The first complete genome sequence of garlic common latent virus occurring in India

Vineeta Yadav¹ · Shahana Majumder¹

Received: 23 November 2018 / Accepted: 18 December 2018 / Published online: 5 January 2019
© Indian Virological Society 2019

Abstract The complete genome sequence of a garlic common latent virus (GarCLV) isolate collected from NHRDF, Karnal, India was determined. The whole genome of this GarCLV isolate consists of 8573 nucleotides, excluding the 3'-end poly-A tail. The sequence is 91–93% similar to rest of the five whole genome sequences available till date. It contains six open reading frames (ORFs) which encode polypeptide of 219 kDa, 25 kDa, 11 kDa, 7 kDa, 32.74 kDa and 15 kDa from ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 respectively. The phylogenetic study based on coat protein sequence demonstrated high diversity present among Indian population of GarCLV.

Keywords Molecular characterization · Garlic common latent virus · Complete genome

Garlic (*Allium sativum*) is one of the most important horticultural crop cultivated all over the world. The area under garlic crop has been increasing globally (<http://www.fao.org/faostat/en/>) however virus diseases of garlic has been causing significant reduction in crop yield and quality throughout, in all garlic growing areas. Garlic crops are often infected by mixture of more than one virus belonging to different taxonomic group which are known as “garlic virus complex” [20]. Garlic common latent virus (GarCLV) infecting garlic, belongs to the genus *Carlavirus* and family

Betaflexiviridae has been reported from all over world [1, 13]. GarCLV does not produce any visible symptom but can decrease the yield up to 50% and deteriorate the quality of the crop [12]. Because of vegetative propagation, these viruses can accumulate in bulb and can get vertically transmitted to next generation. Owing to vegetative propagation of garlic, elimination of these viruses becomes very difficult. Virus free plants of garlic could be generated by tissue culture techniques but they quickly get infected with the virus when planted in open field [10]. To plan an effective management strategy for the virus it is essential that we know their identity and their molecular characteristics. Till date only five complete genome sequence of GarCLV, three from Australia and two from India (Acc. No KX255694, KJ020285), are available in genomic data base. No report of molecular characterization of GarCLV genome has been published yet. In this study, for the first time, we characterized a complete genome sequence of GarCLV isolate from India.

Ten garlic accessions were collected from NHRDF, Karnal. Three accessions viz., G-1, PGS-401, G-21 were found ELISA positive for GarCLV (DAS ELISA kit, Bioreba). Total RNA was extracted from cloves of G-1 isolate by using an RNeasy Plant RNA Kit (QIAGEN) according to manufacturer's instructions. Nine pairs of primers for amplifying overlapping fragments of whole genome were designed using Bioedit version 7.0.2 [6] and were synthesized at Sigma, Bangalore (Supplementary Table 1). First-strand cDNA was synthesized using the protocol discussed in Majumder and Baranwal [13]. 50 µl PCR mixture was prepared by using 1 µl of first strand cDNA, 2.0 µl of forward and reverse primer (100 ng/µl, Sigma Aldrich), 5 µl of 10X Taq Assay buffer (Thermo-scientific), 2 µl dNTPs (10 mM, GeNeiTM), 0.5 µl Taq polymerase (3 U/µl, Thermo-scientific) and final volume

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13337-018-0505-x>) contains supplementary material, which is available to authorized users.

✉ Shahana Majumder
shahana.majumder@sharda.ac.in

¹ Department of Biotechnology, Sharda University, Knowledge Park III, G. Noida 201306, India

adjusted by adding 37.5 µl of sterile distilled water. The temperature profile was 94 °C for 5 min, then 40 cycles of 45 s at 94 °C, 45 s at 50 °C and 1 min at 72 °C and one final extension at 72 °C for 10 min. The conditions were same for all the fragments, except Poly-A fragment where annealing temperature was 45 °C. Three separate PCR product for each fragment were directly sequenced from both directions at Chromous Biotech, Bengaluru. The reads from same fragments were 98.7 to 99.5% similar. Reads with more than 99% similarity in the overlapping regions were chosen to ensure that they were part of the same strain/ genome.

Such reads were aligned and overlapping fragments were collapsed using Bioedit (7.0.7.1). Open reading frames (ORFs) were predicted using ORF finder (www.ncbi.nlm.nih.gov/projects/gorf). Molecular weight of predicted proteins were estimated by using the programs available at <http://www.bioinformatics.org>. The conserved domains in the predicted proteins were explored using conserved protein domain database (CDD) and protein predictor database (<http://www.predictprotein.org>). The analyses were carried out using the nucleotide sequence of complete genome and translated amino acid sequences from predicted ORF regions. Sequences were aligned using Clustal X-2.1, sequence identity percentage was calculated using Bioedit. Phylogenetic analysis was carried out using MEGA6 Maximum parsimony method [18].

RT-PCR amplification produced expected size of products with different primer pairs. The sequences were assembled, to obtain a 8571 nucleotide (nt) long sequence (excluding the poly A tail). The sequence was analysed and ORFs were predicted before submitting to GenBank (accession number KX255694). The 5'UTR of GarCLV isolate was found to be 52 nucleotides long. The ORFs could be translated in-silico into six polypeptides showing similarity to already predicted GarCLV proteins. The complete nt sequence of this isolate shared maximum homology of 93.1% with an Indian (KJ020285.1) followed by 93% with an Australian isolate (JQ899445.1) and 91% with another Australian Isolate (AB004804.1). The ORF1 shows 95 to 98% homology with ORF1 of other GarCLV accessions. The polypeptide encoded by ORF1 was found to have three distinct conserved functional domains viz., methyltransferase, helicases and RNA replicase (Table 1). Methyl transferase domain (53- 290aa) is present in the N terminal part of polypeptide and is usually conserved in all the + sense RNA viruses infecting plants [16]. Center part of the polypeptide encoded by ORF1 contains G-X₂-G-X-G-K-S/T motif, a signal for helicase superfamily 1 (SF1), which is found conserved in helicase of all positive stranded RNA viruses [17]. It shows 96–97% similarity to helicase of other GarCLV isolates. This motif is responsible for binding Nucleoside triphosphate (NTP) during viral RNA

replication [5]. Helicase here contains seven conserved amino acid sequence motifs as is found in SF1 type Helicase [15]. The NTP-dependent helicases are an essential component of replication cycle of single stranded RNA viruses. They help in unwinding of dsRNA structures formed temporarily during RNA transcription and also required for initiation of (early) transcription [9]. The C terminal portion of ORF1 contains a conserved motif S-G-X₃-T-X₃-N-T-X₁₈₋₃₇GDD for replicase with sequence identity of 95% to 98% with replicase of other GarCLV. This region contains five major motifs viz., A (1788aa-1811aa), B (1842aa-1870aa), C (1872aa-1892aa), D (1893aa-1913aa), E (1914aa-1932aa) common among all RNA-dependent RNA polymerase classes which indicate that protein encoded by this region is involved in RNA replication.

Amino acid 993 to 1079 of ORF1, codes for Carlavirus endopeptidase. This peptidase belongs to a family involved in auto-proteolysis of polyproteins formed by carlaviruses. Cys 995 and His 1076 are the catalytic residues. [11]. The polypeptide belongs to family C23 of MEROPS database.

ORFs 2, 3, 4 showed high similarity (94% to 96%) with the triple gene block (TGB) of GarCLV isolates (JQ899445, NC016440, JF320810). The three partially overlapping ORFs are known to encode proteins involved in the cell-to-cell movement of the viruses in plant tissues [8] (Table 1). ORF2 was predicted to encode a 25 kDa polypeptide that contains a GXGKS/T motif, a signal for helicase superfamily I (SF1). This indicates that ORF2 like ORF1 also codes for a helicase belonging to viral helicase I from protein family pfam01443. As this is not embedded in the replicative protein, so may be not involved in replication process [9] rather is essentially a movement protein [2, 15]. Viral movement is an energy consuming process, so involves an energy-dependent step. The NTPase and RNA-binding motif in this helicase indicate that this enzyme is involved in binding to and transport of RNA at the expense of energy. The polypeptide showed about 94 to 95% similarities with Helicase of other GarCLV isolates.

ORF3 has about 96% similarity with TGB2 proteins of other GarCLV isolates. It contains a highly conserved sequence, common to all carlaviruses (Table 1). Predicted TGB2 of this isolate contains two transmembrane helices and a central conserved hydrophilic loop (<http://www.enzim.hu/hmmtop/>) (Supplementary Fig. 1). These transmembrane domains are known to be essential for membrane binding and are known to bind to endoplasmic reticulum [7, 14, 15]. Association of TGB2 proteins with ER may play a role in their formation of viral replication complexes (VRC) common in *Betaflexiviridae*. This membrane anchorage are known to increase in size exclusion limits (SEL) of plasmodesmata thus helping in

Table 1 ORFs position and length of GarCLV isolate with their conserved domain and motif patterns

S.N	UTR/ ORF	Nucleotide position	Length	Proteins encoded by genome	Conserved functional domain/motif pattern/conserved sequence present in the current isolate	References
1	5'UTR		52			
2	ORF1	53-5878	5825	219 kDa polypeptide of 1941 AA residues	Methyltransferase Helicase/G-X 2-G-X-G-K-S/T RNA replicase/S-G-X3-T-X3-N-T-X-18-37GDD	[16] [15, 17]
3	ORF2	993-1079	86	25 kDa polypeptide of 230 AA residues	GXGKS/T	[2, 15]
4	ORF3	6624-6950	326	11 kDa polypeptide of 113 AA residues	Two transmembrane helices and a central conserved hydrophilic loop/VGDX2HXLPHGGXYRDGTKX3Y	[7, 14]
5	ORF4	6429-7135	706	~7 kDa polypeptide of 68 AA residues.	One transmembrane domain/CX5GX8C	[19]
6	ORF5	7227-8117	890	32.74 kDa of 296 AA residues	-KXAAFDXFXNAV-	[17]
7	ORF6	8095-8496	401	15 kDa of 133 AA residues	CHKCGKQSHYGLCSRNTISNMEVEHLIR.	[4]

cell to cell transport of the viral ribonucleoprotein complex [3].

The putative protein encoded by ORF4 showed 96% similarity with TGB3 gene of other GarCLV. This high homology region consists of a conserved sequence, of the format CX5GX8C, a common feature of all carlavirus TGB3. It contains one transmembrane domain (Supplementary Fig. 2) at N terminal and a C terminal which indicates its role in cell to cell movement of the viral nucleoprotein. Report from mutational studies on other *Betaflexiviridae* TGB3 also supports its role in virus movement [19].

The ORF5 contains a domain (Table 1) which is a conserved coat protein sequence common to all Carlaviruses [17]. This protein shows 95–97% similarities with CP of other GarCLVs. Three motifs could be identified in the sequence which belongs to protein families PF00286 (Flexi_CP), PF08358 (Flexi_CP_N) and PF04005 (Hus1-like protein), which are a constant features of many *Carlavirus* coat proteins. The coat protein did not show any strong recombination signals when subjected to RDP4, programme. In our study the cluster dendrogram based on the amino acid sequence of CP shows three major clusters. Indian isolates could be seen clustering in two major clusters. Our isolate is clustering closely with isolates from Brazil and South Korea (Acc. No. AF228416, DQ520092) but is branching out separately as a distinct line in that cluster (Fig. 1). It is not clustering closely with any Indian isolate. It shows highest identity of 96.2% with the South Korean isolate followed by a Brazilian isolate, which further indicates it to be a distinct isolate. This analysis also showed high variability among Indian population which

Fig. 1 Phylogenetic tree based on the predicted amino acid sequences of GarCLV coat protein with corresponding region of other GarCLV isolates

may be due to ease in international trade and import of garlic bulbs from various countries.

ORF6 encodes a 15 kDa cysteine-rich protein with a zinc finger signature motif found in nucleic acid-binding protein family of other carlaviruses (Table 1). This protein

is known to be responsible for transmission by aphids and also for the suppression of gene silencing in host cells of other carlaviruses [4]. It is possible that ORF6 plays a role in suppression of host defense in garlic plants.

Phylogenetic analysis involving the whole genome sequence has failed to indicate any special lineage as only four whole genome sequences of GarCLV are available. Molecular characterization of our isolate confirms its sequence identity with other GarCLV isolates and also shows distinct features which are signature of other carlaviruses. Phylogenetic analysis of coat protein gene sequence of Indian isolates do not display any strong homology among themselves (Fig. 1) rather have more identity to sequences from other parts of the globe. The phylogenetic tree and amino acid based identity matrix indicated high variability present among Indian population pointing towards a polyphyletic evolution. This variability of GarCLV in Indian garlic may have been caused by introduction of this virus from multiple locations owing to international trade.

References

1. Barg E, Lesemann DE, Vetten HJ, Green SK. Identification, partial characterization and distribution of viruses infecting *allium* crops in South and Southeast Asia. *Acta Hort.* 1994;358:251–8.
2. Beck DL, Guilford PJ, Voot DM, et al. Triple gene block proteins of white clover mosaic potyvirus are required for transport. *Virology.* 1991;183:695–702.
3. Benitez-Alfonso Y, Faulkner C, Ritzenthaler C, et al. Plasmodemata: gateways to local and systemic virus infection. *Mol Plant Microbe Interact.* 2010;23:1403–12.
4. Chiba M, Reed JC, Prokhnevsky AI, Chapman EJ, et al. Diverse suppressors of RNA silencing enhance agro-infection by a viral replicon. *Virology.* 2006;346:7–14.
5. Gorbalenya AE, Koonin E. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol.* 1993;3:419–29.
6. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
7. Hsu HT, Chou YL, Tseng YH, et al. Topological properties of the triple gene block protein 2 of Bamboo mosaic virus. *Virology.* 2008;379:1–9.
8. Huisman MJ, Linthorst HJ, Bol JF, et al. The complete nucleotide sequence of potato virus X and its homologues at the amino acid level with various plus-stranded RNA viruses. *J Gen Virol.* 1988;69:1789–98.
9. Kadare G, Haenni AL. Virus-encoded RNA helicases. *J Viro.* 1997;71:2583.
10. Kehr AE, Schaeffer GW. Tissue culture and differentiation of garlic. *Hort Sci.* 1976;11:422–3.
11. Lawrence DM, Rozanov MN, Hillman BI. Autocatalytic processing of the 223-kDa protein of blueberry scorch carlavirus by a papain-like proteinase. *Virology.* 1995;207:127–35.
12. Lot H, Chovelon V, Souche S, Delecalle B. Effects of Onion yellow dwarf and Leek yellow stripe viruses on symptomatology and yield loss of three French garlic cultivars. *Plant Dis.* 1998;82:1381–5.
13. Majumder S, Yadav V, Yakasai MA, Muhammad JY. First report of *Garlic common latent virus* in garlic from Nigeria. *J Plant Pathol.* 2016;98:3.
14. Mitra R, Krishnamurthy K, Blancaflor E, et al. The Potato virus X TGBp2 protein association with the endoplasmic reticulum plays a role in but is not sufficient for viral cell-to-cell movement. *Virology.* 2003;312:35–48.
15. Morozov SY, Solovyev AG. Triple gene block: modular design of a multifunctional machine for plant virus movement. *J Gen Virol.* 2003;84:1351–66.
16. Rozanov MH, Koonin EV, Gorbalenya AE. Conservation of the putative methyltransferase domain: a hallmark of the ‘Sindbis-loke’ supergroup of positive-strand RNA viruses. *J Gen Virol.* 1992;73:2129–34.
17. Song SI, Choi JN, Song JT. Complete genome sequence of *Garlic Latent Virus*, a member of the *Carlavirus* family. *Mol Cells.* 2002;14:205–13.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
19. Verchot-Lubicz J, Torrance L, Solovyev AG, Morozov SY, Jackson AO, Gilmer D. Varied movement strategies employed by triple gene block-encoding viruses. *Mol Plant-Microbe Interact.* 2010;23:1231–47.
20. Walkey DGA, Ebb MJW, Ballooned CJ, Millar A. Production of virus-free garlic (*Allium sativum* L.) and shallot (*A. ascalonicum* L.) by meristem-tip culture. *J Hort Sci.* 1987;62:211–20.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.