

P8: Discovery and partial characterization of a novel virus, tentatively named Grapevine virga-like virus

João Marcos Fagundes Silva¹, Bruna Ribeiro de Melo Pereira¹, Maher Al Rwahnih², Rosana Blawid¹, Tatsuya Nagata¹, Thor Vinicius Martins Fajardo^{3*}

¹Universidade de Brasília, Brasília-DF, Brazil, ²Department of Plant Pathology, University of California, Davis-CA, USA, ³Embrapa Uva e Vinho, Bento Gonçalves-RS, Brazil

*Corresponding author: thor.fajardo@embrapa.br

INTRODUCTION

The families *Bromoviridae* and *Virgaviridae* are composed of viruses with a single-stranded positive sense RNA genome possessing an alpha-like replication complex and a 3'-t-RNA-like structure. The main difference between these two families is that while virions in the family *Virgaviridae* are rod-shaped, those in the family *Bromoviridae* are bacilliform or icosahedral (Adams et al., 2009; Bujarski et al., 2012). Phylogenetic analysis of these viruses further supports the distinction between these two groups (Adams et al., 2009). By using high-throughput sequencing (HTS), we encountered a novel virus, provisionally named Grapevine virga-like virus (GVLV) in three out of seventeen grapevine samples. This virus also possesses an alpha-like replication complex and depending on the genomic region, shows low identity to viruses belonging to either families *Virgaviridae* or *Bromoviridae*. It is, however, more closely related to a newly described, unassigned virus, Citrus virga-like virus (CVLV) (Matsumura et al, 2017). So far, 4,620 nucleotides (nt) have been sequenced, enabling partial characterization of this divergent virus.

MATERIALS AND METHODS

To characterize the viromes of 17 grapevine samples, collected from the south, southeast and northeast regions of Brazil, dsRNA extracts were subjected to HTS on the Illumina HiSeq 2000 platform at Macrogen (Seoul, South Korea) or Eurofins Genomics (Huntsville, USA). Following a typical metagenomic pipeline, we previously identified a novel virus, Grapevine enamovirus-1, infecting four different cultivars. Reanalysis of the data using the most up-to-date viral RefSeq database from the NCBI revealed the presence of GVLV. Reads were trimmed and host derived sequences subtracted with Trimmomatic (Bolger et al., 2014) and BWA (Li and Durbin, 2010) before *de novo* assembly with SPAdes (Bankevich et al., 2012) and taxonomic assignment directly from the reads with the Kaiju webserver (Menzel et al, 2016). Reads that aligned to CVLV in the Kaiju analysis were extracted and *de novo* assembled. Contigs built by SPAdes were subjected to tBlastX (Altschul et al, 1990) searches against the most up-to-date viral RefSeq from NCBI. GVLV was found at very low coverage depth in three different grapevine samples: *Vitis flexuosa* (sample 2M-VF; 12 reads), *V. vinifera* cv. Semillon (sample S16-S; 26 reads) and *V. vinifera* cv. Cabernet Franc (sample S19-CF; 2 reads). In total, eight contigs were assembled for GVLV in the S16-S sample. Blastx searches aligned these contigs to different regions of the alpha-like replication complex of CVLV, with 45-68% of amino acid identity. Four sets of primers were designed to sequence the gaps between these contigs and confirm the infection of GVLV on the S16-S sample. PCR amplicons were sequenced yielding two contigs (GVLV-Met-Hel and GVLV-RdRp). To eliminate ambiguous characters from these sequences, GVLV-Met-Hel and GVLV-RdRp were reassembled with contigs previously built with the CLC Bio workbench assembler (CLC Bio, Qiagen, USA), also extending those sequences to a total of 4,620 nt. Phylogenetic trees for the methyltransferase and partial helicase domains were built by maximum likelihood on MEGA 7 (Kumar et al, 2016). The best-fit substitution model was estimated, and trees were built under the LG + G + I model (Le and Gascuel, 2008) with 5 gamma categories and 1,000 bootstrap replicates. This analysis included viruses from the families *Virgaviridae*, *Bromoviridae* and the genus *Idaeovirus*.

RESULTS AND DISCUSSION

Phylogenetic analysis of the methyltransferase domain positioned GVLV and CVLV as outgroups of the families *Virgaviridae* and *Bromoviridae*, whereas in the case of the partial helicase domain, GVLV and CVLV were both more closely related to the family *Virgaviridae*. This incongruence in the phylogenetic trees when considering

distinct genomic regions suggests the occurrence of early recombination events in the alpha-like replication proteins of the *Virgaviridae*, *Bromoviridae* and related viruses (Codoñer and Elena, 2008). However, it may also reflect inaccuracy in these phylogeny reconstructions since they were based on small genomic regions, especially in the partial helicase tree, that showed the worst bootstrap values. Whether GVLV and CVLV should be included as members of either one of these families depends mostly on the virion particles they form, as well as other genomic features, which description would require knowledge of the full genome sequence. GVLV shows low similarity with viruses in the families *Virgaviridae* and *Bromoviridae*, and a great portion of the putative polymerase shows no similarity to any known virus besides CVLV, indicating that these two viruses may be part a novel group. Based on Blastx alignments, GVLV-Met-Hel and GVLV-RdRp contigs show 26% and 32% identity with *Bacopa chlorosis virus* (*Bromoviridae*; query cover = 48%) and *Rehmannia mosaic virus* (*Virgaviridae*; query cover = 40%), respectively. Attempts to amplify the genomic region located between the GVLV-Met-Hel and GVLV-RdRp contigs have failed, suggesting that they may be located on distinct genomic segments. To further characterize this virus, 3' and 5' rapid amplification of cDNA ends (RACE) and visualization of the viral particles of GVLV by transmission electron microscopy (TEM) are currently underway. Additionally, RNA extracted from semi-purified viral particles will be subjected to HTS.

ACKNOWLEDGEMENTS

This work was supported by CNPq and EMBRAPA (project 02.13.14.002).

REFERENCES

- Adams, M.J., Antoniw, J.F., Kreuze, J. 2009. *Virgaviridae*: a new family of rod-shaped plant viruses. *Arch. Virol.*, 154(12): 1967-1972.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*, 19(5): 455-477.
- Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120.
- Bujarski, J., Figlerowicz, M., Gallitelli, D., Roossinck, M.J., Scott, S.W. 2012. *Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses*. Ed. King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. Elsevier Academic, San Diego. pp. 965-976.
- Codoñer, F.M., Elena, S.F. 2008. The promiscuous evolutionary history of the family *Bromoviridae*. *J. Gen. Virol.*, 89(7): 1739-1747.
- Kumar, S., Stecher, G., Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 33(7): 1870-1874.
- Le, S.Q., Gascuel, O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.*, 25(7): 1307-1320.
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5): 589-595.
- Matsumura, E.E., Coletta-Filho, H.D., Nouri, S., Falk, B.W., Nerva, L., Oliveira, T.S., Dorta, S.O., Machado, M.A. 2017. Deep sequencing analysis of RNAs from citrus plants grown in a citrus sudden death-affected area reveals diverse known and putative novel viruses. *Viruses*, 9(92). doi: 10.3390/v9040092.
- Menzel, P., Ng, K.L., Krogh, A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.*, 7: 11257.