Title: **Molecular Identification and Characterization of Two Rubber Dandelion Amalgaviruses**

Humberto Debat[1*], Zinan Luo[2], Brian J. Iaffaldano[2], Xiaofeng Zhuang[2] and Katrina Cornish[2]

[1] Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), 11 de setiembre 4755, Córdoba, Argentina, X5020ICA.

[2] Department of Horticulture and Crop Science, The Ohio State University, Ohio Agricultural Research and Development Center, Williams Hall, 1680 Madison Avenue, Wooster, OH 44691.

* Corresponding author: Humberto Debat, debat.humberto@inta.gob.ar, Tel: +54 9 351 4973636, Fax: +54 9 351 4974330

Novel Amalgaviruses of Rubber Dandelion

Summary

The *Amalgaviridae* family is composed of persistent viruses that share the genome architecture of *Totiviridae* and gene evolutionary resemblance to *Partitiviridae*. A single *Amalgavirus* genus has been assigned to this family, presenting only four recognized species, corresponding to plant infecting viruses with dsRNA monopartite genomes of ca. 3.4 kb. Here, we present the genomic identification and characterization of two novel viruses detected in rubber dandelion (*Taraxacum kok-saghyz*). The sequenced isolates presented genomes of 3,409 and 3,413 nt long, including two partially overlapping ORFs encoding a putative coat protein and an RNA-dependent RNA polymerase (RdRP). Phylogenetic insights based on the detected virus sequences suggest them to be members of two new species within the *Amalgavirus* genus. Multiple independent RNAseq data suggest that the identified viruses have a dynamic distribution and low relative RNA levels in infected plants. Virus presence was not associated with any apparent symptoms on the plant hosts. We propose the names rubber dandelion latent virus 1 & 2 to the detected amalgaviruses; the first viruses to be associated to this emergent and sustainable natural rubber crop.

Keywords

Amalgavirus, Rubber dandelion, dsRNA virus, deep sequencing, virus discovery

Introduction

Natural rubber is an essential material to the manufacture of 50,000 different rubber and latex products. A steadily increasing demand cannot be met only by the industrial exploitation of the rubber tree (*Hevea brasiliensis*). Viable alternative crops that could be established may supplement the demand, with carbon footprint savings, which is currently supported by diverse synthetic rubbers (Cornish, 2017). Rubber dandelion (*Taraxacum kok-saghyz*) is currently being developed as a sustainable source of natural rubber. Thus, a robust metabolomic, genomic, and transcriptomic characterization should advance in parallel to explore the biological landscape of this important natural resource (Zhang *et al*., 2017). An additional aspect of this process is the exploration of potential microbes that could be associated with rubber dandelion. There are studies reporting the incidence of a virus in the related common dandelion (*T. officinale*), the *Dandelion yellow mosaic virus* (*Secoviridae; Sequivirus* – Bos *et al*., 1983). In addition, *T. officinale* has been described to be a reservoir host of important viral disease agents such as *Tomato ringspot virus* (*Secoviridae; Nepovirus* – Mountain *et al*., 1983), and *Tomato spotted wilt orthotospovirus* (*Tospoviridae; Orthotospovirus* – Groves *et al*., 2002). Interestingly, there are no reports describing viruses associated with rubber dandelion. Several members of a relatively new family of plant viruses have been identified in the last decade (Sabanadzovic *et al*., 2009). *Amalgaviridae* viruses are conformed by a dsRNA genome, containing two overlapping ORFs and an apparent common evolutionary history. Amalgaviruses are persistent, appear to be cryptic, share the genome architecture of *Totiviridae* and gene evolutionary resemblance to *Partitiviridae* (Martin *et al*., 2011). In this report, we present the characterization of two novel tentative members of the *Amalgaviridae* family and the *Amalgavirus* genus, the first viruses associated to *T. kok-saghyz*.

Materials and Methods

The first *T. kok-saghyz* transcriptome and associated reads as described by (Luo *et al*., 2017) were used as input for virus discovery. This transcriptome was produced from total RNA extracted from of 6 month old root samples of *T. kok-saghyz* of individuals of 6 different genotypes characterized by different rubber yields at The Ohio State University, and sequenced by Illumina Hiseq2000, obtaining 357,694,286 paired-end reads 100bp reads (GenBank accession numbers: genotype TK6 (SRR5181667); TK9 (SRR5181665); TK10 (SRR5181664); TK14 (SRR5181663); TK18 (SRR5181662); TK21 (SRR5181661)). The sequenced reads were quality evaluated using the FASTX-Toolkit, with a cut-off

score of 30 (-q). The filtered reads then went through Trinity *de novo* assembly (version 2.2.0) using standard parameters. The filtered and normalized assembly (NCBI Transcriptome Shotgun Assembly (TSA) accession number GFJE00000000) of 55,532 transcripts was assessed by bulk searches using as query the complete virus refseq database available at ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/ in a local server. BLASTX with an expected value of 10e-5 was employed as threshold, and hits were explored by hand. Tentative virus contigs were curated by iterative mapping of reads using Bowtie2 http://bowtie-bio.sourceforge.net/bowtie2/index.shtml virus Fragments Per Kilobase of transcript per Million mapped reads (FPKM) were estimated with Cufflinks 2.2.1 http://cole-trapnell-lab.github.io/cufflinks/releases/v2.2.1/. Virus ORFs were predicted by ORFfinder https://www.ncbi.nlm.nih.gov/orffinder/ translated gene products were assessed by InterPro https://www.ebi.ac.uk/interpro/search/sequence-search and NCBI Conserved domain database v3.16 https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi to predict domain presence and architecture. The 3D structure of the putative coat proteins was determined with the EMBOSS 6.5.7 Tool Garnier http://www.bioinformatics.nl/cgi-bin/emboss/garnier and coiled coil regions were predicted with COILS https://embnet.vital-it.ch/software/COILS_form.html using a MTIDK matrix. Predicted protein similarity plots were generated with Circoletto http://tools.bat.infspire.org/circoletto/ setting as E-value 10e-10. Phylogenetic insights based in predicted virus proteins were generated by MAFTT 7 https://mafft.cbrc.jp/alignment/software/ multiple amino acid alignments and FastTree 2.1.5 maximum likelihood phylogenetic trees computing local support values with the Shimodaira-Hasegawa test http://www.microbesonline.org/fasttree/. The FreeBayes v0.9.18S tool with standard parameters was employed for SNPs prediction https://github.com/ekg/freebayes. Results were integrated and visualized in the Geneious 8.1.9 platform (Biomatters Ltd.).

Results

The first publically-available RNA-Seq based *T. kok-saghyz* transcriptome, which was developed from pools of roots of genotypes with high and low rubber yields (Luo *et al*., 2017), was subjected to bulk BLASTX-NCBI searches using as query the complete virus refseq database. Interestingly, two transcripts presented consistent sequence identity to the *Amalgavirus Southern tomato virus* (Sabanadzovic *et al*., 2009) (50% identity at the aa level; E-value = 0.0) and *Blueberry latent virus* (49% identity at the aa level; E-value = 0.0). The corresponding transcripts were curated by iterative mapping of RNA reads,

which gave a mean coverage support of 49.1 X and 77.7 X, respectively. The curated 3,409 nt and 3,413 nt long sequences were further explored in detail and designated tentatively isolate OH of rubber dandelion latent virus 1 & 2 (RdLV1 & RdLV2).

The RdLV1 determined genome presents a 143 nt 5´UTR, a 97 nt 3´UTR, and two partially overlapping ORFs on the positive strand (Figure 1.A). The predicted ORF1 encodes a 387 aa putative Coat protein (CP). The overlapping ORF2 encodes an 825 aa RdRP with a corresponding RNA_dep_RNAP domain (Pfam: pfam00680, E-value = 1.30e-07) at the 360-544 aa coordinates. Genome position 981 (_$) presents a putative "slippery" sequence of the form ACU_UUU_CGC suggesting a host ribosomal +1 frameshift signal that could induce the generation of a characteristic 1,055 aa, 120 kDa fusion protein (Figure 1.B). This slippery sequence is identical to the reported frameshifting signal of the *Amalgavirus Rhododendron virus A* (Sabanadzovic *et al*., 2010). The RdLV2 genome presents a 171 nt 5´UTR and a 100 nt 3´UTR (Figure 1.A). The predicted ORF1 encodes a 377 aa putative CP. The overlapping ORF2 encodes a 749 aa RdRP with a RNA_dep_RNAP domain (pfam00680, E-value = 2.99e-10) at the 304-473 aa coordinates. Genome position 946 (_$) presents a putative "slippery" sequence CAG_UUU_CGU that could induce the generation of a 1,046 aa, 118 kDa fusion protein (Figure 1.B). The UTR regions of RdLV1 & RdLV2 were A+U rich, as described for amalgaviruses (Sabanadzovic *et al*., 2009), ranging from 53.1 % in the RdLV1 5´UTR to 61 % in the 3´UTR of RdLV2. The putative CP of RdLV1 & RdLV2 were subjected to 3D structure prediction with the EMBOSS 6.5.7 Tool Garnier and coiled coil determination by COILS with a MTIDK matrix. A comparison of these predictions to that of reported *Amalgavirus* members (Figure 1.C) suggests that RdLV1 & RdLV2 present a typical α-helical central region with high probability of coiled coil as part of its tertiary structure, as is prevalent in *Amalgaviridae* (Nibert *et al*., 2016). It is worth mentioning that the predicted forms of potential slippery sequences of RdLV1 & RdLV2 are of the general form UUU_CGN, similar to the experimentally validated sequence of influenza A virus (Firth *et al*., 2012). Theoretically, the ribosome may stall on a slippery sequence, making a pause at a rare codon (such as CGN = R) for which scarce tRNAs might be available. This pause may lead to a movement forward of one nucleotide. Translation resolves on the advanced ribosome in the +1 frame (Figure 1.B). This phenomenon has been predicted to be widespread among most plant amalgaviruses (Nibert *et al*., 2016). RdLV1 & RdLV2 share a 55.9 % genome nt identity and a 49.5 % aa pairwise identity between their predicted RdRPs. Their proposed assignment as separate species is consistent with the species demarcation criteria for the genus *Amalgavirus* proposed by the International

Committee on Taxonomy of Viruses (ICTV), which specifies an amino acid sequence divergence of over 25% at the RdRPs. The structural highlights of RdLV1 & RdLV2 were compared to the ICTV recognized *Amalgavirus* species (Table 1). The predicted genome lengths and architectures, ORFs, UTRs, gene products, protein sizes, and general viral sequence cues are consistent with the proposed assignment of RdLV1 & RdLV2 to the *Amalgavirus* genus. The predicted RdRP of RdLV1 & RdLV2 were employed to glimpse some evolutionary insights of the identified viruses. Maximum-likelihood phylogenetic trees of RdLV1 & RdLV2, and reported amalgaviruses, in the context of related viral families were generated, based on MAFTT protein alignments. The resulting trees evidently place RdLV1 & RdLV2 in a cluster of amalgaviruses, and more distantly associated to new unclassified viruses and members of the *Partitiviridae* and *Totiviridae* families (Figure 2.A). The complete fusion protein (FP) of RdLV1 & RdLV2 was explored in sequence similarity among recognized *Amalgavirus* species (Figure 2.B), and with closely related species (Figure 2.C) using the Circoletto tool (Darzentas, 2010), highlighting a stronger and broader link among the FP of RdLV1 & RdLV2 and reported amalgaviruses. Interestingly, sequence identity robustly falls beyond the *Amalgavirus* genus. Further similarity with a species proposed to be a member of a new genus of fungi derived *Amalgaviridae*, the *Zygosaccharomyces bailii virus Z* (proposed genus Zibavirus - ZbvZ) (Depierreux *et al*., 2016), is consistently low, supporting that both RdLV1 & RdLV2 could be members of the *Amalgavirus* genus.

To confirm the presence of the identified viruses and explore their preliminary prevalence, we investigated six independent root total RNA samples of *T. kok-saghyz* which were individually sequenced by Illumina Hiseq2000 generating over 291 million 100 bp pair end reads, ranging between 5.2 Gb to 6.7 Gb per sample. Interestingly, the presence of the cognate viruses was confirmed in five of the six samples by iterative mapping of sequencing reads to the reference transcripts of RdLV1 & RdLV2 (Figure 2.D). Virus relative RNA levels varied among samples, ranging from 3.69 FPKM for RdLV1 in TK-R14, to 12.11 FPKM for RdLV2 in TK-R6. In addition, in the TK-R18 sample, only RdLV2 was found, and both viruses were absent in TK-R21, suggesting that RdLV presence is dynamic and that mixed infections, whilst common, are not necessary. *De novo* assembly of the raw RNA data and further identification of RdLV isolates on the diverse samples were carried out in order to address preliminary virus diversity. Sequence variants among samples were reduced, presenting a high degree of homogeneity. Overall identity among individuals ranged from 98.3% to 99.4%, which was roughly equivalent to the observed intra-individual identity which ranged between 99.2% and 99.5%. A consistent identity among isolates

was reported for *Blueberry latent virus*, when 35 diverse cultivars were assessed and over 99% sequence identity among isolates was observed (Martin *et al*., 2011). Additionally, SNP were predicted by implementing the FreeBayes tool (Garrinson & Marth, 2012), and 259 variants were identified among the CDS of RdLV1 & RdLV2 (Figure 2.E); 78.37% of the polymorphisms involved the 3rd position of the predicted codon, suggesting a tentative constraint to avoid amino acid changes and thus maintain structure and functional domains of the respective viruses.

Discussion

Recurrent attempts to transmit *Amalgavirus* via grafting and mechanical inoculation have failed. In addition, *Amalgavirus* are very efficiently transmitted vertically via seed (70–90%), and have been associated with symptomless infections in their respective hosts (Sabanadzovic *et al*., 2010). The latter is consistent with our observations on tested rubber dandelions, which could not be linked with symptoms or altered phenotypes. Future studies should explore whether RdLV1 & RdLV2 share the biological properties of persistence and exclude potential horizontal transmission. To our knowledge, there are no reports of interspecific transmission of amalgaviruses, and transmission by potential vectors has not been conclusively ruled out. Even though there are only four species of *Amalgavirus* species recognized by the ICTV, recent reports suggest that the diversity of this family of viruses is much more complex and widespread among plants (Nibert *et al*., 2016). The discovery of potentially cryptic viruses has been hampered by the targeted study of symptomatic organisms which lead to the biased discovery of pathogenic viruses (Geoghegan & Holmes, 2017). Next generation sequencing is unraveling a new multifaceted virosphere paradigm, were viruses are widespread and associated to every organism (Greninger, 2017). The identified RdLV1 & RdLV2 correspond to the first viruses associated with *T. kok-saghyz*. The molecular characterization of these prospective members of the *Amalgaviridae* family is a first step on the path to advance the understanding of the intriguing biology of these potential endophytes and their economically important plant host.

-Nucleotide sequence accession number: The genome sequences of Rubber dandelion latent virus 1 & 2 have been deposited in NCBI GenBank under accession no MF197380 and MF197379

Acknowledgments

This work was supported by the Center for Applied Plant Sciences (CAPS), and the College of Food Agricultural and Environmental Sciences, The Ohio State University, and USDA National Institute of Food and Agriculture (Hatch project 230837).

References

Bos L, Huijberts N, Huttinga H, Maat D.Z. (1983) Further characterization of dandelion yellow mosaic virus from lettuce and dandelion. *European Journal of Plant Pathology*, **89**, 207-222.

Cornish K. (2017) Alternative Natural Rubber Crops: Why Should We Care? *Technology Innovation*, **18**, 244-255.

Darzentas N. (2010) Circoletto: visualizing sequence similarity with Circos. *Bioinformatics*, **26**, 2620–2621.

Depierreux D., Vong M., Nibert M.L. (2016) Nucleotide sequence of *Zygosaccharomyces bailii* virus Z: evidence for + 1 programmed ribosomal frameshifting and for assignment to family Amalgaviridae. *Virus Research*, **217**, 115–124.

Firth A.E., Jagger B.W., Wise H.M., Nelson C.C., Parsawar K., Wills N.M., Napthine S., Taubenberger J.K., Digard P., Atkins J.F. (2012) Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. *Open Biology*, **2**, 120109.

Garrison E., Marth G. (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint*, **1207**, 3907.

Geoghegan J.L., Holmes E.C. (2017) Predicting virus emergence amid evolutionary noise. *Open biology*, **7**, 170189.

Greninger A.L. (2017) A decade of RNA virus metagenomics is (not) enough. *Virus Research,* doi: 10.1016/j.virusres.2017.10.014

Groves R.L., Walgenbach J.F., Moyer J.W., Kennedy G.G. (2002) The role of weed hosts and tobacco thrips, *Frankliniella fusca*, in the epidemiology of *Tomato spotted wilt virus*. *Plant Disease*, **86**, 573-582.

Luo Z., Iaffaldano B.J., Zhuang X., Fresnedo-Ramírez J., Cornish, K. (2017) Analysis of the first Taraxacum kok-saghyz transcriptome reveals potential rubber yield related SNPs. *Scientific Reports*, **7**, 9939.

Martin R.R., Zhou J., Tzanetakis I.E. (2011) Blueberry latent virus: an amalgam of the Partitiviridae and Totiviridae. *Virus Research*, **155**, 175–180.

Mountain W.L., Powell C.A., Forer L.B., Stouffer R.F. (1983) Transmission of tomato ringspot virus from dandelion via seed and dagger nematodes. *Plant disease*, **67**, 867-868.

Sabanadzovic S., Ghanem-Sabanadzovic N.A., Valverde R.A. (2010) A novel monopartite dsRNA virus from rhododendron. *Archives of Virology*, **155**, 1859-1863.

Sabanadzovic S., Valverde R.A., Brown J.K., Martin R.R., Tzanetakis I.E. (2009) Southern tomato virus: the link between the families Totiviridae and Partitiviridae. *Virus Research*, **140**, 130–137.

Nibert M.L., Pyle J.D., Firth A.E. (2016) A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology*, **498**, 201–208.

Zhang Y., Iaffaldano B.J., Zhuang X., Cardina J., Cornish K. (2017) Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biology*, **17**, 34.

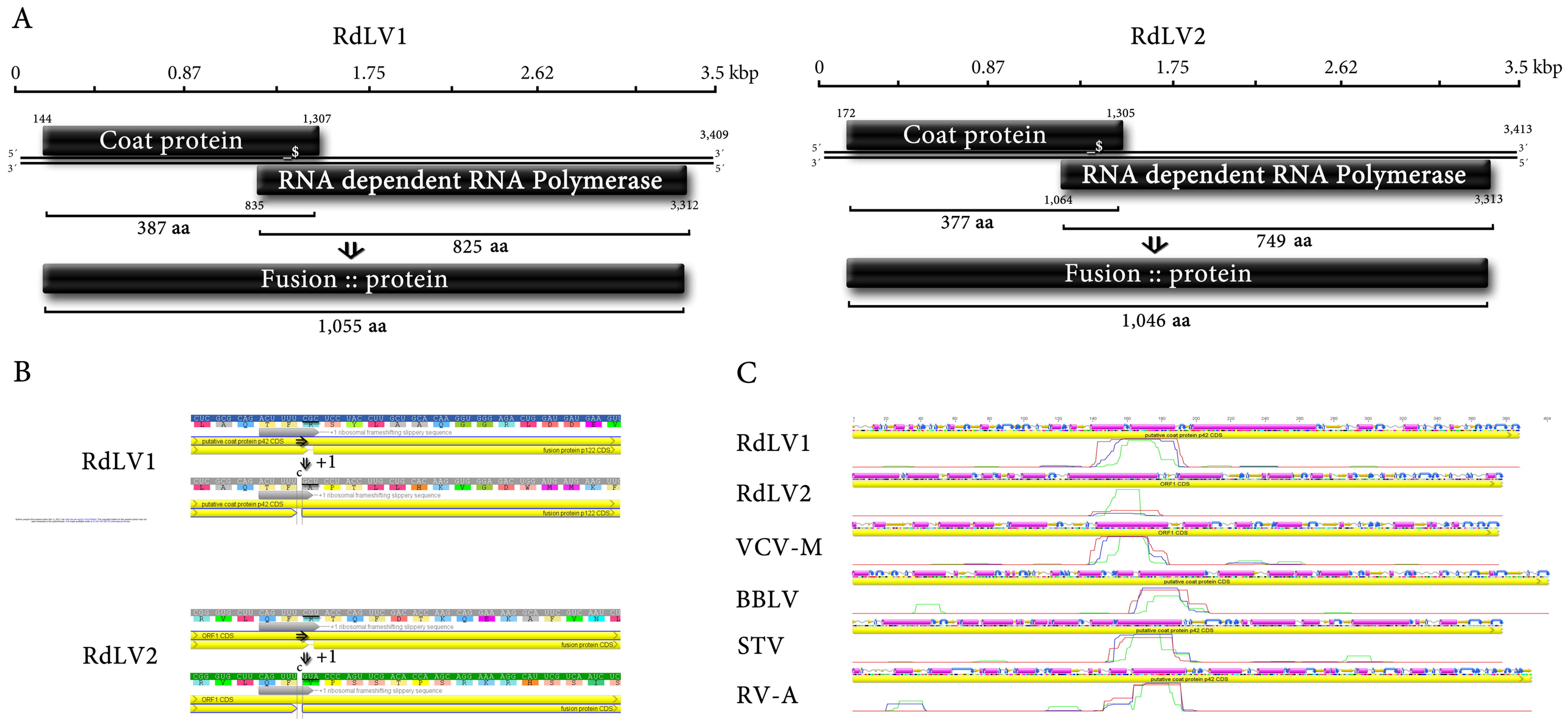
Table 1 Diverse structural highlights of RdLV 1 & 2 in comparison with ICTV recognized *Amalgaviridae* species. GS: Genome size (nt), 5´U: 5´UTR length (nt), OR1: ORF 1 length (nt), OR: Overlapping region (nt), OR2: ORF 2 length (nt), 3´U: 3´UTR length (nt), SLPs: Slippery sequence, SLPp: SLP position, FP: Fusion protein length (aa), RP: RdRP (aa), CP: Putative Coat protein length (aa), RPi, CPi, GSi: RdRP, CP, and complete genome sequence identity of the corresponding *Amalgavirus* in relation to RdLV1.

Figure Legends

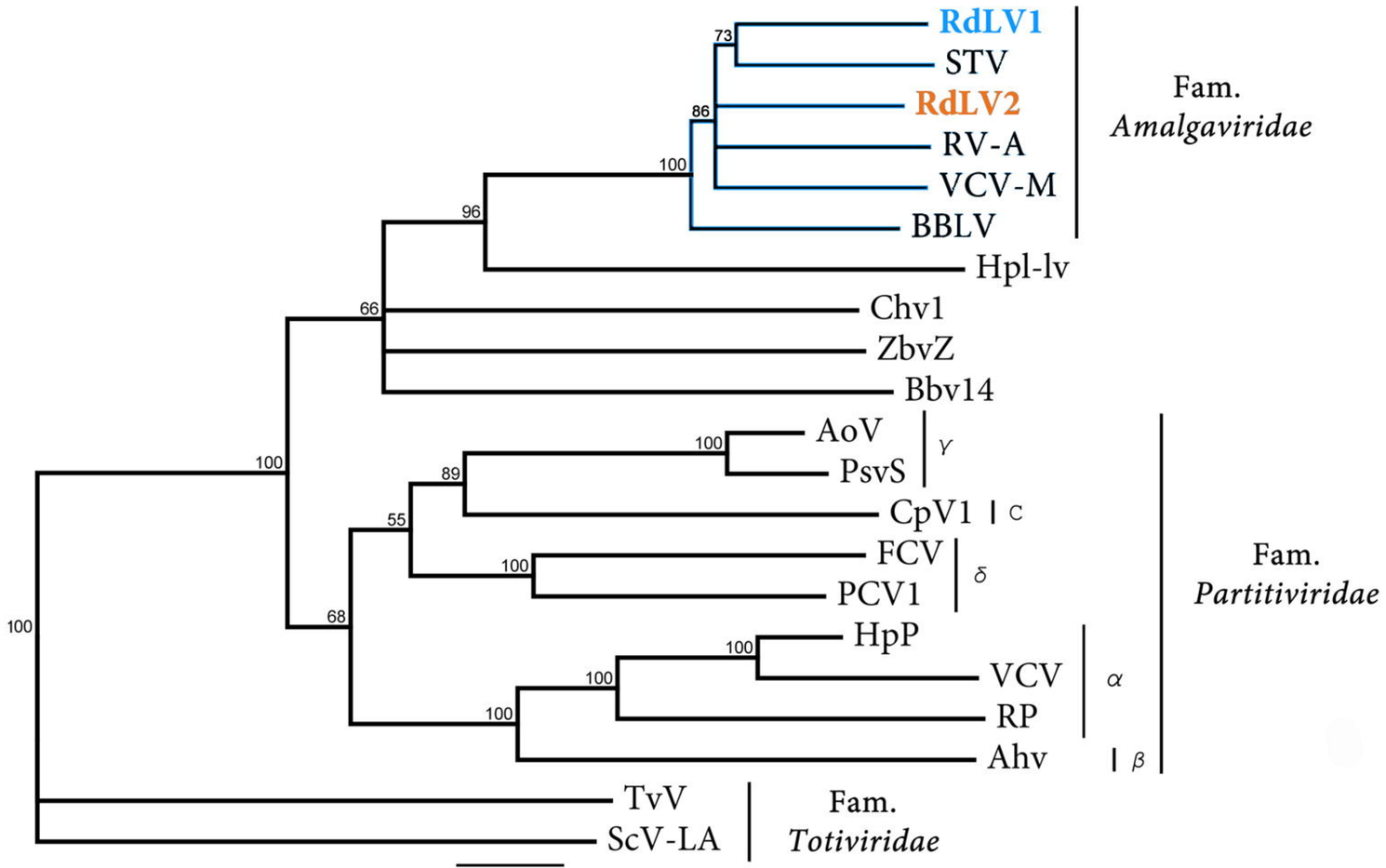
Figure 1. Molecular characterization of RdLV1 & RdLV2 (**A**) Rubber dandelion latent virus 1 & 2 (RdLV1 & RdLV2) linear monopartite dsRNA genome are 3,409 & 3,413 nt long, arranging a translation strategy based in two partially overlapping ORFs. The RdLV1 genome presents a 143 nt 5´UTR and a 97 nt 3´UTR. The predicted ORF1 encodes a 387 aa putative Coat protein. The overlapping ORF2 encodes a 825 aa RNA dependent RNA Polymerase. Genome position 981 (_$) presents a putative "slippery" sequence that could induce the generation of a 120 kDa fusion protein. The RdLV2 genome presents a 171 nt 5´UTR and a 97 nt 3´UTR. The predicted ORF1 encodes a 377 aa putative Coat protein. The overlapping ORF2 encodes a 749 aa RdRP. Genome position 946 (_$) presents a putative "slippery" sequence that could induce the generation of a 118 kDa FP. (**B**) Potential programmed ribosomal frameshifting of RdLV1 & 2. The RdLV1 ACU_UUU_CGC motif and RdLV2 CAG_UUU_CGU motif, of the general form UUU_CGN, are +1 ribosomal frameshifting motif prevalent among most plant amalgaviruses. (**C**) 3D structure prediction of the corresponding Coat proteins of RdLV1 & 2 and of reported amalgaviruses, assessed with the EMBOSS 6.5.7 tool Garnier represented on top, and coiled coil determination by COILS with a MTIDK matrix as a line graphs. Regions of high coiled coil probability are constrained to the typical α-helical central region of the CPs.

Figure 2. Phylogenetic insights and exploratory prevalence of RdLV1 & RdLV2 (**A**) Maximum-likelihood phylogenic tree of the RdRP predicted protein of reported amalgaviruses in the context of related viral families based on a MAFTT multiple alignments. Numbers at the nodes indicate percentage of bootstrap consensus support values obtained for 1000 replicates. Sequence similarity levels of

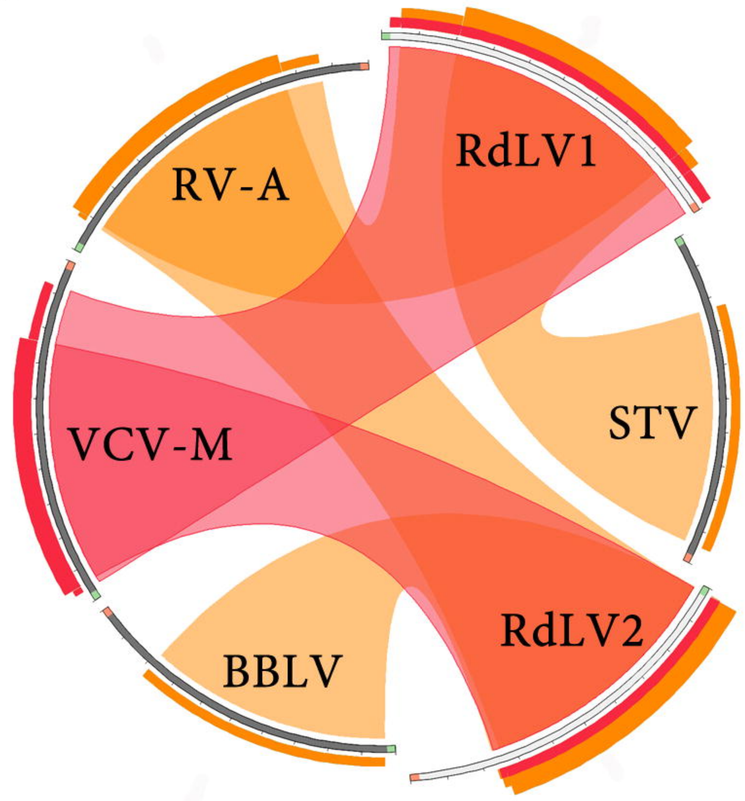
amalgaviruses Fusion Proteins among the *Amalgavirus* genus (**B**) and RdRP proteins of related viruses (**C**) expressed as Circoletto diagrams. FPs or RdRPs are depicted clockwise, and sequence similarity is visualized from blue to red ribbons representing low-to-high sequence identity. (**D**) Virus RNA levels expressed as FPKM of NGS sequenced rubber dandelion total RNA root samples. Values for RdLV1 are depicted in blue columns and values for RdLV2 in orange columns. (**E**) RNAseq based read mapping graphs of RdLV1 and RdLV2 with the 6 combined RNA libraries. Tracks from top to bottom represent coverage per base, sequence identity from red to green (higher), and SNP prediction. GenBank accession numbers and abbreviations for the respective viruses are *Southern tomato virus* (STV, NC_011591), *Rhododendron virus A* (RV-A, NC_014481), *Blueberry latent virus* (BBLV, NC_014593), *Vicia cryptic virus M* (VCV-M, EU371896), *Hubei partiti-like virus 59* (Hplv, APG78262), *Beihai barnacle virus 14* (Bbv14, APG78182), *Zygosaccharomyces bailii virus Z* (ZbvZ, KU200450), *Colletotrichum higginsianum dsRNA virus 1* (Chv1, NC_028242), *Heterobasidion partitivirus P* (HpP, AAK52739), *Radish partitivirus* (AY748911), *Vicia cryptic virus* (VCV, EF173396), *Saccharomyces cerevisiae virus L-A* (ScV-LA, NC_003745), *Penicillium stoloniferum virus S* (PsvS, NC_007539), *Aspergillus ochraceous virus* (AoV, EU118277), *Cryptosporidium parvum virus 1* (CpV1, CPU95995), *Pepper cryptic virus 1* (PCV1, JN117276), *Trichomonas vaginalis virus* (TvV, NC_003824), *Fig cryptic virus* (FCV, NC_015494), *Atkinsonella hypoxylon virus* (NP_604475). α: *Alphapartitivirus* genus, β: *Betapartitivirus* genus, γ: *Gammapartitivirus* genus, δ: *Deltapartitivirus* genus, C: *Cryspovirus* genus.


A
RdLV1
0
0.87
1.75
2.62
3.5 kbp
144
1,307
Coat protein
_$
3,409
5′
3′
3′
5′
RNA dependent RNA Polymerase
835
3,312
387 aa
825 aa
Fusion :: protein
1,055 aa
RdLV2
0
0.87
1.75
2.62
3.5 kbp
172
1,305
Coat protein
_$
3,413
5′
3′
3′
5′
RNA dependent RNA Polymerase
1,064
3,313
377 aa
749 aa
Fusion :: protein
1,046 aa
B
RdLV1
+1 ribosomal frameshifting slippery sequence
putative coat protein p42 CDS
fusion protein p122 CDS
+1
+1 ribosomal frameshifting slippery sequence
putative coat protein p42 CDS
fusion protein p122 CDS
RdLV2
+1 ribosomal frameshifting slippery sequence
ORF1 CDS
fusion protein CDS
+1
+1 ribosomal frameshifting slippery sequence
ORF1 CDS
fusion protein CDS
C
RdLV1
RdLV2
VCV-M
BBLV
STV
RV-A
putative coat protein p42 CDS
ORF1 CDS
ORF1 CDS
putative coat protein CDS
putative coat protein p42 CDS
putative coat protein p42 CDS

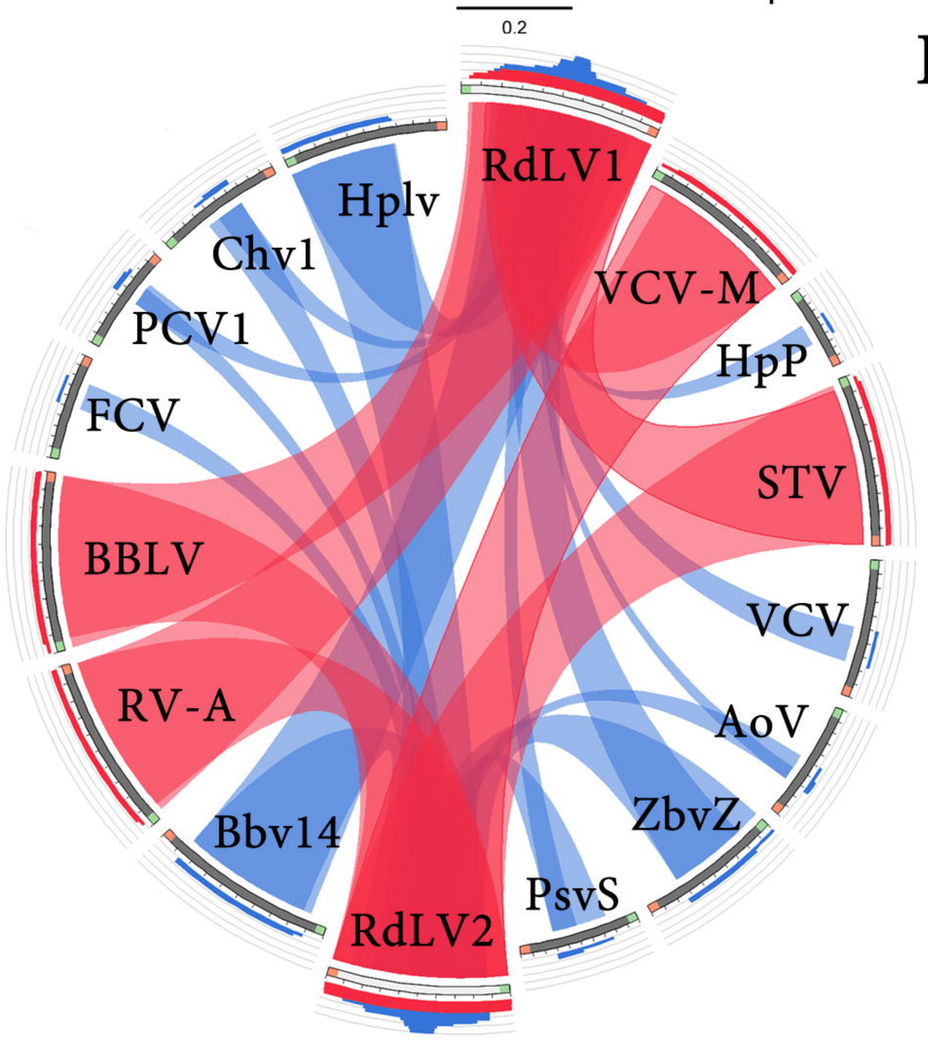
A



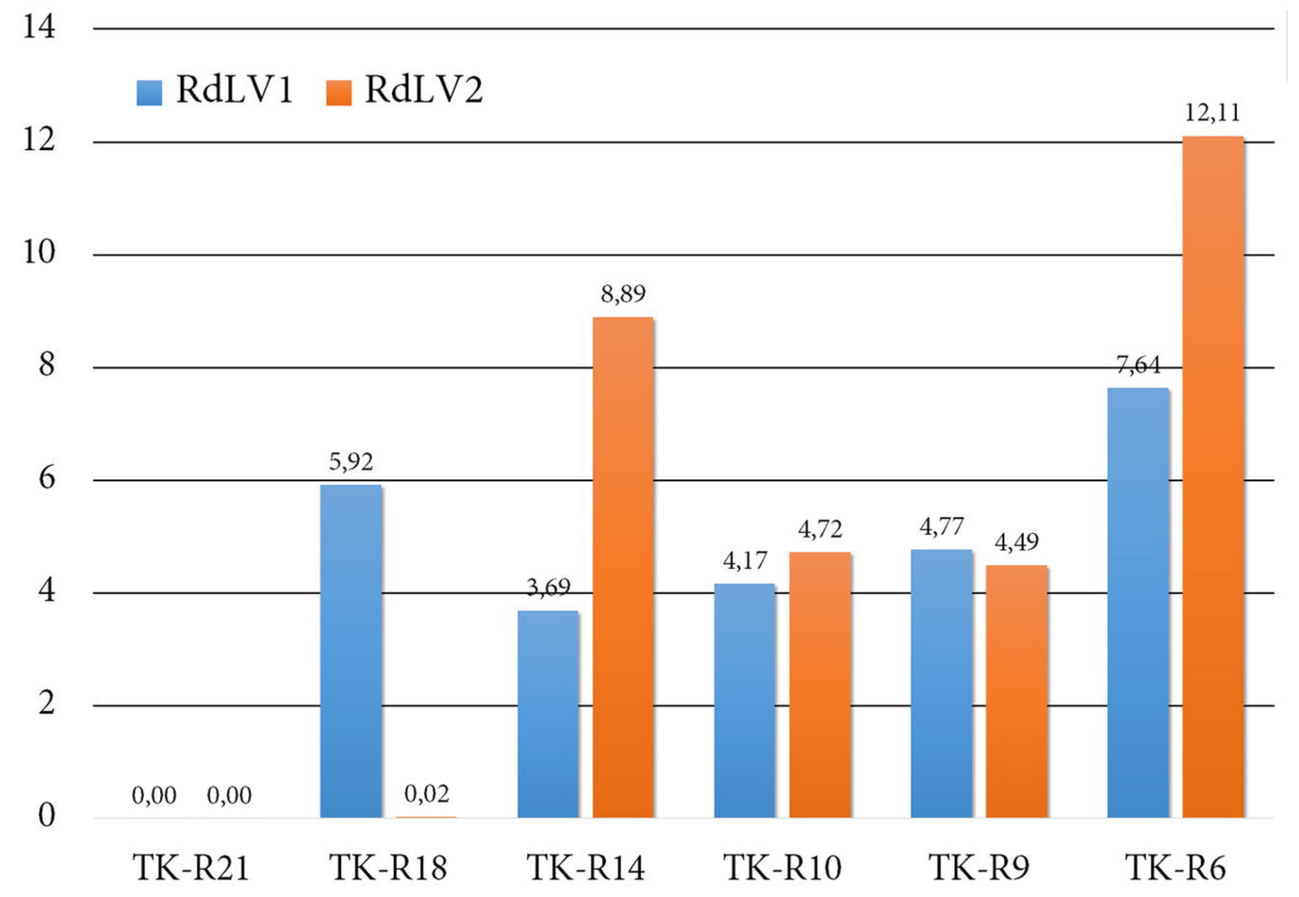
B



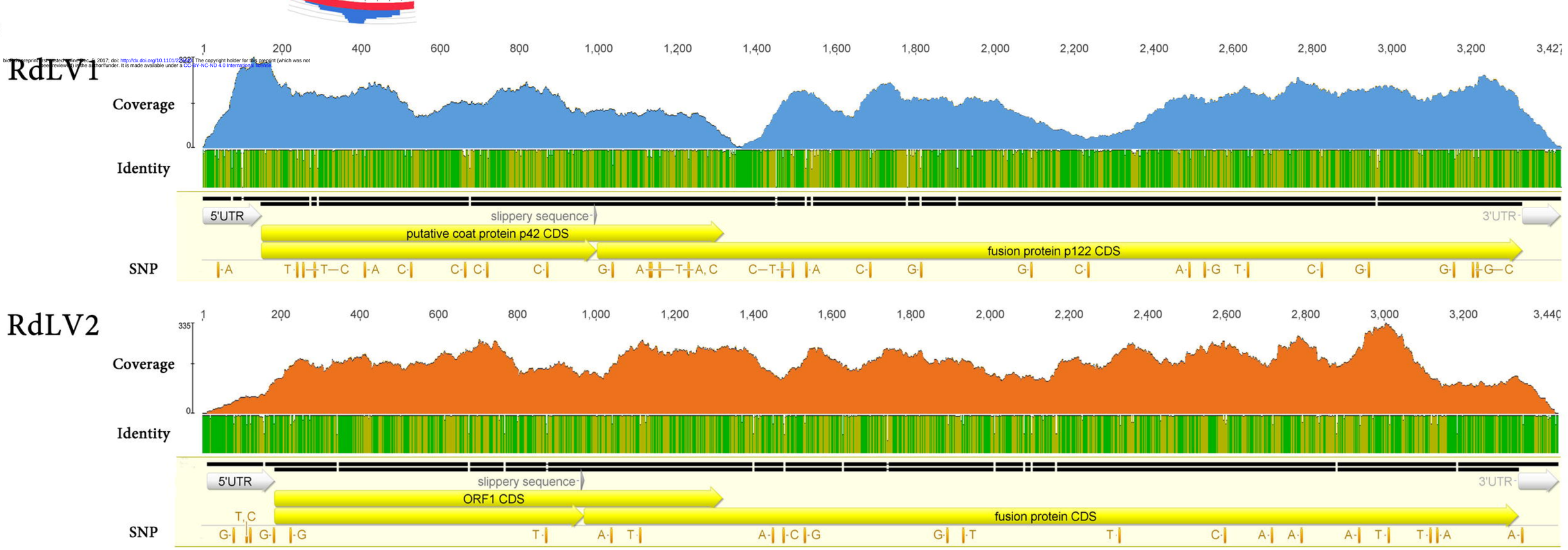
C



D



E



Molecular Identification and Characterization of Two Rubber Dandelion Amalgaviruses

Humberto Debat, Zinan Luo, Brian J. Iaffaldano, Xiaofeng Zhuang and Katrina Cornish

Amalgavirus	GS	5´U	OR1	OR	OR2	3´U	SLPs	SLPp	FP	RP	CP	RPi	CPi	GSi	Accession n.
Rubber dandelion latent virus 1	3,409	143	1,164	473	2,575	97	ACU_UUU_CGC	981	1,055	825	387	-----	-----	-----	MF197380
Rubber dandelion latent virus 2	3,413	171	1,134	242	2,250	100	CAG_UUU_CGU	946	1,046	749	377	49.5	21.5	55.9	MF197379
Southern tomato virus	3,437	137	1,134	233	2,289	110	CUU_AGG_CGU	983	1,063	763	378	49.5	22.0	55.7	NC_011591
Blueberry latent virus	3,431	166	1,128	359	2,397	99	UCU_UUU_CGU	979	1,055	799	376	46.2	19.5	54.8	NC_014593
Rhododendron virus A	3,427	94	1,215	405	2,424	47	ACU_UUU_CGC	1,181	1,078	808	405	48.0	22.1	54.4	NC_014481
Vicia cryptic virus M	3,434	142	1,185	287	2,277	117	ACU_UUU_CGU	1,089	1,058	759	395	47.3	19.7	54.8	EU371896

Table 1 Diverse structural highlights of RdLV 1 & 2 in comparison with ICTV recognized Amalgaviridae species. GS: Genome size (nt), 5´U: 5´UTR length (nt), OR1: ORF 1 length (nt), OR: Overlapping region (nt), OR2: ORF 2 length (nt), 3´U: 3´UTR length (nt), SLPs: Slippery sequence, SLPp: SLP position, FP: Fusion protein length (aa), RP: RdRP (aa), CP: Putative Coat protein length (aa), RPi, CPi, GSi: RdRP, CP, and complete genome sequence identity of the corresponding *Amalgavirus* in relation to RdLV1.