Virus Genes
https://doi.org/10.1007/s11262-021-01867-4

SHORT REPORT

Genome characterization of fig umbra-like virus

Xupeng Wang[1] · Alejandro Olmedo-Velarde[1] · Adriana Larrea-Sarmiento[1] · Anne E. Simon[2] · Alexandra Kong[1] · Wayne Borth[1] · Jon Y. Suzuki[3] · Marisa M. Wall[3] · John Hu[1] · Michael Melzer[1]

Received: 22 January 2021 / Accepted: 11 July 2021

Abstract

The complete genome of a new umbra-like virus from edible fig (*Ficus carica*) was identified by high-throughput sequencing. Based on its similarity to umbra-like virus genome sequences available in GenBank, the proposed name of this new virus is "fig umbra-like virus" (FULV). The genome of full-length FULV-1 consists of 3049 nucleotides organized into three open reading frames (ORFs). Pairwise comparisons showed that the complete nucleotide sequence of the virus had the highest identity (71.3%) to citrus yellow vein-associated virus (CYVaV). In addition, phylogenetic trees based on whole-genome nucleotide sequences and amino acid sequences of the RNA-dependent RNA polymerase showed that FULV forms a monophyletic lineage with CYVaV and other umbra-like viruses. Based on the demarcation criteria of the genus *Umbravirus*, and lack of two umbravirus ORFs, we propose that FULV is a putative new member of the umbra-like virus clade within the family *Tombusviridae*.

Keywords Umbra-like virus · Fig · Hawaii · New species · High-throughput sequencing

The genus *Umbravirus* (family *Tombusviridae*) is composed of single-stranded RNA viruses that lack coat protein genes [1]. Viruses from this genus are known or suggested to require a helper virus from the *Luteoviridae* family for genome encapsidation and vector transmission, and the presence of viruses from this family coinfecting the same host plant is consistent with this observation [2]. The typical umbravirus genome contains four open reading frames (ORFs) [3] (Fig. 1a). ORF1 is a replication-required protein, and a -1 ribosome frameshifting event extends translation through ORF2 to generate the RNA-dependent RNA polymerase (RdRp). ORF3 and ORF4 encode the long-distance and cell-to-cell movement proteins, respectively. Sequences for several umbra-like viruses are also available in GenBank such as papaya virus Q (PpVQ) and papaya meleira virus 2 (PMeV2), which contain only the first two umbravirus ORFs [4, 5] (Fig. 1a). The genome of PMeV2 is encapsulated by the coat protein from PMeV, an unclassified member of the *Totiviridae* family, and together these viruses are the causal agents of papaya meleira disease. PMeV2 has not been shown to be associated with luteovirus infection in papaya plants affected by meleira disease [5]. Notwithstanding, the umbra-like virus from sugarcane (SULV), which contains one additional ORF (ORF5) (Fig. 1a), has been found co-infecting with a luteovirus, sugarcane yellow leaf virus, in sugarcane with mosaic symptoms [6]. The similar umbra-like virus from opuntia (OULV) has been found in aphids with a luteovirus, although not in infected opuntia plants [7] and the related umbravirus-like RNA from citrus that is associated with yellow vein disease (CYVaV) is postulated to use the enamovirus citrus vein enation virus as a helper virus [8].

Edited by Seung-Kook Choi.

Xupeng Wang and Alejandro Olmedo-Velarde have contributed equally to this work.

✉ John Hu
johnhu@hawaii.edu

✉ Michael Melzer
melzer@hawaii.edu

[1] Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, USA

[2] Department of Cell Biology and Molecular Genetics, University of Maryland College Park, College Park, MD, USA

[3] United States Department of Agriculture, Agricultural Research Service, Pacific Basin Agricultural Research Center, Hilo, HI, USA

Published online: 15 September 2021

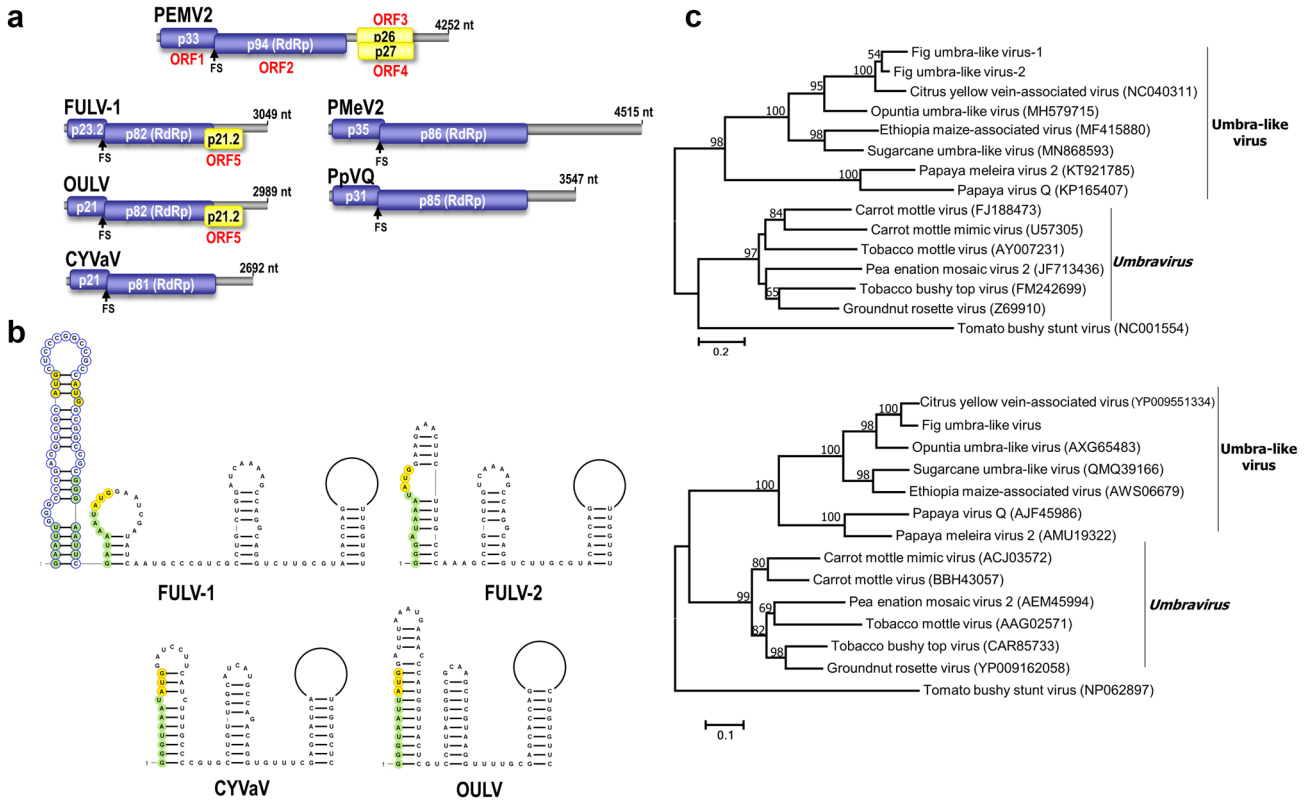

Fig. 1 Genome organization and 5'-terminal sequences of FULV and related umbra-like viruses. **a** Genome organization of FULV compared with umbravirus PEMV2 and other umbra-like viruses. ORF designations are in red. FULV is most closely related to CYVaV, which lacks ORF5 because of two large deletions in the 3'UTR [15]. **b** Proposed structures at the 5' ends of FULV-1 and FULV-2. Structures were determined using known structure of CYVaV as a model [15]. Blue circled bases in FULV-1 are not found in FULV-2 and are of unknown origin. FULV-2 5' end is more similar to those of CYVaV and OULV. Green-shaded residues comprise a CCS (Carmovirus consensus sequence) motif (2–3 G followed by a stretch of A/U), found at the 5' end of nearly all carmoviruses and umbraviruses. Initiation codons are shaded yellow. Note that FULV-1 has three in-frame AUGs, with the 5' proximal one in a very poor Kozak context [19]. **c** Phylogenetic trees constructed using the whole-genome nucleotide sequences (above) and the amino acid sequences of the RdRp sequences (bottom) of select umbraviruses and umbra-like viruses. GenBank accessions of whole-genome nucleotide sequences and amino acid sequences of the RdRp are provided. The tombusvirus tomato bushy stunt virus was used as an outgroup. The numbers above the nodes on the trees indicate statistic bootstrap support value

Edible fig (*Ficus carica*) is a crop that is primarily propagated by cuttings and is therefore prone to the accumulation of viral pathogens. Fig mosaic disease (FMD) is a complex syndrome of putative viral origin that has been commonly reported in areas where fig is grown. Typical symptoms of FMD are foliar chlorosis, deformation, and mosaic patterning; 12 viruses and 3 viroids have been reported from fig trees displaying these symptoms [9]. High-throughput sequencing (HTS) technologies are currently considered the most powerful molecular plant virus detection methods. Different nucleic acid templates, including ribosomal RNA-depleted total RNA and double-stranded RNA (dsRNA), are commonly used in HTS-based plant virus discovery and detection studies [10]. When combined with specific bioinformatic tools, HTS can be used to detect known viruses and discover novel ones. HTS does not rely on any previous knowledge of viral sequences to detect all viruses present in a plant, including those still unknown [11]. In Hawaii, work has been done using HTS to discover new RNA and DNA viruses, and novel virus strains [12–14]. In this study, we report the complete genome sequence of a new umbra-like virus that was found in FMD-diseased fig plants using HTS.

In 2009, fig trees with FMD symptoms (leaf mottling, distortion, severe mosaic) were identified from a nursery on the island of Kauai, Hawaii, USA. A single fig plant displaying these symptoms was collected and dsRNA was extracted from roots and leaves. This dsRNA was used to construct a library for HTS using a 454 Roche sequencer. Approximately, 262 k single-end reads were generated, and after quality control, trimming, and de novo assembly, 1183 contigs were produced. Contig annotation with BLASTX revealed variously sized contigs showing sequence similarity to viruses belonging to the genera *Closterovirus*, *Ampelovirus*, *Badnavirus*, *Endornavirus*, *Totivirus*, *Emaravirus*,

Carlavirus, and *Trichovirus* (Olmedo-Velarde et al., unpublished data). Moreover, BLASTX searches revealed a 3088 bp contig (GenBank accession number MW480893) that resembled CYVaV, OULV, and SULV. 8765 raw reads mapped to this 3088 bp contig. Based on this result, the name "fig umbra-like virus" (FULV) is proposed, with the Kauai isolate designated as FULV-2.

Unfortunately, the sample from the island of Kauai is no longer available. Therefore, new samples from the Urban Garden Center on the island of Oahu, Hawaii, that displayed similar FMD symptoms were collected in October 2018. Based on the umbra-like virus genome sequence from HTS, specific primers were designed to verify the presence of the viral agent in the new fig samples. Of the thirteen samples tested, seven were positive for FULV using RT-PCR and a virus-specific primer set (Table S1). Virus-specific primer sets were additionally employed to validate the virus genome and to obtain complete 5′- and 3′-terminal sequences that were determined by rapid amplification of complementary ends (RACE). RACE was performed using the SMARTer® RACE 5′/3′ Kit (Takara Bio Inc., USA) following the manufacturer's instructions. The 3′ terminus of FULV is not polyadenylated, so poly(A) polymerase (Takara Bio Inc., USA) was used prior to first-strand 3′-cDNA synthesis to establish a 3′ polyadenylated tail. An RNA extract obtained from a single plant was used to generate the entire full-length RNA genome. Additionally, all amplicons were either directly sequenced or cloned into pGEM®-T Easy Vector (Promega, USA) for which at least three independent clones were sequenced.

The complete genome of FULV from the Oahu sample (designated FULV-1) consists of 3049 nucleotides (GenBank accession number MW480892). The ORF finder tool in NCBI showed that FULV-1 is similar to OULV and SULV in containing three ORFs (Fig. 1a). The three ORFs have a nucleotide sequence identity that range from 35.6% to 83.5% with other umbra-like viruses (Table S2). Based on protein homologies and synteny of ORFs with viruses in the *Tombusviridae*, the function of the proteins encoded by ORF1 (23.2 kDa) is likely a replication-required protein, while the ORF1-ORF2 fusion product encodes the RdRp, which is generated by a translational-1 frameshift as was demonstrated for CYVaV [15] and umbravirus pea enation mosaic virus [16]. The function of the protein coded by ORF5 is unknown but contains motifs consistent with some movement proteins [15] (Fig. S1).

Phylogenetic trees were constructed using the sequences of FULV, other umbra-like viral RNAs, and umbraviruses. To infer the phylogeny, whole-genome nucleotide sequences and amino acid sequences of RdRp from FULV homologs were retrieved from GenBank. A Maximum Likelihood method using the model GTR + G + I was employed with 1000 bootstrap repetitions as support value and implemented in MEGA X [17]. The phylogenetic trees show that umbra-like viruses, including the FULV sequences, form a distinct monophyletic clade that branches out from the *Umbravirus* clade (Fig. 1c). FULV sequences were placed most closely to those of CYVaV and OULV.

The nucleotide sequence identity between the sequences obtained by Sanger sequencing from the Oahu sample (MW480892, 3049 bp) and HTS from the Kauai sample (MW480893, 3088 bp) was 86%. In addition to the sequence variation, both FULV-1 and FULV-2 contain inserted sequences not found in the other isolate. For FULV-1, the 5′ end starts with 53 nt not found in FULV-2 or in the related CYVaV or OULV (Fig. 1b). This insert contains two in-frame AUGs and the weak Kozak context of the 5′ proximal AUG suggests that additional information is required before the identity of the ORF1 initiation codon is known. Interestingly, this extra sequence is predicted to fold into a 5′ proximal hairpin, similar to the 5′ hairpins of FULV-2, CYVaV and OULV (Fig. 1b). The 90 nt insert in FULV-2 is located in the 3′ UTR, just past an extended hairpin that has been identified in CYVaV as the 3′ cap-independent translation enhancer [15] (Fig. 2). This molecular diversity in the fig umbra-like virus may have several explanations: different geographic sample origins or these isolates may actually correspond to different strains of the same viral species.

Umbraviruses may be distributed throughout the whole plant in parenchymatic tissue, in contrast to luteoviruses that are restricted to the phloem [18]. Samples of leaf midribs, petioles, and buds were used to extract RNA. From these RNA samples, FULV was detected, but no luteovirus sequences were detected, even when a more sensitive method, nested RT-PCR, was used with universal primer sets for luteovirus detection [18]. It is possible that the concentration of a luteovirus in our samples was too low to be detected, or the luteovirus is genetically distinct and is escaping detection using the universal primer sets, although no luteovirus sequences were found in the HTS dataset. Alternatively, it cannot be excluded that the FULV helper virus may be a virus other than a luteovirus.

Demarcation criteria for a new species in the genus *Umbravirus* are nucleotide sequence identity below 70% and difference in natural host range [3]. FULV-1 shares 72% nucleotide sequence identity overall with CYVaV, and FULV-2 retrieved from HTS shares 69% nucleotide sequence identity overall with CYVaV. Furthermore, the known hosts of FULV and CYVaV are different. For these reasons, we tentatively conclude that the virus present in fig in Hawaii is a putative new member of the umbra-like viral RNA clade within the family *Tombusviridae*, and the name "fig umbra-like virus" is proposed. Further work is needed to determine the correlation between this putative new virus and FMD symptoms. The collection of additional samples is necessary to investigate the distribution of FULV

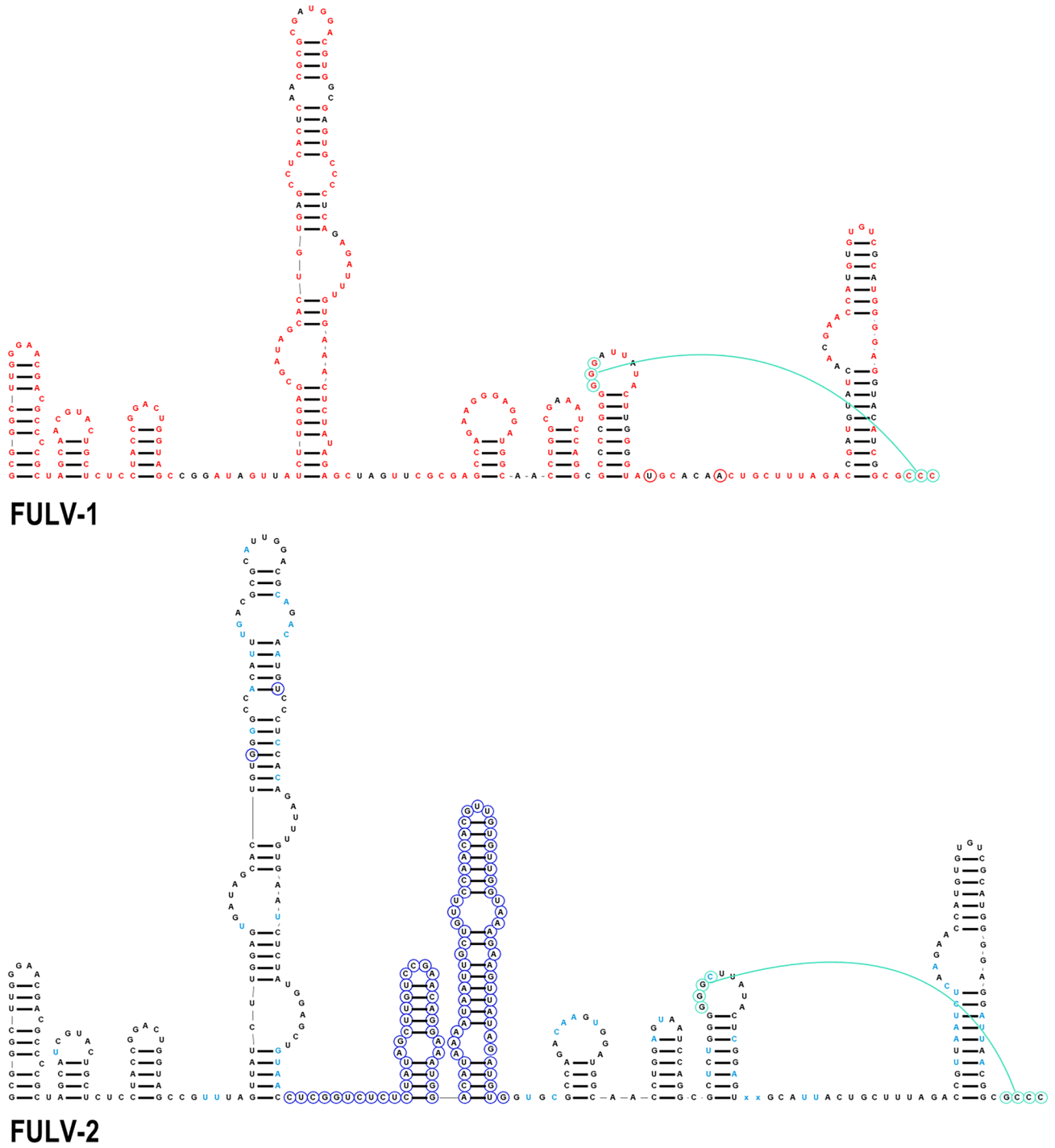

Fig. 2 Proposed structures at the 3' ends of FULV-1 and FULV-2. Structures were determined using known structure of CYVaV as a model [15]. Highly conserved pseudoknot throughout the *Tombusviridae* is denoted by a cyan line. FULV-1 structure: Bases in red are conserved in CYVaV; circled bases are not found in CYVaV. FULV-2 structure: Bases in blue denote sequence differences with FULV-1; circled bases in blue are not found in FULV-1; blue "x" denotes base found in FULV-1 is deleted in FULV-2

and possible involvement in FMD, if any. Correspondingly, detecting luteoviruses or other possible helper viruses could lead to a better understanding of the FULV and fig mosaic disease.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11262-021-01867-4.

Acknowledgements Funding was provided by National Institute of Food and Agriculture (Hatch HAW09025-H[1001478]), Agricultural Research Service (58-5320-4-012), and United States Agency for International Development (BFS-G-1100002 and EEM-G-0004-00013) to JH. Funding was also provided from National Science Foundation MCB-1818229 and United States Department of Agriculture: The NIFA Citrus Disease Research and Extension grant AP18PPQS&T00C223 program to AES.

Author contributions XW, AO-V, and MM performed experiments. XW, AO-V, AES, and MM performed data analysis. XW and AO-V wrote the manuscript. AO-V, JH, and MM designed experiments. JH and MM supervised the study. All authors contributed to the review and modification of the manuscript.

Declarations

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

Research involving human participants and/or animals This study did not include experiments with human or animal participants performed by any of the authors.

References

1. Lim S, Lee S-H, Moon JS (2019) Complete genome sequence of a tentative new umbravirus isolated from Patrinia scabiosaefolia. Arch Virol 164(9):2375–2378. https://doi.org/10.1007/s00705-019-04312-y
2. Yoo RH, Lee S-W, Lim S, Zhao F, Igori D, Baek D, Hong J-S, Lee S-H, Moon JS (2017) Complete genome analysis of a novel umbravirus-polerovirus combination isolated from Ixeridium dentatum. Adv Virol 162(12):3893–3897. https://doi.org/10.1007/s00705-017-3512-0
3. Ryabov EV, Taliansky ME, Robinson DJ, Waterhouse PM, Murant AF, de Zoeten GA, Falk BW, Vetten HJ, Gibbs MJ (2012) Umbravirus. In: King AM, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, London, pp 1191–1195
4. Quito-Avila DF, Alvarez R, Ibarra M, Martin R (2015) Detection and partial genome sequence of a new umbra-like virus of papaya discovered in Ecuador. Eur J Plant Pathol 143(1):199–204. https://doi.org/10.1007/s10658-015-0675-y
5. Sá Antunes TF, Amaral RJV, Ventura JA, Godinho MT, Amaral JG, Souza FO, Zerbini PA, Zerbini FM, Fernandes PMB (2016) The dsRNA virus papaya meleira virus and an ssRNA virus are associated with papaya sticky disease. PLos One 11(5):e0155240. https://doi.org/10.1371/journal.pone.0155240
6. Tahir MN, Bolus S, Grinstead SC, McFarlane SA, Mollov D (2021) A new virus of the family Tombusviridae infecting sugarcane. Adv Virol 166(3):961–965. https://doi.org/10.1007/s00705-020-04908-9
7. Felker P, Bunch R, Russo G, Preston K, Tine JA, Suter B, Xiaohan M, Cushman JC, Yim WC (2019) Biology and chemistry of an Umbravirus like 2989 bp single stranded RNA as a possible causal agent for Opuntia stunting disease (engrosamiento de cladodios)-A Review. J Prof Assoc Cactus Dev 21:1–31
8. Kwon S-J, Bodaghi S, Dang T, Gadhave KR, Ho T, Osman F, Al Rwahnih M, Tzanetakis IE, Simon AE, Vidalakis G (2021) Complete nucleotide sequence, genome organization and comparative genomic analyses of citrus yellow-vein associated virus (CYVaV). Front Microbiol 12:683130. https://doi.org/10.3389/fmicb.2021.683130
9. Preising S, Borges D, Ambrosio MMQ, da Silva WL (2020) A fig deal: A global look at fig mosaic disease and its putative associates. Plant Dis 105(4):727–738. https://doi.org/10.1094/PDIS-06-20-1352-FE
10. Villamor DEV, Ho T, Al Rwahnih M, Martin RR, Tzanetakis IE (2019) High throughput sequencing for plant virus detection and discovery. Phytopathology 109(5):716–725. https://doi.org/10.1094/PHYTO-07-18-0257-RVW
11. Perez-Losada M, Arenas M, Galan JC, Bracho MA, Hillung J, Garcia-Gonzalez N, Gonzalez-Candelas F (2020) High-throughput sequencing (HTS) for the analysis of viral populations. Infect Genet Evol 80:104208. https://doi.org/10.1016/j.meegid.2020.104208
12. Olmedo-Velarde A, Park AC, Sugano J, Uchida JY, Kawate M, Borth WB, Hu JS, Melzer MJ (2019) Characterization of Ti Ringspot-associated virus, a Novel Emaravirus associated with an emerging Ringspot disease of Cordyline Fruticosa. Plant Dis 103(9):2345–2352. https://doi.org/10.1094/PDIS-09-18-1513-RE
13. Zhang J, Dey KK, Lin B, Borth WB, Melzer MJ, Sether D, Wang Y, Wang I-C, Shen H, Pu X, Sun D, Hu JS (2017) Characterization of Canna yellow mottle virus in a New Host, Alpinia purpurata, in Hawaii. Phytopathol 107(6):791–799. https://doi.org/10.1094/PHYTO-04-16-0160-R
14. Wang Y, Wu B, Borth WB, Hamim I, Green JC, Melzer MJ, Hu JS (2017) Molecular characterization and distribution of two strains of Dasheen mosaic virus on Taro in Hawaii. Plant Dis 101(12):1980–1989. https://doi.org/10.1094/PDIS-04-17-0516-RE
15. Liu J, Carino E, Bera S, Gao F, May JP, Simon AE (2021) Structural Analysis and Whole Genome Mapping of a New Type of Plant Virus Subviral RNA: Umbravirus-Like Associated RNAs. Viruses 13(4):646. https://doi.org/10.3390/v13040646
16. Gao F, Simon AE (2016) Multiple cis-acting elements modulate programmed-1 ribosomal frameshifting in Pea enation mosaic virus. Nucleic Acids Res 44(2):878–895. https://doi.org/10.1093/nar/gkv1241
17. Stecher G, Tamura K, Kumar S (2020) Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evol 37(4):1237–1239. https://doi.org/10.1093/molbev/msz312
18. Chomič A, Pearson M, Clover G, Farreyrol K, Saul D, Hampton J, Armstrong K (2010) A generic RT-PCR assay for the detection of Luteoviridae. Plant Pathol 59(3):429–442. https://doi.org/10.1111/j.1365-3059.2010.02282
19. Kozak M, Shatkin AJ (1978) Identification of features in 5′ terminal fragments from reovirus mRNA which are important for ribosome binding. Cell 13(1):201–212. https://doi.org/10.1016/0092-8674(78)90150-2