Rapid Communication

Permutation of the active site of putative RNA-dependent RNA polymerase in a newly identified species of plant alpha-like virus

Sead Sabanadzovic a, *, Nina Abou Ghanem-Sabanadzovic a, Alexander E. Gorbalenya b

a Department of Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA
b Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 4 May 2009
Returned to author for revision 5 June 2009
Accepted 4 August 2009
Available online 29 September 2009

Keywords:
Virus
Evolution
RdRp
Palm sub-domain
Internal permutation
Marafivirus
Tymoviridae
Grapevine
Blackberry

ABSTRACT

To direct the genome synthesis, RNA viruses without a DNA stage in the replication cycle use RNA-dependent RNA polymerase (RdRps). All RdRps have conserved right hand-like shape that includes characteristic A→B→C sequence motifs forming the active site. Recently, the structural permutation of the RdRp active site (C→A→B) has been described in few double-stranded RNA birnaviruses and a subset of positive-stranded RNA tetraviruses distantly related to Picorna-like viruses. Here we describe a permuted RdRp in the newly identified plant alpha-like virus with 6.5 kb-long polyadenylated genome, dubbed Grapevine virus Q (GVQ). The multi-domain layout and sequence similarities place GVQ into the genus Marafivirus of the family Tymoviridae. In contrast to other tymovirids, GVQ has 21 amino acid residues corresponding to the motif C relocated upstream of the motif A in the putative RdRp. This unique sequence characteristic was extensively verified and identified in several GVQ isolates infecting wild and cultivated Vitis and Rubus spp.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Family Tymoviridae is currently composed of three genera (Tymovirus, Marafivirus and Maculavirus), which accommodate viruses reported from cultivated and wild monocotyledonous and dicotyledonous plants as well as one entomovirus (Martelli et al., 2002; Dreher et al., 2005; Katsuma et al., 2005). Genomes of viruses in the family Tymoviridae slightly differ in number and organization of cistrons, but all encode a large polyprotein essential for viral replication (Dreher et al., 2005). This polyprotein contains signature amino acid motifs of viral methyltransferase (MTR), endopeptidase/protease (PRO), helicase (HEL) and RNA-dependent RNA-polymerases (RdRp) as in other “alpha-like” phytoviruses (Goldbach et al., 1991).

Diversed RdRps are the most conserved among virus-encoded proteins and share several conserved signature motifs in a particular order (Poch et al., 1989; Koonin, 1991; Koonin and Dolja, 1993) that underlies remarkable structural conservation. Indeed, all viral RdRps studied to date have conserved “right hand-like” shape with three conserved sub-domains referred to as finger, palm and thumb (Hansen et al., 1997; van Dijk et al., 2004). The palm sub-domain is the most conserved part of viral RdRps and comprises four out of eight conserved motifs described by Koonin (1991) that correspond to motifs A→B→C→D reported by Poch et al. (1989). Motifs A and C contain spatially juxtaposed and conserved Aspartic (Asp) residues involved in a two-metal (Mg2+ and/or Mn2+) mechanism of catalysis (Arnold et al., 1999; Crotty et al., 2003; Ng et al., 2008). Motif B determines whether RNA or DNA will be synthesized by selecting NTPs and dNTPs (Hansen et al., 1997).

Due to their functional importance, it would be expected that the canonical order of motifs is universally conserved across the “RdRp universe.” Surprisingly, Gorbalenya et al. (2002) reported the non-canonical organization of RdRp motifs in viral replicases of two entomoviruses with +ssRNA genomes, Thsea asigna virus (Tav) and Eusprosterna elaeaea virus (EeV), and in two dsRNA viruses (Infectious pancreatic necrosis virus—IPNV and Infectious bursal disease virus—IBDV) belonging to the family Birnaviridae (Table 1s, Supplementary material). In these RdRps, the motif C is located upstream of the motif A to form a non-canonical C→A→B arrangement associated with a unique connectivity of major structural elements underneath of the active site (Gorbalenya et al., 2002). Subsequently the permutation was verified in two studies of the IBDV RdRp

0042-6822/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.08.006
structure (Pan et al., 2007, Garriga et al., 2007). Phylogenetic analysis of RdRps of TaV/EeV/birnaviruses and representative ssRNA viruses (nidoviruses and picorna-like viruses) with canonical polymerases showed that internally permuted RdRps form a monophyletic deeply rooted clade. This clustering is due to extensive similarities outside the permuted motifs rather than the permutation per se (Gorbalenya et al., 2002). The RdRps with the C→A→B order were referred to as “internally permuted” or “non-canonical” and that terminology will be also used in this text.

During a screening for viruses infecting native Vitis and Rubus spp., which was carried out in 2007/2008, we have identified a novel virus, named Grapevine virus Q (GVQ), most resembling members of the genus Marafivirus (family tymoviridae) (Sabanadzovic and Abou Ghanem-Sabanadzovic, 2009). Surprisingly, bioinformatics analysis of the newly determined GVQ genome sequence revealed the C→A→B order of RdRp motifs that is likely to have evolved independently from the previously reported permuted RdRp of TaV/EeV/birnaviruses. These data are detailed below.

Results and discussion

Identification of GVQ

Using the “universal” tymovirid primer set RD that was designed on viral RdRp (Sabanadzovic et al., 2000) a single amplicon was generated in two tested muscadine grapes. The amplicon size was considerably larger than that of the PCR product amplified from the positive controls used in tests—Grapevine fleck virus (GFkV) (Sabanadzovic et al., 2001) and Poinsettia mosaic virus (Bradel et al., 2000), which prompted further studies. The same specimens did not generate amplicons with other primer sets used for the purposes of the survey.

Sequence analyses of multiple clones showed that the amplicons from both muscadines were of the uniform size of 450 bp and 97% identical, indicating that they were generated from very closely related templates present in both samples. These amplicons were 63 nucleotides larger than the sequences of known tymovirids used for primers design (387 bp). Consistently with the primer design, Blast-mediated comparison of the amplicon sequences (that belong to GVQ) with the GenBank/EMBL databases showed most significant similarities with marafiviruses, tymoviruses and maculaviruses of the family Tymoviridae (E values ranging from $2e^{-24}$ to $2e^{-41}$). Accordingly, multiple sequence alignment of the analyzed region involving typical tymovirids identified the 21 amino acid (aa) insertion upstream of the RdRp motif A in the newly determined sequences. We noted that the characteristic 21 aa insertion in the GVQ RdRp had a high similarity with the RdRp motif C universally located downstream in all tymovirus sequences. Combined with the lack of a counterpart to motif C in the expected position of the GVQ genome (see below), this observation indicated that GVQ has the permuted C→A→B order of RdRp sequence motifs (Fig. 1).

Given that the permutation affects the RdRp active site and has not been described for either plant or alpha-like viruses, we extensively verified the GVQ RdRp sequence. We confirmed that the permutation was observed independent of the type of template used (dsRNA or total RNA), denaturation conditions (heat, DMSO, methyl mercuric hydroxide), reverse transcriptases or cloning strategies (RT-PCR based or classic dscDNA synthesis employing primer walking technique). One Vitis and one blackberry source of GVQ (see below), co-infected with typical marafiviruses and/or maculaviruses encoding a canonical RdRp, were used to analyze sequences of the two PCR products. Comparison of sequences co-amplified from the same specimen showed they were clearly different, eliminating the possibility that the longer PCR product is an artifact produced during manipulation/processing (RNA extraction, reverse transcription, PCR) of the nucleic acids of the virus with “canonical” RdRps (Fig 1s; Supplementary material).

Genome sequences

To characterize GVQ, its complete genome sequence was determined using isolate MG-02 (muscadine grape) as a template. Excluding the poly(A) tail, it is composed of 4681 nt and includes two ORFs, a long ORF1 accounting for 96% of the whole genome and a relatively short overlapping putative ORF2 (Fig. 2A). The ORF1 encodes for a 2081-amino acid-long polypeptide of an estimated molecular mass of ca 229.5 kDa (p230). The N-terminal three fourths portion of this polypeptide includes signature motifs of viral MTR (Rozanov et al., 1992), PRO (Gorbalenya et al., 1991), HEL (Gorbalenya et al., 1988) and RdRps (Koonin, 1991; Poch et al., 1989; Koonin and Dolja, 1993) implicated in virus replication. It is predicted to be autoproteolytically processed between the Hel and RdRp domains by the Pro-mediated activity to generate two proteins, with the N-terminal one including MTR-Pro-HEL domains (Bransom and Dreher, 1994; Rozanov et al., 1995). The genome region downstream of the RdRp is likely involved in formation of two carboxy co-terminal coat proteins of the estimated Mr of 23000 and 21000. The genome contains the “tymo/marafbox” signature sequences reported to be a transcription promoter for subgenomic RNA synthesis followed by a CAA “consensus box” (Ding et al., 1990). Amino acid identities of GVQ with the closest known tymovirids approximate 65% in the conserved genome products used in the comparison (MTR, HEL, CP). Computer-assisted sequence analysis revealed the presence of an additional putative ORF, potentially encoding a proline/serine rich protein (total 43% of unknown function with the calculated mass of 27 kDa (p27).

We performed phylogenetic analysis using neighbor-joining, maximum likelihood and Bayesian algorithms. In trees involving the MTR-Pro-HEL and CP domains, GVQ consistently appeared as a sister lineage to GRVFV, forming a major sub-clade within marafiviruses (Figs. 2B, E). The clustering of GVQ with marafiviruses is also evident in the RdRp tree (Fig. 2C), although the relationships of GVQ as well as GRFV and MRFV are not fully resolved in this tree. This uncertainty is not due to the presence of the permuted motif C in the GVQ RdRp, which along with other poorly conserved regions was removed from the alignment submitted to the analysis. Importantly, the trees of marafiviruses, including GVQ, produced for the MTR-Pro-HEL protein and for an unprocessed replicase polyprotein (including MTR-Pro-HEL and RdRp domains) were congruent (Figs. 2B, D). This result indicates that the phylogenetic signal generated by the RdRp domain of marafiviruses is relatively weak. In contrast, the RdRp dominates over the MTR-Pro-HEL protein in the signal produced for some lineages of tymoviruses (Figs. 2B–D). Taken together, these observations imply that the observed difference between the MTR-Pro-HEL and RdRp trees in relation to marafiviruses is relatively minor, indicating that the RdRp motif C permutation is not likely to be associated with another gross evolutionary event in the GVQ lineage.

Virus occurrence and variability

We further proceeded to study the occurrence of GVQ in other grapevine sources. For this analysis we used new degenerate primers that were designed for broad tymovirid detection. Specifically, a pair of primers designed for the RdRp region (TymZ-F and TymZ-R) clearly distinguished between tymovirids with the “canonical” and permuted RdRps (Fig. 3A) by generating amplicons of 344 bp and 407 bp, respectively. Mixed infections were characterized by the presence of both bands (Fig. 3B).

Amplicons of 407 bp were found in two samples of Vitis rotundifolia, single specimens of Vitis aestivalis and in an unknown cultivar of Vitis vinifera. The same primer set was also applied for screening native Rubus germplasm present in the Great Smoky Mountain National Park. Besides the detection of a new “canonical” marafivirus in multiple blackberry accessions (Blackberry virus S, BIVS; Sabanadzovic and Abou Ghanem-Sabanadzovic, in press), two
distinct bands of 344 bp and 407 bp were evident in the specimen GSM-9. Sequence analyses showed that the lower band indeed originated from BlVS genome (Fig. 3B and Fig. 2s; Supplementary material). Analyses of 407 bp long PCR products from all plant specimens showed that they are (nearly) identical to the respective region of the GVQ genome (muscadine isolate).

Another GVQ-specific primer set, designed to the CP region, was used to determine the full sequence of viral coat proteins of several isolates identified in the prior analysis. In these isolates the amino acid conservation varied from 93% to 99% depending on the compared pair (Table 3sand Fig. 3s; Supplementary material).

Biochemical and biological implications of internal permutation in the GVQ RdRp

Alignment of permuted RdRp sequences encoded by GVQ and TaV/EeV shows the colinearity of characteristic motifs C (GDD), A (DX4-5D) and B (GX2-3TXN) (Fig. 3C) implying that the GVQ RdRp is likely to have a fold conserved in the permuted RdRps of other viruses (Gorbalenya et al., 2002, Pan et al., 2007, Garriga et al., 2007). In this fold, major structural elements represented by motifs C, A and B form the active site at one side and are unconventionally, compared to canonical RdRps, interconnected at the opposite side. In the IBDV RdRp, this structural organization enables regulation of the active site access (Garriga et al., 2007). A multiple alignment of RdRps of marafviruses includes only two regions containing gaps, both of which are associated with the GVQ motif C (Fig. 1). Compared to other marafviruses, the motif C in the GVQ RdRp is flanked by extra short sequences of two/three and five residues, respectively, which may be insertions accompanying the motif C permutation (reported in blue in Fig. 1). Interestingly, in silico conversion of the canonical RdRp of poliovirus into the permuted form required insertion of additional residues in the proximity of motif C (Gorbalenya et al., 2002). Thus, it is tempting to speculate...
that the extra residues flanking motif C in GVQ have been accepted to accommodate the new loop connectivity underneath the permuted motif C β-hairpin.

Apart of the common fold, the permuted RdRps of GVQ and other viruses can be contrasted. First, GVQ is a plant virus while all other viruses with permuted RdRps are of the animal origin. All other unique aspects of the GVQ RdRp are related to the phylogenetic position of GVQ among tymoviruses of the alpha-like supergroup (see above) and outside of the lineage formed by TaV/EeV/birnaviruses that has affinity to Picorna-like supergroup. This deep phylogenetic separation of the permuted RdRps, identified previously and reported in this study, is most compatible with parallel origin of the RdRp permutation in two lineages of RNA viruses. Although available data are limited, RdRps in these two lineages seem to initiate the RNA synthesis differently, either relying or not on (protein) priming (Ball, 2007). Thus, the RdRp permutation of the RdRp active site is compatible with functioning of different types of RdRps. An intermediate position of the GVQ lineage in the tymovirus tree

Fig. 2. (A) Schematic representation of GVQ genome with nucleotide coordinates. Boxes depict ORFs and lines represent untranslated genomic regions. (B–E) Phylograms depicting the relationships of Grapevine virus Q with the members of the family Tymoviridae for the MTR-Pro-Hel protein (B), RdRp (C), replicase polyprotein, (D) and coat protein (E). The trees were generated using ML algorithm; similar trees were produced using neighbor-joining method or Bayesian inference. Bootstrap values larger than 50 obtained in 100 replics are indicated at internal branch points. Internal branches with less than 50% support were collapsed. Congruent portions of replicase vs. MTR-Pro-Hel trees and replicase vs. RdRp trees are highlighted. Viruses used to construct trees, acronyms and RefSeq/GenBank accession numbers are: Anagyrus vein yellowing virus (AVYV, NC_011559), Bombyx mori macula-like latent virus (BmMLV, AB186123), Citrus sudden death associated virus (CSDaV, NC_006950), Eggplant mosaic virus (EMV, J04374), Grapevine rupestris vein feathering virus (GRVV, AY128949), Grapevine fleck virus (GFkV, NC_003347), Kennedya yellow mosaic virus (KYMV, NC_001746), Maize rayado fino virus (MRFV, AF265566), Nemesia ring necrosis virus (NeRNV, NC_011538), Oat blue dwarf virus (OBdV, U87832), Ononis yellow mosaic virus (OYMV, J04375), Physalis mottle virus (PhyMV, J16104), Poinsettia mosaic virus (PnMV, NC_002164) and Turnip yellow mosaic virus (TYMV, NC_004063). The definitive/tentative members of the genera Tymovirus, Marafivirus and Maculavirus reported in black, red and blue, respectively. Poinsettia mosaic virus, an unassigned species in the family, is reported in green.
provide an attractive alternative model for further studies of biological properties associated with permuted RdRps, when reverse genetics for this virus is established.

The presence of two bands among PCR products in our analyses was reminiscent of the data published by Shi et al. (2003) who reported two PCR bands of 353 bp and 416 bp in selected grapevine samples using a single primer pair originally designed for the specific detection of GFkV. The authors tentatively assigned bands as variants GFkV353 and GFkV416 of Grapevine fleck virus and did not further investigate the genetic basis of the variability. Comparison of limited amino acid sequences reported in the paper (see Shi et al., 2003; Fig. 2B) with our data showed that GFkV416 is indeed an isolate of GVQ (94% aa identity).

Finally, while reviewing this manuscript before final submission to the publisher, we noticed newly deposited sequences of a marafivirus denominated Grapevine Syrah virus 1 (GSyV-1; Al Rwahnih et al., 2009). Comparison of GVQ and GSyV-1 revealed over 98% sequence similarity including conservation of the permuted RdRp, which was not observed/reported in the original publication describing GSyV-1 (Al Rwahnih et al., 2009). Thus, we propose that GVQ, GSyV-1 and, likely, GFkV416 are different isolates of a single marafivirus species distinguished by the signature organization of RdRp motifs.

Materials and methods

Screening for viruses

Plant materials initially collected for the purpose of screening for possible viruses in native Vitis spp present in the Southeastern United States were tested in RT-PCR with degenerate primers for members of the family Closteroviridae (Tian et al., 1996), Flexiviridae (Dovas and Katis, 2003), Tymoviridae (Sabanadzovic et al., 2000) and members of the genus Nepovirus (DiGiaro et al., 2007) which covered the most common taxa of viruses reported in grapevines.

Virus sources

The primary plant material for this investigation was collected in 2007/2008 from an apparently healthy muscadine (Vitis rotundifolia Michx.) accession MG-02, which tested positive for tymovirids during initial assessment of the viruses infecting Vitis spp. and was used for the complete molecular characterization of the virus. Mature vines of field-grown muscadine were collected and used for double stranded RNA analyses and/or leaf petioles early in the spring for total RNA extractions.

Additionally, a total of 47 samples belonging to native and cultivated grapevines and blackberries were analyzed in this study. Samples positive for GVQ were used to partially sequence viral RdRp and coat protein (CP).

Cloning, sequencing and data analyses

Double-stranded RNAs were extracted from phloem scrapings of the accession MG-02 by selective chromatography through CF-11 columns in the presence of an appropriate buffer containing 16% ethyl alcohol as described (Valverde et al., 1990). In addition, extracted nucleic acids were treated with RNase-free DNase and RNase A (in high salt conditions) in order to eliminate possible traces of DNA and ssRNA molecules prior to use as a template for further molecular work.

Complementary DNA (cDNA) was synthesized applying a slightly modified protocol of Froussard (1992) involving cDNA synthesis and its enrichment via PCR, size selection and directional cloning into an EcoRI-cut pUC119/EcoRI plasmid (TaKaRa, South Korea). Ligation mix was transferred into Escherichia coli DH5α cells and selected plasmids were custom sequenced at MWG Biotech facility (Huntsville, AL, USA).
USA). Initial sequence data were mapped with the help of Lasergene software (DNASTar, Madison, WI) and specific primers were designed in order to generate the rest of genomic data via PCR. For determination of the viral 3′ end, an oligo dt-generated cDNA was amplified using oligo dt and RD1 primer as described (Abou Ghanem-Sabanadzovic et al., 2003). The 5′ end was determined according to the protocol enclosed in the 5′/3′ RACE kit (Roche Diagnostics, USA) using GVQ-specific primers (Table 2s, Supplementary material).

Sequences were analyzed and compared with GenBank using online resources (Altschul et al., 1997; Marchler-Bauer et al., 2007). Alignments of polyprotein sequences containing MTR, Pro, Helicase, and RdRp domains and CP sequences of a selected viruses of the Tymoviridae were produced using MUSCLE (Edgar, 2004) with subsequent removal of columns containing gaps using GBlocks (Castresana, 2000). These alignments were submitted to phylogenetic analyses using different methods/software: Neighbour Joining (Phylip; Felsenstein, 1989), Maximum-Likelihood (PhyML; Guindon and Gascuel, 2003) and Bayesian inference (MrBayes; Huelsenbeck and Ronquist, 1999) applying a 50% majority consensus rule.

RT-PCR

The primer set TymZ (Table 2s; Supplementary material) was applied in a study on GVQ incidence in Vitis and Rubus germplasm. It was designed to amplify 344 bp-long amplicon in canonical marafiviruses and 407 bp in GVQ. Total RNAs were extracted combining the protocols of Foissac et al. (2005) and commercial RNAeasy Plant Minikit (Qiagen (Sabanadzovic and Abou Ghanem-Sabanadzovic, in press), reverse transcribed with and submitted to PCR with primers TymZ under annealing temperature of 52 °C and an extension time of 45 s. Nucleic acids extracted from Poinsettia mosaic virus-infected poinsettias (PmMV; Bradel et al., 2000) and Grapevine fleck virus-infected grapevines (GfKV; Sabanadzovic et al., 2001) were used as positive controls.

In total, 35 samples belonging to different species/cultivars of native and cultivated Vitis spp. were tested with the degenerate primer set TymZ. Furthermore, the same primers were used to test 12 samples of native Rubus germplasm from the Great Smoky Mountains National Park for tymovirus/marafivirus infections as part of the general virus survey.

Another primer set, designed to amplify a 721-nt-long genomic segment encompassing the viral coat protein gene (GVQ-CP set, Table 2s; Supplementary395}material), was used to investigate the variability of GVQ genome among different isolates. PCR conditions were similar to those described for TymZ set.

Further analyses

In order to further check the sequence data in the RdRp region, several additional cloning experiments were performed. In these experiments we adopted several different strategies/conditions: (i) random primer cloning and PCR using different templates (total RNA or dsRNA), reverse transcriptases (Thermoscript, M-MLV or AMV), primer sets; (ii) primer walking technique using virus specific primers (see Fig 1s; Supplementary material); (iii) all generated products were cloned in a proper plasmid vector and sequenced as described. In all tests at least one sample containing canonical marafivirus/maculavirus infections was always processed in the same way as GVQ-infected ones.

Acknowledgments

This work was supported partially by Special Research Initiative funds from the Mississippi Agricultural and Forestry Experiment Station and Discover Life in America Inc granted to S.S. Approved for publication as a Journal Article No J-11620 of the Mississippi Agricultural and Forestry Experiment Station, Mississippi State University.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.virol.2009.08.006.

References


