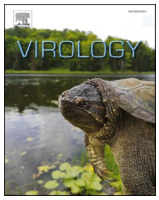
Virology 600 (2024) 110256

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology



From the discovery of a novel arepavirus in diseased arecanut palms (*Areca catechu* L.) in India to the identification of known and novel arepaviruses in bee and plant transcriptomes through data-mining

R. Thava Prakasa Pandian [a,1], Bhavishya [a,1], V. Kavi Sidharthan [b,1], M.K. Rajesh [a,*], Merin Babu [c], Susheel Kumar Sharma [d], B.J. Nirmal Kumar [a], M. Chaithra [e], Vinayaka Hegde [f]

[a] *ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI) Regional Station, Vittal, Karnataka, India*
[b] *Division of Genetics and Tree Improvement, ICFRE-Institute of Forest Biodiversity (IFB), Hyderabad, India*
[c] *ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI) Regional Station, Kayamkulam, Kerala, India*
[d] *ICAR-Indian Agricultural Research Institute (ICAR-IARI), New Delhi, India*
[e] *ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI) Research Centre, Kakikuchi, Assam, India*
[f] *ICAR-Central Plantation Crops Research Institute (ICAR-CPCRI) Kasaragod, Kerala, India*

ARTICLE INFO

Handling Editor: Prof. Xue-Ping Zhou

Keywords:
Necrotic ring spot disease
Areca palm
India
Arepavirus
Novel
Data-mining

ABSTRACT

Arecanut palm is a commercially important plantation crop valued for its nut. In this investigation, we report the discovery of a putative novel arepavirus, named areca palm necrotic ringspot virus 2 (ANRSV2), in necrotic ringspot diseased areca palms in Bantwal, Dakshina Kannada, Karnataka, India through RNA-sequencing and transmission electron microscopy. Further, the presence of ANRSV2 in the diseased samples was confirmed through reverse transcriptase-polymerase chain reaction assays. In addition, by mining public domain transcriptome data for arepaviral sequences, we identified a putative novel arepavirus in *Psychotria rubra*, a non-palm host. We recovered the genome sequences of the areca palm necrotic ringspot virus in honey bees, tomato, *Onobrychis viciifolia,* and *Rhamnus heterophylla*. These findings broaden our comprehension of arepaviral diversity and host range, and suggest an intriguing possibility of pollen-mediated arepaviral transmission that necessitates empirical validation. Further studies are needed to understand the biology of identified putative novel arepaviruses.

1. Introduction

Arecanut palm (*Areca catechu* L.; Arecaceae), also known as betelnut tree, is predominantly cultivated as a commercial plantation crop in India and other Southeast Asian countries. Primarily used for masticatory purposes, arecanut also holds significant medicinal, religious, and socio-economic value in India (Bhat et al., 2024). Additionally, arecanut is a vital component in traditional Chinese medicine practices (Tong et al., 2024), where it is utilized for its various therapeutic properties. Earlier research has highlighted its potential in treating ailments such as digestive and nervous system disorders, and it is also believed to possess anti-inflammatory and anti-microbial properties (Sun et al., 2024; Tong et al., 2024).

Biotic factors, including diseases caused by oomycetes, fungal and bacterial pathogens, and insect pests, play a pivotal role in determining arecanut production and productivity (Chowdappa et al., 2016). Fruit rot disease (caused by *Phytophthora meadii*) is responsible for the major yield loss in the humid tropics of Southern India during the peak southwest monsoon season (Patil et al., 2022). Occurrence of yellow leaf disease and root grub (*Leucopholis* spp.) also causes severe yield loss and death of the palms in this region (Nampoothiri et al., 2000; Mohan et al., 2022). With the changing climate, inflorescence dieback and leaf spot diseases are emerging as serious threats in many arecanut-growing areas in India, especially in the southern and north-eastern regions, resulting in huge crop losses (Pandian et al., 2024). However, no viral disease has been reported in areca palms in India so far.

Recently, areca palm necrotic spindle-spot virus (ANSSV) and areca palm necrotic ringspot virus (ANRSV) causing necrotic spindle-spot and

* Corresponding author.
E-mail address: rajesh.mk@icar.gov.in (M.K. Rajesh).
[1] Authors contributed equally to the work.

https://doi.org/10.1016/j.virol.2024.110256
Received 18 August 2024; Received in revised form 22 September 2024; Accepted 30 September 2024
Available online 2 October 2024

ringspot diseases, respectively, have been reported in areca palms in Hainan, China (Yang et al., 2018, 2019). Though ANSSV and ANRSV are closely related, properties like the high prevalence of ANRSV and the exclusion of ANSSV in ANRSV-infected plants distinguish both viruses (Wang et al., 2021). Both ANSSV and ANRSV have been recognized as members of a new *Arepavirus* genus in the family *Potyviridae* (Inoue-Nagata et al., 2022).

The family *Potyviridae* is the largest of plant-infecting RNA viruses and is known to have members of potential economic importance, which causes significant yield loss in many agricultural and horticultural crops (Revers and García, 2015; Martínez-Turiño and García, 2020). As per the latest report of the International Committee on Taxonomy of Viruses (ICTV), the family *Potyviridae* consists of 12 genera that are distinguished based on the host range, genomic features, and phylogenetic analysis (Inoue-Nagata et al., 2022). Most of the members in this family possess monopartite, positive-sense, single-stranded RNA as the genetic material (Martínez-Turiño and García, 2020) with flexuous filamentous particles (11–20 nm x 680–900 nm) except for bymoviruses, which contain bipartite genome. The potyvirus genome (about 10 Kb long), with a genome-associated protein (VPg) at the 5' end and a 3′ poly(A) tail, possesses a single long open reading frame (ORF) encoding a polyprotein that is cleaved by three viral proteases (P1, HC-Pro, and NIa) into ten smaller functional proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP). Besides, a small frameshift-derived peptide is produced from Pretty interesting Potyviridae ORF (PIPO) (Revers and García, 2015; Martínez-Turiño and García, 2020).

During our field survey, necrotic ringspot symptoms were observed in areca palms growing in Dakshina Kannada District, Karnataka State, India, similar to those induced by ANRSV (Yang et al., 2019). Search for viral particles in one of the diseased samples under TEM revealed the presence of flexuous viral particles like those of arepaviruses (Yang et al., 2018, 2019). To further characterize the associated virus, we resorted to next-generation sequencing (NGS) of isolated total RNA from diseased samples, as this approach can detect multiple known and novel viruses, if any, present in the sample, unlike the conventional virus detection techniques (Sidharthan et al., 2020). Through RNA-sequencing and reverse transcription-polymerase chain reaction (RT-PCR), we confirm the association of a putative new arepavirus, tentatively named areca palm necrotic ring spot virus 2 (ANRSV2), with diseased areca palms in Karnataka, India, showing ringspot symptoms in the present study. Besides, we explored the public domain transcriptome data to identify additional novel arepavirus members and broaden the host range of known arepaviruses.

2. Materials and methods

2.1. Plant materials

Symptomatic leaf samples were collected from the diseased palms in Kundadka and Neralakatte plots and named KUN1 and NER1, respectively, in Bantwal Taluk of Dakshina Kannada district, Karnataka State, India. Leaf samples were also collected from an asymptomatic palm in the Kundadka plot named KUN2. All the collected samples were subjected to RNA-sequencing, while one of the symptomatic samples was subjected to transmission electron microscopy (TEM) imaging.

2.2. Transmission electron microscopy (TEM) studies

TEM was employed to detect virus particles in leaf samples derived from a diseased arecanut palm with typical necrotic ringspot symptoms. Fresh leaf tissue was mixed with sterilized distilled water at a 1:10 ratio (g ml^{-1}) and ground with a pestle and mortar. The mixture was incubated on ice for 5 min, and the supernatant was immediately placed onto 200-mesh copper grids. After a 2-min absorption period, the grids were negatively stained with 2% phosphotungstic acid (pH 7.0) for 1 min. The copper grids were then dried under a tungsten lamp for 30 min and examined using a TEM (JEOL JEM1101, Japan).

2.3. RNA extraction and RNA-seq library construction and sequencing

Total RNA was extracted from two symptomatic (KUN1 and NER1) and one healthy (KUN2) leaf tissue (1 g) using the TRIzol reagent (Invitrogen, U.S.A.). The quantity and purity of the RNA were analyzed using a Tapestation 2200 (Agilent Technologies, U.S.A.). Isolated RNA from the three samples with RNA integrity number more than 6.0 were depleted of ribosomal RNA (rRNA) using Fast select rRNA Plant kit (Qiagen, Germany) as per manufacturer's protocol. Later, the NGS libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England BioLabs, U.S.A.) according to the manufacturer's protocol. Library quality was checked using Qubit 4.0 Fluorometer and using the Agilent 2200 TapeStation System utilizing High Sensitivity D1000 ScreenTape analysis (Agilent Technologies, U.S.A.). Paired-end sequencing reads of libraries were performed using the Illumina NovaSeq 6000 System with a read length of 150 bp (M/s Bionivid Technology [P] Limited, Bengaluru, India).

2.4. Virus identification in NGS data derived from diseased areca palms and their genome recovery

NGS data analysis was performed in the Galaxy server (The Galaxy Community, 2024). Raw reads were trimmed using Trimmomatic (v 0.39) (Bolger et al., 2014) to remove the adapter and poor quality (quality threshold: 30) sequences. Trimmed reads were *de novo* assembled into contigs using rnaviralspades (v 0.39) (Bushmanova et al., 2019). BLASTx analysis of assembled contigs (e-value: 1e-5) was performed against reference viral/viroid protein sequences downloaded from the National Centre for Biotechnology Information (NCBI) viral genomes database (https://www.ncbi.nlm.nih.gov/genome/viruses/) using NCBI BLAST + tool (v 2.14.1) (Cock et al., 2015). Open reading frames (ORFs) (≥300 nt) were predicted in the viral genomes using NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/). Molecular weights (MW) of the encoded proteins were determined using the 'compute pI/Mw' tool in the Expasy server (https://web.expasy.org/compute_pi/), while motifs were predicted using the Motif Search tool (https://www.genome.jp/tools/motif/). Pretty interesting potyviridae ORF (PIPO) and polyprotein cleavage sites were determined in recovered viral genomes and their encoded polyproteins, respectively, compared with known viral genomes and polyprotein sequences. The mean depth of recovered arepaviral genomes was obtained by mapping trimmed reads of the respective library to the recovered genomes using HISAT2 (v 2.2.1) (Kim et al., 2015) followed by visualization using QualiMap (v 2.3) (Okonechnikov et al., 2016).

2.5. Detection of the identified novel arepavirus in diseased areca palm samples through reverse transcription- polymerase chain reaction (RT-PCR)

RT-PCR assay was performed using RNA isolated from necrotic ringspot diseased (six) and asymptomatic (four) areca palm leaf samples, including the samples used for RNA-sequencing, for the detection of identified novel arepavirus. Complementary DNA (cDNA) was synthesized from 1 μg of total RNA using Moloney Murine Leukemia Virus reverse transcriptase and random primers (G-Biosciences, USA). CP-F2/CP-R and CP-F2/CP-R2 primers designed by Yang et al. (2019) were used for the subsequent PCR reactions with the following conditions: pre-denaturation at 94 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min. The PCR products were visualized in 3% agarose gel post electrophoresis. Positive amplicons of two samples obtained using CP-F2/CP-R primers were purified using a DNA gel extraction kit (QIAGEN, Germany), and subjected to bidirectional Sanger sequencing.

2.6. Identification of putative arepavirus-positive SRA libraries and recovery of genomes of arepavirus isolates

Sequences of areca palm necrotic ringspot virus (ANRSV) and areca palm necrotic spindle-spot virus (ANSSV) were used as queries in RNA-dependent RNA polymerase (RdRp) searches in the Serratus explorer (Edgar et al., 2022) to identify putative arepavirus-positive SRA libraries. Further, the identified libraries were downloaded, trimmed, and *de novo* assembled, as mentioned in Section 2.4. Assembled contigs were subjected to BLASTn analysis (e-value: 1e-5) against reference genome sequences of known arepaviruses using NCBI BLAST+. ORF and motif predictions in recovered genomes were performed, as mentioned in Section 2.4.

2.7. Sequence similarity and phylogenetic analyses

Recovered and known arepavirus genomes and their encoded polyprotein were aligned using CLUSTALW and MUSCLE, respectively, in Sequence Demarcation Tool (v 1.2) (Muhire et al., 2014) and sequence identity matrices were obtained. MUSCLE-aligned polyprotein sequences of members of the *Potyviridae* family were subjected to maximum-likelihood (ML) tree construction using the best-fit 'LG + G + I + F' model with 100 bootstrap replicates in MEGA7 (v 7.0.26) (Kumar et al., 2016).

3. Results

3.1. Areca palm necrotic ringspot disease incidence and symptoms

During a field survey in 2021, we observed necrotic ringspot disease in six to seven-years-old arecanut palms (n = 8; Cultivar-Mangala) in farmer's gardens at Vittal-Mudnoor village, Kundadka (12°49′30.8"N, 75°07′25.7"E) and six to seven-years-old palms (n = 11; Cultivar-Mangala) in Kodaje village, Neralakatte (12°49′30.8"N 75°07′25.7″ E) in Bantwal Taluk of Dakshina Kannada district, Karnataka State, India. The disease incidence was 1.6% and 3.2% in the Kundadka and Neralakatte plantations, respectively. In the initial stages, the diseased palms exhibited chlorotic ringspot symptoms on the lower and middle leaves. When the same diseased palms were revisited later, chlorotic ringspots that were observed earlier turned into necrotic ringspots. However, no symptom was observed on the top leaves of the palms (Fig. 1). When the crude sap of one of the diseased areca palm samples was visualized under TEM, flexuous filamentous particles were observed (Fig. 2).

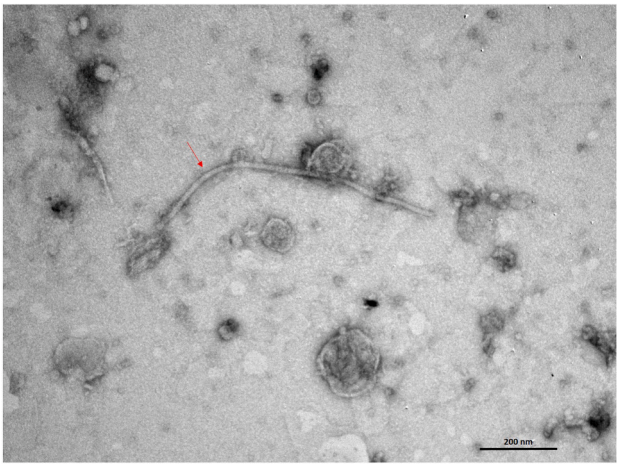


Fig. 2. Flexuous filamentous particle (indicated by red arrow) of ANRSV2 in the crude sap of a diseased areca palm visualized under transmission electron microscope.

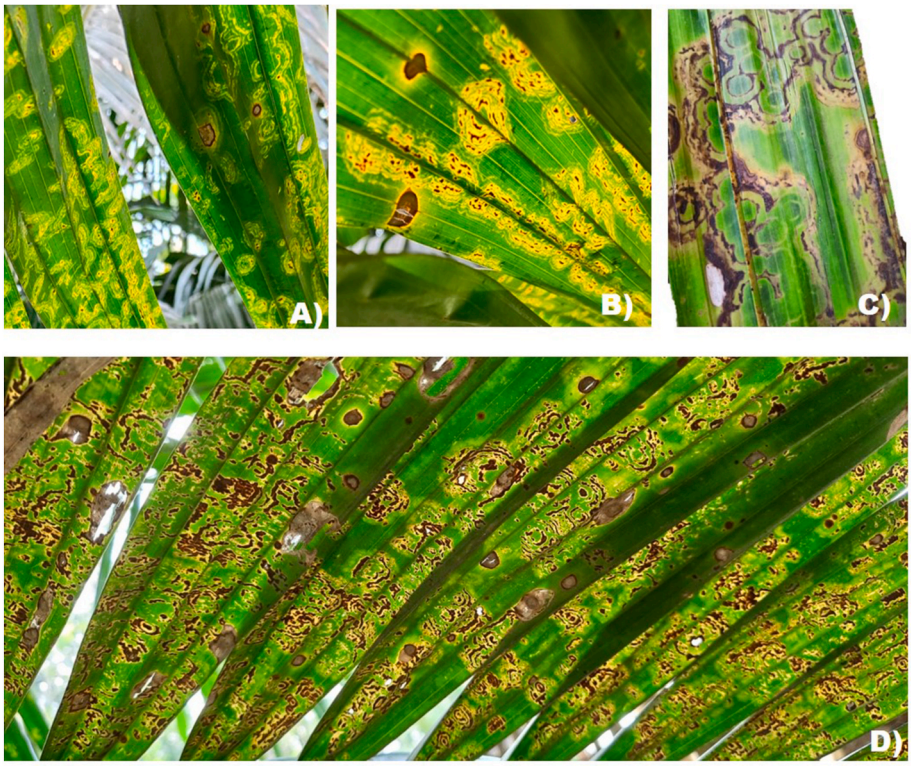


Fig. 1. Arecanut necrotic ring spot disease symptoms; A) Chlorotic ringspots on lower leaves; B) progression of chlorotic ringspots to necrotic ringspots; C) Necrotic ring spots seen on lower leaves; D) Severely infected lower leaves with necrotic ringspots.

3.2. Identification of a novel arepavirus in diseased areca palms and its genome recovery

In total, 45.8, 47.1, and 68.8 million raw reads were obtained from three arceanut palm samples-KUN-1, KUN-2, and NER-1, respectively, of which 38.9, 40.7, and 58.5 million reads survived trimming. BLASTx analysis of *de novo* assembled contigs identified a viral contig of length 9485 nt and 9421 nt in diseased samples KUN-1 and NER-1, respectively, that shared 64.3% and 68.1% sequence identity with the reference polyprotein sequence of ANRSV (YP_010087001.1). However, no plant viral contig was identified in the asymptomatic sample (KUN-2). The identified arepaviral contig from KUN-1 and NER-1 samples were regarded as coding-complete genomes of KUN-1 and NER-1 isolates, respectively, of a putative novel virus tentatively named as areca palm necrotic ringspot virus 2 (ANRSV2). A large ORF (9063 nt) encoding a 345.6 kDa polyprotein (3020 aa) with two helper component (HC) proteinase (PF00851) motifs near the N-terminal region followed by DEAD/DEAH box helicase (PF00270), helicase conserved C-terminal domain (PF00271), potyviridae polyprotein (PF08440), peptidase family C4 (PF00863), viral RdRp (PF00680) and potyvirus coat protein (CP) (PF00767) motifs was predicted in ANRSV2 genomes. Nine putative cleavage sites were determined in the encoded polyprotein that upon processing by viral proteinases would yield ten mature proteins- HC-Pro1 (262 aa), HC-Pro2 (318 aa), P3 (371 aa), 7K (59 aa), CI (649 aa), 9K (83 aa), NIa-VPg (174 aa), NIa-Pro (214 aa), NIb (610 aa) and CP (280 aa) in the order starting from N-terminal region. Of the two HC proteinases determined in ANRSV2 polyprotein, one was located in the HC-Pro1 region while the other was in the HC-Pro-2 region. In HC-Pro1 and HC-Pro2 of ANRSV2, conserved glycine, cysteine, and histidine residues in the putative sites of catalytic activity (Yang et al., 2019) were determined in the HC protease region (Fig. S1). In CI and NIb of ANRSV2, the conserved motif(s) associated with helicase activity- ^{1094}GSGKS(X)$_3$P^{1102} and ^{1187}DEXH^{1190} (Sorel et al., 2014) and replicase activity- ^{2580}GDD^{2582} (Deng et al., 2015), respectively were determined while the conserved residues- H^{1961}, D^{1993}, C^{2061}, H^{2077} in the NIa-Pro active site and the conserved ^{2059}GXCG^{2062} around the 'C' residue in the active site (Adams et al., 2005) were located in NIa-Pro of ANRSV2. In addition to polyprotein ORF, PIPO was determined in the P3 cistronic region in the +2 reading frame of polyprotein ORF (Fig. 3). The conserved $^{2619/2609}$GAAAAAAA$^{2616/2626}$ motif (Chung et al., 2008) was determined near the 5′ region of PIPO. Polyprotein of ANRSV2 isolates shared a maximum of 68.6–68.8% nucleotide and 68.5–68.6% amino acid sequence identities with the respective sequences of ANRSV isolates. In comparison, the polyprotein of both the ANRSV2 isolates shared 95.9% nucleotide and 98.7% amino acid sequence identities with each other (Fig. 4A and B). HC-Pro1, HC-Pro2, P3, 7K, CI, 9K, NIa-VPg, NIa-Pro, NIb and CP of ANRSV2 isolates shared a maximum of 54.6%, 64.4%, 58.0%, 81.4%, 78.3%, 51.8%, 83.3%, 75.7%, 72.8% and 85.7%, amino acid sequence identities, respectively with the corresponding sequences of other arepaviral species (Fig. 4C–L). On the other hand, the PIPO-encoded protein of ANRSV2 isolates, sharing 97.8% sequence identity among each other, shared a maximum of 63.0% sequence identity with the corresponding protein of other arepavirus (Fig. 4M). The mean depth of the recovered ANRSV2 genome was 1839.7x and 1959.6x in the KUN-1 and NER-1 libraries, respectively. Phylogenetic analysis based on polyprotein sequence revealed that ANRSV2 isolates together formed a distinct clade within arepaviruses (Fig. 5). In addition to ANRSV2, contigs of cucumber mosaic virus (CMV) were also identified in NER-1 library, of which two longer contigs of lengths 2064 nt and 1294 nt corresponded to complete RNA3 and partial RNA1 segment, respectively. The presence of ANRSV2 in six diseased samples and its absence in four asymptomatic samples was confirmed through RT-PCR assays (Fig. S2). Sanger sequencing of obtained amplicons (ca. 530 nt) using CP-F2/CP-R primers from two of the diseased samples and BLAST analysis revealed 96.8%–98.0% nucleotide identities of amplicons with the ANRSV2 genomes obtained through RNA sequencing.

3.3. Identification of arepavirus-positive transcriptome libraries through data-mining

Through Serratus RdRp searches, we identified 14 arepavirus-positive transcriptome libraries derived from four plant species (*Onobrychis viciifolia*, *Psychotria rubra*, *Rhamnus heterophylla,* and *Solanum lycopersicum*) and an insect species (*Apis mellifera*). Six and five arepavirus-positive transcriptome libraries of *A. mellifera* and *O. viciifolia*, respectively, were identified, while one each arepavirus-positive transcriptome library of *P. rubra*, *R. heterophylla,* and *S. lycopersicum* was identified.

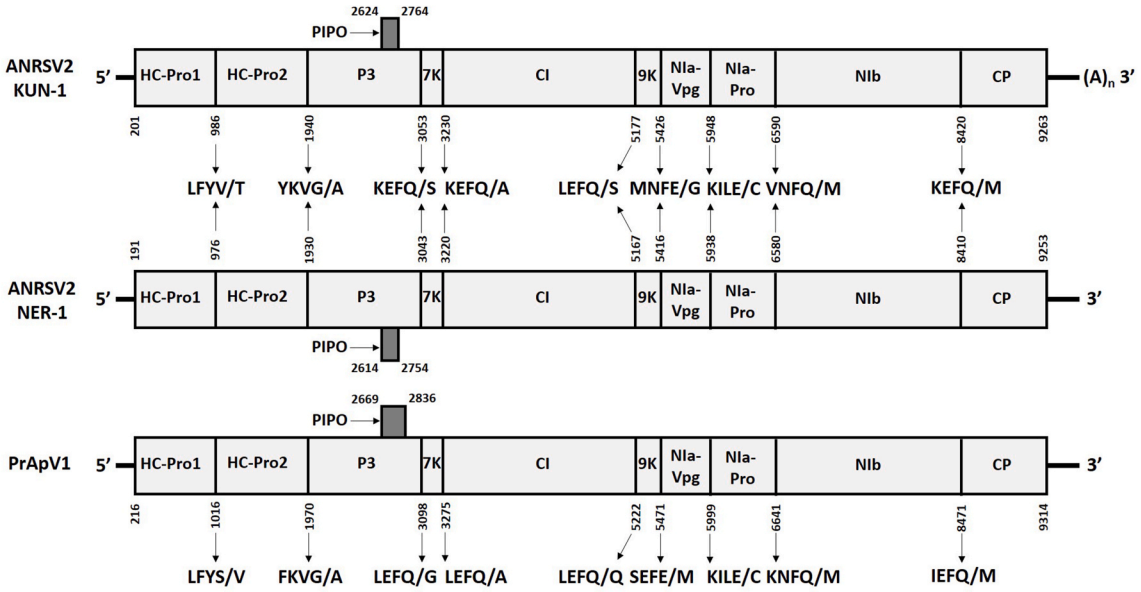


Fig. 3. Genome organization of putative novel arepaviruses ANRSV2 (isolates KUN-1 and NER-1) and PrApV1 identified in this study. Light and dark grey boxes indicate the polyprotein ORF and PIPO, respectively. Putative cleavage sites in the encoded polyprotein and the corresponding sequences are indicated.

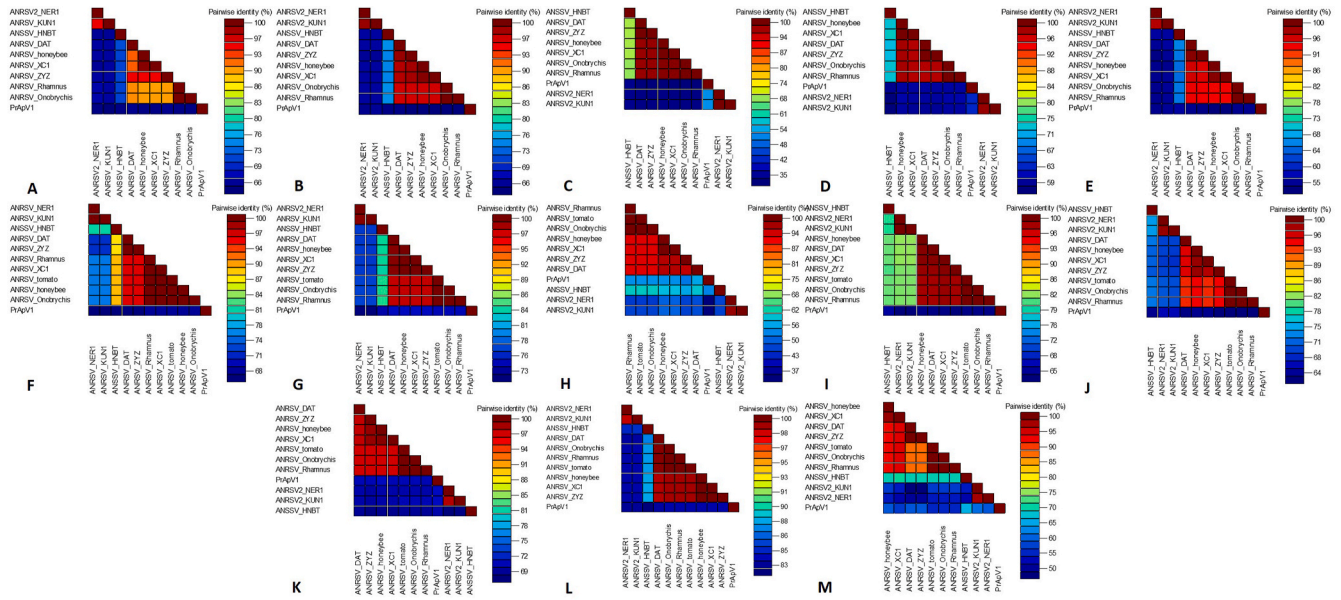


Fig. 4. Sequence identity matrices showing the pairwise sequence identities of known and novel arepaviral isolates of this study with other arepaviral isolates based on polyprotein nucleotide (A) and amino acid (B) sequences, and amino acid sequences of HC-Pro1 (C), HC-Pro2(D), P3 (E), 7K (F), CI (G), 9K(H), NIa-VPg (I), NIa-Pro (J), NIb (K), CP (L) and PIPO-encoded protein (M).

3.4. Identification of a novel arepavirus in Psychotria rubra transcriptome and its genome recovery

We identified a 9479 nt long contig of a putative novel arepavirus, tentatively named as Psychotria rubra arepavirus 1 (PrApV1), that shared 69.4% nucleotide sequence identity with the reference genome of ANSSV (NC_040836.1) in the arecavirus-positive *P. rubra* transcriptome library (SRR9722453) through BLAST analysis of assembled contigs. PrApV1 coding-complete genome contained a single large ORF (9099 nt) encoding a 346.7 kDa polyprotein (3032 aa) with two HC proteinase (PF00851) motifs near the N-terminal region followed by DEAD/DEAH box helicase (PF00270), helicase conserved C-terminal domain (PF00271), potyviridae polyprotein (PF08440), peptidase family C4 (PF00863), viral RdRp (PF00680) and potyvirus coat protein (CP) (PF00767) motifs. Nine putative cleavage sites were determined in PrApV1 polyprotein, which would be processed by viral proteinases to yield ten mature proteins- HC-Pro1 (267 aa), HC-Pro2 (318 aa), P3 (376 aa), 7K (59 aa), CI (649 aa), 9K (83 aa), NIa-VPg (176 aa), NIa-Pro (214 aa), NIb (610 aa) and CP (280 aa) in the order starting from N-terminal region (Fig. 3). Similar to ANRSV2, one each of the two HC proteinase motifs determined in PrApV1 polyprotein was located in HC-Pro1 and HC-Pro2 regions that contained the conserved glycine, cysteine and histidine residues in the putative catalytic sites (Yang et al., 2019) (Fig. S1). Similarly, the conserved motifs- ^{1104}GSGKS(X)$_3$P^{1112} and ^{1197}DEXH^{1200} (Sorel et al., 2014) associated with helicase activity, ^{2592}GDD^{2594} (Deng et al., 2015) associated with replicase activity and the conserved residues- H^{1973}, D^{2005}, C^{2073}, H^{2089} of the NIa-Pro active site along with ^{2071}GXCG^{2074} spanning the 'C' residue (Adams et al., 2005) were found in PrApV1 mature proteins- CI, NIb and NIa-Pro, respectively. Besides, in the P3 cistronic region of the PrApV1 genome, PIPO was determined, whose 5′ end contained the conserved ^{2664}GAAAAAAA^{2671} motif (Chung et al., 2008), in the +2 reading-frame of polyprotein ORF. PrApV1 polyprotein sequence shared a maximum of 67.3% and 67.6% identities at nucleotide and amino acid levels, respectively, with the corresponding sequences of other arepaviruses (Fig. 4A and B). PrApV1 HC-Pro1, HC-Pro2, P3, 7K, CI, 9K, NIa-VPg, NIa-Pro, NIb CP shared a maximum of 54.6%, 64.4%, 58.2%, 69.5%, 76.3%, 57.8%, 67.2%, 67.8% 72.8% and 84.3% amino acid sequence identities, respectively with the corresponding mature protein sequences of other arepaviruses (Fig. 4C–L). In contrast, the PIPO-encoded protein shared a maximum of 67.3% identity with the corresponding protein of ANSSV (Fig. 4M). The mean depth of the PrApV1 genome in the respective *P. rubra* library was 668.9x. Polyprotein sequence-based phylogeny placed PrApV1 in a distinct sub-clade among arepaviruses that were related to ANRSV2 rather than to ANRSV and ANSSV (Fig. 5).

3.5. Recovery of genomes of ANRSV isolates in arepavirus-positive transcriptome libraries

Assembly of reads of the arepavirus-positive *A. mellifera* (SRR8823607), *O. viciifolia* (SRR9656499), *R. heterophylla* (SRR11649679), and *S. lycopersicum* (SRR8835210) libraries (libraries with the highest number of arepaviral reads in cases of *A. mellifera* and *O. viciifolia*) and BLAST analysis revealed the presence of ANRSV contigs in the analyzed libraries. The longest ANRSV contig (6.9 kb–9.5 kb) in each library that shared 84.1%–99.2% nucleotide sequence identities with the reference genome of ANRSV (NC_055501.1) was regarded as the genome sequence of the respective host's ANRSV isolate. Except for the partial genome sequence of the ANRSV tomato isolate, the genome sequences of the remaining isolates were coding-complete. The mean depth of respective ANRSV genome recovered in the analyzed *A. mellifera, O. viciifolia, R. heterophylla,* and *S. lycopersicum* libraries was 185.9x, 877.6x, 85.4x, and 19.4, respectively. Phylogenetic analysis based on polyprotein sequences grouped all the identified ANRSV isolates with the ANRSV isolates of areca palm. However, the ANRSV isolates from *O. viciifolia, R. heterophylla,* and *S. lycopersicum* together formed a distinct sub-clade within the ANRSV group, while ANRSV *A. mellifera* isolates grouped with the ANRSV areca palm isolates (Fig. 5).

4. Discussion

The genus *Arepavirus* in the family *Potyviridae* contains two viruses-ANRSV and ANSSV causing necrotic ring spot and necrotic spindle spot diseases, respectively, in arecanut palms (Yang et al., 2018, 2019; Inoue-Nagata et al., 2022). Both viruses were first identified in diseased areca palms in Hainan, China (Yang et al., 2018, 2019) and have not yet been reported elsewhere. Unlike other potyviruses, arepaviruses encode

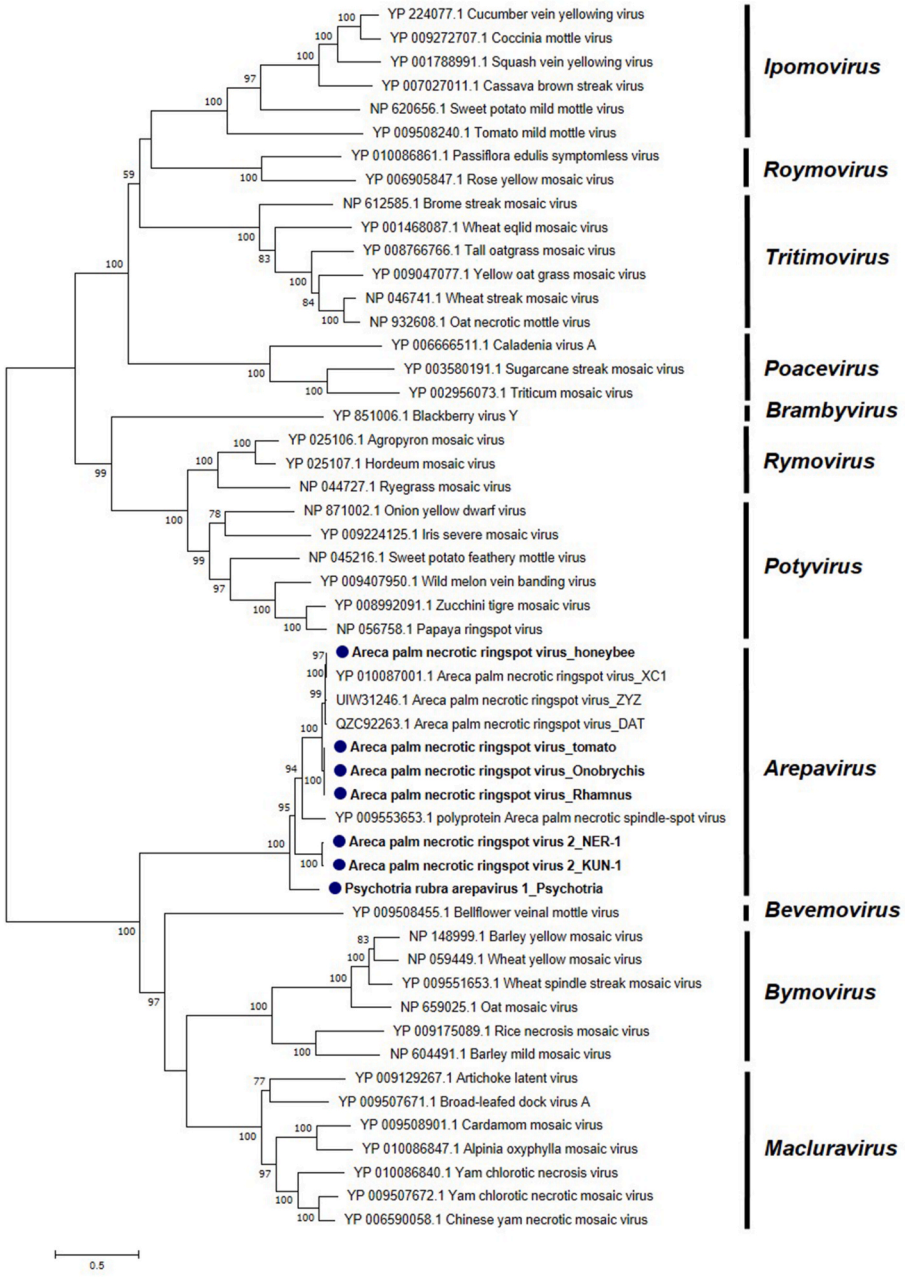


Fig. 5. Phylogenetic tree showing the relationships of identified viruses with other viruses of the family *Potyviridae* based on polyprotein sequences. The tree was constructed using the maximum-likelihood method and LG + G + I + F model with 100 bootstrap replicates. Bootstrap support values more than or equal to 50% are only shown. Viruses/viral isolates identified in this study are shown in bold and indicated by blue dots.

two mature cysteine proteinases designated as HC-Pro1 and HC-Pro2 that exhibit self-cleavage activity (Cui and Wang, 2019; Qin et al., 2020). Though the HC-Pro1 N-terminal region exhibits RNA silencing suppression activity, its function is suppressed by the C-terminal region. On the other hand, both the N-terminal and C-terminal regions of HC-Pro2 exhibit RNA silencing suppression activity, qualifying HC-Pro2 as a viral suppressor protein of the host's RNA silencing. Though functionally divergent, HC-Pro1 and HC-Pro2 are indispensable for arepaviral infection (Qin et al., 2020). Though ANRSV and ANSSV are closely related, they are antagonistic to each other, thus reducing the possibility of their co-infection. ANRSV causes severe disease and is relatively more widespread than ANSSV (Wang et al., 2021).

The genome organization of arepavirus-like sequences identified in diseased areca palm samples in the present study largely resembled ANRSV and ANSSV (Yang et al., 2018, 2019). However, the polyprotein sequence of the identified arepaviral isolates shared <70% nucleotide and amino acid sequence identities with the corresponding sequences of ANRSV and ANSSV isolates. As per the polyprotein sequence-based species demarcation criteria of the family *Potyviridae* (<76% nucleotide and <82% amino acid identity) (Inoue-Nagata et al., 2022), the arepaviral isolates identified in diseased areca palm samples in the present study are regarded as those of a putative new arepaviral species for which the name areca palm necrotic ringspot virus 2 (ANRSV2) is proposed. However, like ANRSV and ANSSV (Yang et al., 2019), CP of ANRSV2 isolates shared >80% amino acid sequence identities with other arepaviruses, which is above the species demarcation criteria (Inoue-Nagata et al., 2022).

The recent increase in Next Generation Sequencing (NGS) projects in a wide range of organisms worldwide has led to the increased deposition of NGS data to the SRA database (NCBI) and its mirror databases. This

offers scope for exploring novel/known viral sequences in the deposited publicly available NGS data, often unnoticed by the original submitters, across the breadth of living organisms (Lauber and Seitz, 2022; Sidharthan and Baranwal, 2024). Such data-centric 'virus-hunting' studies have expanded the host range and genetic diversities of several plant virus groups (Sidharthan and Baranwal, 2024). Because only two arepaviral species infecting arecanut palm hosts are known to date, we searched the publicly available NGS data for arepaviral sequences to expand the host range of known arepaviruses and identify putative novel arepaviruses. Interestingly, we identified ANRSV genome sequences in transcriptomes of *O. viciifolia* (Yin et al., 2022), *R. heterophylla* (Shang et al., 2024), and *S. lycopersicum* (Ding et al., 2019), which expands the potential host range of ANRSV to dicot hosts. Coincidentally, transcriptome data of all these species were derived from samples collected in China. To our surprise, the ANRSV genome was also identified in a transcriptome library derived from head and thorax segments of the newly emerged queen bee (*A. mellifera*), containing less than 10 million paired-end reads (Wei et al., 2019). Aphids generally transmit potyvirids, though plasmodiophorids, pollen, or seeds transmit some members. Unlike several potyvirids, arepaviruses lack the conserved aphid-transmission-associated motifs in HC-Pro and CP, suggesting their potential transmission by non-aphid vectors or other means (Inoue-Nagata et al., 2022). However, the mode of transmission of arepaviruses is yet to be identified. Identification of the ANRSV genome in bee transcriptome in the present study hints at the possible pollen transmission of arepaviruses by pollinators or wind. Roberts et al. (2018) adopted RNA-seq of *A. mellifera* bees for plant virus surveillance in Australia and identified a few potyviruses through that approach. As bees provide pollination services to plants, they serve as unbiased pollen samplers from plants. NGS of nucleic acid pools isolated from bees would help detect plant viruses prevailing in an area at an early stage (Roberts et al., 2018). It is noteworthy that the ANRSV genome was identified in the transcriptome of bees sampled in Nanchang, China, in 2018 (Wei et al., 2019), which coincided with the period (2017–2018) when necrotic ringspot disease in areca palms was first observed in Hainan, China (Yang et al., 2019). Phylogenetically, ANRSV areca palm isolates and honeybee isolates grouped, while ANRSV isolates of dicot hosts formed a distinct group. Notably, ANSSV sequences could not be identified in any of the NGS data available in the SRA database, reiterating the predominance of ANRSV over ANSSV (Wang et al., 2021).

Besides ANRSV, sequence of a novel arepavirus (PrApV1) was identified in the transcriptome of *P. rubra* (family Rubiaceae) sampled from Qingyuan, China (Liu et al., 2020), expanding the arepaviral host range to dicots. PrApV1 genome organization resembled other arepaviruses, and its encoded polyprotein contained all the conserved motifs determined in other arepaviral polyproteins (Yang et al., 2018, 2019). Based on polyprotein-based species demarcation criteria of the family *Potyviridae* (Inoue-Nagata et al., 2022), PrApV1, that shared <68% nucleotide and amino acid sequence identities with the respective polyprotein sequences of other arepaviruses, can be regarded as a putative new member of the genus *Arepavirus*. Like other arepaviruses, CP of PrApV1 also shared >80% amino acid identities with other arepaviral CP sequences, which, however, exceeds the CP sequence-based potyvirid species demarcation criteria (Inoue-Nagata et al., 2022). Nevertheless, unlike other arepaviruses, the PrApV1 CP amino acid sequence is the only mature protein sequence to share >77% identity with the corresponding sequence of other arepaviruses. PrApV1, a virus infecting non-areca palm host, formed a distinct sub-clade of arepaviruses in phylogenetic analyses. Like other data-mining studies (Sidharthan et al., 2024), host names of viruses/viral isolates identified through data-mining in this study were assigned solely based on the metadata provided in the SRA database, which should be regarded cautiously until further validated.

5. Conclusion

We identified, for the first time, a putative novel arepavirus associated with necrotic ringspot disease of areca palms in India. Besides, a putative novel arepavirus sequence was identified in the transcriptome of *P. rubra*. By discovering two putative novel arepaviruses, the current study expanded the genetic diversity and host range of arepaviruses. Identification of ANRSV in transcriptomes of three dicot hosts represents a significant expansion of the ANRSV host range, while its identification in honeybee transcriptomes hints at the possible pollen transmission of ANRSV by bees, which, however, needs to be validated in future studies. Therefore, the present study paves the way for further studies on the biological characterization of identified novel viruses and the genome sequences of identified novel viruses will help develop detection assays.

Data availability

Raw RNA-sequencing data generated in this study are available in NCBI under the Bioproject PRJNA1146205 and BioSample accessions SAMN43088681 (NER-1), SAMN43088682 (KUN-1) and SAMN43088683 (KUN-2). Viral genome sequences obtained from diseased arecanut palms in this study are available in NCBI's GenBank with accession numbers PQ197196 to PQ197201, while the arepaviral sequences obtained through data-mining are available in NCBI's Third Party Annotation database with accession numbers BK068283 to BK068287.

CRediT authorship contribution statement

R. Thava Prakasa Pandian: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Bhavishya:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **V. Kavi Sidharthan:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **M.K. Rajesh:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Merin Babu:** Writing – review & editing, Investigation, Formal analysis. **Susheel Kumar Sharma:** Writing – review & editing, Formal analysis. **B.J. Nirmal Kumar:** Writing – review & editing, Investigation, Formal analysis. **M. Chaithra:** Writing – review & editing, Investigation, Formal analysis. **Vinayaka Hegde:** Writing – review & editing, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Raw RNA-sequencing data generated in this study are available in NCBI under the Bioproject PRJNA1146205 and BioSample accessions SAMN43088681 (NER-1), SAMN43088682 (KUN-1) and SAMN43088683 (KUN-2)

Acknowledgments

The authors thank the Indian Council for Agricultural Research, New Delhi, for funding (ICAR-CPCRI Project No.1000765040). We thank the original submitters of transcriptome data analyzed in the present study and the National Centre for Biotechnology Information (NCBI) for publicizing the data. We also acknowledge the Galaxy platform and the Serratus project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2024.110256.

References

Adams, M.J., Antoniw, J.F., Beaudoin, F., 2005. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. Mol. Plant Pathol. 6 (4), 471–487.

Bhat, R., Sujatha, S., Bhavishya, Priya, U.K., Gupta, A., Uchoi, A., 2024. Arecanut (*Areca catechu* L.). In: Soil Health Management for Plantation Crops: Recent Advances and New Paradigms. Springer Nature Singapore, Singapore, pp. 177–206.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.

Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8 (9), giz100.

Chowdappa, P., Vinayaka Hegde, V.H., Pandian, R.T.P., Chaithra, M., 2016. Arecanut Diseases and Their Management.

Chung, B.Y.W., Miller, W.A., Atkins, J.F., Firth, A.E., 2008. An overlapping essential gene in the Potyviridae. Proc. National Acad. Sci. 105 (15), 5897–5902.

Cock, P.J., Chilton, J.M., Grüning, B., Johnson, J.E., Soranzo, N., 2015. NCBI BLAST+ integrated into Galaxy. GigaScience 4 (1), s13742, 015.

Cui, H., Wang, A., 2019. The biological impact of the hypervariable N-terminal region of potyviral genomes. Annual Review of Virology 6 (1), 255–274.

Deng, P., Wu, Z., Wang, A., 2015. The multifunctional protein CI of potyviruses plays interlinked and distinct roles in viral genome replication and intercellular movement. Virol. J. 12, 1–11.

Ding, J., Zhao, J., Pan, T., Xi, L., Zhang, J., Zou, Z., 2019. Comparative transcriptome analysis of gene expression patterns in tomato under dynamic light conditions. Genes 10 (9), 662.

Edgar, R.C., Taylor, B., Lin, V., Altman, T., Barbera, P., Meleshko, D., Babaian, A., 2022. Petabase-scale sequence alignment catalyses viral discovery. Nature 602 (7895), 142–147.

Inoue-Nagata, A.K., Jordan, R., Kreuze, J., Li, F., López-Moya, J.J., Mäkinen, K., et al., 2022. ICTV virus taxonomy profile: potyviridae 2022. J. Gen. Virol. 103 (5), 001738.

Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12 (4), 357–360.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874.

Lauber, C., Seitz, S., 2022. Opportunities and challenges of data-driven virus discovery. Biomolecules 12 (8), 1073.

Liu, N., Zhang, S., Huang, Y., Cai, H., Zhu, X., 2020. Understory and canopy additions of nitrogen differentially affect carbon and nitrogen metabolism of *Psychotria rubra* in an evergreen broad-leaved forest. Sci. Total Environ. 724, 138183.

Martínez-Turiño, S., García, J.A., 2020. Potyviral coat protein and genomic RNA: a striking partnership leading virion assembly and more. Adv. Virus Res. 108, 165–211.

Mohan, C., Josephrajkumar, A., Thube, S.H., Saneera, E.K., Rajkumar, 2022. Pests and their management in arecanut. Trends in Horticultural Entomology 1529–1544.

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

Nampoothiri, K.U.K., Ponnamma, K.N., Chowdappa, P., 2000. Arecanut Yellow Leaf Disease. CPCRI, Kasaragod, India.

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32 (2), 292–294.

Pandian, R.T.P., Thube, S.H., Rajashekara, H., Madhu, T.N., Diwakar, Y., Santhoshkumar, P., et al., 2024. *Colletotrichum kahawae* subsp. *cigarro* causing leaf spot disease on arecanut, Areca catechu L. in India: a first report. Crop Protect. 179, 106607.

Patil, B., Hegde, V., Sridhara, S., Pandian, R.T.P., Thube, S.H., Palliath, G.K., Jha, P.K., 2022. Multigene phylogeny and haplotype analysis reveals predominance of oomycetous fungus, *Phytophthora meadii* (McRae) associated with fruit rot disease of arecanut in India. Saudi J. Biologic. Sci. 29 (8), 103341.

Qin, L., Shen, W., Tang, Z., Hu, W., Shangguan, L., Wang, Y., et al., 2020. A newly identified virus in the family *Potyviridae* encodes two leader cysteine proteases in tandem that evolved contrasting RNA silencing suppression functions. J. Virol. 95 (1), 10–1128.

Revers, F., García, J.A., 2015. Molecular biology of potyviruses. Adv. Virus Res. 92, 101–199.

Roberts, J.M., Anderson, D.L., Durr, P.A., 2018. Metagenomic analysis of Varroa-free Australian honey bees (*Apis mellifera*) shows a diverse Picornavirales virome. J. Gen. Virol. 99 (6), 818–826.

Shang, H., Xun, L., Miao, T., Chen, C., Lu, Y., Li, B., 2024. Transcriptome analysis provides valuable insights into leaf size variation in *Rhamnus heterophylla*. Agronomy 14 (2), 396.

Sidharthan, V.K., Reddy, V., Kiran, G., Rajeswari, V., Baranwal, V.K., Kumar, M.K., Kumar, K.S., 2024. Probing of plant transcriptomes reveals the hidden genetic diversity of the family Secoviridae. Arch. Virol. 169 (7), 150.

Sidharthan, V.K., Sevanthi, A.M., Jaiswal, S., Baranwal, V.K., 2020. Robust virome profiling and whole genome reconstruction of viruses and viroids enabled by use of available mRNA and sRNA-Seq datasets in grapevine (*Vitis vinifera* L.). Front. Microbiol. 11, 1232.

Sidharthan, V.K., Baranwal, V.K., 2024. Public domain databases: a gold mine for identification and genome reconstruction of plant viruses and viroids. In: Genomics Data Analysis for Crop Improvement. Springer Nature Singapore, Singapore, pp. 247–284.

Sorel, M., García, J.A., German-Retana, S., 2014. The Potyviridae cylindrical inclusion helicase: a key multipartner and multifunctional protein. Mol. Plant Microb. Interact. 27 (3), 215–226.

Sun, H., Yu, W., Li, H., Hu, X., Wang, X., 2024. Bioactive components of areca nut: an overview of their positive impacts targeting different organs. Nutrients 16 (5), 695.

The Galaxy Community, 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res., gkae410 https://doi.org/10.1093/nar/gkae410.

Tong, T., Xu, A., Tan, S., Jiang, H., Liu, L., Deng, S., Wang, H., 2024. Biological effects and biomedical applications of areca nut and its extract. Pharmaceuticals 17 (2), 228.

Wang, Y., Shen, W., Dai, Z., Gou, B., Liu, H., Hu, W., et al., 2021. Biological and molecular characterization of two closely related arepaviruses and their antagonistic interaction in *Nicotiana benthamiana*. Front. Microbiol. 12, 755156.

Wei, H., He, X.J., Liao, C.H., Wu, X.B., Jiang, W.J., Zhang, B., et al., 2019. A maternal effect on queen production in honeybees. Curr. Biol. 29 (13), 2208–2213.

Yang, K., Ran, M., Li, Z., Hu, M., Zheng, L., Liu, W., et al., 2018. Analysis of the complete genomic sequence of a novel virus, areca palm necrotic spindle-spot virus, reveals the existence of a new genus in the family *Potyviridae*. Arch. Virol. 163, 3471–3475.

Yang, K., Shen, W., Li, Y., Li, Z., Miao, W., Wang, A., Cui, H., 2019. Areca palm necrotic ringspot virus, classified within a recently proposed genus Arepavirus of the family *Potyviridae*, is associated with necrotic ringspot disease in areca palm. Phytopathology 109 (5), 887–894.

Yin, H., Zhou, H., Wang, W., Tran, L.S.P., Zhang, B., 2022. Evidence for miRNAs involved in the high-altitude responses of sainfoin (*Onobrychis viciifolia*) grown in the Qinghai-Tibetan plateau. J. Plant Biochem. Biotechnol. 1–12.