321/24

Complete nucleotide sequence, genome organization and phylogenetic analysis of eight open reading frames of grapevine leafroll associated virus 9

Author/s: Raed Alkowni(1) and Adib Rowhani(2)
(2)Department of Biology and Biotechnology, The Arab American University, Jenin, Palestine.
(1)Department of Plant Pathology, University of California, Davis, California, U.S.A.

Place & Date: UCDavis, 10th Feb, 2008

Abstract

The family *Closteroviridae* comprises the longest filamentous stranded RNA plant viruses and the most economically important ones. Even more than 30 viruses have been classified as definitive or tentative closteroviruses only few of them had been fully sequenced Analysis of these genomes has demonstrated a great deal of diversity in gene content and organization within the family *Closteroviridae*. It is believed that Closterovirus genome should have at least 7 ORFs as a "minimal set", two coding for the replication associated with ORF1a (Methyltransferase (MTR), helicase (HEL)) and ORF1b (RNA-dependent RNA polymerase (RdRp), and other 5 ORFs representing a closterovirus-specific gene block that includes HSP70, and two structural proteins, CP and CPm representing closterovirus-specific gene block required for normal movement of the virus, infection through the plant, and virus assembly

Based on the available molecular and biological information, *Closteroviridae* was revised to compromise three genera (Martelli *et al.*, 2002): (i) *Closterovirus*: type species *Beet yellows virus*; (ii) *Ampelovirus*: type species *Grapevine leafroll virus 3*;and (iii) *Crinivirus*: type species *Lettuce infectious yellows virus*. Although, some other species in this family (i.e. GLRaV-7) remained unassigned (Martelli *et al.*, 2002). The new classification of this family was

Viruses belongs to *Ampelovirus*, the newly added genus, having monopartite genome of 16.9–19.5 kb in size, and a major coat protein CP subunits of molecular mass 35–39 kDa. There are two types of genome structure in the genus, typified by GLRaV-3 and *Little cherry virus 2* (LChV-2). With GLRaV-3 and several other sequenced members of the genus, the CPd gene follows the CP gene, whereas with LChV-2, the CPd gene is five ORFs upstream of the CP cistron (Rott and Jelkmann, 2001). *Grapevine leafroll-associated virus 1* (GLRaV-1) shows a further peculiarity in that its CPd gene is duplicated (Fazeli and Rezaian, 2000).

Species of *Ampelovirus* are generally believed to be mealybug transmitted viruses (Martelli *et al.*, 2002). So far only GLRV-3 has been proven to be transmitted by *Planococcus ficus*, *P. longispinus*, *P. calciolariae*, and *Pseudococcus affinis* (**Engelbrecht and kasdorf, 2000**; Petersen and Charles, 1997; Rosciglione and Gugerli,. 1989) and PMWaV-1 and -2 by *Dysmicoccus* spp (Sether et al., 1998). GLRaV-1 was reported to be transmitted by *Heliococcus bohemicus* and *Phenacoccus aceris* (Sforza, et

al., 2000) and GLRaV-5 was by long tailed mealybug *Pseudococcus longispinus* (Golino et al., 2002)

All grapevine isolated closteroviruses (but not one (**Rowhani *et al.*, 2004?**) are associated with leafroll disease complex, the world wide spread disease, and causing yield losses up to 40% (Woodham *et al.*, 1984). Up to now 9 closteroviruses had been isolated from the grapevine and associated to leafroll disease named as Grapevine leafroll associated virus -1 to -9.
Grapevine leafroll associated virus 9, a new member of the family Closteroviridae, was isolated from grapevine showing mild leafroll symptoms (Alkowni et al., 2004). Subsequently, the full length HSP70 sequence was determined, and found to be most similar to GLRaV-5 (78% homology), but serologically are distinct. GLRaV-9 was found in some of the major grapevine varieties grown in Australia (Habili and Rowhani, 2002) as well as in Californian Vineyards (unpublished data).

Methods

Isolation and analysis of GLRaV-9 dsRNA. Viral double stranded RNA (dsRNA) was extracted from mature canes and leaf petioles of GLRaV-9-infected Helena cv.; which is also referred to as isolate GLRV118 (Golino, 1992), by the method of Valverde *et al.* (1990). Isolated dsRNA from infected grapevine, were used as a template for the generation of cDNA libraries (**Alkowni et al., 2004?),** as an electrophoretic transfer to nylon membrane in Northern hybridization, or for approximate determination of full-length size of GLRaV-9 genome. The concentration of dsRNA was estimated the UV fluorescent density of an ethidium bromide-stained dsRNA in comparison with a known concentration of DNA marker using the Kodak 1D Image Analysis Software Upgrade (1X/2X to 3.6 USB, Version 3.6., Scientific Imaging System , Eastman Kodak Company, Rochester, NY, USA).

cDNA synthesis and cloning. Complementary DNA of GLRaV-9 was synthesized from purified dsRNA and a cDNA library was prepared as described by Jelkman *et al.*, (1989) and modified by Zhang and Rowhani, (2000). Briefly, virus dsRNA was denatured by heat or 20 mM methyl mercuric hydroxide at room temperature for 10 min in the presence of 350 ng random primers. The first strand cDNA was synthesized by using Superscript II reverse transcriptase (Invitrogen Life Technologies, Carlsbad, California, USA) at 37 °C for 1 h. The second strand cDNA was obtained by using *E. coli* DNA ligase, *E. coli* DNA Polymerase I and *E. coli* RNase H . Double-stranded cDNA was blunt-ended with T4 DNA polymerase following the manufacturer instructions (Invitrogen). Taq DNA polymerase was used to add additional (A) at the ends of synthesized cDNAs. The dsDNA were ligated to pCR4-TOPO vectors (Invitrogen) using TA cloning kit and an electroporator to transform the dsDNAs to *E. coli* cells (Electromax DH10B strain). Plasmids from successfully transferred clones were

extracted using QIAGEN Miniprep extraction Kit and their inserts were checked by *Eco*RI enzymatic digestion.

Identification and sequencing of selected clones. Plasmid DNA with inserts ranging between 300-2000 bp were selected and initially sequenced in both strands using universal vector primers T3, T7, M13 or M13R followed by step-by step GLRaV-9-specific primer extension (Fig2). The clones were sequenced with The ABI PRISM® 3100 Capillary Electrophoresis Genetic Analyzer (Applied Biosystems and Hitachi, Ltd.), an automated sequencing facility (DBS) at the University of California, Davis. The obtained sequence data were processed by Chromas program (Techneylysium Pty. Ltd.) and compared. Database searching of similarity to any amino acid sequences found in that database were conducted by BLAST search programs of the National Center for Biotechnology Information (NCBI) (Altschul et al., 1997).

Bridging gaps between clones. After sequence assembly, and in order to bridge the gaps between obtained clones, several PCR-based strategies were then used. Pairs of specific primers designed at the end-sequences from selected clones (RA9-25:RA9-x2; RA9-13 : RA-X8; RA10-RA-47) were used to establish the initial sequence contigs to fill the gap among them. RT-PCR was generated using as a template dsRNA or total nucleic acid purified from the virus-infected sap by RNeasy extraction kit (Qiagen). The electrophoresed PCR-amplified products were purified from a low melting point agarose gel by QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA, USA), and sequenced directly or cloned using TA cloning Kit followed the manufacturer recommendation and sequenced on both directions.

Additional PCR with specific primers on total nucleic acid as template were carried out for further checking and verification of the obtained sequence.

Obtaining nucleotide sequences on both termini of GLRaV-9 genome. The clone walking strategy of the viral genome using the specific primers were made further sequence extensions in both directions.
The 5'-end was determined by RACE PCR products (Fig. 2). To obtain the 5′-terminal end sequence, GLRaV-9 dsRNA was denatured by the Methyl Mercuric hydroxide and reverse transcript by Supercript II RT. The purified cDNA by the S.N.A.P. column purification, was concentrated by lyophilization and the entire contents used in the. oligo-dC tailing reaction by TdT enzyme and following the manufacturer's instructions (Invitrogen).. Oligo(dG)15 was then used to prime first-strand synthesis, and the cDNA was amplified with a GLRaV-9 -specific primer. The reaction was repeated 3 times to ensure the obtained results. The RACE cDNA products were cloned and sequenced.

Sequencing and computer-assisted nucleotide and amino acid sequence analysis. Nucleotide Sequence data were initially analyzed using the web-based Genetics Computer Group (GCG) sequence analysis software package SeqWeb 2.1 (University of Wisconsin, Madison, WI, USA). BLASTX, based on the Basic Local Alignment Search Tool algorithm (Altschul et al., 1990, 1994), was used to translate nucleotide sequences to amino acid sequences and compare these putative translation products with the non-

redundant amino acid sequence database at the National Center for Biotechnology Information website (NCBI). The amino acid sequences of other closteroviruses and phylogenetic outgroups were obtained through the Entrez program at the NCBI and compared to other closterovirus sequences available in its databank.

PileUp aligning program (GCG package program) was used to create a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The PILEUP program was used for alignment of amino acid sequences of polypeptide of GLRaV-9 HEL, RdRp, HSP70 and CP, CPm with the corresponding proteins of other closteroviruses (SWISS-PROT database): *Grapevine leafroll associated virus 1* (GLRaV-1) (AF195822); GLRaV-2 (AF039204); GLRaV-3 (AF037268); GLRaV-4 (AF039553); GLRaV-5 (AF233934); GLRaV-7 (Y15987); *Beet yellows virus* (BYV) (BY51931); *Lettuce infectious yellows virus* (LIYV) (NC_003618); *Citrus tristeza virus* (CTV) (CT16304) *Potato yellow vein virus* (PYVV) (AF150984); *Little cherry virus 1* (LChV-1) (Y10237) and *Pineapple mealy bug associated virus 1* (PMWV-1) (AF414119).. Preliminary phylogeny was derived from the distances between pairs of input sequences and application of the UPGMA algorithm (Sneath & Sokal, 1973), which guides the alignment of ancestral sequences. To assess more accurately the relationships among closteroviruses, corresponding HEL and RdRp sequences from XXXX virus (XXX) (acc#....) were used as an outgroup member. In the case of HSP70, Heat shock protein 70 of *Daucus carota* (X60088) was used as an outgroup. To further confirm the tentative phylogenetic relationship, additional phylogenetic analysis was done with the assistance of the Phylogenetic Analysis Using Parsimony (PAUP*4.0) programs (Sinauer Associates, Inc., USA) and Bootstrap was used to obtain a consensus tree for a better assessment of phylogenetic relationships.
The predicted molecular masses of the polyprotein of virus open reading frames (ORFs) were determined by the DNASIS Max program package (Hitachi software Engineering Co., Ltd., UK).

Western Blot analysis. Western blot analyses were carried out as described by Uyemoto *et al.* (1997) to analyze the purified protein from infected grapevine phloem extracts with GLRaV-9, -4 and –5. those viruses are showing similarity among themselves. The proteins transferred to PVDF membrane (Immobilon, Millipore) were incubated at room temperature for 1 hour with the specific polyclonal antiserum. Membranes were washed with TBS-T buffer and stained using the Immuno-Blot kit following the manufacturer protocol (Bio-Rad).

Results and Discussion

Isolation of GLRaV-9 dsRNA and size determination. The yield of GLRaV-9 dsRNA was estimated to be 5-10 ng / 15 g phloem tissue. The low virus titer in infected tissues is also common in closteroviruses (Zhu *et al.*, 1998, Ling *et al.*, 1998, Melzer, *et al.*, 2001). The purified dsRNAs were used for the establishment of the cDNA library, RT-PCR to bridge the gaps among clones as well as for genome walking on both termini of viral

genome. It was found that dsRNA is the only good source for targeting conserved motifs within *hsp70* by using the degenerate primers that enable to amplify that portion of the gene (unpublished data). The extracted dsRNA from GLRaV-9-infected grapevines was analyzed in agarose gel and compared to other known closteroviruses (CTV and GLRaV-3) to predict its size. The consistently identified high molecular mass of the virus dsRNA was estimated to be *ca.* 15 kb (Fig. 1). The size of this virus genome is similar to minimum range of other closterovirus genomes found in grapevine (karasev, 2000).

Cloning and sequencing strategy. Over 50 different clones ranging form 300 to 2500 bp in their size were identified from sequencing cDNA library and were able to give the majority of the virus genome sequences. RT-PCR with primers designed on the obtained sequences from cDNA library was used to fill the gaps and their products were cloned and sequenced on both strands. In many cases RT-PCR was used to check the ORFs starting and stopping codons. The 5'-end was determined by RACE PCR products (Fig. 2). The sequences of seven independent RACE cDNA clones selected to be representative for the 5'-end yielded an identical sequence (*GGGGGTA*ATCTTTTGCTA), with TA in the first and second position, respectively. (Fig 3). Thus, the first nucleotide of the sequence could be designed as N, because it could not be determined if the G residue was part of the viral sequence or complementary of the artificial poly(C) tail added during RACE protocol. Almost 80% of GLRaV-9 genome (12588 bp) had been sequenced and deposited in GenBank (accession no. AY297819).

Genome organization and sequence analysis of GLRaV-9.

The virus genome organization is very helpful for understanding the mechanism of its replication and genes expression (Karasev, 2000). It is also invaluable for drawing the evolutionary development of the virus with its family member. seven open reading frames, that including the replication-associated genes (putative helicase, (putative methyltransferase (MT), and the accessory processing enzymes (papain-like proteases) (ORF1a) and RNApolymerase (ORF1b) was determined by using the web-based Genetics Computer Group (GCG) sequence analysis software package SeqWeb 2.1 (University of Wisconsin, Madison, WI, USA).

The second group of genes includes a six-gene: two genes block encodes a small, 10-kDa and 5- KDa of hydrophobic proteins, respectively; protein a 58-kDa homologue of the cellular HSP70 proteins, a 60-kDa protein, homologue of the cellular HSP90 proteins and a tandem of two structural proteins, a 30-kDa capsid protein (CP) followed by the 23-kDa coat protein minor (CPm). The genome organization of GLRaV-9 is consistent with that expected for a typical monopartite closterovirus.

ORF 1a is gene block that encodes the largest polyprotein (238 KDa), which containing the PRO, MTR and HEL domains. The putative papain-like protease (P- PRO) domain, located in the 5'-region of ORF1a, was identified by amino acids homology with P-PRO domains of other closteroviruses. The putative catalytic *cysteine* and *histidine* residues were predicted at positions 444 and, 487 respectively. The cleavage site of the P-PRO of GLRaV-9 was found at the residue Gly-Ala (504-505) (Fig. 3.A????).

The region of ORF1a immediately downstream of the PRO domain was identified as MTR domains based on homology with other closteroviruses. The amino acid residues spanning the three MTR motifs of GLRaV-9. The C-terminal region of ORF1a contained the six conserved motifs associated with the Superfamily 1 helicase of positive-strand RNA viruses (Gorbalenya & Koonin, 1993). This region also shared significant similarity with helicases of other closteroviruses, most notably that of GLRaV-3.

ORF1b, a sequence homologous to the RdRp gene of other closteroviruses, was 1,194 bp long (including the initiation codon), and potentially encoded a 398 amino acid polypeptide with a predicted molecular mass of 44.6 kDa, calculated for the implied translation product (using the DNASIS Max package program). The nucleotide sequence of ORF1b overlapped with the terminus of the preceding ORF1a at the last 7 nt, including its stop codon (ATGTT*TAG*CGTA).
Like other family members of the closteroviruses, GLRaV-9, ORF1b is thought to be expressed by a +1 ribosomal frameshift (Fig .3-B).

Database searches revealed that this protein has significant similarity with, the RdRps of positive-strand RNA viruses and contains the six conserved motifs common to them. Since there is no sequence available for GLRaV-5 and -4, the most closely related sequence to GLRaV-9 RdRp is PMWaV-1, as sharing similarity of 60% at the amino acid level (Table1).
The nucleotide sequences of the RdRp and HSP70 genes of GLRaV-9 were compared with those of a number of other viruses in the family *Closteroviridae* (Fig.4).

ORF2 starts on the 15th nucleotide after ORF1 ends. It is 255 bp in length and encodes 85 amino acids with a predicted *Mr* of about 10 kDa.

ORF3 (a small ORF unique to *Ampeloviruses*) was 138 nucleotides long, encoding 46 amino acids with a predicted *Mr* of about 5 kDa. The Mean Hydrophobicity profiles are generated for these two proteins based on the general method of Kyte and Doolittle (1982), by using the BioEdit program (Hall, 1999). (Fig. 5-0). The function of Designated this putative protein (p5) is unknown and a search of databases did not identify significantly similar proteins except its counterpart.

ORF4 is the sequence of the putative heat shock protein 70 (HSP70) with a predicted *Mr* of 58 kDa. This gene started immediately on the 7th nucleotide after the previous frame, and contained 1,599 nucleotides. The closest homology of GLRaV-9 HSP70 gene with GLRaV-5 (88%) as well as with GLRaV-4 (87%) was revealed (Table 1). Phylogenetic analysis using the the PileUp program (GCG) with the assistance of PAUP program, clustering GLRaV-9 with GLRaV-5 and –4, PMWaV-1 and PBNSaV all are classified as mealybug transmitted Ampeloviruses tentative species (Fig-5). HSP70 motifs were located between on the amino acid alignment of the GLRaV-9 and compared to other closteroviruses (Fig-5-1).

ORF 5 representing the putative HSP90 protein and overlapping HSP70, with previous gene, as starting on the nucleotide 1582 of it. HSP90 encodes a protein with a *Mr* 60 kDa

that shows similarity to the other heat shock protein 90 of closteroviruses (table-1). The virus coat protein gene comes after that gene, a character considered for the all *Ampeloviruses.*

ORF 6 has 804 amino acids that region encodes genes for the putative coat protein (CP) with an *Mr* of 29 kDa. The comparison for best fit alignment of amino acid sequences revealed that the first half of the coat protein gene sequences of GLRaV-9 at the N-terminal end had 74% identity to GLRaV-5, while the remaining portion was close to identity (Fig-7). Amino acid alignments using PileUp of the deduced nucleotide sequence with other viruses in *Closteroviridae* family revealed that GLRaV-9 clustered with *Grapevine leafroll associated virus 5* (87%) and *Pineapple mealybug wilt-associated virus-1* (56%) .

ORFs 7 downstream of the virus CP stop codon referred as coat protein minor (CPm) and potentially encoded a 207 amino acid protein with a predicted molecular mass of 23 kDa

Acknowledgement:
This work was carried out in the FPS lab at UCDavis, where part of it was done by one of the author within the scholarship Merit program of the IDB.