



Identification and genomic characterization of grapevine Kizil Sapak virus, a novel grapevine-infecting member of the family *Betaflexiviridae*

Maher Al Rwahnih¹ · Olufemi J. Alabi² · Min Sook Hwang¹ · Kristian Stevens^{1,3} · Deborah Golino¹

Received: 21 July 2019 / Accepted: 11 September 2019 / Published online: 15 October 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

A novel virus with a (+) single-stranded RNA genome was detected by high-throughput sequencing (HTS) in a sample of grapevine (*Vitis vinifera*) cv. Kizil Sapak (sample/isolate 127) that originated from Turkmenistan. The complete genome of the virus, tentatively named “grapevine Kizil Sapak virus” (GKSV), is 7,604 nucleotides in length, excluding the poly(A) tail. The genome organization of GKSV, encoded genes, and sequence domains are typical for members of the family *Betaflexiviridae*, specifically those belonging to the subfamily *Trivirinae*. Phylogenetic analysis placed GKSV within the subfamily *Trivirinae*, in the same clade as fig latent virus 1 (FLV-1) but distinct from the clades formed by members of other genera. A comparative analysis of GKSV-127 with the HTS-derived sequences obtained from two additional isolates showed that they are genetic variants of the same virus species. Based on current ICTV species and genus demarcation criteria, and the results of the sequence and phylogenetic analyses, we propose that GKSV and FLV-1 represent a new genus within the subfamily *Trivirinae*.

The family *Betaflexiviridae* is one of five families in the order *Tymovirales*, and it consists of the subfamilies *Quinvirinae* and *Trivirinae* (<https://talk.ictvonline.org/taxon/omy/>), the latter of which is more diverse, including nine genera (*Capillovirus*, *Chordovirus*, *Citrivirus*, *Divavirus*, *Prunevirus*, *Tepovirus*, *Trichovirus*, *Vitivirus* and *Wamavirus*), whereas the former has only three (*Carlavirus*, *Foveavirus* and *Robigovirus*). A common feature of all members of the family is their flexuous filamentous virions of 12–13

nm in diameter and 600–1000 nm in length [1]. The capped (or probably capped) linear, positive-sense, single-stranded (ss) RNA genomes of betaflexiviruses range in length from 5.9 to 9.0 kb and are partitioned into two (*Capillovirus*) to five (*Foveavirus* and *Vitivirus*) open reading frames (ORFs) with different functions. Regardless of their subfamily and genus designations, each member typically codes for a replicase protein (Rep) with a size range of 190–250 kDa, a movement protein (MP) that is either of the “30K” superfamily or triple gene block (TGB) type, and a coat protein (CP) that ranges in size from 18 to 41 kDa [1]. Some members such as vitiviruses, carlaviruses, and trichoviruses may also code for a nucleic-acid-binding protein (NABP), which has been implicated in RNA silencing suppression activity, as documented for grapevine virus A [2–4]. Members of the *Betaflexiviridae* also show variation in their genome organization. For instance, whereas the “30K” type MP of capilloviruses is nested within ORF1, and that of citriviruses and trichoviruses overlap with the Rep, the vitivirus homologs are separated by a hypothetical protein and both proteins are separated by an intergenic non-coding sequence in chordoviruses. Also, viruses of the same genus or different genera can differ in their biological properties, and several members of the *Betaflexiviridae* have been shown to induce distinct

Handling Editor: Ioannis E. Tzanetakis.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00705-019-04434-3>) contains supplementary material, which is available to authorized users.

✉ Maher Al Rwahnih
malrwhahnih@ucdavis.edu

¹ Department of Plant Pathology, University of California, Davis, Davis, CA 95616, USA

² Department of Plant Pathology and Microbiology, Texas A&M AgriLife Research and Extension Center, Weslaco, TX 78596, USA

³ Department of Evolution and Ecology, University of California, Davis, Davis, California 95616, USA

symptoms in their primary and indicator hosts; several others cause asymptomatic infection.

In 2014, a new selection of white table/wine grape (*Vitis vinifera*) cv. Kizil Sapak (sample 127) was received from Turkmenistan for inclusion in the Foundation Plant Services (FPS, University of California, Davis) collection. The vine was grown in a screenhouse and assayed for a panel of grapevine viruses in the FPS pipeline as described by Al Rwahnih et al. [5]. Furthermore, the material was subjected to high-throughput sequencing (HTS) analysis as part of the routine testing procedure at the FPS. Briefly, total nucleic acid (TNA) extracts made from leaf petioles of sample 127 using a MagMax Plant RNA Isolation Kit (Thermo Fisher Scientific) were used as template for cDNA library construction employing a TruSeq Stranded Total RNA with Ribo-Zero Plant Kit (Illumina) as per the manufacturer's protocol. The cDNA library was sequenced using the Illumina NextSeq 500 platform, yielding 25.6 million single and 156.4 million paired-end raw reads, which were filtered and trimmed using Illumina bcl2fastq software. Viral sequences were obtained from SPADES v3.13.0 [6] assemblies of the deep paired-end Illumina sequence (Supplementary Table 1), preprocessed with FLASH2 [7].

The HTS analysis revealed a mixed infection of several grapevine viruses/viroids (data not shown) with one large contig of 7,590 nucleotides (nt) showing a distant relationship (39% to 48% identity; 42% to 63% coverage) with several members of the subfamily *Trivirinae* (family *Betaflexiviridae*) based on tBLASTx searches [8]. The genome sequence of the putative betaflexivirus from sample 127 was extended to completion by 5' and 3' RACE using a FirstChoice RLM-RACE Kit (Thermo Fisher Scientific) and determined to be 7,604 nt in length, excluding the poly(A) tail (GenBank no. MN172165). The virus was tentatively named "grapevine Kizil Sapak virus" (GKSV), since no discernible symptoms were associated with its occurrence.

A search of the GKSV-127 sequence with the program ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) revealed five potential protein-encoding segments, four of which were verified using the SMART BLAST or BLASTP tools and determined to show significant matches to the corresponding proteins of members of the family *Betaflexiviridae*. The 5' untranslated region (UTR) of the

virus is 170 nt long. The predicted ORF1 (nt position: 171-5,339) codes for a 196.7-kDa Rep, which is the typical size for members of the family *Betaflexiviridae* and within the size range for currently described vitiviruses (typically 190-200 kDa; [1]). Analysis of the Rep sequence using the Pfam program [9] led to the identification of conserved domains for methyltransferase (Mtr; nt position: 300-1,166), helicase (Hel; nt position: 2,907-3,644), and RNA-dependent RNA polymerase (RdRp; nt position: 4,113-5,276), all with highly significant E-values (>1.0). A pairwise comparison of this protein with homologs from the family produced the highest level of amino acid (aa) identity at 55.4% with the corresponding sequences of fig latent virus 1 (FLV-1; GenBank no. FN377573), followed by 32.6-33.4% with members of the genus *Trichovirus*. The predicted ORF2 (nt position 5,422-6,279) codes for a 31.7-kDa MP, and its Pfam analysis showed that it belonged to the "30K" superfamily type of MPs. Notably, the Rep and MP of GKSV-127 are separated by an 84-nt non-coding intergenic region (Fig. 1), and this was confirmed by reverse transcription PCR (RT-PCR) amplification of a 1.2-kb DNA fragment with primers designed downstream of the Rep and upstream of the MP, followed by Sanger sequencing of 20 independent recombinant clones (data not shown). The predicted ORF3 (nt position 6,158-6,754) overlaps with the MP and codes for a 22.0-kDa CP (Fig. 1); its Pfam analysis showed that its only matches were to members of the trichovirus CP family (E-value: 5.4e-16). In pairwise comparisons, the CP of GKSV-127 shared the highest level of aa sequence identity, at 35.7%, with the corresponding sequences of FLV-1 (GenBank no. FN377573), and was 24.8-30.1% identical to those of members of the genus *Trichovirus*. The predicted ORF4 of GKSV-127 (nt position 6,784-7,365) is separated from the CP by a 162-nt non-coding intergenic region and codes for a 21.4-kDa protein of unknown function with no significant homology to known proteins based on Pfam analysis. The predicted ORF5 (nt position 7,262-7,570) overlaps with the 21.4-kDa protein and codes for an 11.7-kDa NABP (Fig. 1); its Pfam analysis returned significant homology to the Carla_C4 family of the clan of viral NABP (E-value: 5.8e-08). The 3' UTR of GKSV-127, excluding the poly(A) tail, is 35 nt long.



Fig. 1 Genome organization of grapevine Kizil Sapak virus (GKSV). Five predicted open reading frames (ORFs) are shown as rectangular boxes: replicase (REP; ORF1; 196.7 kDa), nt 171-5339; movement protein (MP; ORF2; 31.7 kDa), nt 5422-6279; coat protein (CP;

ORF3; 22.0 kDa), nt 6158-6754; hypothetical protein (ORF4; 21.4 kDa), nt 6784-7365; and nucleic acid binding protein (NABP, ORF5; 11.7 kDa), nt 7262-7570



Fig. 2 Neighbor-joining phylogenetic trees (1,000 bootstrap replications) depicting the evolutionary relationship of grapevine Kizil Sapak virus (GKSV) to viruses belonging to the nine established genera within the subfamily *Trivirinae* of the family *Betaflexiviridae*. The

trees were derived based on analyses of (A) complete genome nucleotide sequences, (B) replicase gene amino acid sequences, and (C) coat protein gene amino acid sequences of the viruses. The type sequence of GKSV is indicated in bold font with a colored background

The near-complete genome sequences of GKSV were also recovered by HTS from two additional table/wine grape (*V. vinifera*) selections, also received from Turkmenistan in 2014. The sample 132 (white grape cv. Kara Uzum Nuhurskii) produced a single GKSV-specific contig of 7,591 nt (GKSV-132, GenBank no. MN172166), while sample 70 (red grape cv. Black Seedless) yielded a single GKSV-specific contig of 7,500 nt (GKSV-70, GenBank no. MN1721667). Both samples/isolates 132 and 70 are missing about 18 nt in their 5' UTR and more than 30 nt at the 3'-terminal end based on sequence comparisons. Analysis of all three GKSV sequences showed uniformity in the numbers and sizes of their coding genes, except for a truncation in the length of the NABP (ORF5) of GKSV-70 as a result of incomplete sequencing. In pairwise comparisons, the genome of GKSV-127 was 84% and 82% identical to those of GKSV-132 and GKSV-70, respectively, and its Rep shared 82%/90% and 89%/81% nt/aa sequence identity, respectively, with GKSV-132 and GKSV-70. The CP of GKSV-127 shared 94%/91% and 94%/88% nt/aa sequence identity, respectively, with GKSV-132 and GKSV-70, and its MP shared 83%/81% and 84%/83% nt/aa sequence identity, respectively, with GKSV-132 and GKSV-70. The ORF4 of GKSV-127 shared 91%/90% and 73%/77% nt/aa sequence identity, respectively, with GKSV-132 and GKSV-70, and its NABP shared 93%/96% and 86%/89% nt/aa sequence identity, respectively, with both virus isolates. Based on these results, it can be concluded that GKSV isolates 127, 132 and 70 are divergent variants belonging to the same virus species.

Detection primers were designed to confirm the presence of GKSV in the original grapevine sources (samples 127, 132 and 70) via one step RT-PCR. The oligonucleotides GKSV-F (5'-ATGAGATTCACAGGGGAATTCTGT-3') and GKSV-R (5'-CAAGTCCCTGATAACCCTCTGT-3'), which flank a conserved region of the virus Rep and MP genes and amplify a 1,240-bp amplicon, were used for the molecular test. The RT-PCR reaction was performed using SuperScript II Reverse Transcriptase (Life Technologies) and GoTaq (Promega), and the program consisted of 30 min at 52°C, 35 cycles of 30 s at 94°C, 45 s at 55°C, 1 min at 72°C, and a final elongation step of 5 min at 72°C. All three isolates tested positive for GKSV with this assay, and the PCR products were sequenced directly and determined to be GKSV-specific.

The complete genome sequence of the GKSV type isolate (GKSV-127) and its Rep and CP aa sequences were analyzed phylogenetically along with the corresponding sequences of representative members of each of the nine genera within the subfamily *Trivirinae*. Corresponding sequences of the mycoflexivirus botrytis virus F (GenBank no.: AF238884) were included in these analyses as an outgroup. The results

showed a consistent clustering of GKSV within the same clade as FLV-1 and separately from the clade formed by members of other genera within the subfamily *Trivirinae*, regardless of the sequence dataset employed (Fig. 2). This new clade within the subfamily *Trivirinae* is positioned close to the *Trichovirus* clade on the genome and gene-specific trees (Fig. 2).

Species belonging to the same genus within the family are expected to share $\geq 72\%$ nt (or $\geq 80\%$ aa) sequence identity in their Rep or CP, while the cutoff for distinct genera within the family is set at $< 45\%$ nt sequence identity in both genes [1]. Based on these criteria and the results of the sequence and phylogenetic analyses, we propose that GKSV and FLV-1 represent a new genus within the subfamily *Trivirinae*. Since FLV-1 is yet to be fully sequenced, the virus isolate GKSV-127 would be suited as the type species within a new genus tentatively named "*Fivivirus*" (derived from **F**icus-**V**itis-**v**irus). Further work is being initiated to identify the mechanisms of GKSV transmission and assess its host range. Field surveys, employing the molecular assays that were developed, are also underway to determine the incidence of GKSV in other selections introduced to the FPS and the USDA National Clonal Germplasm Repository in Winters, CA, its incidence and prevalence in commercial vineyards in California, and the extent of its genetic diversity.

Acknowledgements Many thanks to Alfredo Diaz-Lara and Reid Soltero Brisbane (UC-Davis, Foundation Plant Services) and Ms. Cecilia Villegas (Texas A&M) for technical assistance.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Adams MJ (2011) Family betaflexiviridae. Virus Taxon Ninth Rep Int Comm Taxon Viruses 920–941
- Chiba M, Reed JC, Prokhnovsky AI et al (2006) Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. *Virology* 346:7–14. <https://doi.org/10.1016/j.virol.2005.09.068>
- Haviv S, Moskovitz Y, Mawassi M (2012) The ORF3-encoded proteins of vitiviruses GVA and GVB induce tubule-like and punctate structures during virus infection and localize to the plasmodesmata. *Virus Res* 163:291–301. <https://doi.org/10.1016/j.virusres.2011.10.015>
- Zhou ZSh, Dell'Orco M, Saldarelli P et al (2006) Identification of an RNA-silencing suppressor in the genome of Grapevine virus A. *J Gen Virol* 87:2387–2395. <https://doi.org/10.1099/vir.0.81893-0>

5. Al Rwahnih M, Daubert S, Golino D et al (2015) Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. *Phytopathology* 105:758–763. <https://doi.org/10.1094/PHYTO-06-14-0165-R>
6. Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
7. Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
8. Altschul SF, Madden TL, Schäffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
9. El-Gebali S, Mistry J, Bateman A et al (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>

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