Probing of plant transcriptomes unveils the *hitherto* hidden genetic diversity of the family *Secoviridae*

V. Kavi Sidharthan (kavi.icfre@gmail.com)
Vijay Prakash Reddy
G. Kiran
V. Rajeswari
V.K. Baranwal
M.Kiran
Sudhir Kumar

Research Article

Keywords: Waikavirus, new, host range, genetic diversity, public domain

Posted Date: October 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3460801/v1

Abstract

Secoviridae family contains single stranded RNA genome-containing viruses that infect plants. In the present study, we mined publicly available plant transcriptomes and identified sixty-one putative novel secoviral sequences in various plant species ranging from bryophytes to trees, which increased the known secoviral diversity by approximately 0.5-fold. Of the identified viral sequences, 13 were monopartite and 48 were bipartite, and sequences of 52 secoviruses were coding-complete and nine were partial. Except for small open reading frames (ORFs) determined in waikaviral genomes and RNA2 of torradoviruses, all the recovered genomes/genome segments contained a large ORF encoding a polyprotein. Based on genome organization, sequence similarity to known members, phylogeny and secovirus species demarcation criteria, all but three identified novel secoviruses were assigned to different secoviral genera- *Cheravirus* (3), *Comovirus* (2), *Fabavirus* (5), *Nepovirus* (29), *Sadwavirus* (3), *Sequivirus* (1), *Stralarivirus* (1), *Torradovirus* (4) and *Waikavirus* (10). Genome organization of two of the identified waika-like viruses resembled that of the recently identified waika-like virus- Triticum aestivum secovirus. Phylogenetic analysis revealed the host-waikavirus co-evolution pattern in a few waika- and waika-like viruses, the increased phylogenetic diversity of nepoviruses and the phylogenetic clustering of waika-like viruses. The study paves way for further studies on understanding the biological properties of identified novel secoviruses.

1. Introduction

The family *Secoviridae*, under the order *Picornavirales*, contains non-enveloped plant-infecting viruses that are small icosahedral particles (diameter: 25 to 30 nm). Members of this family contain positive-sense monopartite or bipartite single stranded RNA genomes of lengths ranging from 9 to 13.7 kb (Fuchs et al., 2022). *Secoviridae* family was erected in 2009 by including the sub-family *Comovirinae* with three genera *Comovirus*, *Fabavirus*, *Nepovirus* and five other genera that were unassigned to a sub-family viz. *Cheravirus*, *Sadwavirus*, *Sequivirus*, *Torradovirus* and *Waikavirus* (Sanfacon et al., 2009). Later, a new genus *Stralarivirus* was created (Dullemans et al., 2020), and the genus *Sadwavirus* was reorganized into three sub-genera *Cholivirus*, *Satsumavirus* and *Stramovirus* (Sanfaçon et al., 2020). Among the secovirids, waikavirues and sequiviruses contain monopartite genomes whilst the other members possess bipartite genomes. Except for the RNA2 of torradoviruses and waikaviral genomes that contain small additional open reading frames (ORFs), genome/genome segments of all secovirids contain a single large ORF encoding a polyprotein that are cleaved by the viral genome-encoded proteases (Sanfaçon et al., 2020). Like other picornaviruses, secovirid RNA1 polyprotein contains the typical 'replication block' made of a type III helicase (Hel), a small viral genome-linked protein (VPg), a picornavirus 3C-like cysteine protease (Pro) and a type I RNA-dependent RNA polymerase (RdRp) (Sanfacon et al., 2009; Sanfaçon et al., 2020). RNA2 polyprotein of bipartite secovirids contains the movement protein (MP) domain followed by the coat protein (CP) domain(s). Whilst the CP domain is located upstream of the Hel domain in polyprotein of monopartite secovirids, the location of a putative MP domain remains unknown. Using the MP and/or adapted CP(s) to mediate their cell-to-cell and long distance movements, secoviruses accomplish successful invasion of plants (Sanfacon et al., 2009). Many secoviruses are important plant pathogens as they cause severe plant diseases in epidemic proportion and secoviruses are generally transmitted by insects, nematodes, seeds or pollen (Sanfacon, 2015; Fuchs et al., 2022).

Data-driven virus discovery (DDVD) studies on targeted plant virus groups like amalgavirus (Nibert et al., 2016; Sidharthan et al., 2022a), ophiovirus (Debat et al., 2023), rhabdovirus (Bejerman et al., 2021), solemovirus (Sidharthan et al., 2022b), tymovirus (Bejerman and Debat, 2022) and varicosavirus (Bejerman et al., 2022) have unveiled the hidden genetic diversity and expanded the host range of each target virus group using the public domain metatranscriptome data. Our earlier transcriptome shotgun assembly (TSA)-based DDVD study on secoviruses (Sidharthan et al., 2022c) identified nine novel secoviral sequences in eight plant species. In a subsequent sequence read archive (SRA)-based DDVD study on waikaviruses (Sidharthan et al., 2023a), we identified twenty-two novel waikaviral sequences and a highly divergent secoviral sequence in transcriptomes of various plant species. On the other hand, exploration of novel plant viral sequences in targeted plant species identified novel secoviral sequences in various plant transcriptomes (Park and Hahn, 2019; Sidharthan et al., 2021; Mifsud et al., 2022; Sidharthan et al., 2023b). These studies reiterate the importance of DDVD studies in sequence discovery of novel viruses, in general, and secoviruses, in particular. However, a comprehensive study on exploration of novel secoviral sequences, *in toto*, in publicly available SRA data has not yet been undertaken. In the present SRA-based DDVD study, we aimed to unlock the hidden genetic diversity of the family *Secoviridae* by probing the plant transcriptome/metatranscriptome-derived SRA data with the hypothesis that novel secoviral sequences would be present in plant transcriptome/metatranscriptome data if the plants sampled for sequencing were infected with novel secoviruses at the time of sampling.

2. Experimental Procedures

2.1. Identification of putative novel secovirus-positive SRA libraries and genome recovery of putative novel secoviruses

For identification of putative novel secovirus-positive SRA libraries, an RdRp search of SRA libraries for the secoviral sequences was performed using the Serratus explorer (alignment identity < 80; score ≥ 50) (Edgar et al., 2022) (accessed on 29 July 2023). Resulting libraries with atleast 1000 secoviral reads and derived from plant transcriptomes/metatranscriptomes were only considered for further analyses. Putative novel

secovirus-positive SRA libraries were imported into the Galaxy Australia server (Community, 2022) and pre-processed using Trimmomatic (Galaxy version 0.36.6) (Bolger et al., 2014) by including the Illuminaclip step to remove the adapter sequences and setting the average quality threshold to 30. In case of SRA libraries derived from same plant species and containing reads of same secovirus group, the library with the highest number of putative novel secoviral reads was imported and pre-processed. Pre-processed reads of each library were *de novo* assembled using rnaSPAdes (Galaxy version 3.15.4) (Bushmanova et al., 2019) and the resulting contigs were subjected to BLASTx analysis (e-value cut-off: 1e-5) against protein sequence database made of sequences of known secoviruses using the National Centre for Biotechnology Information (NCBI) BLAST + tool (Galaxy version 2.10.1) (Cock et al., 2015). Viral contig sharing less than 80 percent sequence identities with known secoviral sequences and approximating the genome/genome segment length of related viruses, with intact ORF(s) predicted using the NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/), was regarded as the coding-complete genome/genome segment of a putative novel secovirus. In cases where the coding-complete genome of a novel seovirus could not be obtained from a putative novel secovirus-positive SRA library, the same library was *de novo* assembled using Trinity (Galaxy version 2.15.1) (Grabherr et al., 2011) followed by BLASTx analysis for coding-complete genome recovery. If the coding-complete genome could not be obtained even after further assembly and multiple putative novel secovirus-positive libraries of a plant species were available, the library with the next highest number of putative novel secoviral reads was pre-processed and *de novo* assembled for genome recovery, and this process is repeated until the coding-complete genome/genome segment was obtained. Further, if coding-complete genome was not obtained, contig with length more than 75% of the approximate genome/genome segment length of related virus group was considered as partial genome/genome segment of a putative novel secovirus.

2.2. Bioinformatic analyses of recovered genomes of putative novel secoviruses

Molecular weight of and Pfam motifs in proteins encoded by recovered putative novel secoviral genomes were predicted using the Expasy (https://web.expasy.org/compute_pi/) and the Motif Search tools (https://www.genome.jp/tools/motif/), respectively. BLASTp analysis of encoded proteins was performed against NCBI 'non-redundant protein sequence' database to obtain the hit with maximum sequence identity at maximum query coverage. Polyprotein sequences of putative novel secoviruses along with related known secoviral sequences were aligned using the MAFFT webserver (version 7) (Katoh et al., 2019) to predict probable cleavage sites in novel secoviral polyprotein sequences. WebLogo3 server (version 2.8.2) (Crooks et al., 2004) was used to obtain sequence logos of conserved domains in the polyprotein sequence alignment of novel secoviruses, whilst the coiled-coils were predicted as described in Sidharthan et al. (2022a). Sequence conservation in the untranslated regions (UTR) was visualized in the aligned sequence using MAFFT MSA Viewer (Yachdav et al., 2016). Using the polyprotein 1 sequences of identified putative novel secoviruses as queries and by considering the palm ID cut-off as ≥ 90% in palm ID searches in Serratus explorer, other plant SRA libraries positive for each queried novel virus were identified.

2.3. Phylogenetic and sequence identity analyses

The conserved proteinase-polymerase (Pro-Pol) region of polyprotein 1 and the polyprotein 2 sequences of known and novel secoviruses, after individual MAFFT alignment, were subjected to maximum-likelihood (ML) tree construction in the IQ-TREE webserver (Trifinopoulos et al., 2016) using the best-fit model and ultrafast bootstrap (1000 replicates). The resulting trees in newick format were imported into MEGA7 (version 7.0.26) for visualization and editing (Kumar et al., 2016). Sequence Demarcation Tool (version 1.2) (Muhire et al., 2014) was used to obtain the percent sequence identities of identified novel secoviruses with other novel/known secoviruses based on Pro-Pol or polyprotein 2 sequences.

3. Results

3.1. Identification of putative novel secoviral sequences and their genome recovery

A total of 61 putative novel secoviral sequences were identified in the transcriptomes of 69 plant species, ranging from bryophytes to large trees, by *de novo* assembly of public domain SRA datasets derived from plants and BLASTx analysis (Table 1). Of the identified viruses, 13 contained monopartite genome, whilst the other 48 contained bipartite genome (Figure S1). Except for the partial genomes of 9 putative novel viruses, the genomes of identified putative novel viruses were coding-complete. Mean depth of recovered viral genome/genome segment by reads of respective virus-identified library ranged from 25.7x to 3,22,865.4x (Table 1). Tentative viral names of identified putative novel viruses and their acronyms are provided in Table 1. Of the plant species in which putative novel secoviral sequences were identified, 64 are angiosperms, of which, 54 are dicots and 10 are monocots. Considering plant families, maximum number of putative novel secoviral sequences were identified in plants belonging to Asteraceae (9) followed by Poacaeae (5) (Table 1). Besides 69 libraries, sequences of putative novel viruses were also identified in other libraries of same or different plant species (Table S1).

Table 1
Summary of putative novel secoviral sequence identification in public domain Sequence Read Archive datasets

Virus name/Isolate name	Virus acronym	Library from where the genome was derived	Genome length (nt) (RNA1/RNA2)	Mean depth (x) (RNA1/RNA2)	Plant Species/Family	Data Reference
Genus: *Cheravirus*						
Corymbium villosum cheravirus	CvCV	SRR6072272	7232/3700	2,413.5/3,925.3	*Corymbium villosum*; Asteraceae (dicot)	Fisher et al., 2018
Lagerstroemia indica cheravirus	LaiCV	SRR1209371	7466/3986	3,958.7/4,893.3	*Lagerstroemia indica*; Lythraceae (dicot)	Zhang et al., 2014
		SRR14560278	7448/3623	2,415.6/1,290.5	*Nephelium lappaceum*; Sapindaceae (dicot)	Zhang et al., 2021b
		SRR8785265	7272/3633	24,460.8/61,665.0	*Pogostemon cablin*; Lamiaceae (dicot)	Chen et al., 2019
Pternopetalum trichomanifolium cheravirus	PtCV	SRR8863745	6469/3803	3,712.3/15,694.2	*Pternopetalum trichomanifolium*; Apiaceae (dicot)	Wen et al., 2020
Genus: Comovirus						
Camellia comovirus	CamCV	SRR10913214	5709/3394	2,322.1/3,387.5	*Camellia sinensis*; Theaceae (dicot)	Zheng et al., 2021
White-flower bittercress comovirus	WfbCV	DRR215738	5972/3562	40,300.1/1,19,696.5	*Cardamine leucantha*; Brassicaceae (dicot)	Araki et al., 2020
Unclassified but related to Comovirus						
Litsea rubescens seco-like virus	LrSV	SRR10063978	5814/4054	144.4/195.3	*Litsea rubescens*; Lauraceae (dicot)	Chen et al., 2020
Genus: Fabavirus						
Camphor tree fabavirus	CtFabV	SRR15881630	6918/6739	290.1/518.1	*Cinnamomum camphora*; Lauraceae (dicot)	Jiang et al., 2022
Grapevine fabavirus 2	GFabV2	SRR10724800	6250/6002	393.7/268.5	*Vitis vinifera*; Vitaceae (dicot)	Shi et al., 2020
Many-flowered stoneseed fabavirus	MsFabV	SRR5013606	5887/3374	43,274.3/62,776.7	*Lithospermum multiflorum*; Boraginaceae (dicot)	Cohen, 2016
Reaumuria songarica fabavirus	RsFabV	SRR1232022	6644/3322	3,413.5/10,595.8	*Reaumuria songarica*; Tamaricaceae (dicot)	Meiling Liu, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences
Squamellaria imberbis fabavirus	SiFabV	SRR14731637	6351/4133	29,604.6/33,633.5	*Squamellaria imberbis*; Rubiaceae (dicot)	Pu et al., 2021
Genus: Nepovirus						
Actinidia nepovirus	AcNV	SRR14212104	7619/5736	5,189.9/10,811.8	*Actinidia cylindrica var. reticulate*; Actinidiaceae (dicot)	Yao et al., 2022
Aloe haircap nepovirus	AhNV	SRR3737504	7144/6454	2,240.9/4,041.8	*Pogonatum aloides*; Polytrichaceae (Bryophyta)	Zhang et al., 2018

*partial genome/genome segment

Virus name/Isolate name	Virus acronym	Library from where the genome was derived	Genome length (nt) (RNA1/RNA2)	Mean depth (x) (RNA1/RNA2)	Plant Species/Family	Data Reference
Genus: *Cheravirus*						
Asian lizard's tail nepovirus	AltNV	SRR14085884	7322/4003	847.2/706.6	*Saururus chinensis*; Saururaceae (dicot)	Zhao et al., 2021b
Bush clock vine seco-like virus	BcvNV	SRR8752193	6756*/3702*	13,333.3/47,803.8	*Thunbergia erecta*; Acanthaceae (dicot)	Morais et al., 2019
Beetleweed nepovirus	BwNV	ERR2040490	7108/3512	1,617.2/5,976.9	*Galax urceolata*; Diapensiaceae (dicot)	1000 Plant (1KP) Transcriptomes Initiative
Begonia plebeja nepovirus	BpNV	ERR2580197	7896/6645	2,225.4/4,616.1	*Begonia plebeja*; Begoniaceae (dicot)	Emelianova et al., 2021
Canberra spider orchid nepovirus	CsoNV	SRR12619952	7467/7015	13,985.7/41,265.0	*Caladenia actensis*; Orchidaceae (monocot)	Peakall et al., 2021
Cederberg Conebush nepovirus	CecNV	ERR6131074	7178/4188	8,114.3/15,488.0	*Leucadendron dubium*; Proteaceae (dicot)	Scharmann et al., 2021
Chinese milk vetch nepovirus	CmvNV	SRR13390673	7770/6382	656.9/1,087.2	*Astragalus sinicus*; Fabaceae (dicot)	Zhao et al., 2021a
Chrysanthemum nepovirus	ChrNV	SRR15173230	7959/5778	3,904.9/4,133.5	*Chrysanthemum indicum*; Asteraceae (dicot)	Wen et al., 2022
Common thyme nepovirus	CtNV	SRR6262808	7944/6844	2,545.8/3,196.8	*Thymus vulgaris*; Lamiaceae (dicot)	Mollion et al., 2018
Coral plant nepovirus	CopNV	SRR11934224	6610/5217	1,07,541.2/3,22,865.4	*Berberidopsis corallina*; Berberidopsidaceae (dicot)	Zhang et al., 2020
Downy ground fern nepovirus	DgfNV	SRR6920708	7714/7408	2,989.6/5,104.5	*Hypolepis punctate*; Dennstaedtiaceae (fern)	Qi et al., 2018
Gentiana ecaudata nepovirus	GeNV	SRR9856867	7262/3741	195.5/405.3	*Gentiana ecaudata*; Gentianaceae (dicot)	Chen et al., 2021b
Habenaria delavayi nepovirus	HdNV	SRR5722143	8036*/5591*	1,392.6/2,681.2	*Habenaria delavayi*; Orchidaceae (monocot)	Zhang et al., 2017a
Hansenia oviformis nepovirus	HoNV	SRR8884092	7922/3898	53,278.7/1,54,539.1	*Hansenia oviformis*; Apiaceae (dicot)	Liu et al., 2023
		SRR16167593	8242/3970	3,306.3/17,885.5	*Angelica sinensis*; Apiaceae (dicot)	Peng et al., 2021
		SRR15534019	8145/4638	25,380.9/2,76,275.4	*Neotrinia splendens*; Poaceae (monocot)	Ren et al., 2022
Homalium kanaliense nepovirus	HkNV	SRR7388624	8238/6145	11,430.2/11,758.1	*Homalium kanaliense*; Salicaceae (dicot)	García de la Torre et al., 2021
Lettuce nepovirus	LetNV	SRR5856141	8143/5069	167.6/899.5	*Lactuca sativa*; Asteraceae (dicot)	Zhang et al., 2017b
Logan nepovirus	LogNV	SRR9715739	8213/6045	106.8/220.7	*Glycyrrhiza uralensis*; Fabaceae (dicot)	Li et al., 2020
		SRR12042885	6906/6099	22,494.4/68,929.0	*Dimocarpus longan*; Sapindaceae (dicot)	Jue et al., 2021
		SRR19395881	7228*/6661	60.7/295.3	*Triticum aestivum*; Poaceae (monocot)	Ma et al., 2022
		SRR15096808	7340*/6049	118.6/204.6	*Stevia rebaudiana*; Asteraceae (dicot)	Sun et al., 2021

*partial genome/genome segment

Virus name/Isolate name	Virus acronym	Library from where the genome was derived	Genome length (nt) (RNA1/RNA2)	Mean depth (x) (RNA1/RNA2)	Plant Species/Family	Data Reference
Genus: *Cheravirus*						
Musa nepovirus	MuNV	SRR16882082	8205/7032	1,844.6/2,857.9	*Musa* hybrid cultivar; Musaceae (monocot)	Rong et al., 2023
Nitraria roborowskii nepovirus	NrNV	SRR10829654	7566/5830	129.4/268.9	*Nitraria roborowskii*; Nitrariaceae (dicot)	Wang et al., 2021a
Pearl millet nepovirus	PmNV	SRR11547884	7808/6613	10,078.8/32,087.7	*Cenchrus americanus*; Poaceae (monocot)	Kumar et al., 2021
Purple sand food nepovirus	PsfNV	ERR2040529	6621*/4285	215.2/338.1	*Pholisma arenarium*; Boraginaceae (dicot)	1000 Plant (1KP) Transcriptomes Initiative
Rhododendron lacteum nepovirus	RhlNV	SRR14827351	7042/3613	5,081.3/28,544.5	*Rhododendron lacteum*; Ericaceae (dicot)	Liu et al., 2022b
Senecio pinnatifolius nepovirus	SepNV	SRR5237254	6949/4991	110.0/289.9	*Senecio pinnatifolius var. latilobus*; Asteraceae (dicot)	Jayasena et al., 2017
Silene diclinis nepovirus	SdNV	ERR4643626	7823/5748	1,236.5/2,516.01	*Silene diclinis*; Caryophyllaceae (dicot)	Muyle et al., 2021
Snowy daisy-bush nepovirus	SdbNV	SRR5237255	7139*/6564	92.1/75.2	*Olearia lirata*; Asteraceae (dicot)	Jayasena et al., 2017
Tibetan peony nepovirus	TipNV	SRR10948769	7822/5522	9,563.8/10,017.6	*Paeonia ludlowii*; Paeoniaceae (dicot)	Wang et al., 2021b
Yunnan pine nepovirus	YpNV	SRR8259252	7113/4342	15,377.3/ 37,596.3	*Pinus yunnanensis*; Pinaceae (Gymnosperm)	Anpei Zhou, Southwest Forestry University
Genus: Sadwavirus (Sub-genus: Cholivirus)						
Dendrobium palpebrae cholivirus	DpCV	SRR6127588	5903/4471	124.7/572.8	*Dendrobium palpebrae*; Orchidaceae (monocot)	Unruh et al., 2018
Yellow sand-verbena cholivirus	YsvCV	SRR6435331	5092*/4870	25.7/74.8	*Abronia latifolia*; Nyctaginaceae (dicot)	Walker et al., 2018
Genus: Sadwavirus (Sub-genus: Stramovirus)						
Chrysanthemum stramovirus	ChrSV	SRR15321558	6970/6672	417.6/317.0	*Chrysanthemum morifolium*; Asteraceae (dicot)	Chirkov et al., 2022
Genus: Sequivirus						
Cowslip sequivirus	CosSV	ERR5762859	10179	131.5	*Primula veris*; Primulaceae (dicot)	Potente et al., 2022
Genus: Stralarivirus						
Beach cabbage stralarivirus	BcSV	SRR7429941	7177/3591	272.6/1,891.8	*Scaevola taccada*; Goodeniaceae (dicot)	Jing Zhang, Hainan Normal University
Genus: Torradovirus						
Lake cress torradovirus	LcTV	DRR075925	7182/5217	8,980.2/11,494.7	*Rorippa aquatica*; Brassicaceae (dicot)	Nakayama et al., 2018
Lophophytum mirabile torradovirus	LmTV	SRR10883507	6676/4862	417.2/ 506.5	*Lophophytum mirabile*; Balanophoraceae (dicot)	Garcia et al., 2021

*partial genome/genome segment

Virus name/Isolate name	Virus acronym	Library from where the genome was derived	Genome length (nt) (RNA1/RNA2)	Mean depth (x) (RNA1/RNA2)	Plant Species/Family	Data Reference
Genus: *Cheravirus*						
Opisthopappus taihangensis torradovirus	OtTV	SRR21075421	7518/4605	51,833.9/2,07,838.7	*Opisthopappus taihangensis*; Asteraceae (dicot)	Liu et al., 2022a
Sesamum torradovirus	SeTV	SRR12153203	7052/4646	20,991.8/1,57,990.6	*Sesamum indicum* x *Sesamum mulayanum*; Pedaliaceae (dicot)	Dutta et al., 2022
		SRR16672419	8070/4668	18,988.8/13,991.5	*Rehmannia glutinosa*; Orobanchaceae (dicot)	Yanqing Zhou, College of Life Science, China
Genus: Waikavirus						
Asian lily-of-the-valley waikavirus	AlWV	SRR11234883	11504*	77.6	*Convallaria keiskei*; Asparagaceae (monocot)	Lu et al., 2020
Euphorbia ebracteolata waikavirus	EeWV	SRR13511993	12829	6,719.4	*Euphorbia ebracteolata*; Euphorbiaceae (dicot)	Zheng et al., 2022
Eureka dunegrass waikavirus	EudWV	SRR16007046	11866	3,481.9	*Swallenia alexandrae*; Poaceae (monocot)	Huang et al., 2022
Gentiana straminea waikavirus	GesWV	SRR13214721	12270	313.4	*Gentiana straminea*; Gentianaceae (dicot)	Chen et al., 2021a
Gypsywort waikavirus	GwWV	ERR6688665	11945	170.0	*Lycopus europaeus*; Lamiaceae (dicot)	Darwin Tree of Life Project Consortium, 2022
Hooked Veilwort waikavirus	HvWV	SRR8202209	8838*	1,020.7	*Metzgeria leptoneura*; Metzgeriaceae (liverwort)	Dong et al., 2019
Pagoda dogwood waikavirus	PdWV	SRR22133646	12089	399.3	*Cornus alternifolia*; Cornaceae (dicot)	Lu et al., 2023
Plumleaf crab apple waikavirus	PcaWV	SRR8146297	12594	343.5	*Malus prunifolia*; Rosaceae (dicot)	Saito et al., 2019
Rubber waikavirus	RuWV	SRR2156988	12483	564.2	*Hevea brasiliensis*; Euphorbiaceae (dicot)	Hurtado et al., 2015
Sweet wormwood waikavirus	SwWV	SRR15595119	11948	72,420.2	*Artemisia annua*; Asteraceae (dicot)	Ma et al., 2021
Unclassified but related to Waikavirus						
Magellanic bog-moss seco-like virus	MbSV	SRR5830074	9898	1,644.9	*Sphagnum magellanicum*; Sphagnaceae (Bryophyta moss)	Kolton et al., 2022
Pacific island silvergrass seco-like virus	PasSV	SRR13299808	9696	249.6	*Miscanthus floridulus*; Poaceae (monocot)	Zhang et al., 2021a
*partial genome/genome segment						

3.2. Identification of putative novel cheraviruses in plant transcriptomes

In transcriptomes of five dicot plant species, three putative novel cheraviruses were identified, of which, LaiCV alone, was identified in three plant species (Table 1). Genomes of identified novel cheraviruses, recovered from five plant species, are bipartite with each genome segment encoding a polyprotein. Polyprotein 1 (232.3–265.5 kDa) encoded by identified cheraviruses contained the motifs- Hel (PF00910) and RdRp (PF00680) whilst polyprotein 2 (121.0–127.6 kDa) contained no predicted motif (Table S2). Though the genome segments of each identified novel cheravirus shared conserved sequences at the 5’ and 3’ UTRs, extensive degree of sequence conservation was observed in 3’ UTR than 5’ UTR (Figure S2). Putative cleavage sites predicted in polyproteins encoded by identified novel cheraviruses are provided in Table S3. Phylogenetic

analysis based on the conserved Pro-Pol and polyprotein 2 sequences revealed the relatedness of LaiCV isolates with Orobanche cernua secovirus, Trillium govanianum cheravirus and Alpine wild prunus virus, and the distinctness of CvCV and PtCV among cheraviruses (Fig. 1).

3.3. Identification of putative novel comoviruses in plant transcriptomes

Sequences of two putative novel comoviruses (CamCV, WfbCV) and a como-like virus (LrSV) with bipartite genomes were identified in transcriptomes of three dicot hosts (Table 1). Each genome segment of identified como/como-like viruses encoded a single large polyprotein. Polyprotein 1 (206.8–212.1 kDa) of identified como/como-like viruses contained Hel (PF00910), 3C cysteine Pro (PF00548), and RdRp (PF00680) motifs, whilst polyprotein 2 (114.9–132.3 kDa) contained large CP (PF02247) and small CP (PF02248) motifs. In addition, a viral MP motif (PF01107) was determined in the polyprotein 2 of WfbCV (Table S2). The genome segments of each identified como/como-like virus shared a considerable degree of sequence conservation at the UTRs (Figure S3). Amino acid residue at the position upstream of scissile bond (P1 protein) in six putative cleavage sites predicted in polyproteins (4 in polyprotein 1 and 2 in polyprotein 1) encoded by each identified comovirus is glutamine (Q). However, putative cleavage sites could not be precisely determined in polyproteins of LrSV (Table S3). Phylogenetic analysis based on Pro-Pol and polyprotein 2 sequences grouped WfbCV with other brassica-infecting comoviruses. On the other hand, CamCV was grouped with Ullucus comovirus 1 in Pro-Pol-based phylogeny, while CamCV fell in a distinct sub-clade within comoviruses based on polyprotein 2-based phylogeny. LrSV along with Artocarpus altilis secovirus formed a distinct sub-clade away from comovirus cluster in Pro-Pol-based phylogeny and within comovirus cluster in polyprotein 2-based phylogeny (Fig. 2).

3.4. Identification of putative novel fabaviruses in plant transcriptomes

Bipartite genome segments, each encoding a polyprotein, of five putative novel fabaviruses were identified in transcriptomes of five dicot plant species (Table 1). Like comoviruses, polyprotein 1 (209.3–238.4 kDa) encoded by identified fabaviruses contained Hel (PF00910), 3C cysteine Pro (PF00548), and RdRp (PF00680) motifs, whilst the polyprotein 2 (111.1–223.5 kDa) contained large CP (PF02247) and small CP (PF02248) motifs. Besides, an additional viral MP motif (PF01107) was predicted in polyprotein 2 encoded by two of the five identified fabaviruses (Table S2). RNAs1 and 2 of each identified fabavirus shared a considerable degree of sequence conservation at the UTRs (Figure S3). Similar to comoviruses, ‘Q’ residue was observed in P1 position in all but one-cleavage sites predicted in the polyproteins encoded by identified fabaviruses (Table S3). Phylogenetic analysis based on Pro-Pol and polyprotein 2 sequences grouped CtFabV, GFabV2 and RsFabV together with Yucca gloriosa secovirus in a distinct sub-clade to other fabaviruses. On the other hand, Pro-Pol and polyprotein 2-based phylogenies revealed the grouping of SiFabV with black pepper virus F, and the distinct sub-clade of MsFabV within fabavirus cluster (Fig. 2).

3.5. Identification of putative novel nepoviruses in plant transcriptomes

Genome sequences of twenty-nine putative novel nepoviruses were identified in transcriptomes of thirty-four plant species, including dicots, a fern, a gymnosperm, monocots and a moss (Table 1). Of these, LogNV and HoNV genomes were identified in transcriptome of four and three plant species, respectively, totalling the number of recovered genomes of putative novel nepoviruses to thirty-four. Genomes of identified nepoviruses were bipartite with each genome segment encoding a polyprotein and coding-complete, except for partial genome segment(s) of six viruses/viral isolates. Polyprotein 1 (198.9–276.2 kDa) encoded by identified nepoviruses contained Hel (PF00910) and RdRp (PF00680) motifs, besides a 3C cysteine Pro (PF00548) motif in a few viruses. On the other hand, polyprotein 2 (118.0–214.2 kDa) of identified nepoviruses contained the motifs- nepovirus CP N-terminal (PF03689), central (PF03391) and C-terminal (PF03688) domains, besides a viral MP (PF01107) motif in a few viruses (Table S2). The conserved ‘LPL’ motif commonly found in MP of nepoviruses (Mifsud et al., 2022) was determined in polyprotein 2 of nepoviruses identified in non-angiospermic plants (data not shown). Though RNAs1 and 2 of most identified nepoviruses shared conserved sequences at the 5’ and 3’ UTRs, extensive degree of sequence conservation was observed in 3’ UTR than 5’ UTR in most viruses (Figure S4). Cleavage sites were predicted in polyproteins of identified nepoviruses and the amino acid residue at P1 position was highly diverse when compared with other secoviruses (Table S3). Phylogenetic analysis based on Pro-Pol and polyprotein 2 sequences placed AltNV, BcvNV, BwNV, GeNV, HoNV and RhlNV with sub-group A nepoviruses, CecNV and YpNV with sub-group B nepoviruses, and AcNV and NrNV with the unclassified green Sichuan pepper nepovirus, while the remaining viruses clustered with sub-group C nepoviruses (Fig. 3).

3.6. Identification of putative novel sadwaviruses in plant transcriptomes

Three putative novel sadwaviral sequences were identified in transcriptomes of two dicot and a monocot plant species (Table 1). Genomes of identified sadwaviruses were bipartite with each genome segment coding for a single large polyprotein. Amongst the recovered genome segments of identified sadwaviruses, RNA1 of YsvCV was partial. Polyprotein 1 (211.0–219.4 kDa) encoded by identified sadwaviruses contained Hel (PF00910) and RdRp (PF00680) motifs whilst polyprotein 2 (123.4–208.8 kDa) contained one or more of the following motifs- calicivirus CP (PF00915), viral MP (PF01107) and large CP (PF02247) (Table S2). Extensive degree of sequence conservation was observed in RNAs1 and 2 at both the 5’ and 3’ UTRs of ChrSV and DpCV and at the 3’ UTR of YsvCV (Figure S5). Amino acid residue at P1 position in four cleavage sites predicted in polyprotein 1 of each identified sadwavirus was ‘Q’ or glutamic acid (E), while the cleavage site could not be precisely determined in polyprotein 2 of identified sadwaviruses (Table S3). Phylogenetic analysis based on Pro-Pol and polyprotein 2 sequences revealed

the relatedness of YsvCV with pineapple secovirus B, DpCV with pineapple secovirus A, and ChrSV with lettuce secovirus 1. YsvCV and DpCV clustered together with choliviruses, whilst ChrSV grouped with stramoviruses (Fig. 1).

3.7. Identification of a putative novel sequivirus in a plant transcriptome

Monopartite genome of a putative novel sequivirus CosSV was identified in a dicot plant transcriptome (Table 1). CosSV genome encoded a single large polyprotein of 344.6 kDa with the motifs- viral MP (PF01107), picornavirus CP (PF00073), Hel (PF00910) and RdRp (PF00680) (Table S2). Six putative cleavage sites in the order N/A, R/G, Q/N, Q/G, S/L and H/M were predicted in CosSV polyprotein (Table S3). Phylogenetic analysis based on Pro-Pol sequences placed CosSV in a distinct sub-clade to other sequiviruses (Fig. 4).

3.8. Identification of a putative novel stralarivirus in a plant transcriptome

Two genome segments of a putative novel stralarivirus BcSV was identified in the transcriptome of a dicot species (Table 1). BcSV genome segments encoded a single large polyprotein. BcSV polyprotein 1 (263.8 kDa) contained Hel (PF00910) and RdRp (PF00680) motifs whilst BcSV polyprotein 2 (117.9 kDa) contained a viral MP motif (PF01107) (Table S2). The UTRs of BcSV genome segments shared conserved nucleotides, but sequence conservation was more evident at the 3' UTR (Figure S5). Putative cleavage sites predicted in BcSV polyproteins are provided in Table S3. Phylogenetic analysis based on Pro-Pol and polyprotein 2 sequences placed BcSV in a distinct sub-clade to other stralariviruses (Fig. 1).

3.9. Identification of putative novel torradoviruses in plant transcriptomes

Four putative novel torradoviral sequences were identified in transcriptomes of five dicot species, of which SeTV alone was identified in two species (Table 1). Genomes of identified torradoviruses were bipartite with each RNA segment encoding a large polyprotein. Besides, an additional ORF was predicted upstream of the polyprotein ORF in RNA2 of identified torradoviruses that encodes a small protein (22.8–24.0 kDa) with no predicted viral motif. Polyprotein 1 (241.2–248.7 kDa) encoded by identified torradoviruses contained Hel (PF00910) and RdRp (PF00680) motifs, whilst the polyprotein 2 (115.7–152.0 kDa) of all identified torradoviruses contained 3A/RNA2 MP family (PF00803) motif (Table S2). RNAs1 and 2 of each identified torradovirus shared a considerable degree of sequence conservation at the UTRs (Figure S5). Three putative cleavage sites were predicted in polyprotein 2 encoded by each identified torradovirus with 'Q' at P1 position in all the predicted cleavage sites (Table S3). Pro-Pol and polyprotein 2-based phylogenies revealed the relatedness of LcTV with fleabane torradovirus, OtTV with burdock mosaic virus, and SeTV with squash chlorotic leaf spot virus and LmTV (Fig. 1).

3.10. Identification of putative novel waikaviruses in plant transcriptomes

Ten putative novel waikaviral sequences were identified in transcriptomes of ten plant species including dicots, monocots and a liverwort (Table 1). Genomes of identified waikaviruses were monopartite and coding-complete, except for two partial genomes of AlWV and HvWV. Identified waikaviral genomes contained a larger ORF (ORF1) that coded for a polyprotein (382.6 – 432.6 kDa) with motifs- waikavirus CP1 (PF12264), Hel (PF00910), tungro spherical virus-type peptidase (PF12381) and RdRp (PF00680) (Table S2). In addition, a small ORF (ORFX) encoding a protein (9.3–10.9 kDa) with no predicted viral motif was determined in +1 frame to ORF1 in identified waikaviral genomes. Coiled-coils were predicted in the polyprotein alignment of identified waikaviruses near the N-terminal region (Figure S6), whilst two transmembrane domains were predicted in ORFX-encoded protein alignment near the N-terminal region (Figure S7). Amino acid residue 'Q' was present in P1 position of predicted cleavage sites in polyproteins encoded by identified waikaviruses (Table S3). Phylogenetic analysis based on Pro-Pol sequences placed the liverwort-infecting HvWV in a distinct sub-clade to other waikaviruses, and grouped PcaWV and RuWV that are identified in tree species, with sub-group 2 waikaviruses, while the remaining viruses were grouped with sub-group 1 waikaviruses. Amongst sub-group 1 waikaviruses, EudWV identified in Poaceous plant species grouped with the Poaceous plant-infecting maize chlorotic dwarf virus (MCDV) (Fig. 4).

Besides novel waikaviral genomes, two novel waika-like viral genomes- MbSV and PasSV were identified in a moss and a monocot plant species, respectively (Table 1). Unlike waikaviral genomes, MbSV and PasSV genomes were relatively shorter and contained only one ORF that encoded a polyprotein of 318.3–336.9 kDa. MbSV genome-encoded polyprotein contained the motifs in the order- Hel (PF00910), tungro spherical virus-type peptidase (PF12381), RdRp (PF00680) and CRPV CP like (PF08762), whilst PasSV genome-encoded polyprotein contained calicivirus CP (PF00915), CRPV CP like (PF08762), Hel (PF00910), tungro spherical virus-type peptidase (PF12381) and RdRp (PF00680) motifs (Table S2). Pro-Pol based phylogeny grouped PasSV with Triticum aestivum waikavirus and these together with MbSV formed a distinct sub-clade away from waikaviruses (Fig. 4).

4. Discussion

Secoviruses are important plant pathogens as many of them cause severe plant diseases in economically important crops. Currently, the International Committee on Taxonomy of Viruses (ICTV) recognizes over 100 species within the family *Secoviridae* (Fuchs et al., 2022). Like other virus groups (Debat et al., 2023), secoviruses are largely reported from economically important crops depicting viral disease symptoms.

Thus, for comprehensive identification of secoviruses across plant species irrespective of economic importance and symptom depiction, public domain transcriptome data derived from various plant species can be probed for secoviral sequences. Earlier DDVD studies targeting selected plant species/secoviral genera have identified twenty-nine novel secoviral sequences in various plant species (Park and Hahn, 2019; Sidharthan et al., 2021; Mifsud et al., 2022; Sidharthan et al., 2023a, b). Our previous TSA-based DDVD study for comprehensive identification of secoviral contigs identified nine novel secoviral sequences. These studies highlight the importance of plant transcriptome data mining in novel secoviral-sequence discovery. Despite the data-mining efforts for secoviral-sequence discovery, a comprehensive SRA-based DDVD study exploring the entire spectrum of secoviral genera in public domain plant transcriptomes is lacking. In the present study, we identified sixty-one putative novel secoviral sequences, approximately 0.5-fold expansion of the known diversity of the family *Secoviridae*, in a wide range of plant species from bryophytes to trees, of which five secoviral sequences were identified in non-angiospermic plants. AhNV identified in a non-vascular plant encoded a MP with the characteristic 'LPL' motif, similar to the ones encoded by nepoviruses of vascular plants (Mifsud et al., 2022). As nepovirus MP facilitates the cell-to-cell movement of viruses in vascular plants (Hily et al., 2021), identification of such MP homolog in yet another nepovirus of non-vascular plant warrants further studies to confirm their function in viruses of non-vascular plants (Mifsud et al., 2022). Interestingly, the sizes of polyproteins and predicted polyprotein cleavage sites of ChrSV, a stramovirus identified in *Chrysanthemum morifolium* transcriptome in this study were similar to those of chrysanthemum sadwavirus (ChSV) identified in diseased *C. morifolium* plants (Chen et al., 2023). Similar pattern of phylogenetic clustering was observed for ChrSV identified in this study and ChSV identified by Chen et al. (2023). Though we speculate ChrSV of this study to be ChSV, the same could not be confirmed due to non-availability of ChSV sequence in public domain. It is worthy of note that in the present study ChrSV was identified in a transcriptome derived from chrysanthemum plants infected with two carlaviruses- chrysanthemum virus B and R (CVB, CVR) (Chirkov et al., 2022), suggesting the co-infection pattern of CVB, CVR and ChrSV. On the other hand, proteins encoded by SeTV isolate Sesamum RNA2 sequences shared > 95% sequence identities with the partial sequences of soybean torrado virus 1 (Rahman et al., 2023), hinting that SeTV sequence could be the full-length sequence of soybean torrado virus 1. However, SeTV isolate Rehmannia RNA2-encoded proteins shared < 80% sequence identity with soybean torrado virus 1. Likewise, AcNV polyprotein 1 sequence of this study shared > 90% identity with a sequence named Paris mosaic virus 1 in GenBank with no associated publication.

Genome organization of identified secoviruses were in agreement with those of known secoviruses (Fuchs et al., 2022), excepting for the two novel waika-like viral sequences- MbSV and PasSV. Like Triticum aestivum secovirus (TaSV) identified in our previous study (Sidharthan et al., 2023b), MbSV and PasSV encoded a single polyprotein that is smaller than the waikaviral polyprotein. PasSV and TaSV, both identified in monocot plants and clustered together in phylogenetic analysis, contained similar motifs in the encoded polyprotein. MbSV also contains motifs similar to those identified in PasSV and TaSV, but the positions of motifs in the polyprotein were reversed in MbSV. Prevalence of such non-canonical reversed genome organization has already been reported in a seco-like virus (Zhang et al., 2023). Considering the genome organization and phylogenetic clustering, MbSV PasSV and TaSV could be regarded as members of a new genus within the family *Secoviridae*. An extensive degree of sequence conservation at the UTRs among the genome segments was observed in most identified bipartite secoviruses. Sequence conservation at the UTRs of genome segments has been reported in members of a few bipartite secoviral genera (Thompson et al., 2017).

Phylogenetic analysis grouped the two tree waikaviruses- PcaWV and RuWV with other tree waikaviruses in sub-group 2. Also, HvWV identified in a non-vascular plant in the current study formed a distinct sub-clade away from waikaviruses of vascular plants. Similarly, MbSV, a waika-like virus identified in non-vascular plant fell apart from waika-like viruses of vascular plants in phylogenetic tree. On the other hand, EudWV and MCDV identified in Poaceous plants were grouped together in phylogenetic analysis, as is the case with PasSV and TaSV. These results further reiterate the host and waika/waika-like virus co-evolution pattern (Sidharthan et al., 2023a). Interestingly, the size of ORFX-encoded proteins of the two tree waikaviruses identified in this study was larger than that of non-tree plant waikaviruses, and this finding is in agreement with our previous study (Sidharthan et al., 2023a). Consistent with our previous findings (Sidharthan et al., 2023a), coiled-coils were observed near the N-terminal region in the polyprotein of identified waikaviruses. Based on phylogenetic grouping with known nepoviruses (Sanfaçon, 2022), identified nepoviruses were grouped into three sub-groups- A, B and C, except AcNV and NrNV that grouped with unclassified nepoviruses. However, viruses in sub-groups A and C were scattered in distinct clades in both Pro-Pol and polyprotein 2-based phylogenetic trees, and further sub-grouping within sub-groups A and C was difficult. Hily et al. (2021) observed scattering of sub-group A and C nepoviruses as distinct clades in ORF2 nucleotide sequence-based phylogenetic tree. Increasing discoveries of novel nepoviruses increase the phylogenetic diversity of nepoviruses, which warrants revisiting of the nepovirus sub-grouping.

Based on the secoviral species demarcation criteria (amino acid sequence identity of < 80% in Pro-Pol region or < 75% in CP sequences) (Fuchs et al., 2022), sixty-one putative novel secoviruses identified in the study are regarded as new secoviral members. Of the identified secoviruses, fifty-eight viruses were assigned to different secoviral genera- *Cheravirus* (3), *Comovirus* (2), *Fabavirus* (5), *Nepovirus* (29), *Sadwavirus* (3), *Sequivirus* (1), *Stralarivirus* (1), *Torradovirus* (4) and *Waikavirus* (10). As per the consensus statement report of Simmonds et al. (2017), viruses identified in this study can be regarded as *bona fide* ones. Like other DDVD studies (Debat et al., 2023), the present study has the following limitations- inability to validate the identified viruses in plant samples and determine the complete end sequences of identified viruses in the present study due to the practical difficulty in getting back to the original samples used for sequencing. Thus, host assignment of identified

viruses in the current study is preliminary and be treated cautiously until further validation. However, the following evidences provided in this study partly addresses its limitations- detection of a few of the identified putative novel viruses in multiple libraries of the same/related plant species, detection of both the genome segments of a bipartite secovirus in the same library and greater depth of recovered genome sequences.

In conclusion, the present study mined public domain plant transcriptomes and identified sixty-one putative novel secoviral sequences in a wide range of plants, thereby broadening the phylogenetic diversity and host range of the family *Secoviridae*. Further studies are needed to validate the identified viruses in respective plant species and understand the biological properties of identified novel viruses.

Declarations

5. Acknowledgements

The authors thank the original submitters of transcriptome data that are analysed in the current study to SRA and NCBI for making the datasets publicly available. We thank Galaxy Australia, a service provided by the Australian Biocommons and its partners and the Serratus. The authors are grateful to the Head, Division of Genetics and Tree Improvement, the GCR and the Director, IFB (ICFRE), Hyderabad for their support.

6. Author contributions

V. Kavi Sidharthan: Conceptualization; methodology; investigation; formal analysis; writing- original draft. **Vijay Prakash Reddy:** Investigation. **G. Kiran:** Investigation. **V. Rajeswari:** Investigation. **M. Kiran:** Investigation. **Sudhir Kumar:** Investigation. **V. K. Baranwal:** Supervision; writing- review and editing.

7. Data availability

The viral genome sequences described in the study are submitted to NCBI.

8. Conflict of interest statement

The authors declare that they have no conflict of interest.

References

Araki, K.S., Nagano, A.J., Nakano, R.T., Kitazume, T., Yamaguchi, K., Hara-Nishimura, I., Shigenobu, S., and Kudoh, H. (2020) Characterization of rhizome transcriptome and identification of a rhizomatous ER body in the clonal plant *Cardamine leucantha*. *Sci Rep* **10:** 13291.

Bejerman, N., and Debat, H. (2022) Exploring the tymovirales landscape through metatranscriptomics data. *Arch Virol* **167:** 1785-1803.

Bejerman, N., Dietzgen, R. G., and Debat, H. (2022) Unlocking the hidden genetic diversity of varicosaviruses, the neglected plant rhabdoviruses. *Pathogens* **11:** 1127.

Bejerman, N., Dietzgen, R.G., and Debat, H. (2021). Illuminating the plant rhabdovirus landscape through metatranscriptomics data. *Viruses*, **13:** 1304.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform* **30:** 2114-2120.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019) rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, **8:** giz100.

Chen, C., Yang, W., Liu, J., Xi, Z., Zhang, L., and Hu, Q. (2021a) Population transcriptomics reveals gene flow and introgression between two non-sister alpine gentians. *Front Ecol Evol* **9:** 638230.

Chen, C.L., Zhang, L., Li, J.L., Mao, X.X., Zhang, L.S., Hu, Q.J., Liu, J.Q., and Xi, Z.X. (2021b) Phylotranscriptomics reveals extensive gene duplication in the subtribe Gentianinae (Gentianaceae). *J Syst Evol* **59:** 1198-208.

Chen, J., Dong, Y., Wang, H., Changnian, M., Cao, L., Shen, L., ... and Fan, X. (2023) Identification and complete genomic sequence of a novel sadwavirus discovered in chrysanthemum (*Chrysanthemum morifolium* Ramat). Research Square. https://doi.org/10.21203/rs.3.rs-3247284/v1

Chen, X., Li, J., Wang, X., Zhong, L., Tang, Y., Zhou, X., Liu, Y., Zhan, R., Zheng, H., Chen, W., and Chen, L. (2019) Full-length transcriptome sequencing and methyl jasmonate-induced expression profile analysis of genes related to patchoulol biosynthesis and regulation in *Pogostemon cablin*. *BMC Plant Biol* **19:** 1-8.

Chen, Y.C., Li, Z., Zhao, Y.X., Gao, M., Wang, J.Y., Liu, K.W., Wang, X., Wu, L.W., Jiao, Y.L., Xu, Z.L., and He, W.G. (2020) The Litsea genome and the evolution of the laurel family. *Nat Commun* **11:** 1675.

Chirkov, S.N., Sheveleva, A., Snezhkina, A., Kudryavtseva, A., Krasnov, G., Zakubanskiy, A., and Mitrofanova, I. (2022) Highly divergent isolates of chrysanthemum virus B and chrysanthemum virus R infecting chrysanthemum in Russia. *PeerJ* **10:** e12607.

Chirkov, S.N., Sheveleva, A., Snezhkina, A., Kudryavtseva, A., Krasnov, G., Zakubanskiy, A., and Mitrofanova, I. (2022) Highly divergent isolates of chrysanthemum virus B and chrysanthemum virus R infecting chrysanthemum in Russia. *PeerJ* **10:** e12607.

Cock, P. J., Chilton, J. M., Grüning, B., Johnson, J. E., and Soranzo, N. (2015) NCBI BLAST+ integrated into Galaxy. *Gigascience*, **4:** s13742-015.

Cohen, J.I. (2016) De novo sequencing and comparative transcriptomics of floral development of the distylous species *Lithospermum multiflorum*. *Front Plant Sci* **7:** 1934.

Community, T.G. (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res* **50**: W345-W351.

Crooks, G.E., Hon,G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14:** 1188-1190.

Darwin Tree of Life Project Consortium. (2022) Sequence locally, think globally: The Darwin tree of life project. Proceedings of the National Academy of Sciences **119:** e2115642118.

Debat, H., Garcia, M. L., and Bejerman, N. (2023) Expanding the Repertoire of the Plant-Infecting Ophioviruses through Metatranscriptomics Data. *Viruses*, **15:** 840.

Dong, S., Zhao, C., Zhang, S., Wu, H., Mu, W., Wei, T., Li N, Wan T, Liu H, Cui J, Zhu. R. (2019). The amount of RNA editing sites in liverwort organellar genes is correlated with GC content and nuclear PPR protein diversity. *Genome Biol Evol* **11:** 3233-9.

Dullemans, A. M., Botermans, M., de Kock, M. J. D., de Krom, C. E., van der Lee, T. A. J., Roenhorst, J. W., ... and van der Vlugt, R. A. A. (2020). Creation of a new genus in the family Secoviridae substantiated by sequence variation of newly identified strawberry latent ringspot virus isolates. *Arch Virol* **165:** 21-31.

Dutta, D., Harper, A., and Gangopadhyay, G. (2022). Transcriptomic analysis of high oil-yielding cultivated white sesame and low oil-yielding wild black sesame seeds reveal differentially expressed genes for oil and seed coat colour. *Nucleus* **65:**151-64.

Edgar, R.C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., ... and Babaian, A. (2022) Petabase-scale sequence alignment catalyses viral discovery. *Nature*, **602:** 142-147.

Emelianova, K., Martínez Martínez, A., Campos-Dominguez, L., and Kidner, C. (2021) Multi-tissue transcriptome analysis of two Begonia species reveals dynamic patterns of evolution in the chalcone synthase gene family. *Sci Rep* **11:** 17773.

Fisher, M.F., Zhang, J., Taylor, N.L., Howard, M.J., Berkowitz, O., Debowski, A.W., Behsaz, B., Whelan, J., Pevzner, P.A., and Mylne, J.S. (2018) A family of small, cyclic peptides buried in preproalbumin since the Eocene epoch. *Plant Direct* **2:** e00042.

Fuchs, M., Hily, J. M., Petrzik, K., Sanfaçon, H., Thompson, J. R., van der Vlugt, R., ... and ICTV Report Consortium. (2022) ICTV virus taxonomy profile: Secoviridae. *J Gen Virol* **103:** 001807.

García de la Torre, V.S., Majorel-Loulergue, C., Rigaill, G.J., Alfonso-González, D., Soubigou-Taconnat, L., Pillon, Y., Barreau, L., Thomine, S., Fogliani, B., Burtet-Sarramegna, V., and Merlot, S. (2021) Wide cross-species RNA-Seq comparison reveals convergent molecular mechanisms involved in nickel hyperaccumulation across dicotyledons. *New Phytol* **229:** 994-1006.

Garcia, L.E., Edera, A.A., Palmer, J.D., Sato, H., and Sanchez-Puerta, M.V. (2021). Horizontal gene transfers dominate the functional mitochondrial gene space of a holoparasitic plant. *New Phytol* **229:** 1701-14.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., ... and Regev, A. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnol* **29:** 644.

Hily, J.M., Poulicard, N., Kubina, J., Reynard, J.S., Spilmont, A.S., Fuchs, M., ... and Vigne, E. (2021) Metagenomic analysis of nepoviruses: diversity, evolution and identification of a genome region in members of subgroup A that appears to be important for host range. *Arch Virol* **166:** 2789-2801.

Huang, W., Zhang, L., Columbus, J.T., Hu, Y., Zhao, Y., Tang, L., Guo, Z., Chen, W., McKain, M., Bartlett, M., and Huang, C.H. (2022) A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C4 photosynthesis. *Mol Plant* **15:** 755-77.

Hurtado Paez, U.A., Garcia Romero, I.A., Restrepo Restrepo, S., Aristizabal Gutierrez, F.A., and Montoya Castano, D. (2015) Assembly and analysis of differential transcriptome responses of *Hevea brasiliensis* on interaction with *Microcyclus ulei*. *PLoS One* **10:** e0134837.

Jayasena, A.S., Fisher, M.F., Panero, J.L., Secco, D., Bernath-Levin, K., Berkowitz, O., Taylor, N.L., Schilling, E.E., Whelan, J., and Mylne JS. (2017) Stepwise evolution of a buried inhibitor peptide over 45 My. *Mol Biol Evol* **34:**1505-16.

Jiang, R., Chen, X., Liao, X., Peng, D., Han, X., Zhu, C., Wang, P., Hufnagel, D.E., Wang, L., Li, K., and Li, C. (2022) A chromosome-level genome of the camphor tree and the underlying genetic and climatic factors for its top-geoherbalism. *Front Plant Sci* **13:** 827890.

Jue, D., Liu, L., Sang, X., Shu, B., Wang, J., Wang, Y., Zhang, C., and Shi, S. (2021) SNP-based high-density genetic map construction and candidate gene identification for fruit quality traits of *Dimocarpus longan* Lour. *Sci Hortic* **284:**110086.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20:** 1160-1166.

Kolton, M., Weston, D.J., Mayali, X., Weber, P.K., McFarlane, K.J., Pett-Ridge, J., Somoza, M.M., Lietard, J., Glass, J.B., Lilleskov, E.A., and Shaw, A.J. (2022) Defining the Sphagnum core microbiome across the north American continent reveals a central role for diazotrophic methanotrophs in the nitrogen and carbon cycles of boreal peatland ecosystems. MBio. **13:** e03714-21.

Kumar, R.R., Bhargava, D.V., Pandit, K., Goswami, S., Shankar, S.M., Singh, S.P., Rai, G.K., Satyavathi, C.T., and Praveen, S. (2021) Lipase–The fascinating dynamics of enzyme in seed storage and germination–A real challenge to pearl millet. *Food chem* **361:**130031.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biol Evol* **33:** 1870-1874.

Li, Y., Chen, C., Xie, Z., Xu, J., Wu, B., and Wang, W. (2020) Integrated analysis of mRNA and microRNA elucidates the regulation of glycyrrhizic acid biosynthesis in *Glycyrrhiza uralensis* Fisch. Int J Mol Sci **21:** 3101.

Liu, H., Chen, W., Chai, Y., Liu, W., Chen, H., Sun, L., Tang, X., Luo, C., Chen, D., Cheng, X., and Wang, F. (2022a). Terpenoids and their gene regulatory networks in *Opisthopappus taihangensis* 'Taihang Mingzhu'as detected by transcriptome and metabolome analyses. *Front Plant Sci* **13:**1014114.

Liu, M.L., Shang, Q.H., Cheng, Y.J., Wang, N., Sa, W., Li, B.G., and Li, Z.H. (2023) Drivers of intraspecific differentiation of an alpine cold-tolerant herb, *Notopterygium oviforme*: Roles of isolation by distance and ecological factors. *J Syst Evol* **61:** 383-98.

Liu, X.W., Wang, Y.H., and Shen, S.K. (2022b) Transcriptomic and metabolomic analyses reveal the altitude adaptability and evolution of different-colored flowers in alpine Rhododendron species. *Tree Physiol* **42:** 1100-13.

Lu, M., Cao, M., Yang, J., and Swenson, N.G. (2023) Comparative transcriptomics reveals divergence in pathogen response gene families amongst twenty forest tree species. bioRxiv. https://doi.org/10.1101/2023.03.06.531373

Lu, Q.X., Gao, J., Wu, J.J., Zhou, X., Wu, X., Li, M.D., Wei, Y.K., Wang, R.H., Qi, Z.C., and Li, P. (2020) Development of 19 novel microsatellite markers of lily-of-the-valley (Convallaria, Asparagaceae) from transcriptome sequencing. *Mol Biol Rep* **47:** 3041-7.

Ma, J., Zhang, M., Lv, W., Tang, X., Zhao, D., Wang, L., Li, C., and Jiang, L. (2022) Overexpression of TaSNAC4-3D in common wheat (*Triticum aestivum* L.) negatively regulates drought tolerance. *Front Plant Sci* **13:** 945272.

Ma, T., Gao, H., Zhang, D., Sun, W., Yin, Q., Wu, L., Zhang, T., Xu, Z., Wei, J., Su, Y., and Shi, Y. (2021). Genome-Wide Analysis of Light-Regulated Alternative Splicing in *Artemisia annua* L. *Front Plant Sci* **12:** 733505.

Mifsud, J. C., Gallagher, R. V., Holmes, E. C., and Geoghegan, J. L. (2022) Transcriptome mining expands knowledge of RNA viruses across the plant kingdom. *J Virol* **96:** e00260-22.

Mollion, M., Ehlers, B.K., Figuet, E., Santoni, S., Lenormand, T., Maurice, S., Galtier, N., and Bataillon, T. (2018) Patterns of genome-wide nucleotide diversity in the gynodioecious plant *Thymus vulgaris* are compatible with recent sweeps of cytoplasmic genes. *Genome Biol Evol* **10:** 239-48.

Morais, E.B., Schönenberger, J., Conti, E., Antonelli, A., and Szoevenyi, P. (2019) Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. *Appl Plant Sci* **7:** e11290.

Muhire, B.M., Varsani, A., and Martin, D.P. (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PloS One* **9**: e108277.

Muyle, A., Martin, H., Zemp, N., Mollion, M., Gallina, S., Tavares, R., Silva, A., Bataillon, T., Widmer, A, Glémin, S., and Touzet P. (2021) Dioecy is associated with high genetic diversity and adaptation rates in the plant genus Silene. *Mol Biol Evol* **38:** 805-18.

Nakayama, H., Sakamoto, T., Okegawa, Y., Kaminoyama, K., Fujie, M., Ichihashi, Y., Kurata, T., Motohashi, K., Al-Shehbaz, I., Sinha, N., and Kimura, S. (2018). Comparative transcriptomics with self-organizing map reveals cryptic photosynthetic differences between two accessions of North American Lake cress. *Sci Rep* **8:**3302.

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

Park, D., and Hahn, Y. (2019) A novel Waikavirus (the family Secoviridae) genome sequence identified in rapeseed (*Brassica napus*). *Acta Virol* **63:** 211-216.

Peakall, R., Wong, D.C., Phillips, R.D., Ruibal, M., Eyles, R., Rodriguez-Delgado, C., and Linde, C.C. (2021) A multitiered sequence capture strategy spanning broad evolutionary scales: Application for phylogenetic and phylogeographic studies of orchids. *Mol Ecol Resour* **21:** 1118-40.

Peng, T., Wang, Y., Yang, T., Wang, F., Luo, J., and Zhang, Y. (2021) Physiological and biochemical responses, and comparative transcriptome profiling of two Angelica sinensis cultivars under enhanced ultraviolet-B radiation. *Front Plant Sci* **12:** 805407.

Potente, G., Léveillé-Bourret, É., Yousefi, N., Choudhury, R.R., Keller, B., Diop, S.I., Duijsings, D., Pirovano, W., Lenhard, M., Szövényi, P., and Conti, E. (2022). Comparative genomics elucidates the origin of a supergene controlling floral heteromorphism. *Mol Biol Evol* **39:** msac035.

Pu, Y., Naikatini, A., Pérez-Escobar, O.A., Silber, M., Renner, S.S., and Chomicki, G. (2021) Genome-wide transcriptome signatures of ant-farmed Squamellaria epiphytes reveal key functions in a unique symbiosis. *Ecol Evol* **11:** 15882-95.

Qi, X., Kuo, L.Y., Guo, C., Li, H., Li, Z., Qi, JI., Wang, L., Hu, Y., Xiang, J., Zhang, C., and Guo, J. (2018) A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Mol Phylogenet Evol* **127:** 961-77.

Rahman, S.U., Domier, L. L., Raza, G., Ahmed, N., McCoppin, N. K., Amin, I., and Mansoor, S. (2023). Metagenomic study for the identification of viruses infecting soybean in Pakistan. *Australas Plant Pathol* **52:** 191-194.

Ren, G., Jiang, Y., Li, A., Yin, M., Li, M., Mu, W., Wu, Y., and Liu, J. (2022) The genome sequence provides insights into salt tolerance of *Achnatherum splendens* (Gramineae), a constructive species of alkaline grassland. *Plant Biotechnol J* **20:** 116-28.

Rong, W., Rollin, J., Hanafi, M., Roux, N., and Massart, S. (2023) Validation of high-throughput sequencing as virus indexing test for Musa germplasm: Performance criteria evaluation and contamination monitoring using an alien control. *PhytoFrontiers™* **3:** 91-102.

Saito, T., Opio, P., Wang, S., Ohkawa, K., Kondo, S., Maejima, T., and Ohara, H. (2019) Association of auxin, cytokinin, abscisic acid, and plant peptide response genes during adventitious root formation in Marubakaido apple rootstock (*Malus prunifolia* Borkh. var. ringo Asami). Acta Physiol Plant **41:** 1-0.

Sanfaçon, H. (2015). Secoviridae: a family of plant picorna-like viruses with monopartite or bipartite genomes. *eLS*, 1-14.

Sanfaçon, H. (2022) Re-examination of nepovirus polyprotein cleavage sites highlights the diverse specificities and evolutionary relationships of nepovirus 3C-like proteases. *Arch Virol* **167:** 2529-2543.

Sanfaçon, H., Dasgupta, I., Fuchs, M., Karasev, A. V., Petrzik, K., Thompson, J. R., ... and Yoshikawa, N. (2020). Proposed revision of the family Secoviridae taxonomy to create three subgenera,"Satsumavirus","Stramovirus" and "Cholivirus", in the genus Sadwavirus. *Arch Virol* **165:** 527-533.

Sanfaçon, H., Wellink, J., Le Gall, O., Karasev, A., Van der Vlugt, R., and Wetzel, T. (2009) Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Arch Virol* **154:** 899-907.

Scharmann, M., Rebelo, A.G., and Pannell, J.R. (2021) High rates of evolution preceded shifts to sex-biased gene expression in Leucadendron, the most sexually dimorphic angiosperms. *Elife* **10:** e67485.

Shi, Z., Halaly-Basha, T., Zheng, C., Sharabi-Schwager, M., Wang, C., Galbraith, D.W., Ophir, R., Pang, X., and Or, E. (2020) Identification of potential post-ethylene events in the signaling cascade induced by stimuli of bud dormancy release in grapevine. *Plant J* **104:** 1251-68.

Sidharthan, V. K., Rajeswari, V., and Baranwal, V. K. (2022c) Analysis of public domain plant transcriptomes expands the phylogenetic diversity of the family Secoviridae. *Virus Genes* **58:** 598-604.

Sidharthan, V. K., Rajeswari, V., Vanamala, G., and Baranwal, V.K. (2022a). Revisiting the amalgaviral landscapes in plant transcriptomes expands the host range of plant amalgaviruses. *Virology*, **577:** 65-73.

Sidharthan, V.K., Kalaivanan, N.S., and Baranwal, V.K. (2021) Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. *Gene*, **786:** 145626.

Sidharthan, V.K., Nagendran, K., and Baranwal, V.K. (2022b). Exploration of plant transcriptomes reveals five putative novel poleroviruses and an enamovirus. *Virus Genes*, **58:** 244-253.

Sidharthan, V.K., Rajeswari, V., and Baranwal, V. K. (2023a). Broadening the host range and genetic diversity of waikaviruses. *Virology*, **582:** 106-113.

Sidharthan, V.K., Vanamala, G., Rajeswari, V., and Baranwal, V.K. (2023b) Identification of a putative novel cholivirus in the transcriptome of *Gymnema sylvestre* R. Br. *Arch Microbiol* **205:** 186.

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., ... and Zerbini, F. M. (2017) Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* **15:** 161-168.

Sun, Y., Zhang, T., Xu, X., Yang, Y., Tong, H., Mur, L.A., and Yuan H. (2021) Transcriptomic characterization of nitrate-enhanced stevioside glycoside synthesis in stevia (*Stevia rebaudiana*) bertoni. *Int J Mol Sci* **22:** 8549.

Thompson, J.R., Dasgupta, I., Fuchs, M., Iwanami, T., Karasev, A. V., Petrzik, K., ... and ICTV Report Consortium. (2017) ICTV virus taxonomy profile: Secoviridae. *J Gen Virol* **98:** 529-531.

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A. and Minh, B.Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44:** W232-5.

Unruh, S.A, McKain, M.R., Lee, Y.I., Yukawa, T., McCormick, M.K., Shefferson, R.P., Smithson, A., Leebens-Mack, J.H, and Pires, J.C. (2018) Phylotranscriptomic analysis and genome evolution of the Cypripedioideae (Orchidaceae). *Am J Bot* **105:** 631-40.

Walker, J.F., Yang, Y., Feng, T., Timoneda, A., Mikenas, J., Hutchison, V., Edwards, C., Wang, N., Ahluwalia, S., Olivieri, J., and Walker-Hale, N. (2018). From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am J Bot* **105:** 446-62.

Wang, J., Su, H., Han, H., Wang, W., Li, M., Zhou, Y., Li, Y., and Li, M. (2021a) Transcriptomics reveals host-dependent differences of polysaccharides biosynthesis in *Cynomorium songaricum*. *Molecules* **27:** 44.

Wang, M., Gao, L., Li, G., Zhou, C., Jian, J., Xing, Z., Wang, Y., Zhang, W., Song, Z., Hu, Y., and Yang, J. (2021b) Interspecific variation in the unsaturation level of seed oils were associated with the expression pattern shifts of duplicated desaturase genes and the potential role of other regulatory genes. *Front Plant Sci* **11:** 616338.

Wen, J., Yu, Y., Xie, D.F., Peng, C., Liu, Q., Zhou, S.D., and He, X.J. (2020) A transcriptome-based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioideae (Apiaceae). *Ann Bot* **125:** 937-53.

Wen, X., Li, J., Wang, L., Lu, C., Gao, Q., Xu, P., Pu, Y., Zhang, Q., Hong, Y., Hong, L., and Huang, H. (2022) The *Chrysanthemum lavandulifolium* genome and the molecular mechanism underlying diverse capitulum types. *Hort Res* **9:** uhab022.

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., ... and Goldberg, T. (2016) MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinform* **32:** 3501-3503.

Yao, X., Wang, S., Wang, Z., Li, D., Jiang, Q., Zhang, Q., Gao, L., Zhong, C., Huang, H., and Liu, Y. (2022) The genome sequencing and comparative analysis of a wild kiwifruit *Actinidia eriantha*. *Mol Hortic* **2:** 13.

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, CH., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., and Ma, H (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-

genome duplications. *Mol Biol Evol* **37:** 3188-210.

Zhang, G., Ge, C., Xu, P., Wang, S., Cheng, S., Han, Y., Wang, Y., Zhuang, Y., Hou, X., Yu, T., and Xu, X. (2021a). The reference genome of *Miscanthus floridulus* illuminates the evolution of Saccharinae. *Nat Plants* **7:** 608-18.

Zhang, G.Q., Liu, K.W., Li, Z., Lohaus, R., Hsiao, Y.Y., Niu, S.C., Wang, J.Y., Lin, Y.C., Xu, Q., Chen, L.J., and Yoshida, K. (2017a) The Apostasia genome and the evolution of orchids. *Nature* **549:** 379-83.

Zhang, L., Kong, H., Ma, H., and Yang, J. (2018) Phylogenomic detection and functional prediction of genes potentially important for plant meiosis. *Gene* **643:** 83-97.

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., Yan, C., Jia, Y., Larkin, R.M., Lavelle, D., and Truco, M.J. (2017b) RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat commun* **8:** 2264.

Zhang, W., Lin, J., Li, J., Zheng, S., Zhang, X., Chen, S., Ma, X., Dong, F., Jia, H., Xu, X., and Yang, Z. (2021b) Rambutan genome revealed gene networks for spine formation and aril development. *Plant J* **108:**1037-52.

Zhang, Y., Ye, Z.X., Feng, X.X., Xu, Z.T., Chen, J.P., Zhang, C.X., and Li, J.M. (2023) Prevalence of Reversed Genome Organizations for Viruses in the Family Iflaviridae, Order Picornavirales. *Microbiol Spect* e04738-22.

Zhang, Z., Wang, P., Li, Y., Ma, L., Li, L., Yang, R., Ma, Y., Wang, S.A., and Wang, Q. (2014) Global transcriptome analysis and identification of the flowering regulatory genes expressed in leaves of *Lagerstroemia indica*. *DNA Cell Biol* **33:** 680-8.

Zhao, W., Zhu, H., Wei, F., Zhou, D., Li, Y., and Zhang, X.X. (2021a) Investigating the involvement of cytoskeletal proteins MreB and FtsZ in the origin of legume-rhizobial symbiosis. *Mol Plant Microbe Interact* **34:** 547-59.

Zhao, Y.H., Zhang, X.M., and Li, D.Z. (2021b) Development of the petaloid bracts of a paleoherb species, *Saururus chinensis*. *Plos One* **16:** e0255679.

Zheng, H., Yu, M.Y., Han, Y., Tai, B., Ni, S.F., Ji, R.F., Pu, C.J., Chen, K., Li, F.Q., Xiao, H., Shen, Y. (2022) Comparative transcriptomics and metabolites analysis of two closely related Euphorbia species reveal environmental adaptation mechanism and active ingredients difference. *Front Plant Sci* **13:** 905275.

Zheng, Y., Wang, P., Chen, X., Yue, C., Guo, Y., Yang, J., Sun, Y., and Ye, N. (2021) Integrated transcriptomics and metabolomics provide novel insight into changes in specialized metabolites in an albino tea cultivar (*Camellia sinensis* (L.) O. Kuntz). *Plant Physiol Biochem* **160:** 27-36.

Figures

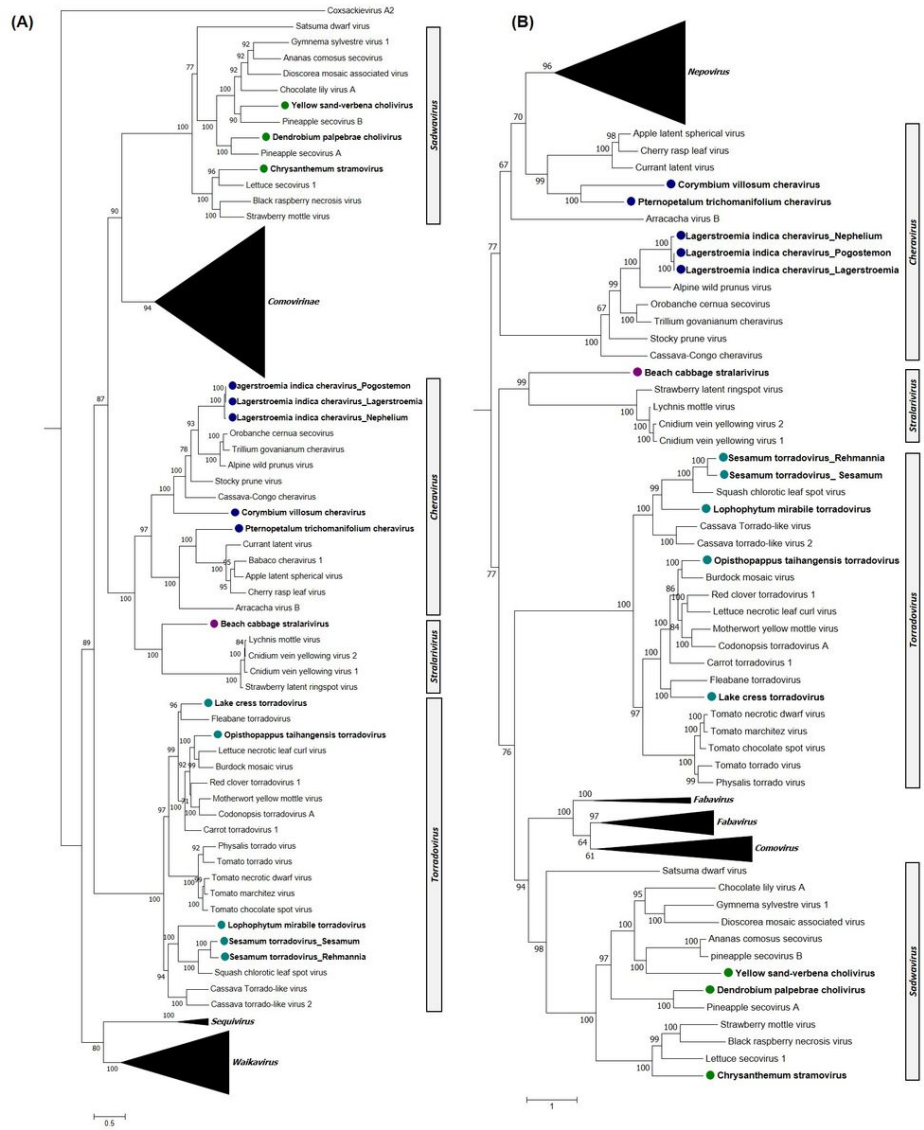

Figure 1

Maximum likelihood trees showing the phylogenetic relationships of identified putative novel chera-, sadwa-, stralari- and torradoviruses with known members based on the conserved Pro-Pol (A) and polyprotein 2 (B) amino acid sequences. Viruses identified in this study are shown in bold. Only bootstrap values >50 are indicated.

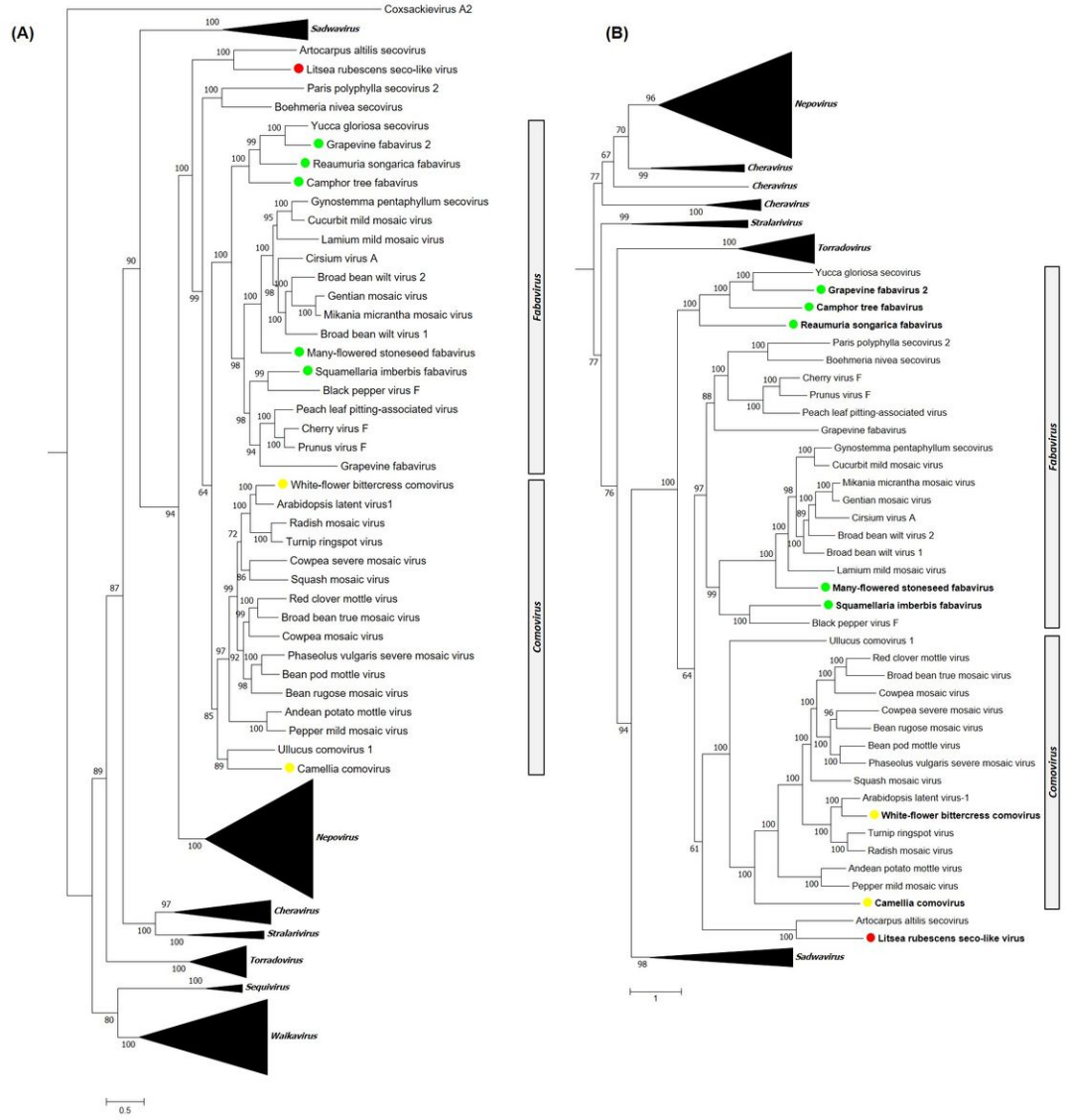

Figure 2

Maximum likelihood trees showing the phylogenetic relationships of identified putative novel como- and fabaviruses with known members based on the conserved Pro-Pol (A) and polyprotein 2 (B) amino acid sequences. Viruses identified in this study are shown in bold. Only bootstrap values >50 are indicated.

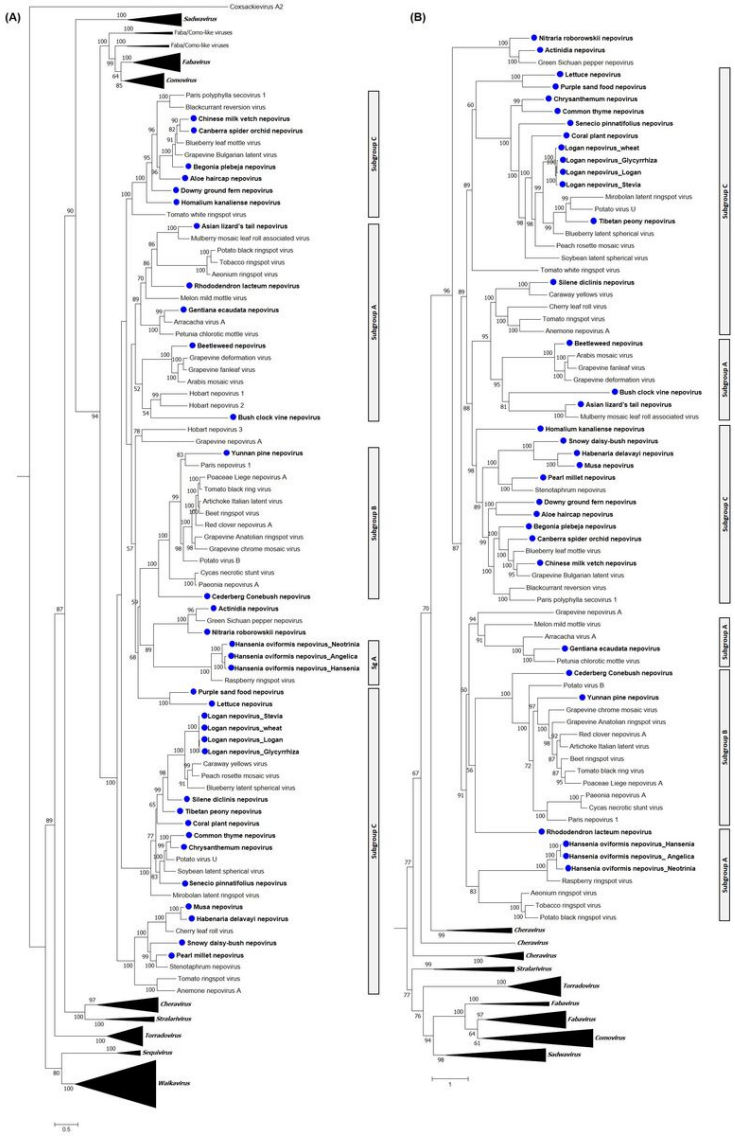

Figure 3

Maximum likelihood trees showing the phylogenetic relationships of identified putative novel nepoviruses with known members based on the conserved Pro-Pol (A) and polyprotein 2 (B) amino acid sequences. Viruses identified in this study are shown in bold. Only bootstrap values >50 are indicated.

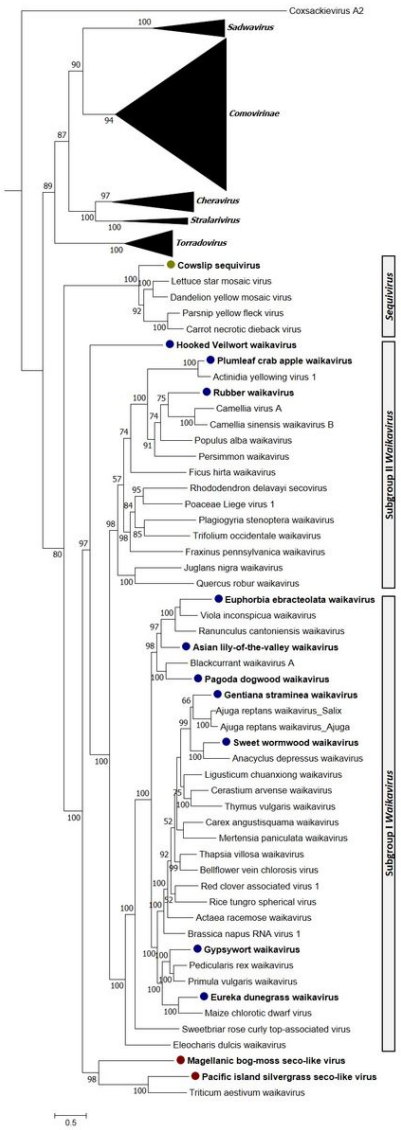

Figure 4

Maximum likelihood tree showing the phylogenetic relationship of identified putative novel sequi- and waikaviruses with known members based on the conserved Pro-Pol amino acid sequences. Viruses identified in this study are shown in bold. Only bootstrap values >50 are indicated.

Supplementary Files

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

- SupplementaryFigures.docx
- SupplementaryTables.xlsx