

Exploration of plant transcriptomes reveals five putative novel poleroviruses and an enamovirus

V Kavi Sidharthan

Institute of Forest Biodiversity

Krishnan Nagendran

Indian Institute of Vegetable Research

Virendra Kumar Baranwal (✉ vbaranwal2001@yahoo.com)



Indian Agricultural Research Institute <https://orcid.org/0000-0002-5251-0905>

Research Article

Keywords: poleroviruses, enamoviruses, *Celmisia lyallii* enamovirus

Posted Date: October 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-977606/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Virus Genes on March 26th, 2022. See the published version at <https://doi.org/10.1007/s11262-022-01896-7>.

Abstract

Transcriptome datasets available in public domain serve as valuable resource for identification and characterization of novel viral genomes. Poleroviruses are economically important plant-infecting RNA viruses belonging to the family *Luteoviridae*. In the present study, we explored the plant transcriptomes available in public domain and identified five putative novel poleroviruses tentatively named as *Foeniculum vulgare* polerovirus (FvPV), *Kalanchoe marnieriana* polerovirus (KmPV), *Paspalum notatum* polerovirus (PnPV), *Piper methysticum* polerovirus (PmPV), *Trachyspermum ammi* polerovirus (TaPV) and a novel enamovirus named as *Celmisia lyallii* enamovirus (CIEV) in *Foeniculum vulgare*, *Kalanchoe marnieriana*, *Paspalum notatum*, *Piper methysticum*, *Trachyspermum ammi* and *Celmisia lyallii*, respectively. Coding-complete genomes (5.56–5.74 kb) of CIEV, KmPV, PnPV, PmPV and TaPV were recovered while only the partial genome of FvPV could be recovered. The genome organization of identified viruses except CIEV is 5′-P0-P1-P2-P3a-P3-P4-P5-3′ while that of CIEV is 5′-P0-P1-P2-P3-P5-3′. Phylogenetic analysis and genome organization revealed the relatedness of FvPV, KmPV, PnPV, PmPV and TaPV to other poleroviruses and CIEV to other enamoviruses.

Main Text

The family *Luteoviridae* includes three important plant pathogenic genera- *Enamovirus*, *Luteovirus* and *Polerovirus* that are distinguished based on their genome organization, replication and expression strategies and genome organization [8]. Members of the family *Luteoviridae* possess monopartite positive-sense single-stranded RNA genomes of lengths 5–6 kb with five or six major Open Reading Frames (ORF) [8, 21]. Poleroviruses are economically important as they affect the quality and quantity of the economic produce [8]. In general, poleroviruses are transmitted by aphids in a persistent circulative manner [8].

In recent times, plant RNA datasets and contigs available in NCBI-Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA) databases, respectively have emerged as one of the major resources for the discovery of putative novel viruses [2, 7, 22, 23]. Such identified novel viruses could be regarded as *bona fide ones* based on the report by Simmonds and colleagues [25] who suggested the incorporation of viruses that were identified only from metagenomic data into the official taxonomy scheme of the International Committee on Taxonomy of Viruses (ICTV) for comprehensive characterization of global virome. Considering the widespread occurrence of poleroviruses in plants and availability of transcriptomes of a large number of plant species in public domain, we hypothesized that the novel poleroviral sequences might be present in publicly available plant transcriptomes. Thus, the present study aimed to explore the publicly available plant transcriptomes for the discovery of novel poleroviral sequences that otherwise would require an expensive Next Generation Sequencing (NGS) based viromic study in each individual plant species.

To identify poleroviral sequences, reference RNA-dependent RNA polymerase (RdRp) sequence (P1-P2) of potato leafroll virus (PLRV) (NP056748.3) was used as query in tBLASTn analysis (expect threshold: 0.05, word size: 6; matrix: BLOSUM62) against TSA database and the search was limited to Viridiplantae (taxid:33090). The resulting hits were further shortlisted as putative poleroviral contigs based on the e-value cut-off (1e-50) and query coverage (>50%). Hits of lengths >4 kb were only considered for reliable identification and to avoid possible misidentification using smaller contigs. Putative poleroviral contigs were further analysed for the presence of intact ORFs using NCBI ORF Finder

(<https://www.ncbi.nlm.nih.gov/orffinder/>). In cases where coding-complete genomes could not be identified, all the available RNA-seq datasets of the corresponding plant species, including the ones from where the contigs originated, were retrieved, trimmed using Trimmomatic v.0.39 [3], assembled using SPAdes v.3.13.1 [5] and subjected to BLASTn analysis (e-value cut-off: 1e-5) against the recovered putative poleroviral contigs and PLRV reference genome (RefSeq) (NC001747) using NCBI BLAST+ v.2.9.0 to obtain coding-complete genomes. Further, the protein sequences coded by the recovered genomes were subjected to molecular weight estimation, motif and transmembrane helix (TMH) prediction using the tools mentioned in [22]. The -1 ribosomal frameshift sites in the recovered genomes were predicted using the KnotInFrame tool (<https://bibiserv.cebitec.uni-bielefeld.de/knotinframe>) while ORF3a was predicted by aligning PLRV RefSeq with recovered poleroviral genomes using CLUSTALW tool in MEGA7 [13]. Conserved domains in the protein sequences of identified viruses were determined after MUSCLE alignment in MEGA7 and visualized in WebLogo 3 (<https://weblogo.berkeley.edu/>) [6]. Phylogenetic tree was constructed using Neighbourhood Joining (NJ) method and Poisson model with 1000 bootstrap replicates after MUSCLE alignment of RdRp sequences (P1–P2) of identified and retrieved poleroviral and enamoviral sequences in MEGA7. Using the recovered coding-complete genomes as queries in MEGABLAST analysis (expect threshold: 0.05, word size: 28), all the available RNA-seq datasets of corresponding plant species were individually searched for the presence of identified poleroviral sequences. Libraries containing at least 10 viral reads were regarded as virus-positive.

A total of five putative novel poleroviruses and an enamovirus tentatively named as *Celmisia lyallii* enamovirus (CIEV), *Foeniculum vulgare* polerovirus (FvPV), *Kalanchoe marnieriana* polerovirus (KmPV), *Paspalum notatum* polerovirus (PnPV), *Piper methysticum* polerovirus (PmPV) and *Trachyspermum ammi* polerovirus (TaPV) were identified in various tissues of *Celmisia lyallii*, *Foeniculum vulgare*, *Kalanchoe marnieriana*, *Paspalum notatum*, *Piper methysticum* and *Trachyspermum ammi*, respectively (Table 1). Interestingly, KmPV and PnPV were also identified in sRNA libraries of respective plant species (Table 1). Coding-complete genomes (5.56–5.70 kb long) of CIEV, KmPV, PnPV, TaPV were obtained directly after initial tBLASTn analysis while the coding-complete genome of PmPV (5.74 kb) could be obtained only after assembly of one of the transcriptome libraries of *P. methysticum*. However, coding-complete genome of FvPV could not be obtained even after assembly of available transcriptome libraries of *F. vulgare*, though two non-redundant contigs (lengths 1.35 kb and 4.20 kb) of FvPV were recovered (Table S1).

The genomes of FvPV, KmPV, PnPV, PmPV and TaPV contains seven ORFs designated as ORF0,1,2,3a,3,4 and 5 that coded for P0, P1, P1–P2, P3a, P3, P4 and P3–P5 proteins, respectively while that of CIEV possessed five ORFs- ORF0,1,2,3 and 5 encoding P0, P1, P1–P2, P3 and P3–P5 proteins, respectively (Fig. 1). P0 protein of the identified viruses ranged from 243–299 aa in size with an estimated molecular weight of 28.0–34.1 kDa and contained a luteovirus P0 motif excepting CIEV. P1 protein of the identified viruses (620–810 aa) with molecular weights ranging from 68.4–90.1 kDa possessed a peptidase S39 motif (Table S1). The conserved H(X₂₅)D(X_{70–80})GXSG domain of S39 peptidase [24] was observed in P1 protein of all the identified viruses (Fig. S1a). ORF1 possess a -1 ribosomal frameshift site at nt position 1998 (CIEV), 1665 (FvPV), 1564 (KmPV), 1586 (PnPV), 1743 (PmPV) and 1495 (TaPV) facilitating the expression of ORF2 of the identified viruses as P1–P2 fusion protein. The heptanucleotide slippery sequence facilitating ribosomal frameshift in CIEV is TTAAAC which is identical to that of grapevine enamovirus 1 (GEV-1) [24]. On contrary, the slippery sequence

of FvPV, KmPV, PnPV, PmPV and TaPV is GGGAAAC which is identical to that of cardamom polerovirus [23]. Similar to other poleroviruses and enamoviruses [7], the knotted structure of 40 nt length was predicted immediately downstream of the slippery sequence in genome sequences of all the identified viruses. P1–P2 protein of identified viruses (1044–1243 aa) with estimated molecular weights of 117.6–139.3 kDa contained a viral RdRp motif in addition to the peptidase S39 motif (Table S1). The conserved GXXXTXXXN(X_{25–40})GDD motif [8] was observed in P1–P2 fusion protein of all the identified viruses (Fig. S1b). Three TMHs were predicted in P1 and P1–P2 proteins of FvPV, PmPV and TaPV while two, four and five TMHs were predicted in P1 and P1–P2 proteins of KmPV, PnPV and CIEV, respectively (Table S1). It is worthy of note that the predicted TMHs are located at the amino terminal regions similar to other poleroviruses possibly facilitating the formation of replication complexes [8]. ORF3a, predicted in all the identified viral genomes except CIEV, encodes protein P3a (44–45 aa) of 4.9–5.0 kDa by a non-canonical start codon. Similar to other poleroviruses [8], a TMH was predicted in P3a of identified viruses containing ORF3a. P3 protein (192–211 aa) of identified viruses with molecular weights 20.7–23.7 kDa contained a luteovirus coat protein domain. Protein P4 (174–196 aa) of 19.6–21.9 kDa containing a putative luteoviral VPg genome linked protein motif was encoded by all the identified viruses except CIEV. ORF5 is possibly expressed via a translational read-through of ORF3 stop codon to produce P3–P5 protein (456–698 aa) of 50.5–78.4 kDa with PLRV read-through protein motif in the identified viruses (Table S1). Length and molecular weight of P3–P5 protein could not be determined for FvPV as its coding-complete genome was not recovered. However a smaller contig (1.35 kb) containing 5' truncated ORF5 sequences encoding a truncated protein with PLRV read-through protein motif was recovered (Table S1). ORF3 stop codon read-through site sequence of all the identified viruses including CIEV is AAUAGGUA which is identical to that of other poleroviruses [9]. The C-rich block with CCNNNN tandem repeat motif commonly found in luteoviruses [4] was determined after 9–12 nt downstream of the read-through site in all the identified viruses except FvPV with one CCCCA motif in each.

BLASTp analysis of encoded proteins revealed the sequence similarities of FvPV, KmPV, PnPV, PmPV and TaPV to other poleroviruses and CIEV to other enamoviruses (Table S1). Phylogenetic tree constructed based on the P1–P2 sequences grouped CIEV with citrus vein enation virus (CVEV), PnPV with cowpea polerovirus 2 (CPPV2) and TaPV with carrot red leaf virus (CtRLV). FvPV was placed in a sister clade to TaPV and CtRLV. KmPV was related to the poleroviruses of beet and turnip while PmPV was distantly related to PnPV, CPPV2 and poleroviruses of maize (Fig. 2).

All the proteins encoded by the identified viruses shared less than 90% sequence identities at maximum query coverage with any of the respective protein sequences of known poleroviruses/enamoviruses except for P3a of PnPV. Members of the family *Luteoviridae* can be regarded as a new species if any of the encoded proteins displayed more than 10% sequence divergence from the corresponding protein sequences of the existing members [21]. Thus, based on the sequence based species demarcation criteria, genome organization, predicted motifs and phylogeny, FvPV, KmPV, PnPV, PmPV and TaPV can be regarded as putative novel members of the genus *Polerovirus* while CIEV can be regarded as a putative novel enamovirus. It is worth mentioning that the reads of few of the identified viruses were discovered in mRNA/sRNA libraries derived from different tissues and/or cultivars of the respective plant species implying that these viruses would possibly be widespread in nature. The plant species in which the novel poleroviruses were identified includes economically important culinary spice crops of the family *Apiaceae*- *F. vulgare* and *T. ammi* [10, 16], a

traditionally important medicinal plant- *P. methysticum* [17], an important grass in natural grasslands of the western Hemisphere- *P. notatum* [18] and a wild relative of an important potted plant- *K. marnieriana* [12]. Further, phylogenetic analysis grouped together the poleroviruses of plants belonging to Apiaceae in a monophyletic clade suggesting their possible origin from a common ancestor. On the other hand, CIEV forms a distinct subgroup of enamoviruses alongwith CVEV and pepper enamovirus. Interestingly, a cytorhabdovirus- *Trachyspermum ammi virus 1* [2] and a putative tymovirus- *Kava virus 1* [1] were identified in the same Bioprojects from where TaPV and PmPV, respectively were discovered in the current study.

In the present study, five putative novel poleroviruses and a novel enamovirus were identified in six plant species from public domain databases. Besides serving as a valuable resource for development of detection assays, the recovered genome sequences of identified viruses will help in better understanding the evolution and genomic features of the identified virus groups. Further studies are needed to understand the biological properties, distribution and economic importance of the identified viruses.

Declarations

Acknowledgements

We thank ICAR for the Advanced Supercomputing Hub for Omics Knowledge in Agriculture (ASHOKA) facility available at ICAR-IASRI, New Delhi, India. The authors are grateful to the Head, Division of Genetics and Tree Improvement, the Director, IFB (ICFRE), Hyderabad, the Head, Division of Plant Pathology and the Director, ICAR-IARI, New Delhi for their support.

Conflict of interest

There is no conflict of interest.

References

1. Bejerman N, Debat H (2021) Exploring the tymovirids landscape through metatranscriptomics data. bioRxiv <https://doi.org/10.1101/2021.07.15.452586>
2. Bejerman N, Dietzgen RG, Debat H (2021) Illuminating the plant rhabdovirus landscape through metatranscriptomics data. *Viruses* 13:1304
3. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
4. 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
5. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience* 8:giz100
6. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190

7. Debat HJ, Bejerman N (2019) Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume-associated enamoviruses and rhabdoviruses. *Arch Virol* 164:1419–1426
8. Delfosse VC, Barrios Barón MP, Distefano AJ (2021) What we know about poleroviruses: Advances in understanding the functions of polerovirus proteins. *Plant Pathol* 70:1047–1061
9. Firth AE, Brierley I (2012) Non-canonical translation in RNA viruses. *J Gen Virol* 93:1385–1409
10. Howyzeh MS, Noori SA, Shariati V, Amiripour M (2018) Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. *Sci Rep* 8:13405
11. Hu Z, Nie Z, Yan C, Huang H, Ma X, Wang Y, Ye N, Tuskan GA, Yang X, Yin H (2021) Transcriptome and degradome Profiling Reveals a Role of miR530 in the Circadian Regulation of Gene Expression in *Kalanchoë marnieriana*. *Cells* 10:1526
12. Kuligowska K, Lütken H, Christensen B, Müller R (2015) Quantitative and qualitative characterization of novel features of *Kalanchoë* interspecific hybrids. *Euphytica* 205:927–940
13. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
14. Ortiz JP, Leblanc O, Rohr C, Grisolia M, Siena LA, Podio M, Colono C, Azzaro C, Pessino SC (2019) Small RNA-seq reveals novel regulatory components for apomixis in *Paspalum notatum*. *BMC Genom* 20:487
15. Ortiz JP, Revale S, Siena LA, Podio M, Delgado L, Stein J, Leblanc O, Pessino SC (2017) A reference floral transcriptome of sexual and apomictic *Paspalum notatum*. *BMC Genom* 18:318
16. Palumbo F, Vannozzi A, Vitulo N, Lucchin M, Barcaccia G (2018) The leaf transcriptome of fennel (*Foeniculum vulgare* Mill.) enables characterization of the t-anethole pathway and the discovery of microsatellites and single-nucleotide variants. *Sci Rep* 8:10459
17. Pluskal T, Torrens-Spence MP, Fallon TR, De Abreu A, Shi CH, Weng JK (2019) The biosynthetic origin of psychoactive kavalactones in kava. *Nat Plants* 5:867–878
18. Podio M, Colono C, Siena L, Ortiz JP, Pessino SC (2021) A study of the heterochronic sense/antisense RNA representation in florets of sexual and apomictic *Paspalum notatum*. *BMC Genom* 22:185
19. Samarth, Lee R, Kelly D, Turnbull MH, Macknight RC, Poole AM, Jameson PE (2021) Molecular control of the floral transition in the mast seeding plant *Celmisia lyallii* (Asteraceae). *Mol Ecol* 30:1846–1863
20. Samarth, Lee R, Song J, Macknight RC, Jameson PE (2019) Identification of flowering-time genes in mast flowering plants using *de novo* transcriptomic analysis. *PloS one* 14:e0216267
21. Shen P, Tian X, Zhang S, Ren F, Li P, Yu YQ, Li R, Zhou C, Cao M (2018) Molecular characterization of a novel luteovirus infecting apple by next-generation sequencing. *Arch Virol* 163:761–765
22. Sidharthan VK, Baranwal VK (2021) Mining of the water hyssop (*Bacopa monnieri*) transcriptome reveals genome sequences of two putative novel rhabdoviruses and a solendovirus. *Arch Virol* 166:1985–1990
23. Sidharthan VK, Kalaivanan NS, Baranwal VK (2021) Identification of two putative novel RNA viruses in the transcriptome datasets of small cardamom. *Plant Gene* 27:100305
24. Silva JM, Al Rwahnih M, Blawid R, Nagata T, Fajardo TV (2017) Discovery and molecular characterization of a novel enamovirus, Grapevine enamovirus-1. *Virus Genes* 53:667–671

25. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R (2017) Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 15:161–168
26. Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJ, Chakrabarty R, Desgagné-Penix I, Haslam TM, Kim YB, Liu E, MacNevin G (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol* 166:122–134

Tables

Table 1. Summary of identification of novel poleroviruses and enamovirus from RNA libraries of various plant species

s	Library/contig from where the genome was derived	Plant Species	Cultivar/Isolate/Genotype	Tissue	SRA accession of virus-positive libraries*	Reference
	GHOT01029599.1	<i>Celmisia lyallii</i>	Hook.f. ecotype	Leaf, Root	SRR9591156 (mRNAome)	[20]
				Leaf	SRR10237409, SRR10237410, SRR10237411, SRR10237412, SRR10237413, SRR10237414, SRR10237415, SRR10237416, SRR10237417 (mRNAome)	[19]
7	GGAC01010024.1 (contig 1), SRR6265712–19 (contig 2)	<i>Foeniculum vulgare</i>	OL2	Leaf	SRR6265712, SRR6265713 (mRNAome)	[16]
			OL1	Leaf	SRR6265714, SRR6265715 (mRNAome)	
			OL164	Leaf	SRR6265716, SRR6265717 (mRNAome)	
			OL9	Leaf	SRR6265718, SRR6265719 (mRNAome)	
V	GIXV01025569.1	<i>Kalanchoe marnieriana</i>	Baijizhiwu	Leaf	SRR13275786 (mRNAome), SRR13275787 (degradome)	[11]
				Epidermal peel	SRR13275777, SRR13275780, SRR13275781, SRR13275782, SRR13275788, SRR13275792, SRR13275794, SRR13275795, SRR13275796 (sRNAome)	
				Mesophyll	SRR13275778, SRR13275779, SRR13275783, SRR13275784, SRR13275785, SRR13275789, SRR13275790, SRR13275791,	

					SRR13275793 (sRNAome)	
7	GIUR01167265.1	<i>Paspalum notatum</i>	Q4117	spikelets	SRR8377175, SRR8377176, SRR8377177, SRR8377180 (mRNAome)	[18]
			C4-4X	spikelets	SRR8377178, SRR8377179, SRR8377181, SRR8377182 (mRNAome)	
			Q4117	Flower	SRR5241464 (sRNAome), SRR3945053 (mRNAome)	[14, 15]
			C4-4X	Flower	SRR5241463 (sRNAome), SRR3945054 (mRNAome)	
V	SRR688037	<i>Piper methysticum</i>	na	Root	SRR8053129, SRR8053131 (mRNAome)	[17]
				Leaf	SRR8053130, SRR8053132 (mRNAome)	
			na	na	SRR610337, SRR688037 (mRNAome)	[26]
7	GHOD01041458.1	<i>Trachyspermum ammi</i>	Arak ecotype	Inflorescence	SRR5137050, SRR5137051 (mRNAome)	[10]

*Libraries containing at least 10 reads are considered as virus positive.

Figures

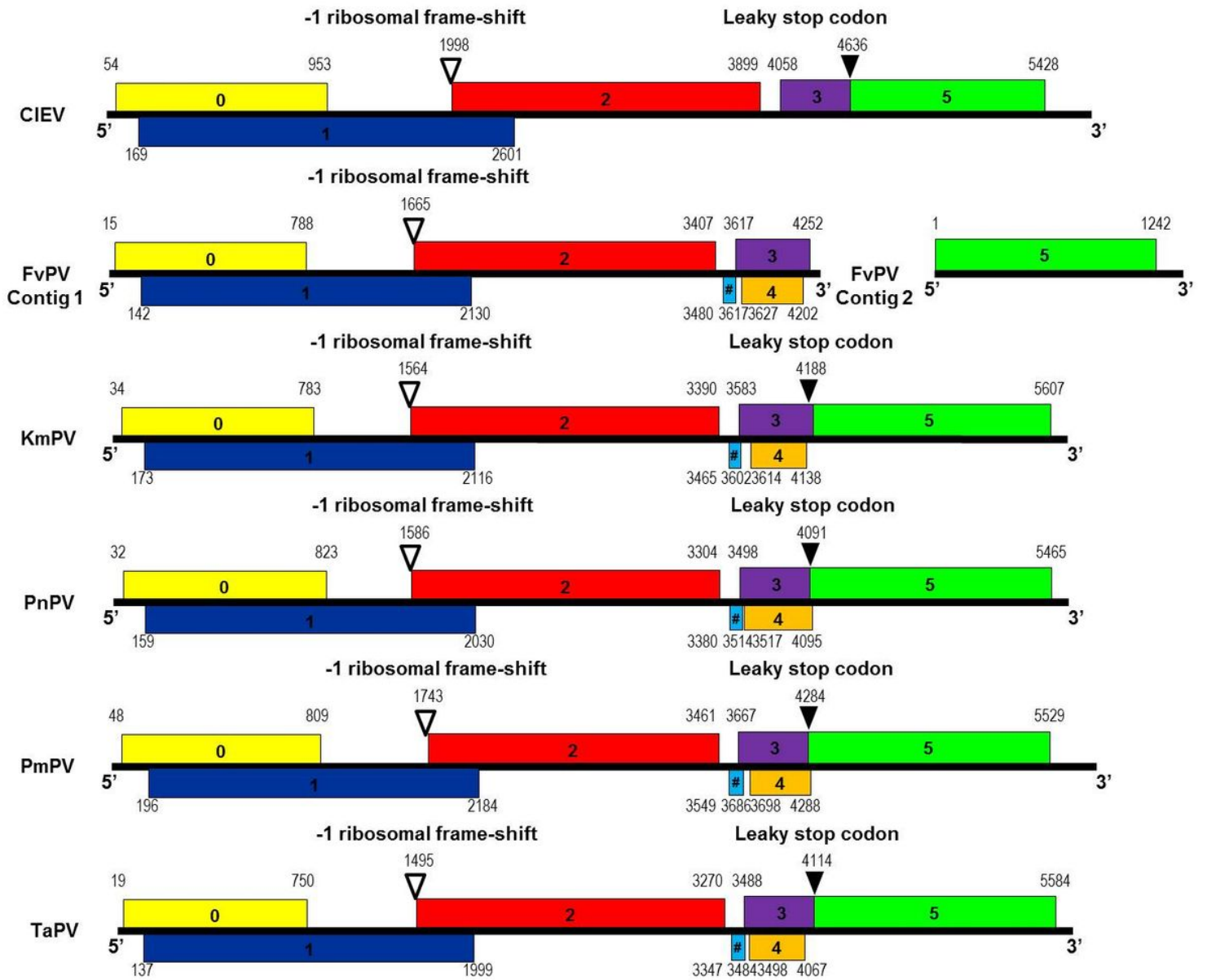


Figure 1

Genome organization of putative novel poleroviruses and enamovirus identified in the current study. Each ORF is shown in a different color. # indicates ORF3a.

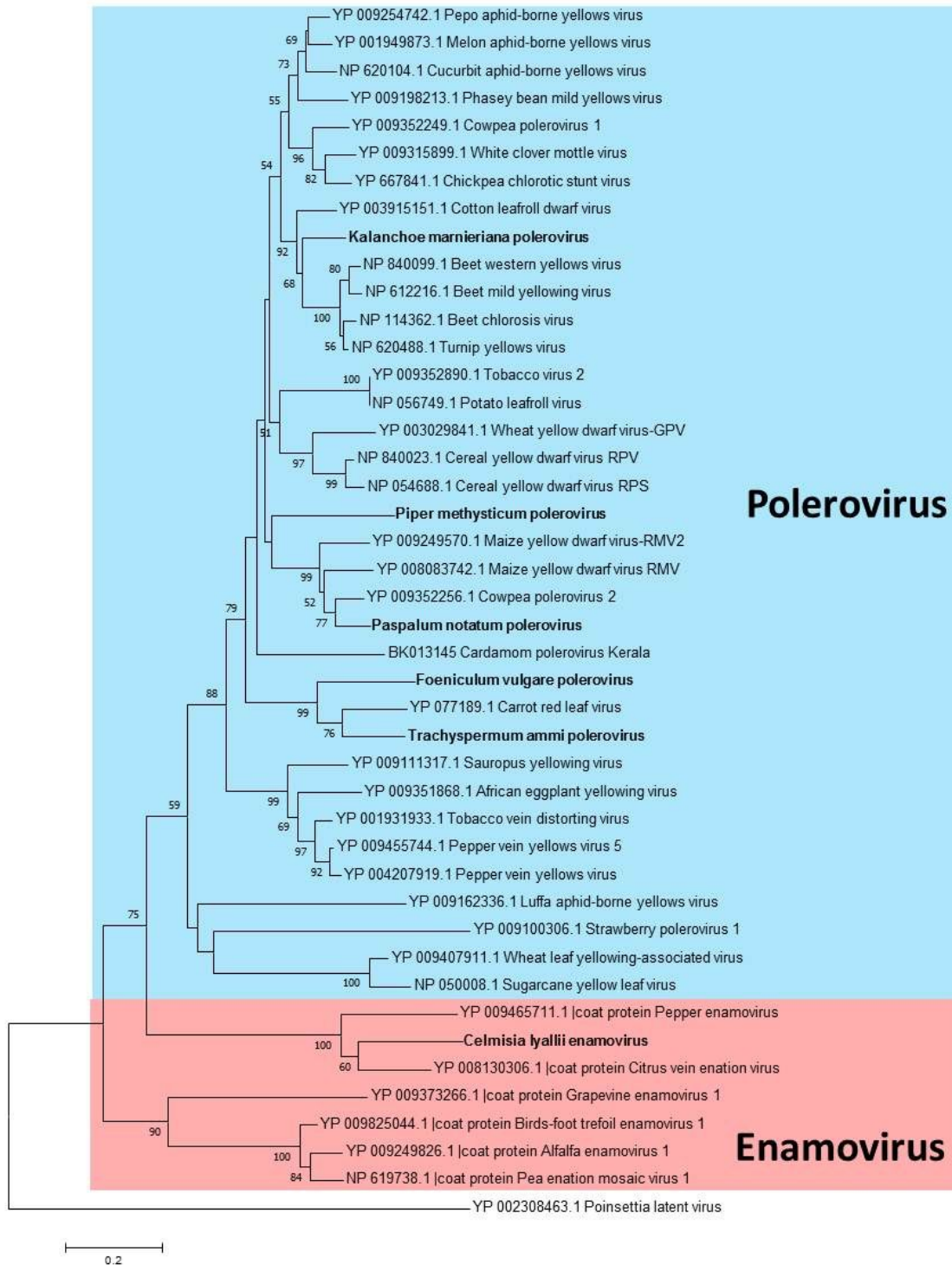


Figure 2

Phylogenetic relationship of identified viruses to other poleroviruses and enamoviruses. Phylogenetic trees were constructed using Neighbourhood-Joining (NJ) method using Poisson model with 1000 bootstrap replicates. Only bootstrap values more the 50% are indicated. Viruses identified in the present study are shown in bold. Poinsettia latent virus was used as outgroup for phylogenetic tree construction.

Supplementary Files

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

- [ESM1.xlsx](#)
- [ESM2.docx](#)
- [ESM3.txt](#)