Identification of novel RNA viruses associated to bird's-foot trefoil (*Lotus corniculatus*)

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:

Humberto J. Debat debat.humberto@inta.gob.ar

Nicolas Bejerman bejerman.nicolas@inta.gob.ar

Highlights

- *Lotus corniculatus* is a highly nutritious forage crop
- We present RNA evidence of a novel enamovirus and nucleorhabdovirus
- The discovered viruses provide insights into the RNA landscape of *L. corniculatus*
- The analysis of public SRA data constitutes an emerging source of novel plant viruses

Abstract

Bird's-foot trefoil (*Lotus corniculatus* L.) is a highly nutritious forage crop, employed for livestock foraging around the world. Despite the agronomical importance of this resilient crop, the related literature is rather scarce. 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 proposed bird's-foot trefoil enamovirus 1 (BFTEV-1) 5,736 nt virus sequence presents a typical 5'-PO-P1-2-IGS-P3-P5-3' enamovirus genome structure. The tentatively named bird's-foot trefoil nucleorhabdovirus (BFTNRV) genome organization is characterized by 13,626 nt long negative-sense single-stranded RNA. BFTNRV presents in its antigenome orientation six predicted gene products in the canonical order 3'-N-P-P3-M-G-L-5'. Phylogenetic analysis suggests that BFTNRV is closely related to *Datura yellow vein nucleorhabdovirus*, and that BFTEV-1 clusters into a monophyletic clade of legumes-associated enamoviruses. The bioinformatic reanalysis of SRA libraries deposited in the NCBI database constitutes an emerging approach to the discovery of novel plant viruses. The RNA viruses reported here provide a first glimpse of the virus landscape of this important crop. Future studies should assess the prevalence of BFTEV-1 and BFTNRV, and unravel whether the infection of these novel viruses is associated to specific symptoms.

Keywords

Lotus corniculatus; bird's-foot trefoil; *Enamovirus*; *Nucleorhabdovirus*; virus discovery

1. Introduction

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) (Escaray et al., 2012). It is a high quality forage that can be grazed or cut for hay or silage, and does not cause bloat in ruminants (Hannaway and Myers, 2004). 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 (McDonald and Susuki, 1983).

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 (Roossinck et al., 2015). 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 (Kim et al., 2018; Nibert et al., 2018). In a recent consensus statement report, Simmonds et al (2017) 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.

2. Materials and Methods

2.1. *Lotus corniculatus* transcriptome dataset

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 (2013), 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).

2.2. Transcriptome assembly, virus discovery and annotation

The 26,492,952 2x90 nt raw reads from SRA: SRS271068 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. The 55,473 obtained transcripts 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. 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. Virus annotation was instrumented as described in Debat (2017), 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. Pseudoknots were

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/). Importin-α dependent nuclear localization signals were predicted using cNLS Mapper available at http://nls-mapper.iab.keio.ac.jp/ and nuclear export signals were predicted using NetNES 1.1 available at www.cbs.dtu.dk/services/NetNES/.

2.3 Phylogenetic analysis

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 (BFTNRV) or G-INS-I (BFTEV-1). 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.

2.4 Data availability

The bird's-foot trefoil enamovirus 1 (BFTEV-1) and bird's-foot trefoil nucleorhabdovirus (BFTNRV) RNA sequences determined in this study are available in GenBank under accession numbers MH614261-MH614262, respectively [not yet released, included as supplementary data of this manuscript submission].

3. Results and discussion

3.1. Identification of novel RNA virus sequences in bird's-foot trefoil transcriptome data

De novo transcriptome assembly of SRA library SRS271068 (Wang et al., 2013) resulted in 55,473 transcripts, which were subjected to BLASTX searches against a refseq virus protein database. A 6,524 nt transcript obtained a significant hit (E-value = 0, identity = 60 %) with the L protein of datura yellow vein nucleorhabdovirus (DYVV; Dietzgen et al., 2015). Further similarity analyses resulted in the detection of three more DYVV–like transcripts of 3,953 nt, 2,354 nt and 1,432 nt which were extended and polished by iterative mapping of raw reads and subsequently reassembled into a 13,626 nt long RNA contig with overall 57,7 % sequence similarity with DYVV. The resulting RNA sequence, reconstructed with a total of 17,656 reads (mean coverage = 116.4X, reads per million (RPM) = 666.4) was further explored in detail (see section **3.2**). In addition, a 5,622 nt transcript obtained a significant hit (E-value = 0, identity = 74 %) with the P1-P2 fusion protein of Alfalfa enamovirus 1 (AEV-1; Bejerman et al., 2016). 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 resulting transcript was further assessed (see section **3.3**).

3.2. *Lotus corniculatus* nucleorhabdovirus

The tentatively named bird's-foot trefoil nucleorhabdovirus (BFTNRV) 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; Wu et al., 2018), DYVV and sonchus yellow net nucleorhabdovirus (SYNV; Goodin et al., 2001) (**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 (Dietzgen et al., 2017) given that no accessory genes are encoded by the BFTNRV genome. Like all plant rhabdoviruses, BFTNRV 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 (**Table 2**). Interestingly, BFTNRV consensus "gene junction" region sequence 3´-AUUCUUUUUGGUUGUA-5´ is identical to that of BCaRV, DYVV and SYNV (**Table 2**). All BFTNRV-encoded proteins contain a classical mono or bi-partite nuclear localization signal (NLS) (Dietzgen et al., 2017). Using cNLS mapper, 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**). NetNES also 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 (Dietzgen et al., 2015). Amino acid sequence comparisons between the deduced BFTNRV proteins and the corresponding sequences of other nucleorhabdoviruses revealed high similarity of BFTNRV to BCaRV, DYVV and SYNV. BFTNRV 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 (**Table 1**). A phylogenetic tree based on the complete amino acid sequence of the polymerase protein shows that BFTNRV clusters together with other viruses within the genus *Nucleorhabdovirus* (**Fig.1B**). BFTNRV appears to have a close evolutionary relationship with BCaRV, DYVV and SYNV, which share a similar genomic organization to that described for BFTNRV (Dietzgen et al., 2015; Heaton et al., 1989; Wu et al., 2018). It is tempting to suggest that BFTNRV, 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 BFTNRV should be considered a member of a new virus species in the genus *Nucleorhabdovirus*.

3.3. *Lotus corniculatus* enamovirus

The assembled sequence of bird's-foot trefoil enamovirus 1 (BFTEV-1) 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 (Fusaro et al., 2012). BFTEV-1 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. Figure 1A**). Interestingly, this motif was not identified in the non-legume infecting enamoviruses. The F-box-like motif is involved in silencing suppression (Fusaro et al., 2012), 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 (Miller et al., 1997), 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 (Demler and De Zoeten, 1991). In addition, an H-type highly structured (Free energy = -12.80) 40 nt pseudoknot was detected immediately following the slippery sequence, as expected (**Supp. Fig. 2B**). 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) (Demler and De Zoeten, 1991). 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 *Luteovirus*, *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 putative 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). BFTEV-1 presents the canonical stop-codon context for read-through of enamovirus (UGA-GGG; Firth and Brierley, 2012). 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 (Brown et al., 1996). 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 (Demler and De Zoeten, 1991). 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. Figure 1.B**). It is tempting to speculate that this motif could be involved in the transmission of BFTEV-1; 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) (Miller et al., 1997) in P5 (P3-P5 residues 221-424; E-value, 1.16e-48) were identified. These functional and structural results indicate that BFTEV-1 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 putative P4 movement protein, and luteoviruses lack a P0 gene (Domier, 2012). Therefore, BFTEV-1 appears to be an enamovirus because its genomic structure, characterized by presenting a P0 gene, and not encoding a P4 protein. The BFTEV-1 ORFs were compared to the predicted ORFs of AEV-1, PeMV-1, grapevine enamovirus-1 (GEV-1; Silva et al., 2017) and those of citrus vein enation virus (CVEV; Vives et al., 2013), in order to determine nucleotide and deduced amino acid sequence identities (**Table 3**). 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 3**). 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* (Domier et al., 2012) and also in *Enamovirus*. Consequently, BFTEV-1 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*, BFTEV-1 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. AEV-1 is associated to a complex of viruses responsible of the alfalfa dwarf disease (Bejerman et al., 2016). It is interesting that BFTEV-1 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.

4. Conclusions

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* (BFTNRV) and *Enamovirus* (BTFTEV-1). The RNA viruses reported here provide for the first time a glimpse of the virus landscape of this important crop. Future studies should assess the prevalence of BFTEV-1 and BFTNRV, 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.

Funding

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

References

Bejerman, N., Giolitti, F., de Breuil, S., Trucco, V., Dietzgen, R.G., Lenardon, S. (2016). Complete genome sequence of a new enamovirus from Argentina infecting alfalfa plants showing dwarfism symptoms. *Archives of Virology*, 161:2029-2032.

Brown, C. M., Dinesh-Kumar, S. P., & Miller, W. A. (1996). Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. *Journal of virology*, 70(9), 5884-5892.

Debat, H. J. (2017). An RNA Virome associated to the Golden orb-weaver Spider Nephila clavipes. *Frontiers in microbiology*, *8*, 2097.

Demler, S. A., & De Zoeten, G. A. (1991). The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *Journal of general virology*, 72(8), 1819-1834.

Dietzgen, R. G., Innes, D. J., & Bejerman, N. (2015). Complete genome sequence and intracellular protein localization of Datura yellow vein nucleorhabdovirus. *Virus research*, *205*, 7-11.

Dietzgen, R. G., Kondo, H., Goodin, M. M., Kurath, G., & Vasilakis, N. (2017). The family Rhabdoviridae: mono-and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus research*, *227*, 158-170.

Domier, L.L. (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.

Escaray, F.J., Men_endez, A.B., Garriz, A., Pieckenstain, F.L., Estrella, M.J.,Castagno, L.N., Carrasco, P., Sanjuan, J., Ruiz, O.A. (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.

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *Journal of General Virology*, 93(7), 1385-1409.

Fusaro, A. F., Correa, R. L., Nakasugi, K., Jackson, C., Kawchuk, L., Vaslin, M. F., and Waterhouse, P. M. (2012). The Enamovirus P0 protein is a silencing suppressor which inhibits local and systemic RNA silencing. *Virology*, 426:178-187.

Goodin, M. M., Austin, J., Tobias, R., Fujita, M., Morales, C., & Jackson, A. O. (2001). Interactions and nuclear import of the N and P proteins of sonchus yellow net virus, a plant nucleorhabdovirus. *Journal of virology*, 75(19), 9393-9406.

Heaton, L. A., Hillman, B. I., Hunter, B. G., Zuidema, D., & Jackson, A. O. (1989). Physical map of the genome of sonchus yellow net virus, a plant rhabdovirus with six genes and conserved gene junction sequences. *Proceedings of the National Academy of Sciences*, *86*(22), 8665-8668.

Hannaway, D. B., Myers, D. (2004). Birdsfoot trefoil (*Lotus corniculatu*s L.). Oregon State University. Species Selection Information System.

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.

McDonald, J.G., Susuki, M. (1983). Occurrence of alfalfa mosaic virus in Prince Edward Island. *Canadian Plant Disease Survey*, 63:47-50.

Nibert, M. L., Vong, M., Fugate, K. K., & Debat, H. J. (2018). Evidence for contemporary plant mitoviruses. *Virology*, 518, 14-24.

Roossinck MJ, Martin DP, Roumagnac P. (2015). Plant virus metagenomics: advances in virus discovery. *Phytopathology*, 105: 716–727.

Silva, J. M. F., Al Rwahnih, M., Blawid, R., Nagata, T., & Fajardo, T. V. M. (2017). Discovery and molecular characterization of a novel enamovirus, Grapevine enamovirus-1. *Virus genes*, 53(4), 667-671.

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., ... & Hull, R. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3), 161.

Vives, M. C., Velázquez, K., Pina, J. A., Moreno, P., Guerri, J., & Navarro, L. (2013). Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. *Phytopathology*, 103(10), 1077-1086.

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(3-4), 131-139.

Wu, L. P., Yang, T., Liu, H. W., Postman, J., Li, R. (2018). Molecular characterization of a novel rhabdovirus infecting blackcurrant identified by high-throughput sequencing. *Archives of Virology*, 162, 2493–4.

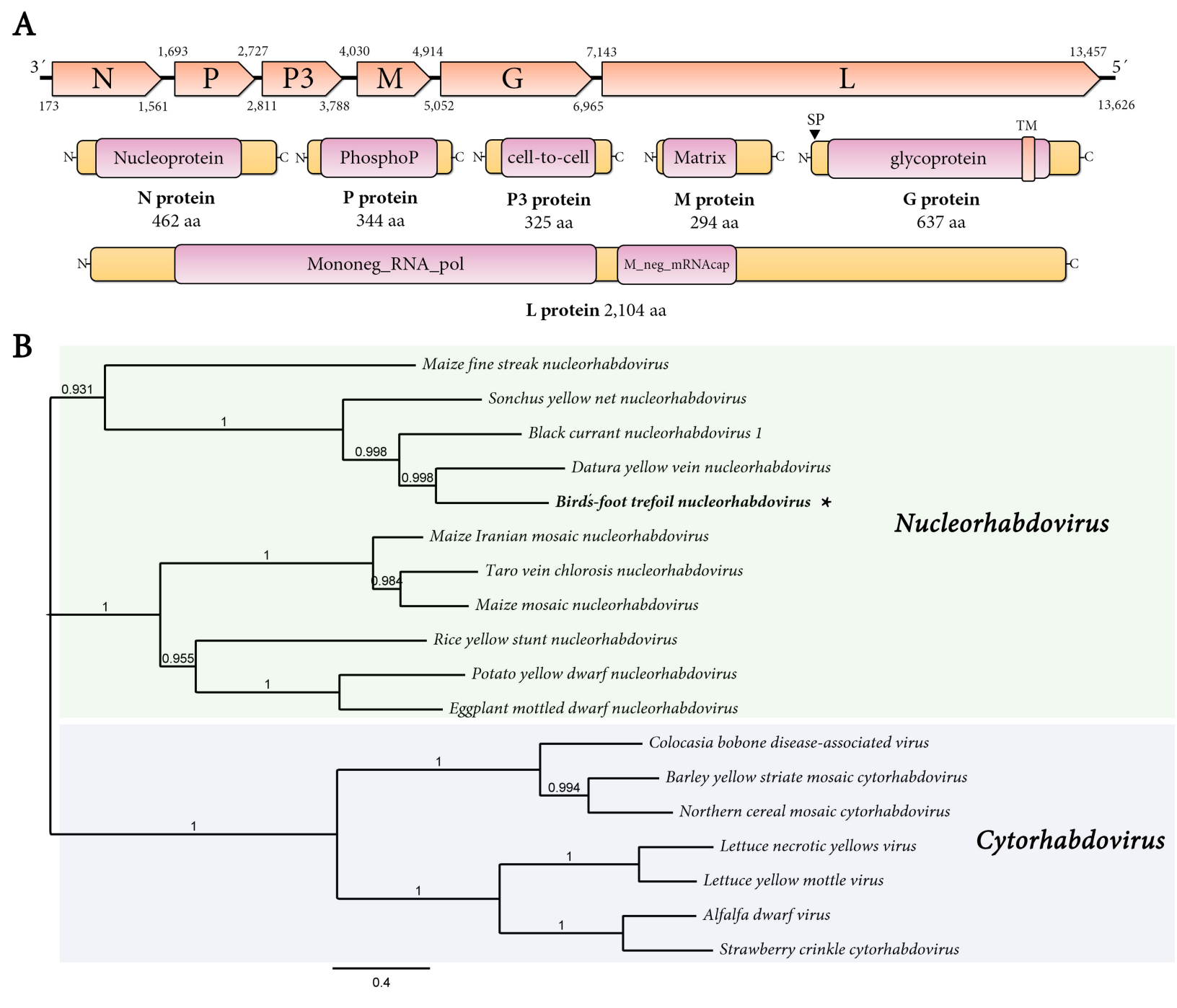
Figure legends

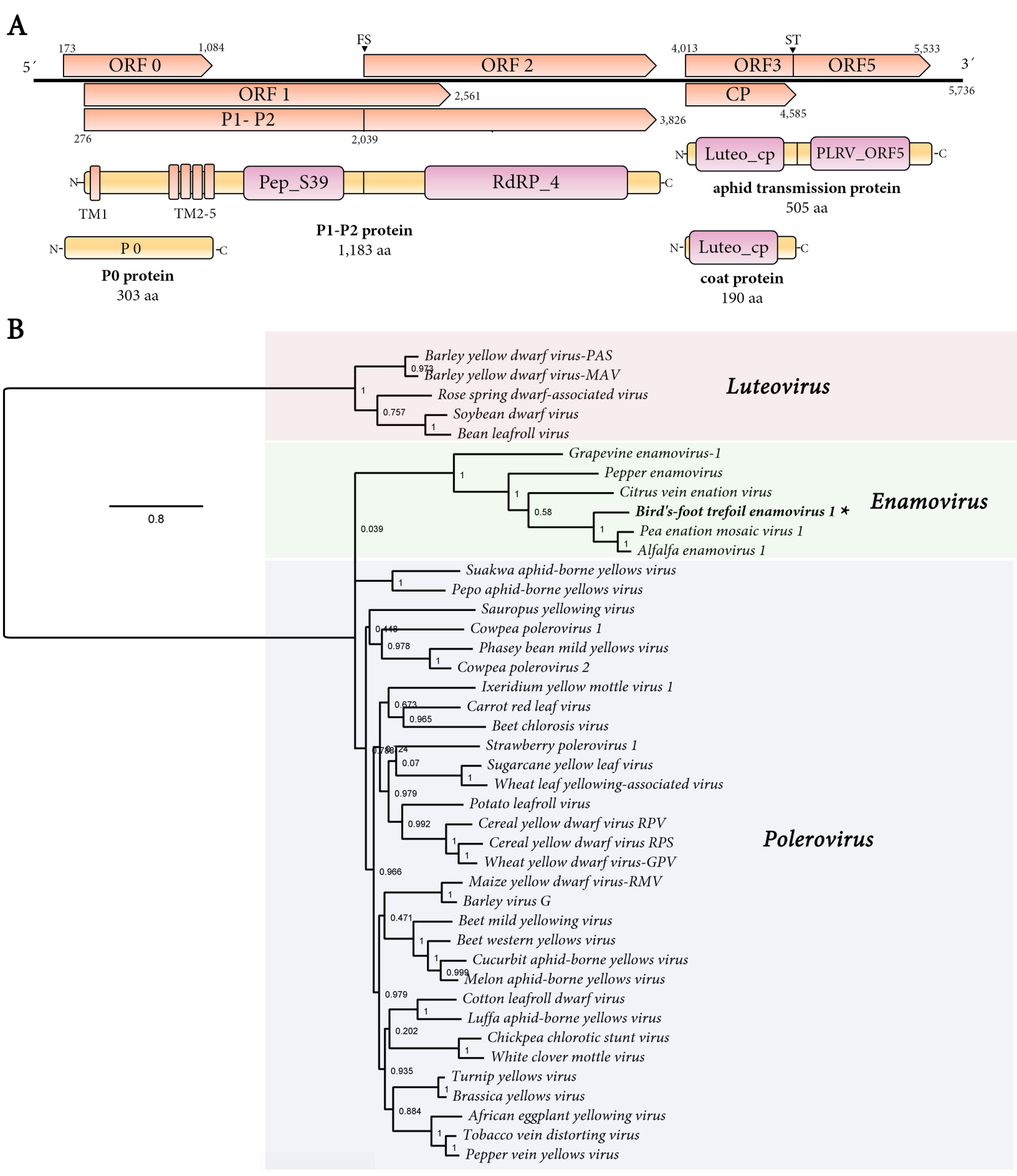
Figure 1. Structural characterization and phylogenetic insights of bird's-foot trefoil nucleorhabdovirus (BFTNRV) (**A**) Genome graphs depicting architecture and predicted gene products of BFTNRV. The predicted coding sequences are shown in orange arrow rectangles, start and end coordinates are indicated. 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).

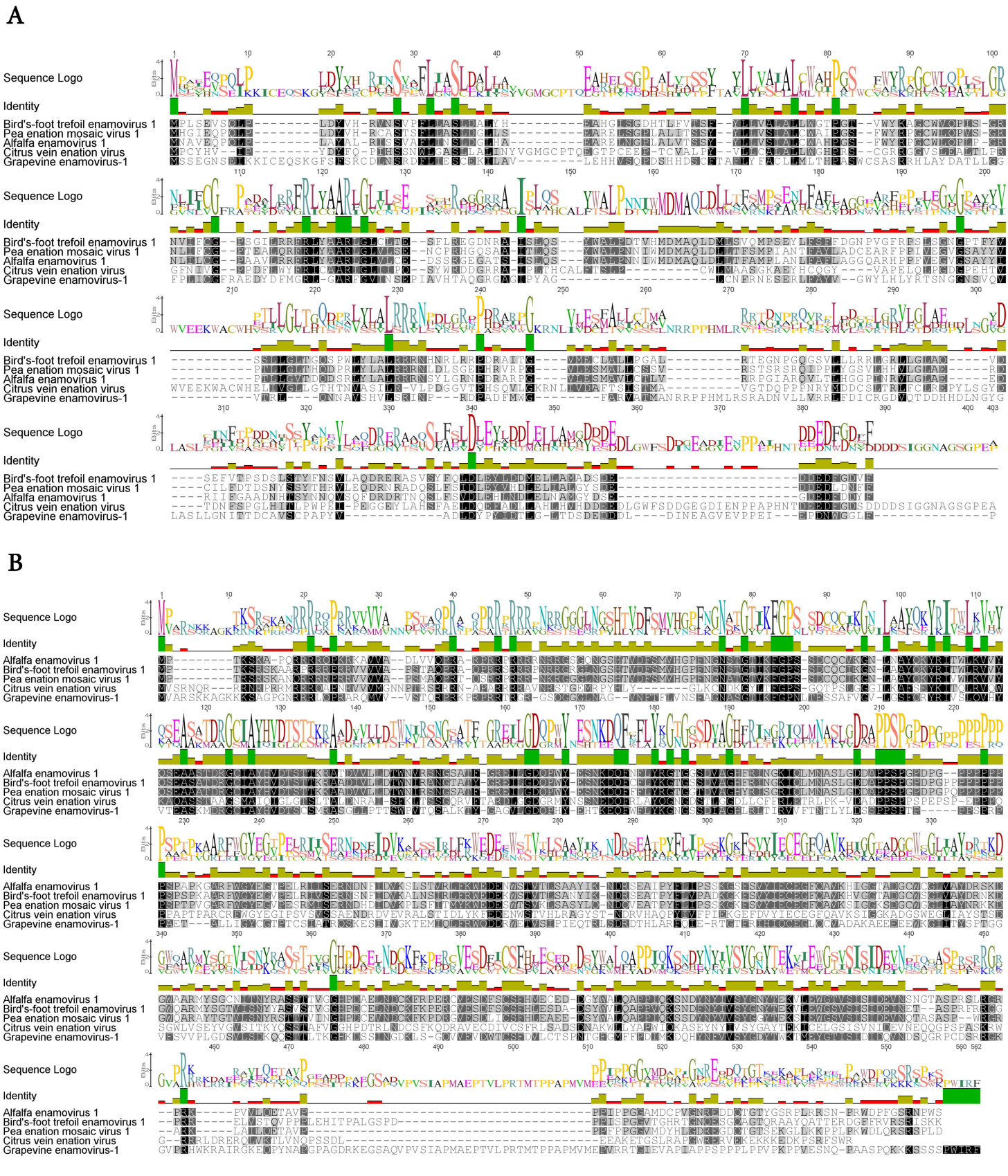
Figure 2. Structural characterization and phylogenetic insights of bird's-foot trefoil enamovirus 1 (BFTEV-1) (**A**) Genome graphs depicting architecture and predicted gene products of BFTEV-1. The predicted coding sequences are shown in orange arrow rectangles, start and end coordinates are indicated. 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: 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 BFTEV-1 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 1**.

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 immediately downstream the slippery sequence.


A
3′
N
P
P3
M
G
L
5′
173
1,561
1,693
2,727
2,811
3,788
4,030
4,914
5,052
6,965
7,143
13,457
13,626
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
Bird's-foot trefoil nucleorhabdovirus ⋆
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
1
0.955
1
1
0.994
1
1
1
1
0.4


A
173
1,084
FS
4,013
ST
5,533
5′
ORF 0
ORF 2
ORF3
ORF5
3′
ORF 1
2,561
CP
5,736
4,585
P1- P2
3,826
276
2,039
N
Luteo_cp
PLRV_ORF5
C
Pep_S39
RdRP_4
TM1
TM2-5
aphid transmission protein
505 aa
P1-P2 protein
1,183 aa
P 0
P0 protein
303 aa
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
Bird's-foot trefoil enamovirus 1 *
Pea enation mosaic virus 1
Alfalfa enamovirus 1
Polerovirus
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


A
Sequence Logo
Identity
Bird's-foot trefoil enamovirus 1
Pea enation mosaic virus 1
Alfalfa enamovirus 1
Citrus vein enation virus
Grapevine enamovirus-1
B
Sequence Logo
Identity
Alfalfa enamovirus 1
Bird's-foot trefoil enamovirus 1
Pea enation mosaic virus 1
Citrus vein enation virus
Grapevine enamovirus-1

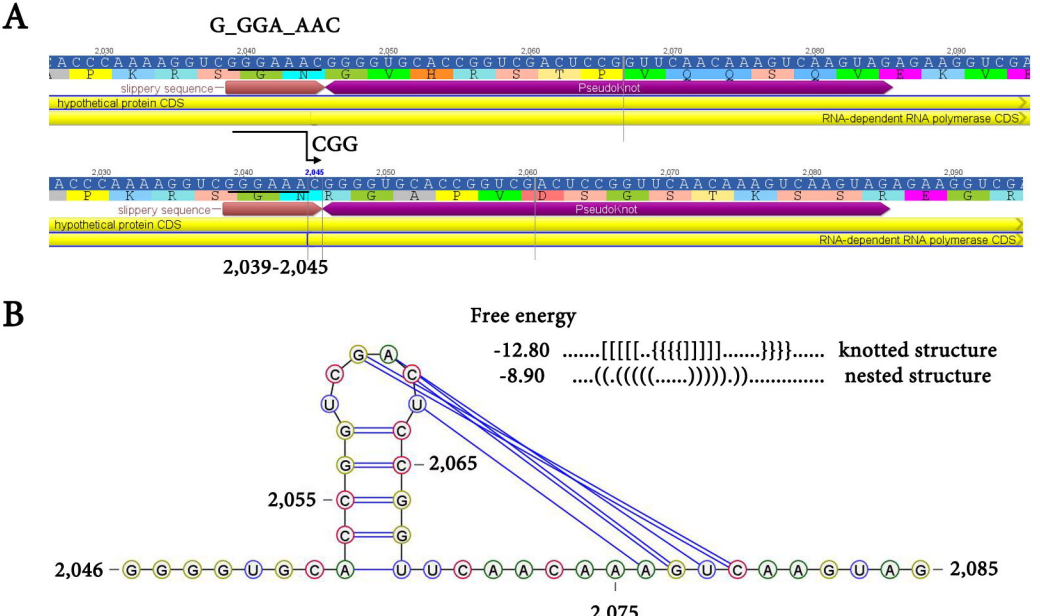

A
G_GGA_AAC
2,030
2,040
2,050
2,060
2,070
2,080
2,090
slippery sequence
PseudoKnot
hypothetical protein CDS
RNA-dependent RNA polymerase CDS
CGG
2,045
2,039-2,045
B
Free energy
-12.80 .....[[[[[..{{{{]]]]].....}}}}..... knotted structure
-8.90 ....((.(((((.....)))))).)..... nested structure
2,065
2,055
2,046
2,085
2,075

Table 1. Summary of BFTNRV genome encoded proteins.

ORF	Gene	Putative function	Lenght (aa)	MW (kD)	IEP[a]	Predicted NLS [cNLS Mapper score]	BLASTP e-value	Identity DYVV (%)	Identity BCaRV (%)	Identity SYNV (%)
1	N	Nucleocapsid protein	462	52.1	8.9	Bipartite [7.5]	0.0	58.4	50.1	46.4
2	P	Phosphoprotein	344	38.3	4.5	Bipartite [3.4]	3,00E-50	34.3	27.6	23.0
3	P3	Movement protein	325	36.7	8.7	Bipartite [3.4]	6,00E-98	42.6	28.9	22.2
4	M	Matrix protein	294	32.6	8.2	$_{231}$LKKTLKRRKIA$_{241}$ [9]	2,00E-54	32.6	32.0	27.7
5	G	Glycoprotein	637	71.1	6.8	Bipartite [3.3]	0.0	46.7	39.2	37.0
6	L	Polymerase	2,104	240.1	7.7	$_{1638}$SSSKRKRTSI$_{1647}$[9.5]	0.0	50.1	48.9	43.2

[a]IEP, isoelectric point pH.

Table 2. BFTNRV genome predicted intergenic regions.

	Poly-adenylation	Intergenic spacer	Transcription start	
3' leader	UUUCUUUU	GUA	UUGUA	N gene
N gene	AUUCUUUUU	GG	UUGUA	P gene
P gene	AUUCUUUUU	GG	UUGUA	P3 gene
P3 gene	AUUCUUUUU	GG	UUGAU	M gene
M gene	AUUCUUUUU	GG	UUGAA	G gene
G gene	AUUCUUUUU	GG	UUGAA	L gene
Consensus	(A/U)UUCUUUU(U)	G(G/U)(A)	UUG(U/A)(A/U)	

Table 3. Amino acid (aa)/nucleotide (nt) sequence identities (%) of BFTEV-1 ORFs and gene products to those of enamoviruses.

Virus/Gene	ORF0	ORF1	ORF1-2	ORF3	ORF3-5
AEV-1	48.5/61.5	51.6/60.4	62.0/66.7	85.3/75.4	73.2/70.4
PeMV-1	50.8/61.7	48.8/59.3	60.3/65.7	82.6/77.1	69.1/69.0
CVEV	23.0/34.7	22.2/35.1	35.7/44.8	26.6/41.7	37.7/44.9
GEV-1	15.7/35.2	21.1/31.8	30.7/42.2	44.0/47.2	30.9/40.7

Abbreviations: AEV-1, Alfalfa enamovirus 1; PeMV-1, pea enation mosaic virus-1; CVEV, citrus vein enation virus; GEV-1, grapevine enamovirus-1.