

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

diego fernando quito-avila (✉ dquito@espol.edu.ec)

Escuela Superior Politecnica del Litoral <https://orcid.org/0000-0002-6170-1113>

Edison Reyes-Proaño

Escuela Superior Politécnica del Litoral: Escuela Superior Politecnica del Litoral

Alma Mendoza

INIAP: Instituto Nacional Autonomo de Investigaciones Agropecuarias

Paolo Margaria

DSMZ: Leibniz Institut - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH

Wulf Menzel

DSMZ: Leibniz Institut - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH

Sayanta Bera

University of Maryland

Anne E Simmon

University of Maryland

Research Article

Keywords: capsidless, Zea mays, Sorghum halepense, Tombusviridae, Umbravirus, MULV, JgULV, ulaRNA

Posted Date: April 26th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1547579/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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, organized in four open reading frames (ORFs). Genome comparisons showed 58% identity between each other, whereas identities to the closest related viruses ranged from 28 to 60%. Phylogenetic inferences using the full genome, or the amino acid sequence of the RNA-dependent-RNA-polymerase (RdRp) placed both sequences in a clade sharing a most recent ancestor with ulaRNAs from sugarcane and maize, suggesting a monophyletic grass-infecting lineage. Terminal regions exhibit features common to umbraviruses and ulaRNAs.

Full Text

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 genes for coat protein (CP) and therefore depend on co-infecting viruses, typically members of the family that previously was called *Luteoviridae*, 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 analyses and predicted genomic secondary structures [12]. Class 1 includes ulaRNAs of ~ 4.5 kb in length with unusually long 3′-untranslated regions (UTRs). This class is typified by papaya virus Q (PpVQ) and papaya meleira virus-2 (PMeV2), which have been found in Ecuador, Brazil, Mexico and most recently Australia, and babaco virus Q (BabVQ) reported from Ecuador [7, 13, 14]. Class 2 comprises smaller ulaRNAs of ~ 2.7 to 3 kb reported from 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 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 and characterize the complete sequence of two new class-2 ulaRNAs found in maize (*Zea mays*) and Johnsongrass (*Sorghum halepense*). Leaf tissue showing mild-to-moderate mosaic were collected in Santa Ana, a representative maize production area in the 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, and pool 3: Johnsongrass) were a composite of ten (pool 1) or six (pools 2 and 3) individual total RNA preparations, mixed in equal amounts to a final concentration of 4 µg/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 the PureLink RNA extraction Kit (Life Technologies). The three pooled samples were subjected to DNase treatment, depleted of the host ribosomal-RNA fraction and subjected to library preparation according to the Illumina Nextera XT DNA Library Prep Kit. The libraries were sequenced as paired-end reads (2 x 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 assemble contigs, which were screened by BLASTn and BLASTp against a virus reference database for virus discovery, reconstruction of virus genomes and taxonomic assignment.

Bioinformatics analyses revealed the presence of several virus contigs from 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 3rd of 2021). Closest hits included EMaV (Acc. No. MF415880), SULV (Acc. No. MN868593), FULV (Acc. No. MW480892-3), CYVaV (Acc. No. MT893741), OULV (Acc. No. MH579715) and strawberry virus A (StVA, Acc. No. MK211273-5), with amino acid (aa) identities in the range 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, giving 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 validate 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 specific primers designed on terminal genomic regions.

The complete genomic sequence of the ulaRNA assembled from the yellow maize sample consists of 3,053 nt (GenBank Acc. No. OM937759), whereas the one from Johnsongrass consists of 3,025 nt (OM937760). For consistency in ulaRNA naming, we will refer to the new ulaRNA from maize as maize umbra-like virus (MULV) and 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 nt positions of 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 from ORF1. However, both contain the same heptameric ribosomal FS sequence (GGGUUUU), which is conserved with other Class 2 ulaRNAs and 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 contains the capacity for a tombusvirid-wide long-distance RNA:RNA interaction with sequence near the 3' terminus (Fig 1D). This strongly suggests that translation of ORF2 is *via* a -1 ribosomal FS. Interestingly, MULV and previously identified EMaV have unique ORF1 termination codons (UAG) two codons upstream 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 that have only a single ORF that partially overlaps with the end of the 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 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 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 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 [15], 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.

Phylogenetic analyses using both the complete genome and the amino acid sequence of the RdRp showed that MULV and JgULV form a clade with 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, nucleotide and amino acid sequence identities between MULV, JgULV and their closest relatives, strongly suggest there are two distinct Class 2 ulaRNA lineages. Interestingly, WULV

forms a clade with StVA, supporting the existence of ulaRNA Class 3 [12], however their additional ORFs are unrelated.

The 5' UTR in JgULV is 9 nt, 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 contain two short terminal hairpins and an extended downstream third structure (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 *Tombusviridae* have been extensively studied, and different step-loop structures have been shown to play key roles in replication and translation. Virtually, all members of the *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) [16-18]. 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, 17, 19]. Most Class 2 ulaRNAs, including MULV and JgULV, contain similarly placed hairpins but lack the capacity to form the 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 member of the *Luteoviridae*, for genome encapsidation and plant-to-plant transmission by vectors [5]. Luteovirids have been incidentally found (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 proven to be the capsid assistor of PMeV-2 [14] (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 where 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) (Acc. No. OK018180) from Mexico became available in the NCBI GenBank. The TULV sequence shares 99% nt 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 its molecular characterization exist at the time of submission.

Declarations

Acknowledgements

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

Compliance with ethical standards

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.

Funding

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

Conflict of interest declaration

The authors declare no conflict of interest.

Author contributions

Study conception and design: D.F. Quito-Avila. Material preparation and data collection were performed by E.G. Reyes-Proañó and A. Mendoza. HTS and bioinformatics analyses were performed under the supervision of P. Margaria and W. Menzel. Sequence and phylogenetic analyses were performed by D.F. Quito-Avila, S. Bera and A.E. Simon. The first draft of the manuscript was written by D.F. Quito-Avila and A.E. Simon; all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

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

References

1. Villamor DEV et al (2019) High Throughput Sequencing For Plant Virus Detection and Discovery. *Phytopathology*® 109(5):716–725
2. Minicka J et al (2020) High-Throughput Sequencing Facilitates Discovery of New Plant Viruses in Poland. *Plants* 9(7):820
3. Simmonds P et al (2017) Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 15(3):161–168
4. Lommel SA, Sit TL (2008) *Tombusviruses*, in *Encyclopedia of Virology (Third Edition)*, B.W.J. Mahy and M.H.V. Van Regenmortel, Editors. Academic Press: Oxford. p. 145–151
5. Ryabov EV, Taliansky ME (2020) *Umbraviruses (Calusvirinae, Tombusviridae)*, in *Reference Module in Life Sciences*. Elsevier
6. Tahir MN et al (2021) A new virus of the family Tombusviridae infecting sugarcane. *Archives of Virology*
7. Cornejo-Franco JF et al (2021) An umbra-related virus found in babaco (*Vasconcellea × heilbornii*). *Arch Virol* 166(8):2321–2324
8. Wang X et al (2021) Genome characterization of fig umbra-like virus. *Virus Genes* 57(6):566–570
9. Koloniuk I et al (2022) Identification and Characterization of a Novel Umbra-like Virus, Strawberry Virus A, Infecting Strawberry Plants. *Plants* 11(5):643
10. Kwon S-J et al (2021) *Complete Nucleotide Sequence, Genome Organization, and Comparative Genomic Analyses of Citrus Yellow-Vein Associated Virus (CYVaV)*. *Frontiers in Microbiology*, **12**
11. Felker P, Russo BR, Preston G, Tine K, Suter JA, Xiao-han B, Cushman M, Yim JC (2019) *WC, 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): p.31
12. 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
13. 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
14. 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
15. Redila CD, Prakash V, Nouri S (2021) *Metagenomics Analysis of the Wheat Virome Identifies Novel Plant and Fungal-Associated Viral Sequences*. *Viruses*, **13**(12)
16. Liu J, Simon AE (2022) *Identification of Novel 5' and 3' Translation Enhancers in Umbravirus-Like Coat Protein-Deficient RNA Replicons*. *J Virol*, : p.e0173621
17. Simon AE (2015) 3'UTRs of carmoviruses. *Virus Res* 206:27–36

18. Simon AE, Miller WA (2013) 3' cap-independent translation enhancers of plant viruses. *Annu Rev Microbiol* 67:21–42
19. McCormack JC et al (2008) Structural domains within the 3' untranslated region of Turnip crinkle virus. *J Virol* 82(17):8706–8720

Figures

Figure 1

Characterization of two new umbravirus like associated RNAs (ulaRNAs). A. Genome organization of new ulaRNAs maize umbra-like virus (MULV) and Johnsongrass umbra-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 primers 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 with 24 representative viral species belonging to umbravirus, betacarmovirus and, tombusvirus genera, and six tombusvirus like associated RNAs (tlaRNAs). Maximum likelihood phylogenetic trees were constructed based on the amino acid sequences of RdRp. Branch numbers indicate bootstrap support in the percentage out of 1000 replicates. The scale bar denotes nucleotide/protein substitutions per site. The tree was mid-point rooted. 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 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). Ribosome slippery site is 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. Sequence that participates in a long-distance interaction (LDI) with the 3' end is 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).

Figure 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., 2021,

2022. Pseudoknot 1 (Ψ_1)-connecting residues are in blue. The sequence that participates in the long-distance interaction (LDI) with the recoding site is shaded gray. The 3' cap-independent translation enhancer (CITE), I-shaped structure (ISS)-like structure (the ISSLS), is labeled. Sequences in the 3' CITE that are shared by all Class 2 ulaRNAs are shaded green and circled. Other conserved residues between these ulaRNA 3' CITEs are shaded green.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SuppTable1.docx](#)
- [SuppTable2.docx](#)
- [ulaRNAsubmittedsequences.txt](#)