Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume associated enamoviruses

Archives of virology –Brief report

Humberto J. Debat [1]*, Nicolas Bejerman[12*]

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

[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

* Correspondence:

Corresponding authors:

Bejerman Nicolas, bejerman.nicolas@inta.gob.ar Tel: +54 9 351 4973636, Fax: +54 9 351 4974330

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

Abstract

Bird's-foot trefoil (*Lotus corniculatus*) is a nutritious forage crop, employed for livestock foraging around the world. Here, we report the identification and characterization of two novel viruses associated with bird's-foot trefoil. Virus sequences with affinity to enamoviruses (ssRNA (+); *Luteoviridae; Enamovirus*) and nucleorhabdoviruses (ssRNA (-); *Rhabdoviridae; Nucleorhabdovirus*) were detected in *L. corniculatus* transcriptome data. The tentatively named birds-foot trefoil associated virus 1 (BFTV-1) genome organization is characterized by 13,626 nt long negative-sense ssRNA. BFTV-1 presents in its antigenome orientation six predicted gene products in the canonical order 3′-N-P-P3-M-G-L-5′. The proposed birds-foot trefoil associated virus 2 (BFTV-2) 5,736 nt virus sequence presents a typical 5′-PO-P1-2-IGS-P3-P5-3′ enamovirus genome structure. Phylogenetic analysis suggests that BFTV-1 is closely related to *Datura yellow vein nucleorhabdovirus*, and that BFTV-2 clusters into a monophyletic cluster of legumes-associated enamoviruses. This sub-clade of highly related and co-divergent legume associated viruses provides insights on the evolutionary history of the enamoviruses. The bioinformatic reanalysis of SRA libraries deposited in the NCBI database constitutes an emerging approach to the discovery of novel plant viruses which should be important for both quarantine purposes and disease management.

Keywords: *Lotus corniculatus*; bird's-foot trefoil; *Enamovirus*; *Nucleorhabdovirus*; virus discovery, quarantine, disease management

From an agronomical point of view, bird's-foot trefoil (*Lotus corniculatus* L.) is the most important *Lotus* species. This plant is a perennial legume mainly grown for fodder production in temperate regions and is considered one of the major forage legumes after lucerne (*Medicago sativa*) and white clover (*Trifolium repens*) [9]. It is a high quality forage that can be grazed or cut for hay or silage, and does not cause bloat in ruminants [13]. Intriguingly, to date the only virus which has been identified to affect this crop was the widely distributed alfalfa mosaic virus (AMV), in a field study of Prince Edward Island, Canada, more than 35 years ago [16].

Numerous novel viruses, many of them not inducing any apparent symptoms, have been identified from different environments using metagenomic approaches, which revealed our limited knowledge about the richness of a continuously expanding plant virosphere, widespread in every potential host assessed [18]. Genomic RNA molecules of these plant RNA viruses are often inadvertently co-isolated with host RNAs and their sequences can be detected in plant transcriptome datasets [15, 17]. In a recent consensus statement report, Simmonds et al [20] provide that viruses that are known only from metagenomic data can, should, and have been incorporated into the official classification scheme of the International Committee on Taxonomy of Viruses (ICTV). Thus, the analysis of public data constitutes an emerging source of novel *bona fide* plant viruses, which allows the reliable identification of new viruses in hosts with no previous record of virus infections. Here, we analyzed a transcriptomic dataset corresponding to bird's-foot trefoil, available at the NCBI SRA database, which resulted in the identification and characterization of a novel enamovirus and nucleorhabdovirus associated with this crop.

The raw data analyzed in this study corresponds to an RNAseq NGS library (SRA: SRS271068), associated to NCBI Bioproject PRJNA77207. As described by Wang et al [22], it is derived of Illumina

Hiseq 2000 sequencing of total RNA isolated from fresh flowers, pods, leaves, and roots from *L. corniculatus* collected in Qinling Mountain, Shaanxi Province, China (BioSample: SAMN00759026). The 26,492,952 2x90 nt raw reads from the SRA were pre-processed by trimming and filtering with the Trimmomatic tool as implemented in http://www.usadellab.org/cms/?page=trimmomatic, the resulting reads were assembled *de novo* with Trinity v2.6.6 release with standard parameters. *De novo* transcriptome assembly of SRA library SRS271068 [22] resulted in 55,473 transcripts, which were subjected to bulk local BLASTX searches (E-value < 1e-5) against a refseq virus protein database available at ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/viral.1.protein.faa.gz. The resulting hits were explored by hand. A 6,524 nt transcript obtained a significant hit (E-value = 0, identity = 60%) with the L protein of datura yellow vein nucleorhabdovirus (DYVV; [6]). Further similarity analyses resulted in the detection of three more DYVV–like transcripts of 3,953 nt, 2,354 nt and 1,432 nt. The tentative virus contigs were curated by iterative mapping of reads using Bowtie2 http://bowtie-bio.sourceforge.net/bowtie2/index.shtml which was also employed for mean coverage estimation and reads per million (RPMs) calculations. The transcripts extended and polished by iterative mapping of raw reads were subsequently reassembled into a 13,626 nt long RNA contig with overall 57,7% sequence similarity with DYVV. This resulting RNA sequence was reconstructed with a total of 17,656 reads (mean coverage = 116.4X, reads per million (RPM) = 666.4. To further advance in the characterization of this sequence, virus annotation was instrumented as described in Debat [4], briefly, virus ORFs were predicted with ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) domains presence and architecture of translated gene products was determined by InterPro (https://www.ebi.ac.uk/interpro/search/sequence-search) and the NCBI Conserved domain database v3.16 (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Further, HHPred and HHBlits as implemented in https://toolkit.tuebingen.mpg.de/#/tools/ was used to complement annotation of divergent predicted proteins by hidden Markov models.

The tentatively named birds-foot trefoil associated virus 1 (BFTV-1) genome organization is characterized by 13,626 nt long negative-sense single-stranded RNA (GenBank accession number MH614262) and contains six open reading frames (ORFs) in the anti-genome, positive-sense orientation (**Fig. 1A**). BLASTP searches identified these ORFs as potentially encoding: a nucleocapsid protein (N; ORF1, 462 aa), phosphoprotein (P; ORF2, 344 aa), movement protein (P3; ORF3, 325 aa), matrix protein (M; ORF 4, 294 aa), glycoprotein (G; ORF 5, 637 aa), and an RNA dependent RNA polymerase (L; ORF 6, 2104 aa), based on highest sequence identity scores with nucleorhabdoviruses, in particular with black currant-associated nucleorhabdovirus (BCaRV; [23]), DYVV and sonchus yellow net nucleorhabdovirus (SYNV; [12]) (**Table 1**). The coding sequences are flanked by 3' leader (l) and 5' trailer (t) sequences revealing a genome organization of 3' l–N–P–P3–P4-M–G–L–t 5' (**Fig.1A**), which is the basic organization described for plant rhabdoviruses [7] given that no accessory genes are encoded by the BFTV-1 genome. Like all plant rhabdoviruses, BFTV-1 genes are separated by intergenic "gene junctions" regions, which are composed of the polyadenylation signal of the preceding gene, a short intergenic region, and the transcriptional start of the following gene (**Supp. Table 1**). Interestingly, BFTV-1 consensus "gene junction" region sequence 3´-AUUCUUUUUGGUUGUA-5´ is identical to that of BCaRV, DYVV and SYNV (**Supp. Table 1**). All BFTV-1-encoded proteins contain a classical mono

or bi-partite nuclear localization signal (NLS) [7]. Importin-α dependent nuclear localization signals were predicted using cNLS Mapper available at http://nls-mapper.iab.keio.ac.jp/, the scores predicted an exclusively nuclear localization for N, M and L proteins, while for P, P3 and G the scores suggested localization to both the nucleus and cytoplasm (**Table 1**). Nuclear export signals were predicted using NetNES 1.1 available at www.cbs.dtu.dk/services/NetNES/ which predicted a leucine-rich nuclear export signal in BFTNRV N protein near amino acid position 350 (data not shown), suggesting both nuclear import and export. Similar results were reported for the predicted products of DYVV [7]. Amino acid sequence comparisons between the deduced BFTV-1 proteins and the corresponding sequences of other nucleorhabdoviruses revealed high similarity of BFTV-1 to BCaRV, DYVV and SYNV. BFTV-1 N, G and L proteins were the most identical to that encoded by BCaRV, DYVV and SYNV sharing a 37.0-58.4% similarity, whereas P, P3 and M were the more divergent (**Supp. Table 2**).

Phylogenetic insights based on predicted replicase proteins were generated by MAFTT 7 https://mafft.cbrc.jp/alignment/software/ multiple amino acid alignments (BLOSUM62 scoring matrix) using as best-fit algorithm E-INS-i (BFTV-1) or G-INS-I (BFTV-2). The aligned proteins were subsequently used as input for FastTree 2.1.5 (http://www.microbesonline.org/fasttree/) maximum likelihood phylogenetic trees (best-fit model = JTT-Jones-Taylor-Thorton with single rate of evolution for each site = CAT) computing local support values with the Shimodaira-Hasegawa test (SH) and 1,000 tree resamples. The obtained tree shows that BFTV-1 clusters together with other viruses within the genus *Nucleorhabdovirus* (**Fig.1B**). BFTV-1 appears to have a close evolutionary relationship with BCaRV, DYVV and SYNV, which share a similar genomic organization to that described for BFTV-1 [8, 13, 23]. It is tempting to suggest that BFTV-1, BCaRV, DYVV and SYNV may represent the most ancestral clade within nucleorhabdoviruses, since no accessory genes are present in their genomes. Taken together our results suggest that BFTV-1 should be considered a member of a new virus species in the genus *Nucleorhabdovirus*.

In addition to the nucleorhabdovirus, the BLASTX searches of the obtained transcriptome showed a 5,622 nt transcript which obtained a significant hit (E-value = 0, identity = 74%) with the P1-P2 fusion protein of Alfalfa enamovirus 1 (AEV-1; [2]). Mapping of the SRA raw reads extended the contig into a 5,736 nt RNA sequence (mean coverage = 61.6X, reads per million (RPM) = 147.1) with an overall identity of 69.2% with AEV-1. The assembled sequence of bird's-foot trefoil virus 2 (BFTV-2) consists of 5,736 nt (GenBank accession number MH614261) and presents a typical 5′-PO-P1-2-IGS-P3-P5-3′ enamovirus genome structure (**Fig.2A**).

The first ORF, ORF 0, consists of 909 nt encoding a putative P0 of 303 aa with a calculated molecular mass of 33.8 kDa. In enamoviruses, P0 has been shown to function as an RNA silencing suppressor [11]. BFTV-2 P0 is more similar to that of legume-infecting enamoviruses AEV-1 and pea enation mosaic virus-1 (PEMV-1), where an F-Box-like motif located on the N-region of the predicted protein was identified (**Supp. Fig. 1A**). Interestingly, this motif was not identified in the non-legume infecting enamoviruses. The F-box-like motif is involved in silencing suppression [11], therefore it is tempting to speculate that other encoded proteins, rather than P0, may be involved in silencing suppression of the non-legume infecting enamoviruses. Nevertheless, proper assays should be employed to test this hypothesis. The second ORF, ORF1, which contains 2,283 nt, is predicted to be expressed by a ribosomal

leaky scanning mechanism for a protein P1 (761 aa, 83.9 kDa). The third ORF, ORF2, which is translated by a -1 ribosomal frameshift from ORF 1 overlaps ORF1 at its 5´ end and is predicted to produce an ORF1-ORF2 fusion protein, P1-P2 (1183 aa, 131.4 kDa). The canonical motif for a −1 frameshift site is X_XXY_YYZ. A putative slippery sequence was detected at position 2,039 of the type G_GGA_AAC (**Supp. Fig. 2A**), identical to that of PEMV-1 [5]. In addition, an H-type highly structured (Free energy = -12.80) 40 nt pseudoknot was detected eight nt downstream (seven nt spacer) immediately following the slippery sequence (**Supp. Fig. 2B**) which was predicted with the KnotInFrame tool available at https://bibiserv2.cebitec.uni-bielefeld.de/knotinframe and visualized with the VARNA 3.93 applet (http://varna.lri.fr/). It is worth noting that the frameshifting pseudoknot in other polero/enamoviruses, including PEMV-1, is seven nt from the end of the shifty site (6 nt spacer), given that the spacing between the slippery sequence and pseudoknot is determinant [1] further studies should confirm the tentative prediction of this report. P1 and P1-P2 have a putative involvement in virus replication, while P1 is considered a serine-like protease, and the frameshift region (P2) of the P1-P2 protein is thought to contain the core domains of the viral RNA-dependent RNA polymerase (RdRP) [5]. A serine-protease-like domain (peptidase S39, pfam02122, P1 residues 314-515; E-value, 3.75e-42) in P1, and an RdRP domain (RdRP-4, pfam02123) resembling those from members of the genera *Polerovirus*, *Rotavirus* and *Totivirus* in P2 (P1-P2 residues 734-1116; E-value, 2.26e-61) were found when these protein sequences were analyzed (**Fig. 2A**). The fourth ORF, ORF3, consists of 570 nt, encoding a predicted coat protein P3 (190 aa, 21.2 kDa). The fifth ORF, ORF5, is a putative in-frame read-through product of ORF3 encoding a fusion protein P3-P5 (506 aa, 56.1 kDa). BFTV-2 presents the canonical stop-codon context for read-through of enamovirus (UGA-GGG; [9]). Moreover, between the 4,601-6,665 nt coordinates, 15 nt downstream the stop codon, 11 CCNNNN tandem repeat motif units were found, which are associated with ORF 3 stop codon read-through [3]. Furthermore, signatures in a cytidine-rich domain immediately after the stop codon (5'C-rich) and a branched stem-loop structure 684 nt downstream of the CP stop codon (CD-DRTE), which are required for ORF3 readthrough [24], were also identified in the BFTV-2 genomic sequence (**Supp. Fig. 3A-B**). Both BFTV-2 5'C-rich and CD-DRTE regions showed high identity with that of AEV-1, with the expected complementary bases of the two regions identical with the exception of a G-to-A SNP at the CD-DRTE, which is a non-complementary base in both BFTV-2 and AEV-1. P3 is the coat protein (CP), whereas the CP read-through extension (P5) of P3-P5 is thought to be an aphid transmission subunit of the virus [5]. While the CP region of the predicted P3-P5 protein is more conserved in enamovirus, some emerging motifs could be observed, such as a proline rich stretch immediately following the read-through region (**Supp. Fig. 1.B**). It is tempting to speculate that this motif could be involved in the transmission of BFTV-2; nonetheless, mutagenesis analysis coupled with virus transmission assays should be conducted to test this hypothesis. When these protein sequences were analyzed in detail, a luteovirid CP domain (Luteo_coat, pfam00894) in P3 (P3 residues 55-188; E-value, 1.63e-70) and a polerovirus read-through protein domain (PLRV_ORF5, pfam01690) in P5 (P3-P5 residues 221-424; E-value, 1.16e-48) were identified.

The functional and structural results indicate that BFTV-2 represents the genome sequence of a virus that should be taxonomically classified in the family *Luteoviridae*. Viruses in the family *Luteoviridae* contain a single-stranded positive-sense RNA genome and are classified into three genera, *Enamovirus*,

Luteovirus and *Polerovirus*. Unlike poleroviruses, enamoviruses do not encode a P4 movement protein, and luteoviruses lack a P0 gene [8]. Therefore, BFTV-2 appears to be an enamovirus because its genomic structure, characterized by presenting a P0 gene, and not encoding a P4 protein. The BFTV-2 ORFs were compared to the predicted ORFs of AEV-1, PEMV-1, grapevine enamovirus-1 (GEV-1; [19]) and those of citrus vein enation virus (CVEV; [21]), in order to determine nucleotide and deduced amino acid sequence identities (**Table 1**). The maximum nt sequence identity for any gene CDS was 75.4, 77.1, 47.2 and 44.9%, respectively, whereas the maximum aa sequence identity for any gene product was 82.6, 85.3, 44.0 and 37.7%, respectively (**Table 1**). Therefore, the differences in aa sequence identity for each gene product were greater than 10%, which is one of the criteria used by the ICTV to demarcate species in the genera *Polerovirus* and *Luteovirus* [8] and also in *Enamovirus*. Consequently, BFTV-2 may belong to a new species in the genus *Enamovirus*. Further, in a phylogenetic tree based on the P1-P2 fusion protein aa sequence of viruses of the family *Luteoviridae*, BFTV-2 clustered with AEV-1 and PEMV-1 in a monophyletic clade of legumes-associated enamoviruses, within the enamovirus complex (**Fig. 2B**), which suggest that all legume-infecting enamoviruses share a common ancestor. Further assessment of the genetic distance of each gene product of all reported enamoviruses (**Supp. Fig. 4**) indicates an evident high similarity among the legume associated enamoviruses, including BFTV-2. It is interesting that BFTV-2 in phylogenetic trees (**Fig. 2B**) clusters more closely with AEV-1 than PEMV-1. Unlike PEMV-1 which was first reported in pea, AEV-1 has been detected only in alfalfa, a member of the *Faboideae* sub-family of legumes, shared with bird´s foot trefoil. Thus, these legume viruses appear to consistently co-diverge among them. Additional legume viruses would be helpful to entertain whether the evolutionary path of enamoviruses is characterized by co-divergence. AEV-1 is associated to a complex of viruses responsible of the alfalfa dwarf disease [2]. It is interesting that BFTV-2 was detected in concert with a nucleorhabdoviruses. Future studies should focus on determining if there might be any synergistic effect on the host caused by the co-infection of enamoviruses and nuclerhabdoviruses: a subtle interaction scenario that may have important implications in management of associated diseases. For instance, the umbravirus PEMV-2 provides the movement protein [5, 8] function lacking in the enamovirus PEMV-1; as no umbraviruses were identified in the bird's-foot trefoil dataset, therefore it is tempting to speculate that perhaps the identified nucleorhabdovirus could be providing the movement functions to complement and thus allow in concert the movement and replication of BFTV-2.

In summary, the analysis of public SRA data constitutes an emerging source of novel plant viruses. Using this approach, we report the identification and molecular characterization of two novel viruses associated with bird's-foot trefoil. Our analyses unveiled that these viruses may be considered novel species of the genera *Nucleorhabdovirus* (BFTV-1) and *Enamovirus* (BFTV-2). The RNA viruses reported here provide for the first time a glimpse of the virus landscape of this important crop. More importantly, these novel viruses provide evidence of a sub-clade of highly related and co-divergent legume associated enamoviruses as new clues on the understanding of the evolutionary history of this genus of viruses. Future studies should assess the prevalence of BFTV1 and BFTV-2, and unravel whether the infection of these novel viruses is associated to specific symptoms.

Acknowledgments

We would like to express a sincere gratitude to the generators of the underlying data used for this work: Dr. Ying Wang and Dr. Zhezhi Wang. By following open access practices and supporting accessible raw sequence data in public repositories available to the research community, they have promoted the generation of new knowledge and ideas.

Compliance with ethical standards

Conflict of interest: All authors declare that they have no conflict of interest

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Barry JK, Miller WA (2002) A− 1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. Proc Natl Acad Sci U S A, 99:11133-11138.
2. Bejerman N, Giolitti F, de Breuil S, Trucco V, Dietzgen RG, Lenardon S (2016) Complete genome sequence of a new enamovirus from Argentina infecting alfalfa plants showing dwarfism symptoms. Arch Virol 161:2029-2032.
3. Brown CM, Dinesh-Kumar SP, Miller WA (1996) Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. J Virol 70: 5884-5892.
4. Debat HJ (2017) An RNA Virome associated to the Golden orb-weaver Spider Nephila clavipes. Front Microbiol 8:2097.
5. Demler SA, De Zoeten GA (1991) The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. J Gen Virol 72:1819-1834.
6. Dietzgen RG, Innes DJ, Bejerman N (2015) Complete genome sequence and intracellular protein localization of Datura yellow vein nucleorhabdovirus. Virus Res 205:7-11.
7. Dietzgen RG, Kondo H, Goodin MM, Kurath G, Vasilakis N. (2017) The family Rhabdoviridae: mono-and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. Virus Res 227:158-170.
8. Domier LL (2012) Family Luteoviridae, in: King, A.M.Q, Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds), Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Oxford, pp. 1045-1053.
9. Escaray FJ, Menendez AB, Garriz A, Pieckenstain FL, Estrella MJ, Castagno LN, Carrasco P, Sanjuan J, Ruiz OA (2012) Ecological and agronomic importance of the plant genus Lotus. Its application in grassland sustainability and the amelioration of constrained and contaminated soils. Plant Science 182:121e133.
10. Firth AE, Brierley I (2012) Non-canonical translation in RNA viruses. J Gen Virol 93:1385-1409.

11. Fusaro AF, Correa RL, Nakasugi K, Jackson C, Kawchuk L, Vaslin MF, Waterhouse PM (2012) The Enamovirus P0 protein is a silencing suppressor which inhibits local and systemic RNA silencing. Virol 426:178-187.
12. Goodin MM, Austin J, Tobias R, Fujita M, Morales C, Jackson AO (2001) Interactions and nuclear import of the N and P proteins of sonchus yellow net virus, a plant nucleorhabdovirus. J Virol, 75:9393-9406.
13. Hannaway DB, Myers D (2004) Birdsfoot trefoil (*Lotus corniculatus* L.). Oregon State University. Species Selection Information System.
14. Heaton LA, Hillman BI, Hunter BG, Zuidema D, Jackson AO (1989) Physical map of the genome of sonchus yellow net virus, a plant rhabdovirus with six genes and conserved gene junction sequences. PNAS 86:8665-8668.
15. Kim H, Park D, Hahn Y (2018) Identification of novel RNA viruses in alfalfa (Medicago sativa): an Alphapartitivirus, a Deltapartitivirus, and a Marafivirus. Gene 638:7-12.
16. McDonald JG, Susuki M (1983) Occurrence of alfalfa mosaic virus in Prince Edward Island. Canadian Plant Disease Survey 63:47-50.
17. Nibert ML, Vong M, Fugate KK, Debat HJ (2018) Evidence for contemporary plant mitoviruses. Virol 518:14-24.
18. Roossinck MJ, Martin DP, Roumagnac P (2015) Plant virus metagenomics: advances in virus discovery. Phytopathol 105: 716–727.
19. Silva JMF., Al Rwahnih M, Blawid R, Nagata T, Fajardo TVM (2017) Discovery and molecular characterization of a novel enamovirus, Grapevine enamovirus-1. Virus genes 53: 667-671.
20. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Hull, R (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nature Reviews Microbiology 15: 161.
21. Vives MC, Velázquez K, Pina JA, Moreno P, Guerri J, Navarro L (2013) Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. Phytopathol 103:1077-1086.
22. Wang Y, Hua W, Wang J, Hannoufa A. Xu Z, Wang Z (2013) Deep sequencing of Lotus corniculatus L. reveals key enzymes and potential transcription factors related to the flavonoid biosynthesis pathway. Molecular genetics and genomics 288:131-139.
23. Wu LP, Yang T, Liu HW, Postman J, Li R (2018) Molecular characterization of a novel rhabdovirus infecting blackcurrant identified by high-throughput sequencing. Arch Virol 162: 2493–4.
24. Xu Y, Ju HJ, DeBlasio S, Carino EJ, Johnson R, MacCoss M, Heck MC, Miller WA, Gray SM (2018) A stem-loop structure in Potato leafroll virus ORF5 that is essential for read through translation of the coat protein ORF stop codon 700 bases upstream. J. Virol. 92: e01544-17.

Figure 1. Structural characterization and phylogenetic insights of bird's-foot trefoil associated virus 1 (BFTV-1) (**A**) Genome graphs depicting architecture and predicted gene products of BFTV-1. The predicted coding sequences are shown in orange arrow rectangles, start and end coordinates are indicated. Gene products are depicted in curved yellow rectangles and size in aa is indicated below. Predicted

domains or HHPred best-hit regions are shown in curved pink rectangles. Abbreviations: N, nucleoprotein CDS; P, phosphoprotein CDS; P3, putative cell-to-cell movement protein CDS; M, matrix protein CDS; G, glycoprotein CDS; L, RNA dependent RNA polymerase CDS; TM, trans-membrane domain; SP, signal peptide. (**B**) Maximum likelihood phylogenetic tree based on amino acid alignments of the L polymerase of BFTNRV and other plant rhabdoviruses. The tree is rooted at the midpoint; nucleorhabdovirus and cytorhabdovirus clades are indicated by green and blue rectangles, respectively. The scale bar indicates the number of substitutions per site. Node labels indicate FastTree support values. The viruses used to construct the tree, and their accession numbers are as follows: black currant nucleorhabdovirus (MF543022), alfalfa dwarf virus (KP205452), barley yellow striate mosaic cytorhabdovirus (KM213865), colocasia bobone disease associated-virus (KT381973), datura yellow vein nucleorhabdovirus (KM823531), eggplant mottled dwarf nucleorhabdovirus (NC_025389), lettuce yellow mottle virus (EF687738), lettuce necrotic yellows virus (NC_007642); maize fine streak nucleorhabdovirus (AY618417), maize Iranian mosaic nucleorhabdovirus (DQ186554), maize mosaic nucleorhabdovirus (AY618418), northern cereal mosaic cytorhabdovirus (AB030277), potato yellow dwarf nucleorhabdovirus (GU734660), rice yellow stunt nucleorhabdovirus (NC_003746); sonchus yellow net virus (L32603), taro vein chlorosis virus (AY674964) and strawberry crinkle cytorhabdovirus (MH129615).

Figure 2. Structural characterization and phylogenetic insights of birds-foot trefoil virus 2 (BFTV-2) (**A**) Genome graphs depicting architecture and predicted gene products of BFTV-2. The predicted coding sequences are shown in orange arrow rectangles, start and end coordinates are indicated. Gene products are depicted in curved yellow rectangles and size in aa is indicated below. Predicted domains or HHPred best-hit regions are shown in curved pink rectangles. Abbreviations: CP, coat protein CDS; P1-P2, RNA dependent RNA polymerase fusion protein CDS; FS, -1 ribosomal frameshifting signal; TM1-5, trans-membrane domains; ST, signal of translation read-through of a UGA stop codon; Pep_S39, peptidase S39 domain; RdRP_4, RdRP domain; Luteo_cp luteovirid coat protein domain; PLRV_ORF5, polerovirus readthrough protein domain. (**B**) Maximum likelihood phylogenetic tree based on amino acid alignments of the P1-P2 polymerase of BFTV-2 and other enamoviruses. The *Luteovirus* genus was used as tree root; *Enamovirus*, *Polerovirus* and *Luteovirus* clades are indicated by green, blue, and pink rectangles, respectively. The scale bar indicates the number of substitutions per site. Node labels indicate FastTree support values. The viruses used to construct the tree and their accession numbers are provided as **Supplementary Table 3**.

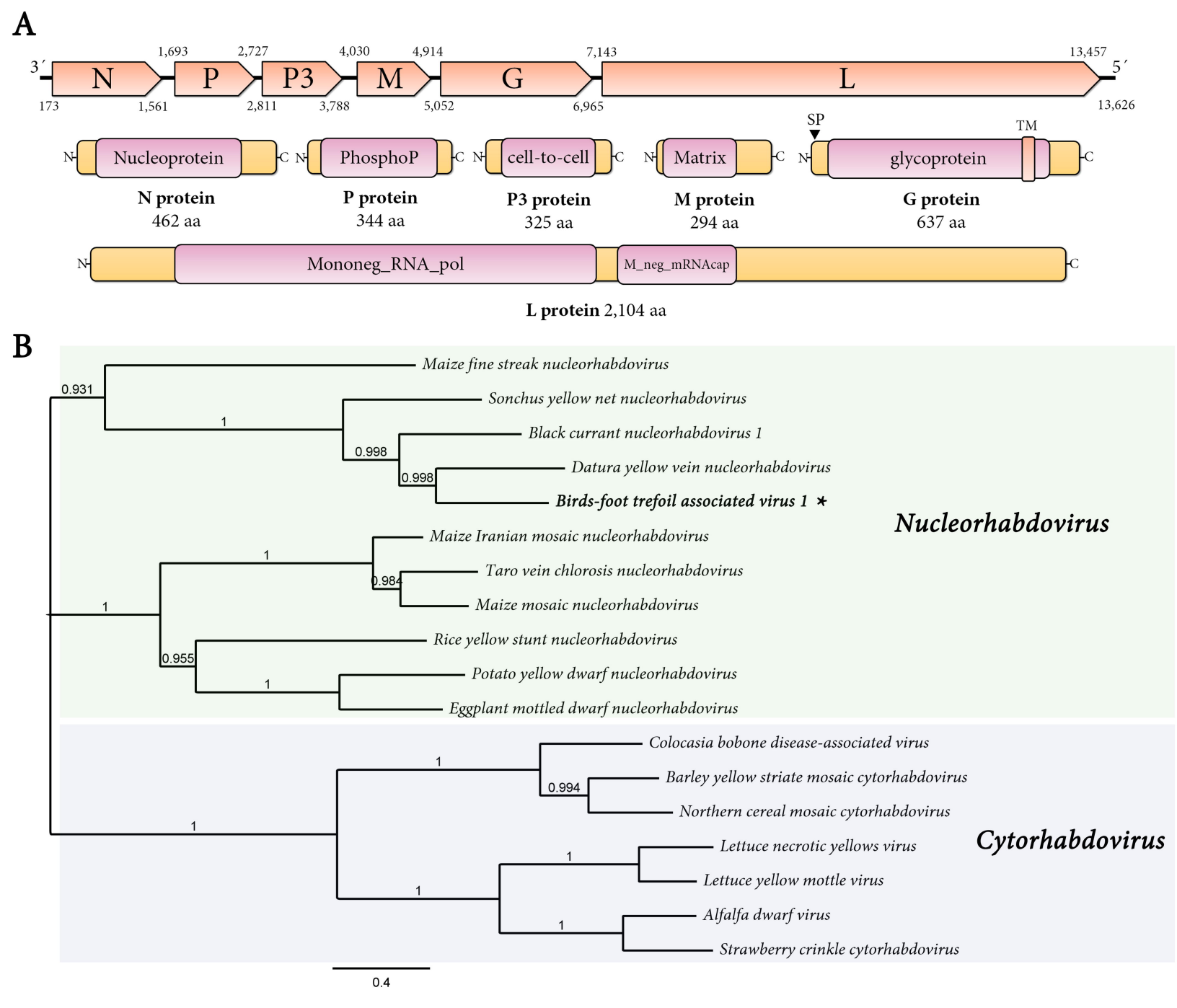
Supplementary Figure 1. Amino acid sequence alignments of P0 protein of enamoviruses (**A**) and P3-P5 readtrough aphid transmission predicted protein of enamoviruses (**B**). Sequence similarity is depicted form white to black and sequence logo expressed as bits per position.

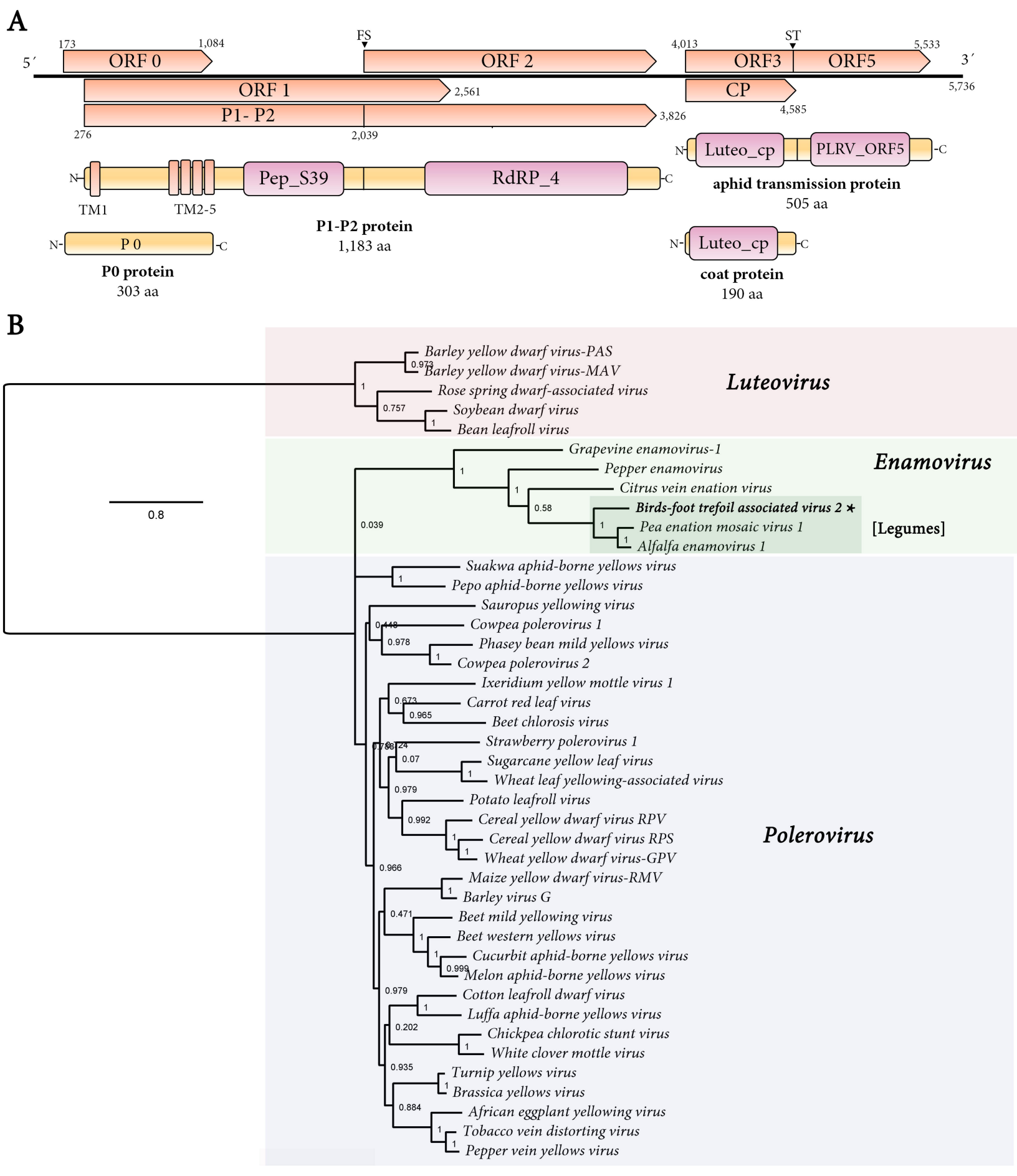
Supplementary Figure 2. Ribosomal frameshifting prediction of P1-P2 CDS region. (**A**) Canonical motif for a −1 frameshift site of the form X_XXY_YYZ. The putative slippery sequence was detected at position 2,039 of the type G_GGA_AAC, identical to that of pea enation mosaic virus-1 (PEMV-1). (**B**)

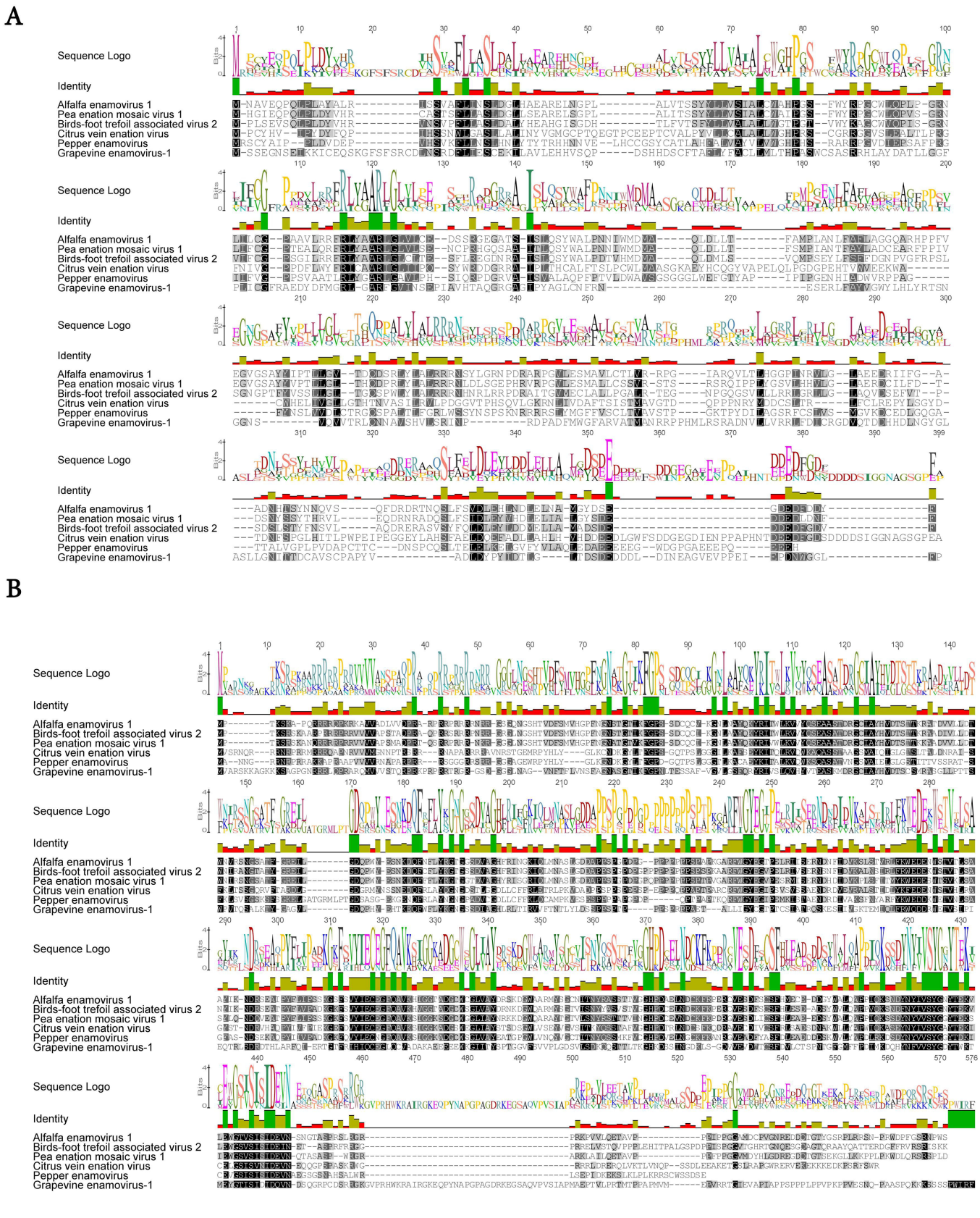
H-type highly structured (Free energy = -12.80) 40 nt pseudoknot detected starting 8 nt downstream the slippery sequence.

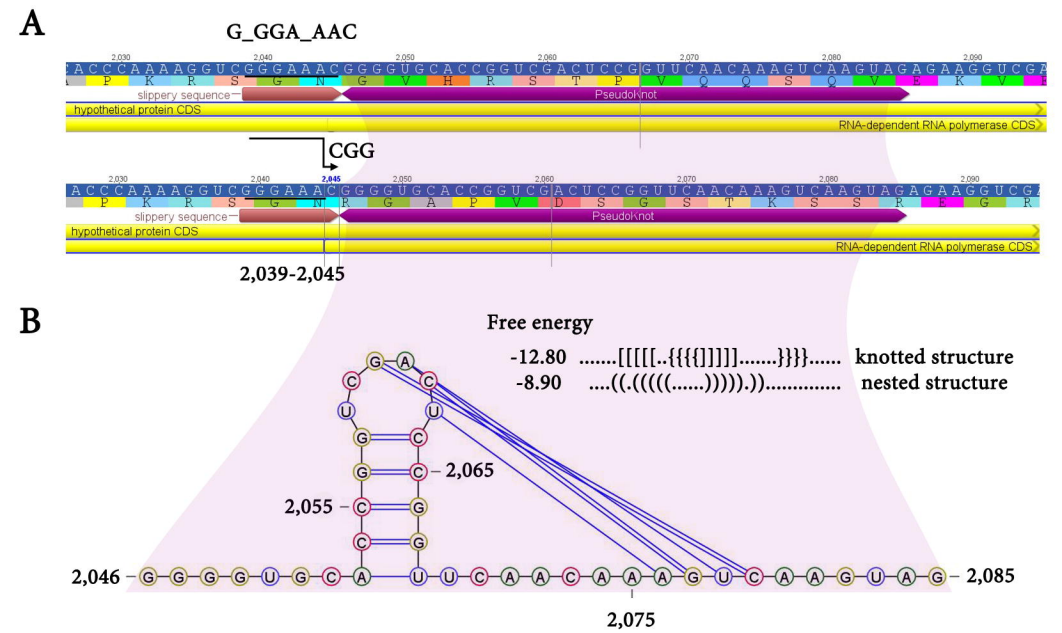
Supplementary Figure 3. CCNNNN oligomer repeats and long-range tentative base pairing between the 5' end of C-rich domain adjacent to the leaky stop codon of ORF3 (**A**) and the distal readthrough element (DRTE) of BFTV-2 and other legume associated enamoviruses (**B**). CCNNNNs are indicated by arrows, the distal element motif (DE) is depicted by a pink rectangle, complementary bases are underlined in blue.

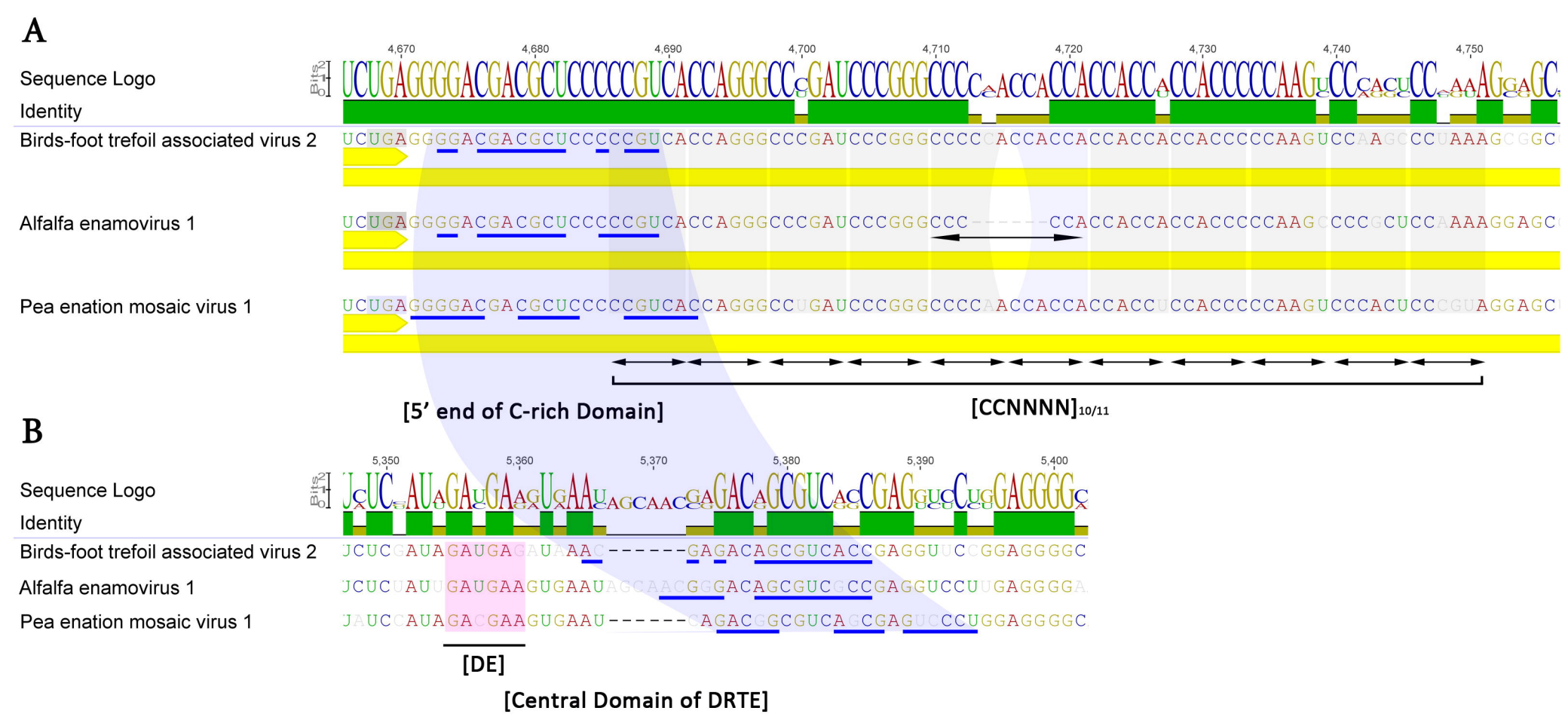
Supplementary Figure 4. Genetic distances (% aa identity) of predicted gene products of BFTV-2 and reported enamoviruses support a sub-clade of legume associated enamoviruses (circled in red). Values are also presented as colored heatmaps ranging from 0% (white) to 100% (black) identity. Abbreviations: AEV-1, alfalfa enamovirus 1; PEMV-1, pea enation mosaic virus-1; CVEV, citrus vein enation virus; PEPV, pepper enamovirus; GEV-1, grapevine enamovirus-1.


A
3′
N
P
P3
M
G
L
5′
173
1,561
1,693
2,811
2,727
3,788
4,030
5,052
4,914
6,965
7,143
13,626
13,457
Nucleoprotein
PhosphoP
cell-to-cell
Matrix
SP
TM
glycoprotein
N protein
462 aa
P protein
344 aa
P3 protein
325 aa
M protein
294 aa
G protein
637 aa
Mononeg_RNA_pol
M_neg_mRNAcap
L protein 2,104 aa
B
Maize fine streak nucleorhabdovirus
Sonchus yellow net nucleorhabdovirus
Black currant nucleorhabdovirus 1
Datura yellow vein nucleorhabdovirus
Birds-foot trefoil associated virus 1 *
Maize Iranian mosaic nucleorhabdovirus
Taro vein chlorosis nucleorhabdovirus
Maize mosaic nucleorhabdovirus
Rice yellow stunt nucleorhabdovirus
Potato yellow dwarf nucleorhabdovirus
Eggplant mottled dwarf nucleorhabdovirus
Nucleorhabdovirus
Colocasia bobone disease-associated virus
Barley yellow striate mosaic cytorhabdovirus
Northern cereal mosaic cytorhabdovirus
Lettuce necrotic yellows virus
Lettuce yellow mottle virus
Alfalfa dwarf virus
Strawberry crinkle cytorhabdovirus
Cytorhabdovirus
0.931
1
0.998
0.998
1
0.984
1
0.955
1
1
1
1
1
1
0.994
0.4


A
5′
3′
173
1,084
ORF 0
FS
ORF 2
4,013
ST
5,533
ORF3
ORF5
ORF 1
2,561
CP
5,736
4,585
P1- P2
3,826
276
2,039
N
Luteo_cp
PLRV_ORF5
C
aphid transmission protein
505 aa
Pep_S39
RdRP_4
TM1
TM2-5
P1-P2 protein
1,183 aa
P 0
P0 protein
303 aa
Luteo_cp
coat protein
190 aa
B
0.8
Luteovirus
Barley yellow dwarf virus-PAS
Barley yellow dwarf virus-MAV
Rose spring dwarf-associated virus
Soybean dwarf virus
Bean leafroll virus
Enamovirus
Grapevine enamovirus-1
Pepper enamovirus
Citrus vein enation virus
Birds-foot trefoil associated virus 2 ★
Pea enation mosaic virus 1
Alfalfa enamovirus 1
[Legumes]
Suakwa aphid-borne yellows virus
Pepo aphid-borne yellows virus
Sauropus yellowing virus
Cowpea polerovirus 1
Phasey bean mild yellows virus
Cowpea polerovirus 2
Ixeridium yellow mottle virus 1
Carrot red leaf virus
Beet chlorosis virus
Strawberry polerovirus 1
Sugarcane yellow leaf virus
Wheat leaf yellowing-associated virus
Potato leafroll virus
Cereal yellow dwarf virus RPV
Cereal yellow dwarf virus RPS
Wheat yellow dwarf virus-GPV
Maize yellow dwarf virus-RMV
Barley virus G
Beet mild yellowing virus
Beet western yellows virus
Cucurbit aphid-borne yellows virus
Melon aphid-borne yellows virus
Cotton leafroll dwarf virus
Luffa aphid-borne yellows virus
Chickpea chlorotic stunt virus
White clover mottle virus
Turnip yellows virus
Brassica yellows virus
African eggplant yellowing virus
Tobacco vein distorting virus
Pepper vein yellows virus
Polerovirus


A
Sequence Logo
Identity
Alfalfa enamovirus 1
Pea enation mosaic virus 1
Birds-foot trefoil associated virus 2
Citrus vein enation virus
Pepper enamovirus
Grapevine enamovirus-1
B
Sequence Logo
Identity
Alfalfa enamovirus 1
Birds-foot trefoil associated virus 2
Pea enation mosaic virus 1
Citrus vein enation virus
Pepper enamovirus
Grapevine enamovirus-1


A
G_GGA_AAC
slippery sequence
Pseudoknot
hypothetical protein CDS
RNA-dependent RNA polymerase CDS
CGG
2,039-2,045
B
Free energy
-12.80 .....[[[[[..{{{{]]]]].....}}}}..... knotted structure
-8.90 ....((.(((((.....)))))).))..... nested structure
2,065
2,055
2,046
2,085
2,075


A
Sequence Logo
Identity
Birds-foot trefoil associated virus 2
Alfalfa enamovirus 1
Pea enation mosaic virus 1
[5' end of C-rich Domain]
[CCNNNN]10/11
B
Sequence Logo
Identity
Birds-foot trefoil associated virus 2
Alfalfa enamovirus 1
Pea enation mosaic virus 1
[DE]
[Central Domain of DRTE]

P0

	AEV-1	PEMV-1	BFTV-2	CVEV	PEPV	GEV-1
Alfalfa enamovirus 1		66.3%	48.5%	15.7%	22.1%	12.5%
Pea enation mosaic virus 1	66.3%		50.8%	16.2%	22.1%	13.3%
Birds-foot trefoil a. virus 2	48.5%	50.8%		18.9%	21.5%	14.4%
Citru vein enation virus	15.7%	16.2%	18.9%		26.4%	14.7%
Pepper enamovirus	22.1%	22.1%	21.5%	26.4%		15.4%
Grapevine enamovirus 1	12.5%	13.3%	14.4%	14.7%	15.4%	

P1-2

	AEV-1	PEMV-1	BFTV-2	CVEV	PEPV	GEV-1
Alfalfa enamovirus 1		79.2%	61.7%	37.4%	35.8%	29.1%
Pea enation mosaic virus 1	79.2%		60.2%	35.7%	35.1%	29.0%
Birds-foot trefoil a. virus 2	61.7%	60.2%		35.8%	35.7%	29.6%
Citru vein enation virus	37.4%	35.7%	35.8%		34.8%	29.2%
Pepper enamovirus	35.8%	35.1%	35.7%	34.8%		29.6%
Grapevine enamovirus 1	29.1%	29.0%	29.6%	29.2%	29.6%	

CP

	AEV-1	BFTV-2	PEMV-1	CVEV	PEPV	GEV-1
Alfalfa enamovirus 1		73.4%	72.9%	37.9%	37.5%	31.9%
Birds-foot trefoil a. virus 2	73.4%		69.2%	37.1%	38.9%	30.1%
Pea enation mosaic virus 1	72.9%	69.2%		37.1%	38.0%	30.7%
Citru vein enation virus	37.9%	37.1%	37.1%		54.2%	23.4%
Pepper enamovirus	37.5%	38.9%	38.0%	54.2%		23.0%
Grapevine enamovirus 1	31.9%	30.1%	30.7%	23.4%	23.0%	

P3-5

	AEV-1	BFTV-2	PEMV-1	CVEV	PEPV	GEV-1
Alfalfa enamovirus 1		73.4%	72.9%	37.9%	37.5%	31.9%
Birds-foot trefoil a. virus 2	73.4%		69.2%	37.1%	38.9%	30.1%
Pea enation mosaic virus 1	72.9%	69.2%		37.1%	38.0%	30.7%
Citru vein enation virus	37.9%	37.1%	37.1%		54.2%	23.4%
Pepper enamovirus	37.5%	38.9%	38.0%	54.2%		23.0%
Grapevine enamovirus 1	31.9%	30.1%	30.7%	23.4%	23.0%	