



Two new umbravirus-like associated RNAs (ulaRNAs) discovered in maize and johnsongrass from Ecuador

Diego F. Quito-Avila^{1,2} · Edison G. Reyes-Proañó¹ · Alma Mendoza³ · Paolo Margaria⁴ · Wulf Menzel⁴ · Sayanta Bera⁵ · Anne E. Simon⁵

Received: 11 April 2022 / Accepted: 22 May 2022 / Published online: 12 July 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Two new umbravirus-like associated RNAs (ulaRNAs) were found, respectively, in maize and Johnsongrass samples from Ecuador. The complete sequences consist of 3,053 and 3,025 nucleotides, respectively, and contain four open reading frames (ORFs). Their genome sequences were 58% identical to each other and 28 to 60% identical to the most closely related viruses. Phylogenetic analysis using full genome sequences and amino acid sequence of the RNA-dependent-RNA polymerase (RdRp) placed both sequences in a clade sharing the most recent common ancestor with ulaRNAs from sugarcane and maize, suggesting that they belong to a monophyletic grass-infecting lineage. Their terminal regions exhibit features common to umbraviruses and ulaRNAs.

In the last decade, high-throughput sequencing (HTS) has enabled the discovery of a considerable number of plant viruses from different hosts [1, 2], contributing to understanding the evolutionary pathways of several taxonomic groups [3]. One of the groups for which several potential new members have been reported is the family *Tombusviridae*, comprising 18 genera of single-stranded positive-sense RNA viruses (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202005192). The genome organization of tombusviruses differs across genera, except for

open reading frame 2 (ORF2), the viral RNA-dependent-RNA-polymerase (RdRp), which is translated through ribosomal readthrough of ORF1 in most cases, or a -1 ribosomal frameshift (FS) in umbra- and dianthoviruses [4, 5]. A major genomic distinction is made for members of the genus *Umbravirus*, which lack the gene for the coat protein (CP) and therefore depend on coinfecting viruses, typically members of the genus *Polerovirus* in the family *Solemoviridae*, for genome encapsidation and plant-to-plant transmission by vectors [5].

In recent years, several viral RNAs sharing significant phylogenetic relationships with the RdRp of umbraviruses have been found in several plants. Although the absence of CP genes is a common characteristic of these viral RNAs, they have unique features that distinguish them from “true” umbraviruses. The term ‘umbra-like virus’ or ‘umbravirus-like associated RNAs (ulaRNAs)’ was coined to group these viral RNA entities [6–11]. Three classes of ulaRNAs have been categorized based on RdRp analysis and genomic secondary structure prediction [12]. Class 1 includes ulaRNAs that are ~ 4.5 kb in length with unusually long 3′ untranslated regions (UTRs). This class is typified by papaya virus Q (PpVQ), papaya meleira virus 2 (PMeV2), and babaco virus Q (BabVQ), which have been found in Ecuador (PpVQ and BabVQ) [13, 14], Brazil (PMeV2) [15], and Mexico (PMeV-Mx) [16]. Recently, an umbra-like virus, related to PpVQ and PMeV2, has been reported in papaya plantings from Australia [17]. Class 2 comprises smaller ulaRNAs

Handling Editor: Massimo Turina .

✉ Diego F. Quito-Avila
dquito@espol.edu.ec

¹ Facultad de Ciencias de la Vida, Escuela Superior Politécnica del Litoral, ESPOL, Km 30.5 Vía Perimetral Campus Gustavo Galindo, Guayaquil, Ecuador

² Centro de Investigaciones Biotecnológicas del Ecuador, CIBE, Escuela Superior Politécnica del Litoral, Km 30.5 Vía Perimetral Campus Gustavo Galindo, Guayaquil, Ecuador

³ INIAP, Estación Experimental Portoviejo, Km 12, Vía a Santa Ana, Manabí, Ecuador

⁴ Plant Virus Department, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7 B, Braunschweig, Germany

⁵ Department of Cell Biology and Molecular Genetics, University of Maryland College Park, College Park, MD 20742, USA

of ~ 2.7 to 3 kb that have been reported in opuntia (opuntia umbra-like virus, OULV), sugarcane (sugarcane umbra-like virus, SULV), fig (fig umbra-like virus, FULV), maize (Ethiopian maize associated virus, EMaV), and citrus (citrus yellow vein associated virus, CYVaV) [6, 8, 10, 11]. A recently proposed third class (class 3) is typified by strawberry virus A (StVA), a 3.2 kb-ulaRNA sharing a most recent ancestor with those in class 1 [9].

Here, we report the characteristics and complete genome sequences of two new class-2 ulaRNAs found in maize (*Zea mays*) and johnsongrass (*Sorghum halepense*). In August of 2021, leaf tissue samples showing mild-to-moderate mosaic were collected in Santa Ana, a representative maize production area in Manabí province of Ecuador (GPS coordinates: -1.123533, -80.414250). Samples were collected from two commercial cultivars, a yellow type 'Trueno' and a white type 'INIAP-543', and from johnsongrass, which was the most prevalent grass weed in the area at the time of sampling.

A virus discovery analysis was conducted by HTS on three total-RNA pools. Pooled samples (pool 1, yellow maize; pool 2, white corn; pool 3, johnsongrass) were a composite of 10 (pool 1) or six (pools 2 and 3) individual totalRNA preparations, mixed in equal amounts totaling 4 µg per sample. After pooling, aliquots of each RNA sample were stored individually at -80 °C for later analysis. Total RNA was extracted from ~100 mg of fresh leaf tissue using a PureLink™ RNA Mini Kit (Life Technologies). The three pooled samples were subjected to DNase treatment, depleted of the host ribosomal RNA fraction using an Illumina Ribo-Zero Plant Kit and subjected to library preparation using an Illumina Nextera XT DNA Library Prep Kit. The libraries were sequenced as paired-end reads (2 × 150 bp) on an Illumina NextSeq2000 instrument at the Leibniz Institute DSMZ. A total of 38.2, 64.8, and 42.1 million raw reads were obtained from RNA pools 1, 2, and 3, respectively.

Raw reads were analyzed in Geneious Prime v. 2022.0.1 (Biomatters) using a bioinformatics pipeline developed in house to subtract host sequences and to assemble contigs, which were screened by BLASTn and BLASTp against a virus reference database for virus discovery, reconstruction of virus genome sequences, and taxonomic assignment.

Bioinformatics analysis revealed the presence of several virus contigs in each sample, most of which corresponded to previously reported viruses belonging to different genera (Online resource 1). However, two contigs of 2,908 and 2,746 nt in length, obtained from pools 1 and 3, respectively, were distantly related to known ulaRNAs (NCBI BLAST analysis date: November 3, 2021). The closest hits included EMaV (accession no. MF415880), SULV (accession no. MN868593), FULV (accession no. MW480892-3), CYVaV (accession no. MT893741), OULV (accession no. MH579715), and strawberry virus A (StVA, accession no.

MK211273-5), with amino acid (aa) identity values in the range of 38–65% for the RdRp (35–64 % protein coverage).

The 2,908-nt-long contig (pool 1) was assembled from a total of 2,040 reads, with an average sequencing depth of 106x, whereas the 2,746 nt-long contig (pool 3) was constructed from 972 reads, with an average sequencing depth of 54x (Fig. 1A). Pairwise alignments between the two contigs showed 58% identity at the nucleotide level and 60.5% identity when the deduced RdRp aa sequences were compared, indicating that the sequences represented two distinct ulaRNAs. Reverse transcription (RT)-PCR was used to confirm the presence of each ulaRNA in the original RNA preparations. Primers were designed using the consensus sequence of each assembly from the region with the highest coverage (Fig. 1A). Amplicons of the expected size were detected in one RNA sample from each group (Online resource 2). The 5' and 3' ends of each contig were verified by rapid amplification of cDNA ends (RACE), using a 5'/3' RACE Kit, 2nd Generation (Roche, Germany) and specific primers designed based on the terminal genomic regions.

The complete genomic sequence of the ulaRNA assembled from the yellow maize sample consists of 3,053 nt (GenBank accession no. OM937759), whereas the one from johnsongrass consists of 3,025 nt (accession no. OM937760). For consistency in ulaRNA naming, we will refer to the new ulaRNA from maize as maize umbra-like virus (MULV) and the one from johnsongrass as johnsongrass umbra-like virus (JgULV).

The genomes of both viruses contain four ORFs organized in a similar manner, with minor variations in each ORF (Fig. 1A). ORF1 encodes a protein of 195 aa (22 kDa) for which no function was predicted. ORF2 is located after a stretch of 50 (MULV) or 170 (JgULV) nt downstream from ORF1. However, both contain the same heptameric ribosomal FS sequence (GGGUUUU), which is conserved in other class 2 ulaRNAs and in those of umbraviruses (consensus: GGAUUUU) (Fig. 1C). In addition, both MULV and JgULV can form structures similar to those of CYVaV in this region, including a hairpin that has the capacity for a tombusvirid-wide long-distance RNA:RNA interaction with a sequence near the 3' terminus (Fig. 1D). This strongly suggests that translation of ORF2 occurs via a -1 ribosomal FS. Interestingly, MULV and the previously identified EMaV have unique ORF1 termination codons (UAG) two codons upstream of the position of the termination codon found in all other class 2 ulaRNAs (UGA), including JgULV. Frameshifting would result in a fused protein of 717 aa (82.5 kDa) and 674 aa (76.5 kDa) for MULV and JgULV, respectively. The non-overlapping region of the fusion protein contains conserved viral RdRp domains (pfam clan number: CL0027).

Unlike class 2 dicot-infecting ulaRNAs, which have only a single ORF that partially overlaps with the end of the

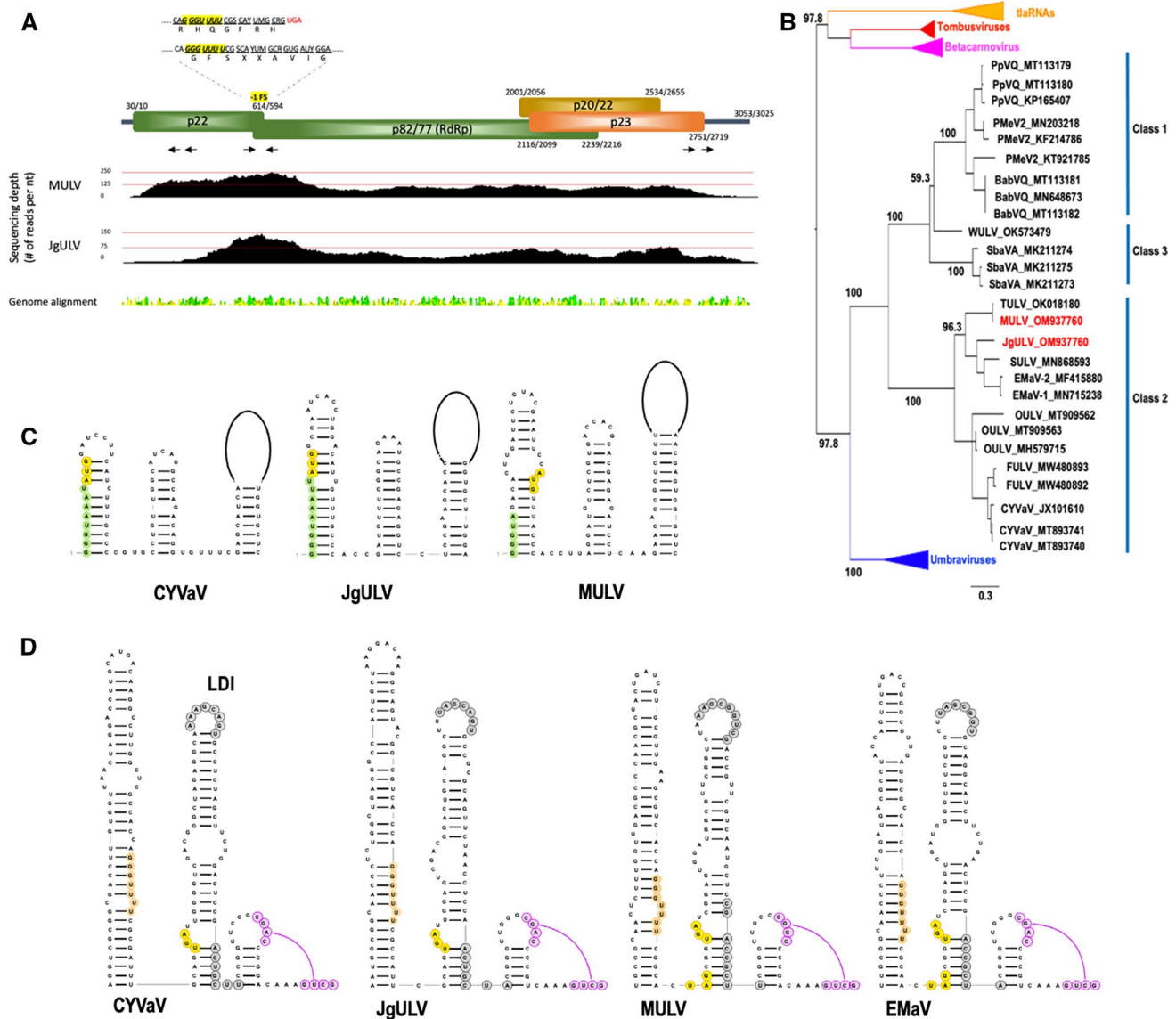


Fig. 1 Characterization of two new umbravirus-like associated RNAs (ulaRNAs). (A) Genome organization of the new ulaRNAs maize umbravirus-like virus (MULV) and johnsongrass umbravirus-like virus (JgULV), showing the hypothetical proteins. Nucleotide positions are shown for each open reading frame (ORF). The slash (/) differentiates nt positions or protein molecular weight (kDa) for MULV or JgULV (MULV/JgULV). The predicted ribosomal frameshift (FS) sequence is highlighted (yellow) at the end of ORF1, with the corresponding deduced amino acid translation at the FS site. Arrows indicate primer sites for detection and amplification of terminal sequences. Graphical representations of the high-throughput sequencing depth for each viral contig and pairwise alignment between MULV and JgULV (green = highly conserved, yellow = less conserved areas) are shown below. (B) Phylogenetic relationship of ulaRNAs to 24 representative umbraviruses, betacarmoviruses, and tomsbusviruses and six tomsbusvirus-like associated RNAs (tlaRNAs). The maximum-likelihood phylogenetic tree was constructed based on amino acid sequences of the RNA-dependent RNA polymerase (RdRp) using the WAG with frequency substitution model previously inferred using jModelTest in

MEGA X. Branch numbers indicate bootstrap support as a percentage of 1000 replicates. The scale bar denotes amino acid substitutions per site. The tree is rooted at the midpoint. The name of each ulaRNA isolate is followed by an underscore and its accession number. (C) Structures at the 5' end of citrus yellow vein associated virus (CYVaV) (determined by RNA structure probing) are also found in JgULV and MULV. Start codons are shaded yellow and the CCS (carmovirus consensus sequence) is shaded green. (D) Structures at the ribosome recoding site of CYVaV that are also found in JgULV, MULV, and all other class 2 ulaRNAs such as Ethiopian maize associated virus (EMaV). The ribosome slippery site is shown in orange, stop codons are in yellow, and an H-type pseudoknot (AES, unpublished) is in pink. Note that EMaV and MULV have an upstream UAG stop codon not found in any other class 2 ulaRNA. The sequence that participates in long-distance interaction (LDI) with the 3' end is shown in gray, and the interacting sequence is also found at the base of the hairpin (also in gray) and likely pairs with the terminal loop in an alternative structure (AES, manuscript in preparation).

RdRp ORF (absent in CYVaV because of two deletions), MULV and JgULV have two additional putative ORFs (ORFs 3 and 4) arranged in an out-of-frame overlapping configuration similar to those of umbraviruses but without the intervening intergenic region (Fig. 1A). The hypothetical protein encoded by ORF3 consists of 178 aa (20.4 kDa) and 200 aa (22.6 kDa) in MULV and JgULV, respectively, sharing 25% aa sequence identity. BLAST alignments did not reveal any homologues to this protein. The hypothetical product of ORF4 is a protein of 212 aa (23.6 kDa) and 207 aa (23 kDa), for MULV and JgULV, respectively, sharing 48% aa sequence identity, and 44–48% identity with the single ORF orthologs of 21–22 kDa from FULV, SULV, OULV, and EMaV. The recently reported wheat umbra-like virus (WULV), a new ulaRNA of 3.5 kb [18], has one ORF overlapping at the end of ORF2 and is suggested to have an additional ORF starting 48-nt apart from the termination codon of the previous ORF. However, this second ORF is in frame, with no intervening termination codons, and thus, its identity as a separate ORF requires further examination. Interestingly, SULV also contains a fourth ORF that partially overlaps with the class 2 orthologue, similar to MULV and JgULV.

Maximum-likelihood phylogenetic trees, constructed in MEGA X [19] using the complete nucleotide (Online resource 3) or amino acid sequences (Fig. 1B) of the RdRp, showed that MULV and JgULV form a clade with the class 2 ulaRNAs SULV and EMaV, suggesting a grass-infecting common ancestor for this lineage. A sister clade was formed by CYVaV, OULV, and FULV, within which CYVaV and FULV exhibit a closer relationship (Fig. 1B). Although demarcation criteria have not yet been established for ulaRNAs, the nucleotide and amino acid sequence identity values obtained when comparing MULV, JgULV, and their closest relatives strongly suggest that there are two distinct class 2 ulaRNA lineages.

The 5' UTR in JgULV is 9 nt in length, including a canonical "carmovirus consensus sequence (CCS; G₂₋₃A/U₄₋₉), found at the 5' ends of all carmoviruses and nearly all ulaRNAs and umbraviruses. MULV has an extended 5' UTR of 29 nt, which is unique among class 2 ulaRNAs, with the exception of FULV-1, which was reported to have a highly unusual 5' UTR that requires additional verification [8]. As with all class 2 ulaRNAs (with the exception of FULV-1), the 5' region of both new ulaRNAs contains two short terminal hairpins and an extended downstream third structure, according to secondary structure predictions for CYVaV using a combination of Selective 2' Hydroxyl Acylation analyzed by Primer Extension (SHAPE) structure probing, computational predictions, and phylogenetic analysis [20] (Fig. 1C).

MULV and JgULV have 306- and 302-nt 3' UTRs, respectively, similar to other class 2 ulaRNAs. The 3' regions

of CYVaV and other members of the family *Tombusviridae* have been studied extensively, and different step-loop structures have been shown to play key roles in replication and translation. Virtually all members of the family *Tombusviridae* have two 3'-terminal hairpins (designated as H5 and Pr for carmoviruses and umbraviruses) that are connected by a four-nucleotide pseudoknot that includes the 3'-terminal residues (Fig. 2) [20–22]. Many umbraviruses and carmoviruses contain two hairpins just upstream of H5 (designated as H4a and H4b), which, along with H5 and two pseudoknots, form a TSS-type 3' cap-independent translation enhancer (CITE) [10, 21, 23]. Most class 2 ulaRNAs, including MULV and JgULV, contain similarly placed hairpins but lack the capacity to form pseudoknots. In CYVaV, the 3' CITE was identified as a novel I-shaped structure (ISS)-like structure (ISSLS), with several critical stretches of perfectly conserved class 2 residues (Fig. 2, green with orange circles) that are also conserved in MULV and JgULV. Several regions of additional conservation among MULV, JgULV, and EMaV were also evident, especially in a lower supporting stem. Our findings evidence the diversity in genomic sequence, size, and organization of ulaRNAs, anticipating the existence of new classes of these RNA entities.

Lastly, an important biological feature of "true" umbraviruses is their association with a capsid-assistor virus, typically a polerovirus, for genome encapsidation and plant-to-plant transmission by vectors [5]. Poleroviruses have been found incidentally (e.g., no formal experiments have been conducted to demonstrate their capsid-lender nature) for SULV, OULV, CYVaV, and StVA [6, 9–11]. For the papaya-infecting ulaRNAs, an unusual dsRNA totivirus-like virus has been shown to be the capsid assistor of PMeV-2 [15] (Quito-Avila, unpublished). In this study, we found the polerovirus maize yellow dwarf virus (MYDV) in samples from the three RNA pools. However, MYDV was not detected in the two samples in which MULV and JgULV were found. A possible explanation could be that the respective host cannot be systemically infected by the helper virus, while class 2 ulaRNAs are capable of independent systemic movement, which likely involves the use of host movement proteins (Liu et al., manuscript submitted). Further studies are needed to determine the natural transmission of MULV and JgULV and their potential involvement in disease.

It should be noted that, at the time this manuscript was being prepared, a nucleotide sequence recorded as Teosinte-associated umbra-like virus (TULV) (accession no. OK018180) from Mexico became available in the NCBI GenBank database. The TULV sequence shares 99% nt sequence identity with MULV but is missing 5' terminal residues and has additional sequence beyond the 3' end sequence conserved with all other class 2 ulaRNAs. We propose that TULV represents a Mexican isolate of MULV. No formal publications about the discovery of TULV or

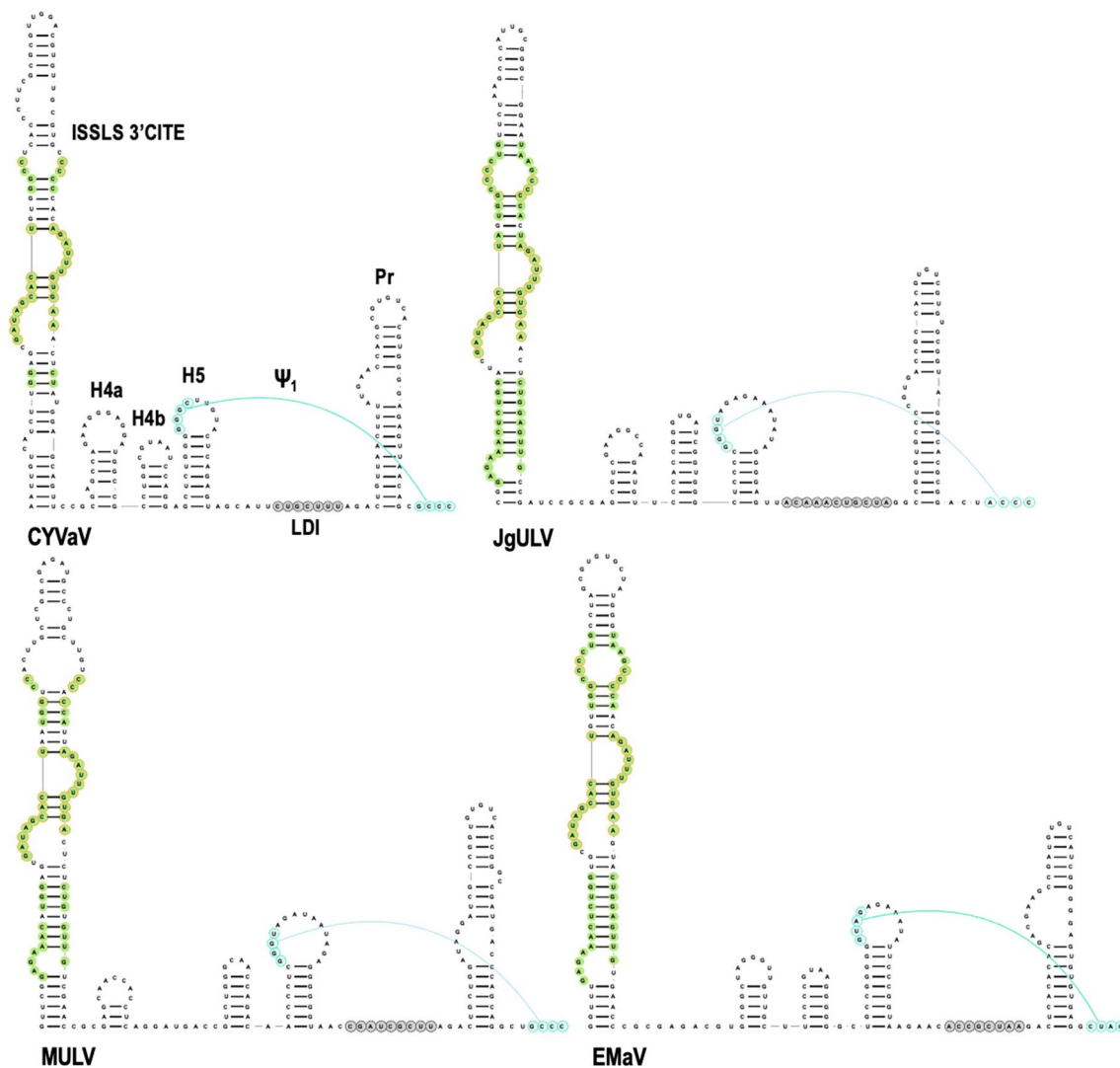


Fig. 2 3'-proximal structures of CYVaV, JgULV, MULV, and EMaV. Structures at the 3' end of CYVaV were determined by RNA structure probing. Names of hairpins are as found in Liu et al. [12, 20]. Pseudoknot 1 (Ψ_1)-connecting residues are shown in blue. The sequence that participates in the long-distance interaction (LDI) with the recod-

ing site is shaded gray. The 3' cap-independent translation enhancer (CITE) and the I-shaped structure (ISS)-like structure (the ISSLS), are labeled. Sequences in the 3' CITE that are shared by all class 2 ulaRNAs are shaded green and circled. Other residues that are conserved between these ulaRNA 3' CITEs are shaded green.

its molecular characterization were available at the time of submission.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00705-022-05525-4>.

Acknowledgements The contribution of Viola Weichert and Dennis Knierim (Leibniz Institute DSMZ) in library preparation and bioinformatics analysis is deeply acknowledged.

Author contributions Study conception and design: DFQ-A. Material preparation and data collection were performed by EGR-P and AM. HTS and bioinformatics analyses were performed under the supervision of PM and WM. Sequence and phylogenetic analyses were performed by DFQ-A, SB, and AES. The first draft of the manuscript was written

by DFQ-A and AES; all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Funding was received from the European Virus Archive Global (EVAg) project, through the European Union's Horizon 2020 research and innovation program under grant agreement no. 871029. Additional funding was received from the National Science Foundation EAGER 1912025 to AES, the National Science Foundation MCB-1818229 to AES, and the USDA NIFA Emergency Citrus Disease Research and Extension Program 2020-08455 to AES.

Data availability The genomic sequences of the two viruses reported here have been deposited in the GenBank database under the accession numbers OM937759 and OM937760.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors. Plant samples were collected under Genetic Resource Access Permit # MAE-DNB-CM-2018-0098 granted by the Department of Biodiversity of the Ecuadorean Ministry of the Environment.

References

- Villamor DEV et al (2019) High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109(5):716
- Minicka J et al (2020) High-throughput sequencing facilitates discovery of new plant viruses in Poland. *Plants* 9(7):820
- Simmonds P et al (2017) Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 15(3):161–168
- Lommel SA, Sit TL (2008) Tombusviruses. In: Mahy BWJ, Van Regenmortel MHV (eds) *Encyclopedia of virology*, 3rd edn. Academic Press, Oxford, pp 145–151
- Ryabov EV, Taliansky ME (2020) Umbraviruses (Calusvirinae, Tombusviridae). Reference module in life sciences. Elsevier, Amsterdam
- Tahir MN et al. (2021) A new virus of the family Tombusviridae infecting sugarcane. *Arch Virol* 166(3):961–965. <https://doi.org/10.1007/s00705-020-04908-9>
- Cornejo-Franco JF et al (2021) An umbra-related virus found in babaco (*Vasconcellea × heilbornii*). *Arch Virol* 166(8):2321–2324
- Wang X et al (2021) Genome characterization of fig umbra-like virus. *Virus Genes* 57(6):566–570
- Koloniuk I et al (2022) Identification and characterization of a novel umbra-like virus, strawberry virus A, infecting strawberry plants. *Plants* 11(5):643
- Kwon S-J et al (2021) Complete nucleotide sequence, genome organization, and comparative genomic analyses of citrus yellow-vein associated virus (CYVaV). *Front Microbiol* 12:1371
- Felker PBR, Russo G, Preston K, Tine JA, Suter B, Xiao-han 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* 21:31
- Liu J et al (2021) Structural analysis and whole genome mapping of a new type of plant virus subviral RNA: Umbravirus-like associated RNAs. *Viruses* 13(4):646
- Quito-Avila DF et al (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
- Cornejo-Franco JF et al (2020) Exploring the virome of *Vasconcellea × heilbornii*: the first step towards a sustainable production program for babaco in Ecuador. *Eur J Plant Pathol* 157(4):961–968
- Sá Antunes TF et al (2016) The dsRNA virus Papaya Meleira virus and an ssRNA virus are associated with papaya sticky disease. *PLoS ONE* 11(5):e0155240
- Zamudio-Moreno E et al (2015) Early diagnosis of a Mexican variant of Papaya meleira virus (PMeV-Mx) by RT-PCR. *Genet Mol Res* 14(1):1145–1154
- Campbell P (2018) New test to offer early detection of papaya sticky disease. Papaya press, p. 1. [https://australianpapaya.com.au/website/wp-content/uploads/2018/05/PAPAYAPRESS-MAY.pdf\(1\)](https://australianpapaya.com.au/website/wp-content/uploads/2018/05/PAPAYAPRESS-MAY.pdf(1))
- Redila CD, Prakash V, Nouri S (2021) Metagenomics analysis of the wheat virome identifies novel plant and fungal-associated viral sequences. *Viruses* 13(12):2457. <https://doi.org/10.3390/v13122457>
- Kumar S et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35(6):1547–1549
- Liu J, Simon AE (2022) Identification of Novel 5' and 3' translation enhancers in Umbravirus-like coat protein-deficient RNA replicons. *J Virol* 96(7):e0173621. <https://doi.org/10.1128/jvi.01736-21>.
- Simon AE (2015) 3'UTRs of carmoviruses. *Virus Res* 206:27–36
- Simon AE, Miller WA (2013) 3' cap-independent translation enhancers of plant viruses. *Annu Rev Microbiol* 67:21–42
- McCormack JC et al (2008) Structural domains within the 3' untranslated region of Turnip crinkle virus. *J Virol* 82(17):8706–8720

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