogenetic and diversity analyses of *Garlic common latent virus* based on the TGB and CP gene sequence

Adyatma Irawan Santosa, Filiz Ertunc*

Department of Plant Protection, Faculty of Agriculture, Ankara University, Ankara, Turkey

**Corresponding author: adyatma_santosa@yahoo.com*

Citation: Santosa A.I., Ertunc F. (2021): Phylogenetic and diversity analyses of *Garlic common latent virus* based on the TGB and CP gene sequence. Plant Protect. Sci., 57: 179–187.

Abstract: *Garlic common latent virus* (GarCLV, *Carlavirus*), in co-infection, often worsens the severity of other *Allium* viruses on the garlic (*Allium sativum* Linnaeus). The nucleotide (nt) and amino acid (aa) sequences of the TGB and CP genes were analysed to get the first deep insight into the genomic variations and population structure of GarCLV. Global recombinant-free isolates were clustered into three clades in both the nt-based phylogenetic trees of TGB and CP. The clade 3 isolates shared low similarity percentages among themselves, as well as to the clade 1 and 2 isolates. Most major aa changes in the CP were observed on its 5' and 3' ends. Clade 2 obtained the lowest *S*, η, *k*, and π values for both the TGB and CP, which indicated low variations among its isolates. Both TGB and CP have undergone a negative selection, with CP being under stronger negative pressure than TGB. Neutrality tests estimated the non-significant negative values to all clades for TGB and CP, except Tajima's *D* for clade 2 of the TGB. The results of the K_s^*, K_{st}^*, Z^*, and *Snn* tests suggested that all three phylogroups were divergent to each other for both TGB and CP. The high F_{st} on all the clade comparisons for both TGB and CP showed a large gene flow among three clades.

Keywords: molecular variability; population structure; selection pressure; similarity percentages

Allium crops often come under threat of serious heavy yield losses caused by virus infections belonging to different taxonomic groups (Lot et al. 1998; Shahraeen et al. 2008). The *Garlic common latent virus* (GarCLV, *Carlavirus*) is found globally in garlic (*Allium sativum* Linnaeus), its main host (Shahraeen et al. 2008; Parrano et al. 2012). The virus was also reported to infect onions (*Allium cepa* Linnaeus) and leeks (*Allium ampeloprasum* Linnaeus) (Shahraeen et al. 2008; Santosa & Ertunc 2020).

Complete GarCLV genome sequences are approx. 8 600 bp long, and have six open reading frames (ORFs). ORF2, ORF3, and ORF4 are a triple gene block (TGB) which encodes proteins that facilitate virus movement from cell-to-cell (Yadav & Majumder 2019). ORF5 contains a coat protein (CP) encoding sequence (Yadav & Majumder 2019). Earlier phylogenetic analyses on the complete nucleotide (nt) sequence of the CP region showed that most GarCLV isolates were divided into two major groups only, which had relatively low diversities between them (Parrano et al. 2012; Pramesh & Baranwal 2013; Torrico et al. 2015). However, the presence of a third phylogroup that was predicted by Parrano et al. (2012) has been proven recently by Santosa and Ertunc (2020). Isolates belong to the third group seem to be much less common in the nature and have higher genetic variability than the isolates in the two other groups, indicating that they were products of a recent evolution of isolates from the two other groups.

Studies on GarCLV, particularly on its molecular characteristics, still occur less than on other *Allium* viruses (Yadav & Majumder 2019), perhaps due

Supported by the Scientific Research Projects Coordinator of Ankara University, Project No. 18H0447001.

https://doi.org/10.17221/149/2020-PPS

to the fact that virus infections in garlic were considered harmless since they are so common, and do not produce clear symptoms. The virus, however, is now regarded as quite important, since it is often involved in the "garlic viral complex" with other viruses such as the *Leek yellow stripe virus* (LYSV, *Potyvirus*) and the *Onion yellow dwarf virus* (OYDV, *Potyvirus*) which increases the severity of the other virus infections (Fajardo et al. 2001). The recent addition of a sequence of many new isolates from different countries and hosts to the National Center for Biotechnology Information (NCBI) GenBank allowed for the first detailed global genetic diversity and population structure studies of GarCLV, which are presented in this study.

MATERIAL AND METHODS

GarCLV isolates and recombination analysis on their TGB and CP genes. The TGB and CP genes coding sequences were analysed in this study. The nucleotide sequence of the TGB genes, which has three slightly overlapping ORFs (2, 3 and 4), is 1 182 bp long (encoding three proteins with a total of 406 aa), has a covering position of 5 981–7 162, while the CP gene is 960 bp long (encoding a protein of 319 aa), has a covering position of 7 196–8 155 in the GarCLV genome; reference isolate: accession no. JF320810.

Thirty-five isolates which have the complete nt sequences of the TGB and CP were obtained from the NCBI GenBank and used in the recombination analysis on each of the TGB and CP genes. Forty-one isolates which have only full CP sequences were also included (for a total of seventy-six isolates) in the CP gene analysis (Table 1).

All the sequences were aligned together using ClustalW (version 1.6) with the default parameters in MEGA7 software. The possible recombination event on each of the TGB and CP genes was analysed using the Recombination Detection Program (version RDP v.4.56). Only a recombination event that was supported by five or more of RDP, GENECONV, BootScan, MaxChi, ChiMaera, SiScan, and 3Seq algorithms with a Bonferroni-corrected *P*-value of < 0.05 was considered credible (Martin et al. 2015). The recombinant isolates were omitted from any further analyses of this study.

Phylogenetic tree and identity percentage analyses on TGB and CP genes. The Neighbour-Joining algorithm was applied in the MEGA7 software, with uniform rates among the sites and the complete deletion of the gaps/missing data, to construct a phylogenetic tree based on each of the TGB and CP nt

Table 1. Origins, hosts and accession numbers of the GarCLV isolates/strains analysed in this study

Origin	Host	Available genome region sequence	Accession numbers	References
Argentina	garlic	CP	KJ801305 – KJ801307, KJ124845 – KJ124848	Torrico et al. (2015)
Australia	garlic	TGB and CP	JF320810, JQ899445	Wylie et al. (2012)
		CP	MH686303, MH686304	–
Brazil	garlic	CP	AF228416	Fajardo et al. (2001)
China	garlic	TGB and CP	MN059105 – MN059112, MN059114 – MN059121, MN059123 – MN059129, MN059132 – MN059140	–
		CP	HQ873853 – HQ873855	Parrano et al. (2012)
Germany	garlic	CP	AB004805, X81138	Tsuneyoshi et al. (1998)
India	garlic	TGB and CP	KX255694	Yadav & Majumder (2019)
		CP	KF010516, JQ818255, JQ818256, JQ818257	Pramesh & Baranwal (2013)
Netherlands	garlic	CP	AB004804	Tsuneyoshi et al. (1998)
Poland	garlic	CP	KF862692 – KF862703	Chodorska et al. (2014)
South Korea	garlic	CP	DQ520092	–
Turkey	onion	CP	MN070134, MN070135, MN102094	Santosa & Ertunc (2020)
USA	garlic	CP	GQ475420, GQ475421, HQ873862, HQ873863	Parrano et al. (2012)

CP – coat protein; TGB – triple gene block

https://doi.org/10.17221/149/2020-PPS

sequences (Hall 2013). The Tamura 3-parameter with 1 000 bootstrap replicates was used to test the statistical significance of the isolate clusters (Tamura 1992). The nucleotide (nt) and amino acid (aa) identity percentages of each of the TGB and CP genes were determined using the Sequence Demarcation Tool (SDT) (version 1.2) (Muhire et al. 2014).

Amino acid variations in TGB and CP genes. The differences in alignment of each of the TGB and CP genes aa sequence were examined using BioEdit (version 7.2.5) (Hall 1999). In the CP comparison, several isolates, which were highly similar to other isolates in their respective clade, were selected, then the aa positions in the selected isolates that show major variations among the different clades were listed in a table.

Genetic diversity and polymorphism analysis. The confidence intervals of the number of haplotypes (*h*), haplotype diversity (*Hd*), number of variable sites (*S*), total number of mutations (η), average number of nt differences between the sequences (*k*), nt diversity (per site) (π), total number of synonymous sites (*SS*), total number of non-synonymous sites (*NS*), and the ratio of non-synonymous to synonymous nt diversity (ω = *dN*/*dS*) on the compared TGB and CP nt sequences were calculated using DnaSP (version 6.12.03) (Rozas et al. 2017). The gene is under positive (diversifying), neutral, and negative (purifying) selection when the ω ratio is > 1, = 1 and < 1, respectively (Rozas et al. 2017).

Neutral selection analysis. Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F** statistical tests (without outgroup, window length: 100 sites and step size: 25 sites) were performed using DnaSP (Rozas et al. 2017). The differences between the number of segregating sites and the average number of nucleotide differences became the basis for Tajima's *D* test (Tajima 1989). The differences between the number of singletons and the total number of mutations are the basis for Fu and Li's *D** test, whilst the differences between the number of singletons and the average number of nucleotide differences between the pairs of sequences are the basis for Fu and Li's *F** test (Fu & Li 1993).

Gene flow and genetic differentiation among populations. The K_s*, K_{st}*, *Z**, *Snn* and F_{st} values (Hudson et al. 1992; Hudson 2000) which determine the genetic differentiation on the TGB and CP nt sequences among phylogroups were calculated using DnaSP (Rozas et al. 2017). K_{st}* will be near zero if there is no genetic differentiation (null hypothesis) (Tsompana et al. 2005). A smaller *Z** means a smaller genetic differentiation among the population (Hudson et al. 1992). 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) (Hudson 2000). The null hypothesis in K_s*, K_{st}*, *Z**, and *Snn* is rejected by a significant *P* value (Hudson et al. 1992; Hudson 2000; Tsompana et al. 2005). F_{st} ranges between the exact same population (value of 0) to fully distinct populations (value of 1) (Hudson et al. 1992; Tsompana et al. 2005). $F_{st} > 0.25$, in most cases, indicates a large gene flow and a big genetic differentiation in the tested populations (Gao et al. 2016).

RESULTS

Recombination analysis on nucleotide sequences of the TGB and CP genes. Recombinant events in the CP gene of three isolates were detected by at least five independent methods (Table 2). The analyses did not detect any recombinant event in the TGB genes.

Phylogenetic tree and identity percentage analyses on the TGB and CP genes. Both phylogenetic trees clustered all the isolates into three lineages (clades). Each isolate that provided TGB and CP gene nt sequences was clustered in the identical clade in both phylogenetic trees [Figure S1 – See the electronic supplementary material (ESM)].

The identity analysis on the mostly Chinese isolates showed that the nt and aa sequences of the TGB genes were rather low conserved. In both the TGB and CP comparisons, the nt and aa similarity percentages among the isolates in clade 3 were observed to be lower than that of among isolates in clade 1 and 2. The clade 3 isolates also had high nt and aa differences to the clade 1 and 2 isolates. In both the TGB and CP analyses, the isolates generally had a higher aa than nt similarity percentages to the other isolates in their respective clade and to the isolates in the other clades (Table 3).

Amino acid variations in the TGB and CP genes. The major aa differences in CP gene among the three clades were mainly at the 5' and 3' ends. The changes in the clade 3 isolates that differentiated them to the clade 1 and 2 isolates occurred at position 14, 18, 26, 27, 29, 112, 233, 240, and 296. The changes in clade 2 occurred at position 15, 47, 224, 229, and 282. The changes in clade 1 occurred at position 12, 16, and 19. Isolates no. KF862692 and KF862701, which belong to clade 1 according to the nt based

https://doi.org/10.17221/149/2020-PPS

Table 2. Recombination analysis of the GarCLV CP gene nucleotide sequences

Recombinant isolate (origin)	Parents: major (origin)/ minor (origin)	Breakpoints* (begin-end)	RDP-implemented method (*P*-value)
KF862693 (Poland)	JQ899445 (Australia)/ KF862695 (Poland)	256-326/889-29 (278-939)	R (3.024×10^{-5}) B (2.005×10^{-2}) M (1.47×10^{-7}) C (1.42×10^{-8}) S (9.009×10^{-5}) 3S (1.58×10^{-11})
KF862702 (Poland)	KF862692 (Poland)/ JQ899445 (Australia)	751-17/163-246 (882-221)	R (1.543×10^{-3}) G (3.53×10^{-6}) B (1×10^{-8}) M (2.261×10^{-2}) C (4.199×10^{-9})
KJ020285 (India)	JQ818257 (India)/ KF010516 (India)	370-446/901-6 (428-914)	R (1.555×10^{-7}) G (1.073×10^{-13}) B (7.016×10^{-6}) M (1.719×10^{-11}) S (6.294×10^{-11}) 3S (7.596×10^{-15})

*Position in alignment; R – RDP; G – GENECOV; B – BootScan; M– MaxChi; C – Chimaera; S – SiScan; 3S – 3Seq

phylogenetic tree, had the same aa residues with the isolates in clade 2 and/or 3 instead of with the isolates in clade 1 at position 12, 15, 16, 19, 27, 112, 224, and 282 (Table 4). The major aa changes in the TGB genes of the clade 3 isolates occurred in a scattered way along its coding region (Figure S2 in ESM).

Genetic diversity and polymorphism analysis. In the TGB genes' comparison, there were twenty-one different haplotypes among the thirty-five tested isolates. The *S*, η, *k*, and π values were observed to be the highest in clade 3. The highest ω was obtained in clade 2, while the lowest ω belonged to clade 1 (Table 5). In the CP gene analysis, fifty-seven distinct haplotypes were detected among the seventy-three tested isolates. Clade 2 had the lowest *S*, η, *k*, and π. Clade 1 had the highest *S* and η whilst clade 3 had the highest *k* and π. The maximum ω was obtained by clade 1, while the minimum ω belonged to clade 2 (Table 5).

Neutral selection analysis. Non-significant negative values of Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F** statistical tests were observed in most clades

Table 3. Similarity percentage among the GarCLV phylogroups

Phylogroup	Similarity (%)			
	TGB		CP	
	nucleotide	amino acid	nucleotide	amino acid
All	73–100	76.4–100	77.1–100	83.4–100
Clade 1	89.9–100	93.1–100	84.9–100	85.6–100
Clade 2	97.5–100	97–100	97.7–100	97.2–100
Clade 3	82.6–100	88.9–100	78.9–100	91.8–100
Clade 1 *vs* 2	86–87.7	89.2–92.6	87.8–94	86.5–98.7
Clade 1 *vs* 3	73–74.6	76.4–79.3	77.1–84.6	83.4–94.7
Clade 2 *vs* 3	73.7–75.2	77.3–81.5	79.7–84	89.3–94

TGB – triple gene block; CP – coat protein

https://doi.org/10.17221/149/2020-PPS

Table 4. Major amino acid differences in the coat protein of the GarCLV isolates

Phylogroups	Isoltaes	Amino acid position*																								
		3	11	12	14	15	16	18	19	20	21	22	26	27	29	30	46	47	112	224	229	233	240	282	296	308
Clade 1	AF228416	V	S	R	L	A	S	R	G	D	A	E	N	D	A	V	L	C	A	A	A	T	T	A	N	I
	JF320810	T	.	.	C	.	.	.	S	E	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.
	AB004805	T	.	.	C	.	.		S	E	V	T	.	.	.	.	V	H	.	.	.	A	.	.	.	.
	AB004804	T	.	.	C	.	.	.	S	E	V	T	.	.	.	.	V	.	.	.	.	.	.	.	.	.
	MN059126	T	.	.	C	.	.		S	E	.	.	.	.	.	.	A	.	.	.	.	.	.	.	D	.
	MN059125	T	.	.	C	.	.	.	S	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	KF862692	T	A	C	C	S	L	.	R	E	.	.	H	E	.	M	V	.	S	T	.	.	.	T	.	V
	KF862701	T		C	C	S	L	.	R	E	.	.	.	.	I	M	V	.	S	T	.	P	.	T	.	.
Clade 2	GQ475420	.	.	C	C	S	L	.	R	E	.	.	.	.	.	.	V	L	.	T	T	.	.	T	.	.
	HQ873855	.	.	C	C	S	L	.	R	E	.	.	.	.	.	.	V	L	.	T	T	.	.	T	.	.
	MN059140	m	.	C	C	S	L	.	R	E	.	.	.	.	.	I	V	.	.	T	T	.	.	T	.	.
	MN059116	T	.	C	C	S	L	.	R	E	.	.	.	.	.	I	E	L	.	T	T	.	.	T	.	.
	MN059112	T	.	C	C	S	L	.	R	E	.	.	.	.	.	.	V	L	.	T	T	.	.	T	.	.
Clade 3	MN070315	.	.	C	.	.	L	S	R	.	.	.	R	.	V	I	V	L	.	.	.	A	A	.	.	.
	X81138	T	A	C	C	.	L	.	R	E	.	.	R	E	T	I	A	L	.	.	.	A	A	.	.	V
	KF010516	.	.	C	.	.	L	S	R	E	T	.	R	E	T	I	.	.	S	.	.	A	.	.	D	.
	MN059129	T	A	C	S	.	L	G	R	E	T	.	R	E	.	.	.	.	S	.	.	A	A	.	D	.
	MN059127	T	.	C	T	.	L	S	R	E	.	.	R	E	T	.	V		S	.	.	A	A	.	D	.
	MN059128	T	.	C	T	.	L	S	R	E	.	.	R	E	T	.	V		S	.	.	A	A	.	D	.

*Amino acid changes that appeared to be minor were excluded; The dots indicate the aa residues identical to the CP of GarCLV accession no. AF228416

https://doi.org/10.17221/149/2020-PPS

Table 5. Summary of the genetic diversity and polymorphism analyses of the GarCLV TGB and CP genes from the different populations

Phylogroups	*N*	*h*	*Hd*	*S*	η	*k*	π	*SS*	*NS*	*dS*	*dN*	ω
TGB												
All	35	21	0.928	480	617	122.936	0.10401	273.28	905.72	0.24745	0.06109	0.2469
Clade 1	13	5	0.538	191	202	45.064	0.03813	271.40	907.60	0.10093	0.01950	0.1932
Clade 2	19	14	0.965	84	92	14.374	0.01216	274.35	904.65	0.02554	0.00814	0.3187
Clade 3	3	2	0.667	206	206	137.333	0.11619	274.67	904.33	0.29153	0.06338	0.2174
CP												
All	73	57	0.983	467	643	92.623	0.09658	223.45	730.55	0.34809	0.01935	0.0556
Clade 1	41	33	0.956	351	422	70.965	0.07392	223.17	733.83	0.26260	0.01686	0.0642
Clade 2	26	19	0.975	73	75	11.335	0.01182	224.05	729.95	0.03790	0.00390	0.0069
Clade 3	6	5	0.933	304	367	150.533	0.15681	225.53	731.47	0.56172	0.03096	0.0551

N – number of isolates; *h* – number of haplotypes; *Hd* – haplotype diversity; *S* – number of variable sites; η – total number of mutations; *k* – average number of nucleotide differences between sequences; π – nucleotide diversity (per site); *SS* – number of synonymous sites analysed; *NS* – total number of non-synonymous sites analysed; *dS* – synonymous nucleotide diversity; *dN* – non-synonymous nucleotide diversity; ω – *dN*/*dS*; TGB – triple gene block; CP – coat protein

Table 6. Summary of the demography test statistics between the GarCLV TGB and CP gene populations

Phylogroups	Tajima's *D*	Fu and Li's *D**	Fu and Li's *F**
TGB			
All	-0.68523^{ns}	-0.06387^{ns}	-0.33534^{ns}
Clade 1	-1.41283^{ns}	-1.31276^{ns}	-1.53498^{ns}
Clade 2	−1.88128*	-0.46965^{ns}	-1.85434^{ns}
Clade 3	NA	NA	NA
CP			
All	-1.04996^{ns}	-1.85010^{ns}	-1.82238^{ns}
Clade 1	-1.04639^{ns}	-2.15115^{ns}	-2.08724^{ns}
Clade 2	-1.64290^{ns}	-2.29517^{ns}	-2.45524^{ns}
Clade 3	-0.41267^{ns}	-0.13860^{ns}	-0.21700^{ns}

*$P < 0.05$; ns – not significant; NA – not available, at least four isolates were needed in the calculation; TGB – triple gene block; CP – coat protein

Table 7. Genetic differentiation estimates for the lineages of the GarCLV, based on the TGB and CP gene sequences comparison

Comparison	$^{\alpha}K_{S^*}$	$^{\alpha}K_{st^*}$	*P* value	$^{\alpha}Z^*$	*P* value	*Snn*	*P* value	$^{\beta}F_{st}$
TGB								
Clade 1 *vs* 2	2.4671	0.3434	0.0000***	4.5379	0.0000***	1.0000	0.0000***	0.8045
Clade 1 *vs* 3	2.3614	0.3079	0.0030**	3.3592	0.0030**	1.0000	0.0050**	0.6995
Clade 2 *vs* 3	2.6585	0.2151	0.0010**	4.1768	0.0010**	1.0000	0.0010**	0.7511
CP								
Clade 1 *vs* 2	3.3943	0.1614	0.0000***	6.0780	0.0000***	1.0000	0.0000***	0.5793
Clade 1 *vs* 3	4.0792	0.0528	0.0000***	5.7912	0.0000***	0.9575	0.0000***	0.4153
Clade 2 *vs* 3	2.7078	0.1892	0.0000***	4.9376	0.0000***	0.9688	0.0000***	0.5586

*,** and *** levels of significance at 0.05, 0.01 and 0.001, respectively; ns – not significant; $^{\alpha}K_{s^*}$, K_{st^*}, Z^* and *Snn* – test statistics of the genetic differentiation; $^{\beta}F_{st}$ – coefficient of the gene differentiation, which measures the inter-population diversity

https://doi.org/10.17221/149/2020-PPS

for both the TGB and CP comparisons. The only significantly negative value of Tajima's *D* was obtained in clade 2 for the TGB genes (Table 6).

Gene flow and genetic differentiation among populations. The K_s^*, K_{st}^*, Z^*, and *Snn* values obtained from each of the TGB and CP gene comparisons among their respective clades were all significantly different, which means these clades are distinct to each other, and deserved to be clustered separately. The *Snn* values were also very high (0.9574–1.0000) in all the comparisons. The lowest F_{st} values in the TGB and CP genes comparisons were obtained from clade 1 *vs* 3 (0.6995 and 0.4152, respectively), which indicated low genetic differences between the related populations (Table 7).

DISCUSSION

Molecular variation studies on the GarCLV, thus far, have mostly been limited to the construction of phylogenetic trees based on the CP gene sequences which divide the isolates into two or three phylogroups (Parrano et al. 2012; Pramesh & Baranwal 2013; Torrico et al. 2015; Santosa & Ertunc 2020). Moreover, variations in the TGB genes have never been reported before. Therefore, the nt and aa sequences of the TGB and CP genes of the recombinant-free GarCLV isolates were observed to better understand the molecular diversities and population structure of the virus.

Torrico et al. (2015) indicated recombinant events on accession no. KF862693, KF862694, KF862701, KF862702, and KF862703. However, this study, which involved many more isolates, only detected credible recombinant signals on the CP genes of KF862693 and KF862702. Recombinancy was also found on the CP of KJ020285. These three isolates with multiple parents were not included in the further analyses, to give more accurate phylogenetic relationships among the tested isolates.

The results of the Neighbour-Joining analysis which clustered each isolate with the known TGB and CP sequences into the identical clade in both phylogenetic trees raised the possibility to predict the other gene characteristics even when only the sequence of one of the genes is available. They also suggested that the evolutionary processes in the TGB and CP were simultaneous and in same direction. MN070134, MN070135, and MN102094, the only non-garlic isolates available in the GenBank, did not create any distinct clade, probably because the *Carlavirus* TGB and CP protein functions were not associated with host specificity (King et al. 2011). The clade 1 isolates, which appeared to be ancestors of clade 2 and 3, were shown to have spread to diverse regions of the world.

The clade 3 isolates have high genomic differentiations with the clade 1 and 2 isolates in both the TGB and CP. In the TGB and CP comparisons, many of the nt changes in the clade 3 isolates generated synonymous substitutions since these isolates shared a low nt, but retained relatively high aa identities to the isolates in clade 1 and 2. Nevertheless, all the isolate variations in the CP were still higher than the molecular threshold for the *Carlavirus* species' demarcation (less than 72% nt or 80% aa identity) (King et al. 2011). The major aa changes in the CP were mostly present on the 5' and 3' ends, while the middle region remained conserved (Table 4). In the TGB, especially on the clade 3 isolates, the aa changes occurred throughout its coding region (Figure 2S in ESM). The high nt and aa diversities among the isolates currently belonging to clade 3 indicated that each of them likely had evolved locally from their respective ancestor in clade 1 or 2. They also predicted the construction of even more phylogroups when the sequence of new isolates that do not fit into clade 1 and 2 become available in the future.

The obtained data on the global GarCLV population showed a very high *Hd* for both the TGB and CP, which means there were many different haplotypes among the tested isolates. The lowest *S*, η, *k*, and π values based on both the TGB and CP analyses were all found in clade 2, proving low diversities among its isolates. Many of the clade 2 isolates were reported as the same haplotypes and, thus far, were exclusively found in China and California (USA) (Parrano et al. 2012), indicating the heavy genetic exchanges between the two regions. The highest *k* and π values obtained by clade 3 for both the TGB and CP, which showed high diversities among its isolates, were in line with the phylogenetic tree and identity percentage analyses results.

The ω values for all the isolates in the TGB and CP genes comparisons were lower than 1, suggesting that the TGB and CP experienced a negative selection, the CP (0.0556) was under a stronger negative pressure than the TGB (0.2469). A stronger negative pressure on the CP than on the TGB was also observed on the *Potato virus M* (PVM), another *Carlavirus* (Ge et al. 2014). The clade 1 and 2 isolates were under more intense negative

https://doi.org/10.17221/149/2020-PPS

evolutionary constraints compared to the isolates in the other clades in the TGB and CP, respectively. Among the clade 1 and 2 isolates, the purifying selections were more vigorous in the CP (0.0642 and 0.0069, respectively) than in the TGB (0.1932 and 0.3187, respectively), which resulted in more diverse nt changes in the TGB than the CP coding sequence of their offspring (clade 3 isolates). The generated nonsynonymous substitutions in the nt sequences of the TGB likely negatively affected the clade 3 isolates' fitness, which could help to explain the clade 3 isolates scarcity in nature.

Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F** tests assigned non-significant negative values to all the clades and overall populations for both the CP and TGB. Hence, the results demonstrated a low-frequency polymorphism and positive selection (population expansions) on those populations. The calculation of Tajima's *D* on clade 2 of the TGB was more convincing, as it was supported by a significant *P* value.

The results of the K_s*, K_{st}*, *Z**, and *Snn* tests, which were all supported by significant *P* values, revealed that all three phylogroups were significantly different from each other for both the TGB and CP comparisons. All the *Snn* values of the phylogroup comparisons reached a high point (> 0.95), supporting the suggestion that there are high genetic variations among the clades. The high F_{st} values (> 0.25) on all the clade comparisons in both the TGB and CP showed a large gene flow (genetic variations) among the compared clades. However, the obtained F_{st} showed a larger variability between clade 1 *vs* 2 than the other comparisons for both the TGB and CP, which contradicted the other results of this study. This might be due to that clade 1 consisted of isolates from geographically distant regions. A report on PVM clustered its populations not only according to phylogroups, but also to the population origins, so that F_{st} among them could be more accurately estimated (Ge et al. 2014).

CONCLUSION

The discussion presented in this paper shed some light on the molecular diversities and population structure of the GarCLV. The virus isolates were divided into three lineages according to the TGB and CP gene analyses. Further analyses are still needed, especially on newly sequenced isolates in the future that do not belong to clade 1 and 2.

REFERENCES

Chodorska M., Paduch-Cichal E., Kalinowska E., Szyndel M.S. (2014): First report of *Onion yellow dwarf virus*, *Garlic common latent virus* and *Shallot latent virus* on garlic in Poland. Plant Disease, 98: 858. doi: 10.1094/PDIS-11-13-1160-PDN

Fajardo T.V.M., Nishijima M., Buso J.A., Torres A.C., Ávila A.C., Resende R.O. (2001): Garlic viral complex: Identification of *Potyviruses* and *Carlavirus* in central Brazil. Fitopatologia Brasileira, 26: 619–626.

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

Gao F., Lin W., Shen J., Liao F. (2016): Genetic diversity and molecular evolution of *Arabis mosaic virus* based on the CP gene sequence. Archives of Virology, 161: 1047–1051.

Ge B., He Z., Zhang Z., Wang H., Li S. (2014): Genetic variation in *Potato virus M* isolates infecting pepino (*Solanum muricatum*) in China. Archives of Virology, 159: 3197–3210.

Hall B.G. (2013): Building phylogenetic trees from molecular data with MEGA. Molecular Biology and Evolution, 30: 1229–1235.

Hall T.A. (1999): BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41: 95–98.

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

Hudson R.R., Boos D.D., Kaplan N.L. (1992): A statistical test for detecting geographic subdivision. Molecular Biology and Evolution, 9: 138–151.

King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J. (2011): Virus taxonomy. In: Ninth Report of the International Committee on Taxonomy of Viruses. San Diego, Elsevier: 1221–1234.

Lot H., Chovelon V., Souche S., Delecolle B. (1998): Effects of Onion yellow dwarf and Leek yellow stripe viruses on symptomatology and yield loss of three French garlic cultivars. Plant Disease, 82: 1381–1385.

Martin D.P., Murrell B., Golden M., Khoosal A., Muhire B. (2015): RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution, 1: 1–5.

Muhire B.M., Varsani A., Martin D.P. (2014): SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE, 9: e108277. doi: 10.1371/journal.pone.0108277

Parrano L., Afunian M., Pagliaccia D., Douhan G., Vidalakis G. (2012): Characterization of viruses associated with garlic plants propagated from different reproductive tissues from Italy and other geographic regions.Phytopathologia Mediterranea, 51: 549–565.

Pramesh D., Baranwal V.K. (2013): Molecular characterization of coat protein gene of *Garlic common latent virus*

https://doi.org/10.17221/149/2020-PPS

isolates from India: An evidence for distinct phylogeny and recombination. Virus Genes, 47: 189–193.

Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., Sánchez-Gracia A. (2017): DnaSP 6: DNA sequence Ppolymorphism analysis of large data sets. Molecular Biology and Evolution, 34: 3299–3302.

Santosa A.I., Ertunc F. (2020): Identification, molecular detection and phylogenetic analysis of four viruses infecting *Allium cepa* in Ankara province, Turkey. Journal of Plant Diseases and Protection, 127: 561–569.

Shahraeen N., Lesemann D.E., Ghotbi T. (2008): Survey for viruses infecting onion, garlic and leek crops in Iran. Bulletin OEPP/EPPO Bulletin, 38: 131–135.

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

Tamura K. (1992): Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Molecular Biology and Evolution 9: 678–687.

Torrico A.K., Celli M.G., Conci L.R., Conci V.C. (2015): Incidence of *Garlic common latent virus* in Argentina, and phylogenetic and recombination analyses of isolates. Pesquisa Agropecuária Brasileira, 50: 363–373.

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

Tsuneyoshi T., Matsumi T., Deng T.C., Sako I., Sumi S. (1998): Differentiation of *Allium* carlaviruses isolated from different parts of the world based on the viral coat protein sequence. Archives of Virology, 143: 1093–1107.

Wylie S.J., Luo H., Li H., Jones M.G.K. (2012): Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. Archives of Virology, 157: 271–284.

Yadav V., Majumder S. (2019): The first complete genome sequence of *Garlic common latent virus* occurring in India. VirusDisease, 30: 311–314.

Received: October 11, 2020
Accepted: February 27, 2021
Published online: May 19, 2021