

Virome of riverside phytocommunity ecosystem of an ancient canal

Shixing Yang

Jiangsu University

Tongling Shan

Chinese Academy of Agricultural Sciences

Yan Wang

Jiangsu University

Jie Yang

Jiangsu University

Xu Chen

Jiangsu University

Yuqing Xiao

Jiangsu University

Zhenqiang You

Hangzhou Medical College

Yumin He

Jiangsu University

Min Zhao

Jiangsu University

Juan Lu

Jiangsu University

Zijun Yang

Jiangsu University

Ziyuan Dai

Jiangsu University

Qi Liu

Jiangsu University

Yuxin Yao

Jiangsu University

Xiang Lu

Jiangsu University

Hong Li

Jiangsu University

Rui Zhou

Jiangsu University

Wang Li

Jiangsu Taizhou People's Hospital

Chenglin Zhou

Jiangsu Taizhou People's Hospital

Xiaochun Wang

Jiangsu University

Quan Shen

Jiangsu University

Hui Xu

The Affiliated Hospital of Jiangsu University

Xutao Deng

Vitalant Research Institute

Eric Delwart

University of California San Francisco

Wen Zhang (✉ z0216wen@yahoo.com)

Jiangsu University <https://orcid.org/0000-0002-9352-6153>

Research

Keywords: Plant virome; Phytocommunity; Virus host switching; Cross-species infection; co-infection.

Posted Date: May 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25620/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background The virus community in plants in a local plant ecosystem has remained largely unknown. In this study, we investigated the virus community in these wild and cultivated plants in Zhenjiang ancient canal ecosystem. Results using viral metagenomic approach, we investigated the viral community in leaf tissues of 161 plant species belonging in 38 different orders in a local riverside plant ecosystem. We discovered 251 different plant-associated virus genomes which included 88 DNA and 163 RNA viruses belonging to 27 different virus families, orders or unclassified virus groups. The identified viruses include some that are sufficiently divergent to comprise new genera, families, or even orders. Our data indicated that some groups of viruses known to infect non-plant organisms had host switching to infecting plants. Cross-species infection and co-infection of viruses were common in this plant ecosystem. Conclusions these data present a view of the viral community in plants present in a local plant ecosystem which is more diverse than that depicted in current classification of plant viruses and provide a solid foundation for studies in virus ecology and evolution in plants.

Background

Much effort has been devoted to studying viruses associated with economically important or symptomatic plants which only comprise a minute fraction of all plant species, suggesting that a large gap exists in our overall understanding of viral diversity, evolution, and ecology in uncultivated plants[1]. It is therefore necessary to study viruses existing in wild plants, whether symptomatic or asymptomatic, to gain a more objective view of virus populations in plant, which will undoubtedly discover novel or even so-call unclassified viruses and provide more information on viral evolution and diversity. High-throughput DNA sequencing coupled with viral metagenomics approaches also makes it possible to identify highly divergent viral genomes in wild plants. Comparing viruses in different plant species within a natural plant ecosystem can also improve our understanding of virus transmission amongst different plant species.

The Zhenjiang ancient canal was created during the Qin Dynasty over 20 centuries ago, and is 16 kilometers long with an average width of 40 meters. The Dingmao section of the canal is 2 km long with lots of wild plants, some landscape plants and crops on both sides. In this study, we investigated the virus community in these wild and cultivated plants in this ecosystem.

Results

Overall view of the virome. We performed a large-scale viral metagenomics survey of potential plant leaf-associated viruses in 161 plant species belonging in 38 different orders, 7 classes (Coniferopsida, Cycadopsida, Dicotyledoneae, Filicopsida, Ginkgopsida, Magnoliopsida, and Monocotyledoneae), and 4 phyla (Angiospermae, Gymnospermae, Pteridophyta, and Tracheophyta) existing in a riverside plant ecosystem (Fig. 1a, Supplementary Table 1, and Supplementary Data 1 and 2). Among the 161 species sampled, 89 belong to wild plant, and 72 are cultivated types. For each species three leaf tissue samples

from three different individual plants were collected. After crushing material with a mortar and pestle, supernatants of the 3 leaves from 3 different individual plants in the same plant species were mixed into a sample pool for viral metagenomics library construction. After virus nucleic acid particles enrichment using filtration (removing eukaryotic and bacterial cell-sized particles) and DNase and RNase treatment (digesting unprotected nucleic acid), total nucleic acid were then extracted and then organized into 161 libraries for Illumina HiSeq 2500 sequencing (Supplementary Table 1). In total, 50,586,188 paired-end reads were generated and binned by barcodes and quality-filtered, leaving high-quality sequence reads which were *de novo* assembled within each barcode. The resulting sequence contigs and unassembled reads were compared with the viral reference database and the GenBank non-redundant protein database using a BLASTx search with an E value cut-off of $<10^{-5}$. Among the 161 libraries, 147 libraries contain sequences showing significant similarities to known viruses with viral reads consisting of 0.04% to 97.93% of the total unique reads, where 52 libraries contained $>50\%$ of viral sequence reads (Fig. 1a, Supplementary Table 1). From these plants, 34 different groups of viruses were detected, including viruses belonging in 26 families, 1 genus (Botybirnavirus) unclassified in family, and 7 unclassified groups including circular replication-associated protein encoding single-stranded DNA virus (CRESS DNA virus), Parvo-like virus, Hepe-like virus, Noda-like virus, Permutotetra-like virus, Rhabdo-like virus, Sobemo-like virus, unclassified members of Picornavirales order, and unclassified members of the Riboviria domain (Fig. 1b, Supplementary Table 1). Comparison of the percentage of viral reads against the total unique reads and the number virus types in each library showed no significant difference between wild and cultivated plant samples (Supplementary Data 3 and 4, Supplementary Table 1), suggesting that in the local plant ecosystem different cultivation modes of plants had no discernable effect on susceptibility to virus infection. From these viral sequences, 251 virus strains generated complete genome ($n=202$) or nearly complete genome sequences ($n=49$), including 5 RNA virus strains belonging to segmented viruses (Fig. 1b). BLASTx search using nucleotide sequences of the 251 virus strains revealed that 61 of them shared $<40\%$ amino acid sequence identities with their best matches in GenBank (Fig. 1c, Supplementary Table 2), suggesting many of the virus strains discovered here are highly divergent from previously known viruses, and could be considered as new virus families or orders. To phylogenetically analyze these viruses derived from plants, amino acid sequence of the most conserved regions, including RNA-dependent RNA polymerase (RdRp) domains for virus belonging in *Riboviria*, replication proteins (Rep) for CRESS DNA virus, and non-structural protein (NS) for parvovirus, were used in phylogenetic analysis.

Expanding plants as new host of some viruses. Since our sample processing only involves the leaf of plants, and the samples were carefully washed with double distilled water (ddH₂O) three times before sample treatment, we assume most of the viruses characterized here were from leaf tissues instead of from other organisms on the plant leaf surface. All of the collected samples were leaves from healthy appearing plants. We detected members of 34 different groups of viruses and fully characterized the genomes from 27 of them, including 12 groups with viruses not previously reported from plants (Fig 2, see Supplementary Data 5-19 for detailed phylogenies).

The *Dicistroviridae* family, within the order *Picornavirales*, is a group of viruses currently composed of 3 genera, whose natural hosts are invertebrates, including aphids, leafhoppers, flies, bees, ants, and silkworms[2]. Here, we assembled 23 genomes from 9 different species of plants, where 7 virus strains were grouped into three previously classified genera while the other 16 strains were clustered into a separate group genetically far from the three known genera (Fig 2, Supplementary Table 1, see Supplementary Data 5). These dicistroviruses in the separate group showed typical genome organizations of dicistroviruses except that 10 of the 14 strains showed no cricket paralysis virus (CRPV) capsid superfamily domain in the capsid protein (Supplementary Data 6). Based on RdRp protein sequences of the 16 strains in the separate cluster, they shared <50% similarities to their best BLASTp matches in GenBank which were all phylogenetically located outside of the new clade, suggesting they might belong to a new genus in family *Dicistroviridae*. In arthropods, infection acquisition and transmission of dicistrovirus is prominently accomplished by ingestion and spread from the alimentary canal. In alimentary canal the virus generally replicates in epithelial cells of the gut and is subsequently shed into the gut lumen, being accumulated in the feces which is often an important infectious source[2,3]. Based on the transmission pattern of dicistrovirus in arthropod, the infection of dicistroviruses in plants may occur when virus-contaminated feces are shed onto the plant leaf surface by insects.

In three different species of plants, three divergent iflavirus strains were discovered and their complete genomes generated, all of which clustered within genus iflavirus based on phylogenetic analysis (Fig 2, Supplementary Data 7). The family *Iflaviridae* is a member in the order *Picornavirales*, which have also all been isolated from arthropods. Although vertical and sexual transmission has been reported among invertebrates for some iflaviruses, the most common route of infection for iflaviruses is through ingestion of virus-contaminated food sources[4,5]. Spread of iflaviruses in plants may therefore also occur through contaminated feces of arthropods.

We also identified 12 marnavirus strains from 5 different species of plants that shared 30%-60% sequence identities based on pairwise comparison of polyprotein sequence and showed typical genome organization of *Marnaviridae* (Supplementary Data 8). Based on BLASTx searches, two of the 12 plant marnavirus strains were also related to viruses from non-marine samples. Phylogenetic analysis including reference marnaviruses and the BLASTx matching viruses from non-marine samples revealed that the 12 plant marnaviruses grouped well into the cluster of genus marnavirus within *Marnaviridae*, which indicated the marnaviruses group includes closely related viruses from plants and two strains from fish and mollusk (Fig. 2, Supplementary Data 9), respectively. *Marnaviridae* is a newly defined virus family in order *Picornavirales*, the currently characterized representative member being *Heterosigma akashiwo* RNA virus, isolated from *Heterosigma akashiwo* algae in ocean water[6]. Closely related viruses have been identified in ocean marine environments [7]. Our data suggest that plants are capable of hosting some members in family *Marnaviridae* or their cellular hosts.

In 3 different plant species, we acquired 6 virus stains with complete genomes showing significant sequence similarity to parvovirus-like hybrid virus (PHV) and 2 viruses showing close relationship to

densovirus (Fig. 2, Supplementary Data 10). These plant PHV genomes were linear with length of 3.6-4.0-kb containing two major forward-direction ORFs encoding the replication and capsid proteins (Supplementary Data 11), which is characteristic of viruses in family *Parvoviridae*. The 6 PHVs detected in plants were grouped in two different clusters, sharing sequence similarities of 50%-67% to other PHVs based replication protein sequence. PHV is a type of highly divergent DNA virus which was recently discovered and phylogenetically located at the interface between the *Parvoviridae* and *Circoviridae*[8,9]. Although this virus was first detected in Chinese patients with seronegative (non-A-E) hepatitis and subsequently discovered in a wide range of clinical samples, sharing ~99% nucleotide and amino acid identity with each other[8], it was eventually traced to contaminated silica-binding spin columns used for nucleic acid extraction[9]. The silica matrix is generally generated by diatoms (algae), belonging to microscopic water plants, detecting PHV in silica-binding spin columns might be the initial evidence that plants can serve as the hosts of PHV. Our data further confirm that plants (or diatoms within them) are capable of hosting PHVs.

Besides the above four groups of viruses with multiple divergent stains found here in plant tissues, another 4 groups of viruses, not previously reported in plants, including noda-like virus, Permutotetra-like virus, Yanvirus-like virus, and Chuvirus-like virus, were also detected here (Fig. 2, Supplementary Data 12-15). These viral groups were recently reported from invertebrates meta-transcriptomes, and vertebrates and environment samples[10–12]. Discovering these viruses in plant leaf samples suggests that plants may also be the natural hosts for some members of these recently described clades. Bastrovirus was previously only detected in feces of mammals (including human) and mosquito, shows a distant relationship to astroviruses[13,14]. Here, a species of plant (*Solanum melongena*) was positive for virus genome sequence showing 25% RdRp sequence similarity to that of bastrovirus (Fig. 2, Supplementary Data 16). Detecting this divergent bastrovirus-like virus in plants may imply bastrovirus originates from plant and /or that its diverse members can infect widely different hosts including vertebrates, invertebrates and plants. Another species of plant was positive for hepe-like virus, which have been reported in mammals, invertebrates, protists, and different environments [12,15–17]. This hepe-like virus strain from plant was well grouped with other hepe-like viruses from different type of organism and environment samples and shared similar genome organization (Fig. 2, Supplementary Data 17), suggesting this type virus may also parasitize plants. Two types of viruses, botybirnavirus and narna-like virus, which were considered to be viruses of fungi[18,19] and more recently *Caenorhabditis* nematodes[20], were detected in two species of plants, respectively (Fig. 2, Supplementary Data 18 and 19). The botybirnavirus showed high sequence identity (96.4%) to fungi batybirnavirus based on RdRp protein sequence. The two narna-like virus strains from 2 different species of plants shared 99.9% nucleotide sequence identity and identical based on RdRp protein sequence, and were divergent from previous narna-like viruses.

Divergent viruses in plants. For these 12 groups of viruses, first reported in plants here, some genomes were so divergent from their closest identifiable relatives using BLASTx they may ultimately qualify as members of new genera or even new families (Fig. 2). For example, for the 23 dicistrovirus genomes, 7 of which grouped well into previously defined genera, the other 16 strains seem to form a separate clade

which could be designated a new genus in the *Dicistroviridae* family. The same conclusion could also apply to some genomes in the groups of noda-like, hepe-like, and bastrovirus-like viruses and in the *Marnaviridae* family.

Another 23 divergent RNA viral genomes whose closest relatives are in the *Picornavirales* order were characterized. Phylogenetic analysis based on RdRp sequences of the 6 defined families and the best matches of the 23 strains in GenBank showed that they were grouped into 8 different clusters which were genetically distinct from the defined 6 families in the order *Picornavirales* (Fig 3, Supplementary Data 20).

Tombusviridae is a large family of plant viruses that is currently composed of more than 76 species divided among 3 subfamilies and 16 genera. Here, we acquired 21 genomes showing sequence similarity to members of the *Tombusviridae*. Seven genomes were genetically close to defined genera while the other 14 were highly divergent and seemed to form several distinct genera (Fig 3, Supplementary Data 21). Four different virus strains belonging to family *Luteoviridae* were also detected in plants here, 3 of which closely clustered with different defined genera, with the remaining forming a single deeply rooted separate branch, which may belong to a putative new genus clustering outside the genus luteovirus (Fig 3, Supplementary Data 22). Four partitivirus strains were characterized in three different species of plants, all of them were putative new species within three different genera of *Partitiviridae* (Fig 3, Supplementary Data 23). Seven virus genomes identified here also showed sequence similarity to sobemo-like viruses which were recently discovered from arthropods using meta-transcriptomics [15]. Although these plants sobemo-like viruses phylogenetically grouped together with invertebrate sobemo-like viruses they were genetically distinct and sharing 30%-62% amino acid sequence similarities to each other (Fig 3, Supplementary Data 24). Two plant rhabdo-like viruses also showed a close relationship to recently discovered invertebrate derived rhabdo-like viruses (Fig 3, Supplementary Data 25). Last one divergent RNA genome showed a distant relationship to three genomes belonging to an unclassified member of the *Riboviria* domain, all from wastewater or soil samples, consistent with a plant origin (Fig 3, Supplementary Data 26).

Plant CRESS virus. CRESS DNA virus is the informal name of several groups of single-stranded (ss) DNA viruses that have circular and replication-associated protein encoding genome, which show high diversity and abundance in various habitats[21,22]. Although there are currently several established CRESS DNA virus families including *Bacillidnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Microviridae*, *Nanoviridae* and *Smacoviridae*, a large number of novel CRESS DNA viruses have been discovered recently and have not been formally classified, for which the hosts are currently unknown [22–24]. Among these well-defined CRESS DNA virus families, *Geminiviridae* and *Nanoviridae* are two plant-infecting members, which also help the replication and package of a satellite virus: *Alphasatellitidae*, another type of circular ssDNA genome[25]. Here, from plant leaves we acquired 79 circular genomes, among which 7 were genetically close to *Geminiviridae*, 9 grouped well into the family *Genomoviridae*, 7 clustered closely to known sequences of *Alphasatellitidae*, 15 belong to new divergent members in family *Microviridae* presumably from bacteria, with the remaining 41 showing significant sequence similarity to unclassified CRESS DNA viruses (Fig. 4).

Among the 7 CRESS DNA viruses belonging in family *Geminiviridae*, 2 of them fell well into the cluster of the genus begomovirus, being closely to sweet potato leaf curl virus, a monopartite geminivirus. The other 5 were not grouped into any known genus in family *Geminiviridae* but deeply clustered outside of all known geminiviruses, suggesting these 5 novel geminiviruses might belong to new genus (genera) in *Geminiviridae* (Fig.4, Supplementary Data 27). Viruses in the family *Genomoviridae* have been frequently found to be associated with a variety of samples ranging from fungi to animal sera [26], indicating that genomoviruses are widespread as well as abundant in the environment. Here, 9 complete genomes of genomovirus, divergent from previous known members in that family, were characterized in 7 different plant species, which phylogenetically clustered into 5 different groups, including two identical genomes detected in two different plant species (Fig 4, Supplementary Data 28). Currently, the hosts of the large majority of CRESS-DNA viruses remain unknown except for one replicating in both fungi[27] and an insect[28]. Detecting genomoviruses in leaf samples from different species of plant may suggest plants or an internal plant-dwelling organism, may host some members in the family *Genomoviridae*.

We also discovered 7 divergent complete circular genomes in a single species of plant, which showed sequence identities of 38%-58% to previous known genomes of members in *Alphasatellitidae* based on amino acid sequence of encoded Rep protein. The 7 alphasatellites had genome sizes ranging from 1309 to 1503 nucleotides, which were divergent from each other and grouped into 4 different clusters composed of previous defined alphasatellites based on phylogenetic analysis of their Rep protein (Fig 4, Supplementary Data 29). Alphasatellites are circular ssDNA components which are generally associated with *Nanoviridae* or some members in *Geminiviridae*, however, we did not detect geminivirus or nanovirus sequence in this species of plant, but discovered a divergent CRESS DNA virus genome that showed the highest Rep protein sequence similarity of 60.7% to an unclassified CRESS DNA virus, temperate fruit decay-associated virus[29], suggesting this type of CRESS DNA virus may infects plant and serves as helper virus for alphasatellites.

Apart from the 3 groups of viruses within classified CRESS DNA virus families, other 41 unclassified CRESS DNA viruses were also discovered from different species of plant. These CRESS DNA viruses were so divergent from each other, we phylogenetically analyzed them in 6 different phylogenetic trees (Fig 4, Supplementary Data 30-35), where each of them includes strains identified here, their best matches in GenBank, and the representative members in known CRESS DNA virus families and other unclassified CRESS DNA viruses, using fewer sequences in each sequence alignment so as to include as large as possible number of conserved amino acid sites in the phylogenetic analysis. Based on Rep proteins sequences, these unclassified CRESS DNA viruses shared sequence similarities 26%-61% to their best matches, where 6 of them grouped with CRESS DNA viruses from feces of mammals, 3 of them with CRESS DNA viruses from invertebrates, 13 sequences with CRESS DNA viruses identified from environmental samples (mainly wastewater), 8 strains with CRESS DNA virus from fish species, one with plant-associated CRESS DNA virus, while the remaining 10 sequences were too divergent to cluster with any known viruses and were included in CRESSV group 6 in Fig.4 (Fig 4, Supplementary Data 30-35). Considering that most of the CRESS DNA viruses characterized in the present study best matched unclassified CRESS DNA genomes from environmental samples, mammalian feces, and arthropods, it is

possible that most of these unclassified CRESS DNA viruses infect plants and were contaminants in feces or the gut content of arthropods.

Fifteen genomes showing sequence similarity to viruses in family *Microviridae* were detected in three different species of plants, 12 of which were from a single species of wild plant, *Kummerowia striat* (Supplementary Data 36). Many studies have demonstrated the ubiquity of *Microviridae* genomes across habitats (marine, freshwater, wastewater, sediment) and global regions (Antarctic to subtropical), especially those related to the *Gokushovirinae* lineage [30–33], which infect obligate intracellular parasites, members of the bacterial genera *Chlamydia*, *Bdellovibrio* and *Spiroplasma* [34].

Cross-species infection and co-infection of plant viruses. Other than through seed dispersal most plants are immobile; hence plant virus transmission is often assisted by others organisms [35,36]. Here, we investigated the virome in plant leaves collected in a single ecosystem, which includes interactions amongst plants, water, soil, air, insects and a multitude of micro-organisms providing favorable conditions for cross-species transmission. Using viral metagenomics, we detected the viral nucleic acids and determined 251 (nearly) complete viral genomes, allowing us to compare genome sequences from different species of plants and estimate whether cross-species transmission might occur for some viruses. Our results indicated cross-species transmission might have occurred for 9 groups of viruses. 24 genomes belonging in family *Potyviridae* were found in 17 different species of plant, all in the genus potyvirus (Supplementary Table 1, Supplementary data 37). Among the 9 groups of potyviruses, 2 groups were composed of 10 and 5 genomes, respectively, sharing 99%-100% sequence RdRp protein identities within each group, suggesting possible cross-species transmission (Supplementary data 37). We then compared the 10 and 5 genome sequences respectively in these 2 groups and found that the 10 genomes shared 94.6%-100% and the 5 genomes shared 94.8%-100% sequence identities (including several pairs of identical sequences) (Fig. 5), suggesting some strains of these potyviruses may be capable of cross-species transmission. Our data also showed that some dicistroviruses might be plant-infecting virus. Here we acquired 22 complete genomes of dicistrovirus from 10 different species of plants, of which 6 pairs presented possible cross-species transmission in plants as pair of genome sequences shared >94.9% identity, including one pair of identical sequences (Fig 5, Supplementary data 5). Putative cross-species transmissions were also observed with unclassified CRESS virus including 5 pairs of identical genomes derived from different species of plants (Supplementary Data 38). Two groups of marnaviruses showing >99% genomic sequence identity, and other 5 pairs of different viruses including geminivirus, genomovirus, luteovirus, parvo-like virus, and sobemo-like virus, from different putative host species showed 92.5%-99.8% sequence identities based on complete genome sequence (Supplementary Data 38).

We marked the accurate sampling sites for each plant species which makes it possible to measure the geographical distance of different species of plants involved in the cross-species transmission of a certain virus so as to infer whether geographical distance of the host plants have effect on the cross-species transmission. Our data indicated that cross-species transmission of potyviruses might be associated with their geographical distance as the genetically very close genomes were mainly from the

same sampling site (Fig. 5). The same phenomena were also observed for the marnavirus, unclassified CRESS DNA virus genomes, luteovirus, and parvo-like virus. For example, all the 5 marnavirus genomes involved in putative cross-species transmission were from a single sampling site and 9 of the 11 CRESS DNA genomes were from the same sampling location (Fig. 5, Supplementary Data 38). However, the remaining several groups of viruses with properties of cross-species transmission (closely related genomes from different plants) including dicistrovirus, geminivirus, genomovirus, and sobemo-like virus seem to have no relationship to the geographical distance of their hosts' location (Supplementary Data 38). The different effect of geographical distance on the cross-species transmission may reflect the different transmission potential of these viruses, for example, geographical distance had no effect on the cross-species transmission of dicistroviruses suggested that the spread of this virus might be assisted by arthropods. Our data also indicated that most of the putative viral cross-species infection in this ecosystem occurred across different levels of plant classification. For instances, the 10 closely related potyvirus genomes were characterized from plants belonging to 7 different orders within 2 different classes (Fig. 5), suggesting a wide host range.

Co-infection of hosts by two or more plant viruses is common in both agricultural crops [37,38] and natural plant communities [39]. In the present study, apart from cross-species infection, co-infection of plant viruses was also commonly observed, where 73 out of 161 (45.3%) libraries contained >3 different virus types (or families) (Supplementary Table 1), suggesting co-infection of viruses existed in nearly half of the plants in this ecosystem as each library consisted of samples from three different individual plant. Considering the same virus families or type in a single library may contain different virus strain or type, the rate of co-infection is likely to be higher than 45.3%. Among the 251 genomes we acquired from these plants, some genomes were from the same libraries which allows us investigate the co-infection of certain viruses in specific species of plants. As shown in Fig. 6, PCR screening of different virus genomes in 7 different species of plants revealed that most of (20/21) the individual plant contained >2 different types of virus, where one plant species of *Forsythia suspensab* even carried 16 viruses belonging in 12 different families. The wide presence of apparent viral co-infections in these plants in a single ecosystem may lead to interactions between viruses that could influence disease development in individual plants.

Other plant viruses. Apart from these viruses mentioned above, many types of typical plant viruses belonging in the *Bromoviridae*, *Closteroviridae*, *Comoviridae*, and *Botourmiaviridae* families and *Tymovirales* order were also detected in several species of plants. These plant viruses were genetically close to previously described viruses (Supplementary data 39-43), indicating typical plant virus infections were readily detected in this plant ecosystem.

Discussion

The common perception that plant viruses are primarily pathogens results from the focus given to agricultural plant health. Emerging diseases have garnered most attention because of damage to economically important food and ornamental plant species. Important examples of viruses that are responsible for well-studied emerging diseases include cassava-infecting begomoviruses (in the

Geminiviridae family)[40], closteroviruses causing grapevine leafroll disease[41], luteoviruses such as barley yellow dwarf virus[42] and sobemoviruses such as rice yellow mottle virus[43]. Relatives of all these pathogenic viruses were detected in this study in apparently healthy plants from diverse families or orders. The relatively unbiased sequencing of viral genomes within entire environments as performed here is changing the perspective of viruses from agents of disease to common components of ecosystems, as the plant tissue samples studied were all from apparently healthy plants.

The data in the present study also revealed that several viruses such as dicistrovirus, iflavirus, marnavirus, noda-like, and parvo-like viruses, which have not been reported in plants were detected here in leaf tissues. Among these viruses, dicistrovirus, iflavirus, and noda-like virus are generally hosted by arthropods[44,45]. Detection of these genomes in plants indicated that insects may have vectored them between plants. The closest non-plant-infecting relatives of some genomes from plants reported here tended to infect arthropods or fungi. Currently plant-infecting viruses may therefore have evolved from viruses that once infected non-plant organisms (or vice versa). Further, the hypothesis that many plant and vertebrate viruses may have originated from arthropod viruses is also plausible as some viruses infect arthropods can also infect plants. For example, flock house virus (in the *Nodaviridae* family) infects arthropods but can also systemically infect plants when it is complemented with the movement proteins of either tobacco mosaic virus or red clover necrotic mosaic virus (both of which are plant viruses)[46].

The cross-species transmission of viruses from one host species to another is responsible for the majority of emerging infections, both in animal and plant populations[47–49]. Decades of inventorying, tracking and analyzing of plant viruses showed that the emergence of new diseases is driven by adaptive viral evolution in response to novel ecological conditions[50,51], including the introduction of viruses and vectors to new areas, the intensification of agriculture and urbanization, and ecological changes in response to changing climatic conditions. Our data showed that a number of genomes from viral families not known to infect plants are indeed present in plants. Furthermore several genetically close or identical viruses were detected in plants from in different species, suggesting cross-plant species transmission. In addition, it is noteworthy that a small number of viruses showing genetic relationship to viruses previously found in mammalian feces (e.g. bastrovirus and hepe-like virus) were also detected in plant tissues, possibly indicating their plant tropism.

Conclusions

Our study presents an overview of the virus community existing in leaf tissues from plants in a local plant ecosystem, which is more diverse than that depicted in current classification of plant viruses. Virus types and viral reads in wild and cultivated plants were compared, showing no difference between two groups. Cross-species infection and co-infection of viruses were common in this plant ecosystem. 251 different plant-associated virus genomes were fully characterized and phylogenetically classified into 27 different virus families, orders or unclassified virus groups. This study provides a solid foundation for studies in virus ecology and evolution in plants and will be helpful for identification of newly emerging, possibly pathogenic, viruses of plants.

Materials And Methods

Sample collection and preparation

The goal of this study was to investigate the virome of plant species in an ancient canal ecosystem in Zhenjiang City, Jiangsu Province, China. The Zhenjiang Ancient Canal has a history of more than 2000 years. It runs through the whole town of Zhenjiang from southeast to northwest and is 16 kilometers long with an average width of 40 meters. By investigating the riparian vegetation of the ancient canal, the Dingmao section of the canal was chosen to study as a representative section. It is 2 km long with lots of wild plants, some landscape plants and crops on both riversides. In total of 161 plant species belonging to 38 different order, 6 classes (Coniferopsida, Cycadopsida, Dicotyledoneae, Filicopsida, Ginkgopsida, and Monocotyledoneae), and 3 phyla (Angiospermae, Gymnospermae, and Pteridophyta) were collected in this area for this study. The sampling sites for each plant species are labeled on the map with numbers corresponding to plant library numbers (Supplementary Table 1, Supplementary Data 1 and 2). Among those plant species, 72 are wild plants and 89 are cultivated plants including landscape plants and crops. During sampling, 3 leaves from three different individual plants belonging to the same species were respectively collected into disposable materials, before this step, distilled water (ddH₂O) was used to clean the dust and other non-plant organisms on the leaf surface. Before viral metagenomic analysis, about 0.1g leaf tissue sample of each plant was grounded using steel balls and re-suspended in 1mL of phosphate-buffered saline (PBS) and vigorously vortexed for 5 min. The grounded samples were then frozen and thawed three times on dry ice. The supernatants were then collected after centrifugation (10 min, 15,000×g) and stored at -80°C until use. Host species identification was initially identified using APP “PictureThis” which is online plant encyclopedia and plant identifier, and future confirmed by experienced field biologists.

Viral metagenomic analysis

About 300 µL supernatant from each of the three different plant samples in the same species was mixed into one sample pool and filtered through a 0.45-µm filter and centrifuged at 120,000g for 20 minutes at 4°C to remove eukaryotic and bacterial cell-sized particles. Un-encapsidated nucleic acids were then digested by DNase and RNase at 37 °C for 60 min[52–55]. Total nucleic acids were extracted as a mixed RNA/DNA solution using QiaAmp Mini Viral RNA kit (Qiagen) according to the manufacturer’s protocol. 161 libraries were constructed using Nextera XT DNA Sample Preparation Kit (Illumina). For bioinformatics analysis, paired-end reads of 250 bp generated by MiSeq were debarcoded using vendor software from Illumina. An in-house analysis pipeline running on a 32-node Linux cluster was used to process the data. Reads were considered duplicates if bases 5 to 55 were identical and only one random copy of duplicates was kept. Clonal reads were removed and low sequencing quality tails were trimmed using Phred quality score ten as the threshold. The unique read number of each library was shown in Table 1. Adaptors were trimmed using the default parameters of VecScreen which is NCBI BLASTn with specialized parameters designed for adapter removal. The cleaned reads were de novo assembled within each barcode using the ENSEMBLE assembler[56]. Contigs and singlets reads are then matched against

a customized viral proteome database using BLASTx with an E value cutoff of $<10^{-5}$, where the virus BLASTx database was compiled using NCBI virus reference proteome (<ftp://ftp.ncbi.nih.gov/refseq/release/viral/>) to which was added viral protein sequences from NCBI nr fasta file (based on annotation taxonomy in Virus Kingdom). Candidate viral hits are then compared to an in-house non-virus non-redundant (NVNR) protein database to remove false-positive viral hits, where the NVNR database was compiled using non-viral protein sequences extracted from NCBI nr fasta file (based on annotation taxonomy excluding Virus Kingdom). Contigs without significant BLASTx similarity to viral proteome database are searched against viral protein families in vFam[57] database using HMMER3 to detect remote viral protein similarities[58–60]. A web-based graphical user interface was developed to present users with the virus hits, along with taxonomy information and processing meta-information. The genome coverage of the target viruses were analyzed by Geneious v11.1.2[61].

Confirmation and extension of virus genomes

Viral contigs which might be from the same genome but without assembled overlaps were merged using the software Geneious v11.1.2 and primers bridge contigs were then designed[61]. Gaps were filled by (RT-)PCR and Sanger sequencing. To confirm the assembly results of a full genome, reads were *de novo* assemble back to the full length genome using the low sensitivity/fastest parameter in Geneious 11.1.2. For genomes with novel structures, we verified the complete or near complete viral genome by designing overlapping primers based on the assembled sequences. For those viruses that firstly isolated from plants, we used PCR and Sanger sequencing to verify it's accurate based on the assembled sequences.

Confirmation of viral co-infection

In 7 libraries including pt065, pt067, pt110, pt111, pt112, pt119 and pt151, which have far more than three different virus strains, showed evident co-infection in individual plant. To investigate the presence status of different viral strain in three individual plants from the same library, PCR and Sanger sequencing were performed using specific primers designed based on the conserved domain sequences of these viruses.

Phylogenetic analysis of viruses

Through analyzed the protein sequences obtained in this study, we divide them into three categories including RNA viruses, Parvovirus-like viruses and CRESS DNA viruses. To infer the phylogenetic relationships, protein sequences of reference strains belonging to RNA viruses, Parvovirus-like viruses, and CRESS DNA viruses were downloaded from the NCBI GenBank database. For RNA viruses, the phylogenetic tree was constructed based on the RNA-dependent RNA polymerase (RdRp), for parvovirus-like viruses, the phylogenetic tree was constructed based on nonstructural protein (NS), for the CRESS DNA viruses, the phylogenetic tree was constructed based on the replication-associated protein (Rep) except for *Microviridae* viruses whose major capsid protein was used for the phylogenetic tree construction. The related protein sequences were firstly aligned using alignment program implemented in the CLC Genomics Workbench 10.0, the alignment result was further optimized using MUSCLE in MEGA v7.0[62] and MAFFT v7.3.1 employing the E-INS-I algorithm[63]. Sites containing more than 50% gaps

were temporarily removed from alignments. Bayesian inference trees were then constructed using MrBayes v3.2[64]. The Markov chain was run for a maximum of 1 million generations, in which every 50 generations were sampled and the first 25% of Markov chain Monte Carlo (mcmc) samples were discarded as burn-in. The approximate family/genus of viruses that obtained in this study was determined through the above tree, further constructed the detailed trees point at each virus family that are relatively closely related to the viruses discovered here using the same method. Maximum Likelihood trees were also constructed to confirmed all the Bayesian inference trees using software Mega v7.0[62] or PhyML v3.0[65].

Virus genome annotation

Putative viral open reading frames (ORFs) were predicted by Geneious v11.1.2 with built-in parameters (Minimum size: 300; Genetic code: Standard; Start codons: ATG)[61], further were checked through comparing to related viruses by Blastp in NCBI. The annotations of these ORFs were based on comparisons to the Conserved Domain Database. Potential exon and intron of Genomovirus were predicted by Netgenes2 at <http://www.cbs.dtu.dk/services/NetGene2/>.

Quality control in the nucleic acid manipulation

Standard precautions were used for all procedures to prevent the cross-sample contamination and nucleic acid degradation. Mainly, aerosol filter pipet tips were used to reduce the possibility of sample cross contamination, and all the materials (including microcentrifuge tubes, pipet tips, etc.) which directly contacted with nucleic acid samples were RNase and DNase free. The nucleic acid samples were dissolved in DEPC treated water. In order to exclude the possibility of contamination with nucleic acids of parvovirus-like hybrid virus (PHV) and micriovirus present in the laboratory or from Qiagen nucleic acid extraction kits, the samples positive for the two types of virus were chosen and the nucleic acid were re-extracted using Trizol reagent (Invitrogen). PCR using primers specific to those viruses confirmed their presence in the original biological samples. As a control, a library was also constructed using ddH₂O as sample which generated 13,228 raw reads and contained no viral reads based on BLASTx searching.

Abbreviations

CRESS DNA virus: circular replication-associated protein encoding single-stranded DNA virus; RdRp: RNA-dependent RNA polymerase; Rep: replication proteins; NS: non-structural protein; CRPV: cricket paralysis virus; PHV: parvovirus-like hybrid virus

Declarations

Acknowledgements

We thank Yiming Liu at Tropical Crops Genetic Resources Institute in Chinese Academy of Tropical Agricultural Sciences for identification of plant species. This work was supported by National Key

Research and Development Programs of China No. 2017YFC1200201, Jiangsu Provincial Key Research and Development Projects No. BE2017693 and Independent Project of Chengdu Research Base of Giant Panda Breeding No. 2020CPB-C11.

Authors' contributions

WZ conceived, designed and supervised the study. SY, YH, MZ, JL, ZY, ZD, and XL collected plant samples. SY, YW, JY, and XC performed experiments. SY, XD, and TS analysed data. YW, JY, XC, YX, HL, RZ, QL, and WL took part in data sorting and analysis. WZ and SY wrote the manuscript. WZ, ED, ZY, CZ, XW, QS, and HX revised and edited the manuscript. All authors read the manuscript and agreed to its contents.

Availability of data and materials

All complete or partial viral genome obtained in this study were deposited in GenBank with the accession numbers MN722411-MN722420, MN723593-MN723599, MN729612-MN729623, MN728806-MN728814, MN724250-MN724258, MN814305-MN814321, MN831436-MN831448, MN823661-MN823692, MN841281-MN841303, MN832441-MN832474, MN862333-MN862357, MN891787-MN891825, MT067617-MT067623 and MT134328 (See detailed information in Supplementary Table 2). The raw sequence reads generated here were deposited into the Sequence Read Archive of GenBank database and the accession nos. are shown Supplementary Table 2.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Roossinck MJ. Symbiosis versus competition in plant virus evolution. *Nat. Rev. Microbiol.* 2005;3:917–24.
2. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, Herrero S, et al. ICTV Virus Taxonomy Profile: Dicistroviridae. *J. Gen. Virol.* 2017;98:355–6.
3. Bonning BC, Miller WA. Dicistroviruses. *Annu. Rev. Entomol.* 2010;55:129–50.
4. Geng P, Li W, de Miranda JR, Qian Z, An L, Terenius O. Studies on the transmission and tissue distribution of *Antheraea pernyi* iflavivirus in the Chinese oak silkworm *Antheraea pernyi*. *Virology.*

- 2017;502:171–5.
5. Ryabov E V., Fannon JM, Moore JD, Wood GR, Evans DJ. The Iflaviruses Sacbrood virus and Deformed wing virus evoke different transcriptional responses in the honeybee which may facilitate their horizontal or vertical transmission. *PeerJ*. 2016;4:e1591.
 6. Tai V, Lawrence JE, Lang AS, Chan AM, Culley AI, Suttle CA. CHARACTERIZATION OF HaRNAV, A SINGLE-STRANDED RNA VIRUS CAUSING LYSIS OF HETEROSIGMA AKASHIWO (RAPHIDOPHYCEAE)1. *J. Phycol.* John Wiley & Sons, Ltd; 2003;39:343–52.
 7. Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. *Nature*. 2003;424:1054–7.
 8. Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, et al. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. *Proc. Natl. Acad. Sci.* 2013;110:10264–9.
 9. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. *J. Virol.* 2013;87:11966–77.
 10. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539–43.
 11. Zhang Y-Z, Shi M, Holmes EC. Using Metagenomics to Characterize an Expanding Virosphere. *Cell*. 2018;172:1168–72.
 12. Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, et al. The evolutionary history of vertebrate RNA viruses. *Nature*. 2018;556:197–202.
 13. Oude Munnink BB, Cotten M, Canuti M, Deijns M, Jebbink MF, van Hemert FJ, et al. A Novel Astrovirus-Like RNA Virus Detected in Human Stool. *Virus Evol.* 2016;2:vew005.
 14. Bauermann F V., Hause B, Buysse AR, Joshi LR, Diel DG. Identification and genetic characterization of a porcine hepe-astrovirus (bastrovirus) in the United States. *Arch. Virol.* 2019;164:2321–6.
 15. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539–43.
 16. Wu N, Zhang P, Liu W, Wang X. Sogatella furcifera hepe-like virus: First member of a novel Hepeviridae clade identified in an insect. *Virus Res.* 2018;250:81–6.
 17. Purdy MA, Harrison TJ, Jameel S, Meng X-J, Okamoto H, Van der Poel WHM, et al. ICTV Virus Taxonomy Profile: Hepeviridae. *J. Gen. Virol.* 2017;98:2645–6.
 18. Zhai L, Yang M, Zhang M, Hong N, Wang G. Characterization of a Botybirnavirus Conferring Hypovirulence in the Phytopathogenic Fungus *Botryosphaeria dothidea*. *Viruses*. 2019;11:266.
 19. Sahin E, Akata I. Viruses infecting macrofungi. *VirusDisease*. 2018;29:1–18.
 20. Richaud A, Frézal L, Tahan S, Jiang H, Blatter JA, Zhao G, et al. Vertical transmission in *Caenorhabditis nematodes* of RNA molecules encoding a viral RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 2019;116:24738–47.

21. Zhao L, Rosario K, Breitbart M, Duffy S. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses: Ubiquitous Viruses With Small Genomes and a Diverse Host Range. *Adv. Virus Res.* 2019;103:71–133.
22. Kazlauskas D, Varsani A, Koonin E V, Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun.* 2019;10:3425.
23. Krupovic M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr. Opin. Virol.* 2013;3:578–86.
24. Tisza MJ, Pastrana D V, Welch NL, Stewart B, Peretti A, Starrett GJ, et al. Discovery of several thousand highly diverse circular DNA viruses. *Elife.* eLife Sciences Publications, Ltd; 2020;9.
25. Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, Moriones E, et al. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Arch. Virol.* 2018;163:2587–600.
26. Krupovic M, Ghabrial SA, Jiang D, Varsani A. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch. Virol.* 2016;161:2633–43.
27. Yu X, Li B, Fu Y, Jiang D, Ghabrial SA, Li G, et al. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107:8387–92.
28. Liu S, Xie J, Cheng J, Li B, Chen T, Fu Y, et al. Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proc. Natl. Acad. Sci. U. S. A.* 2016;113:12803–8.
29. Basso MF, da Silva JCF, Fajardo TVM, Fontes EPB, Zerbini FM. A novel, highly divergent ssDNA virus identified in Brazil infecting apple, pear and grapevine. *Virus Res.* 2015;210:27–33.
30. Roux S, Krupovic M, Poulet A, Debroas D, Enault F. Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. *Dutilh BE, editor. PLoS One.* 2012;7:e40418.
31. Quaiser A, Dufresne A, Ballaud F, Roux S, Zivanovic Y, Colombet J, et al. Diversity and comparative genomics of Microviridae in Sphagnum- dominated peatlands. *Front. Microbiol.* 2015;6:375.
32. Tucker KP, Parsons R, Symonds EM, Breitbart M. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J.* 2011;5:822–30.
33. Hopkins M, Kailasan S, Cohen A, Roux S, Tucker KP, Shevenell A, et al. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. *ISME J.* 2014;8:2093–103.
34. Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, et al. Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J. Bacteriol.* 2002;184:1089–94.
35. Roossinck MJ. Plants, viruses and the environment: Ecology and mutualism. *Virology.* 2015;479–480:271–7.
36. Chen Y, Evans J, Feldlaufer M. Horizontal and vertical transmission of viruses in the honey bee, *Apis mellifera*. *J. Invertebr. Pathol.* Academic Press; 2006;92:152–9.

37. Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, et al. Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* 2018;12:173–84.
38. Muthukumar V, Melcher U, Pierce M, Wiley GB, Roe BA, Palmer MW, et al. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* 2009;141:169–73.
39. Nasir A, Caetano-Anollés G. A phylogenomic data-driven exploration of viral origins and evolution. *Sci. Adv.* 2015;1:e1500527.
40. McCallum EJ, Anjanappa RB, Gruissem W. Tackling agriculturally relevant diseases in the staple crop cassava (*Manihot esculenta*). *Curr. Opin. Plant Biol.* 2017;38:50–8.
41. Almeida RPP, Daane KM, Bell VA, Blaisdell GK, Cooper ML, Herrbach E, et al. Ecology and management of grapevine leafroll disease. *Front. Microbiol.* 2013;4:94.
42. Walls J, Rajotte E, Rosa C. The Past, Present, and Future of Barley Yellow Dwarf Management. *Agriculture. Multidisciplinary Digital Publishing Institute;* 2019;9:23.
43. Pinel-Galzi A, Traoré O, Séré Y, Hébrard E, Fargette D. The biogeography of viral emergence: rice yellow mottle virus as a case study. *Curr. Opin. Virol.* 2015;10:7–13.
44. Junglen S, Drosten C. Virus discovery and recent insights into virus diversity in arthropods. *Curr. Opin. Microbiol.* 2013;16:507–13.
45. Calisher CH, Higgs S. The Discovery of Arthropod-Specific Viruses in Hematophagous Arthropods: An Open Door to Understanding the Mechanisms of Arbovirus and Arthropod Evolution? *Annu. Rev. Entomol. Annual Reviews ;* 2018;63:87–103.
46. Dasgupta R, Garcia BH, Goodman RM. Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes. *Proc. Natl. Acad. Sci. U. S. A.* 2001;98:4910–5.
47. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579:265–9.
48. Ng LFP, Hiscox JA. Coronaviruses in animals and humans. *BMJ.* 2020;368:m634.
49. Ohshima K, Akaishi S, Kajiyama H, Koga R, Gibbs AJ. Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. *J. Gen. Virol.* 2010;91:788–801.
50. Fargette D, Konaté G, Fauquet C, Muller E, Peterschmitt M, Thresh JM. Molecular Ecology and Emergence of Tropical Plant Viruses. *Annu. Rev. Phytopathol.* 2006;44:235–60.
51. Jones RAC. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* 2009;141:113–30.
52. Zhang W, Li L, Deng X, Kapusinszky B, Pesavento PA, Delwart E. Faecal virome of cats in an animal shelter. *J. Gen. Virol.* 2014;95:2553–64.
53. Zhang W, Li L, Deng X, Blümel J, Nübling CM, Hunfeld A, et al. Viral nucleic acids in human plasma pools. *Transfusion.* 2016;56:2248–55.

54. Siqueira JD, Dominguez-Bello MG, Contreras M, Lander O, Caballero-Arias H, Xutao D, et al. Complex virome in feces from Amerindian children in isolated Amazonian villages. *Nat. Commun. Nature Publishing Group*; 2018;9:4270.
55. Zhang W, Yang S, Shan T, Hou R, Liu Z, Li W, et al. Virome comparisons in wild-diseased and healthy captive giant pandas. *Microbiome*. 2017;5:90.
56. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, et al. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res*. 2015;
57. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. Tse H, editor. *PLoS One*. 2014;9:e105067.
58. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*. 2010;11:431.
59. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 2009;23:205–11.
60. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29-37.
61. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. Oxford University Press; 2012;28:1647–9.
62. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. 2016;33:1870–4.
63. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res*. 2013;41:W22-8.
64. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol*. Oxford University Press; 2012;61:539–42.
65. Guindon S, Gascuel O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Rannala B, editor. *Syst. Biol*. Oxford Academic; 2003;52:696–704.

Figures

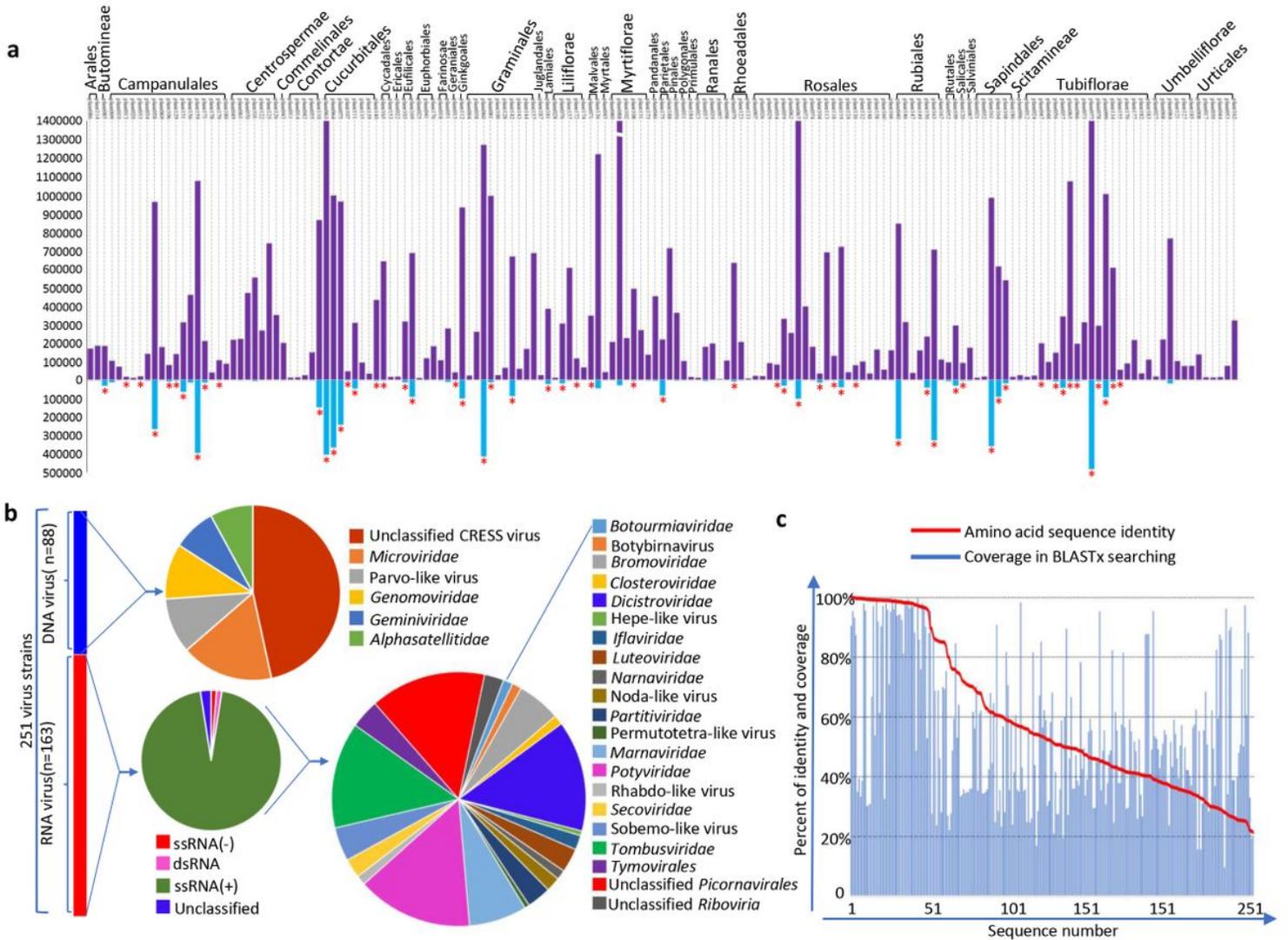


Figure 1

Identification of viruses in different species of plants. **a**, The abundance of plant-associated viruses in different species of plant. The top graph shows the total number of unique reads in each library. The library IDs are shown on top of each bar, while the host Orders are shown above the bar graph. The bottom graph shows the number of virus hits passed NR filter in viral metagenomic bioinformatic analysis. The red asterisk shows those libraries from which we have acquired complete or nearly complete genome of viruses. **b**, The number and diversity of plant-associated viruses. The left histogram shows the numbers of DNA viruses (blue bar) and RNA viruses (red bar). The right pie charts show the virus classification identified in this study. **c**, The amino acid sequence identity and coverage of plant-associated viruses with the best matched virus strains in BLASTx searching based on the 251 complete genome sequence acquired in plant species.

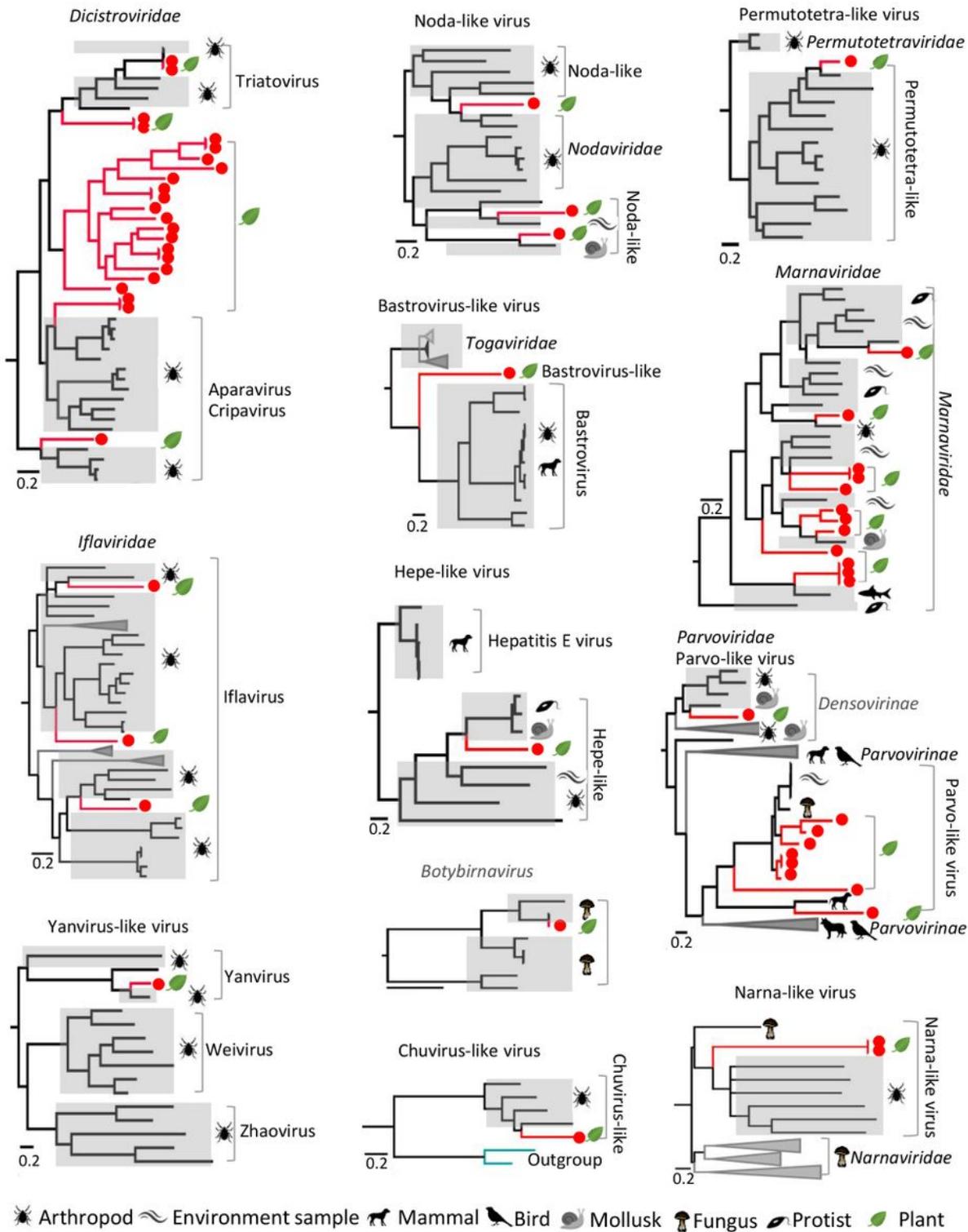


Figure 2

Phylogenies of viral genomes identified from plants. Twelve Bayesian inference trees were constructed using MrBayes v3.2 based on virus RdRp domain of RNA viruses or NS protein of parvovirus-like viruses, within each tree, the viruses found in this study are marked with red line. Hosts are indicated with different silhouette of mammal, bird, arthropod, plant leaf, or waves standing for virus environmental

source. The name of the virus family or genus is shown on the right side of each cluster. Each scale bar indicates 0.5 amino acid substitutions per site.

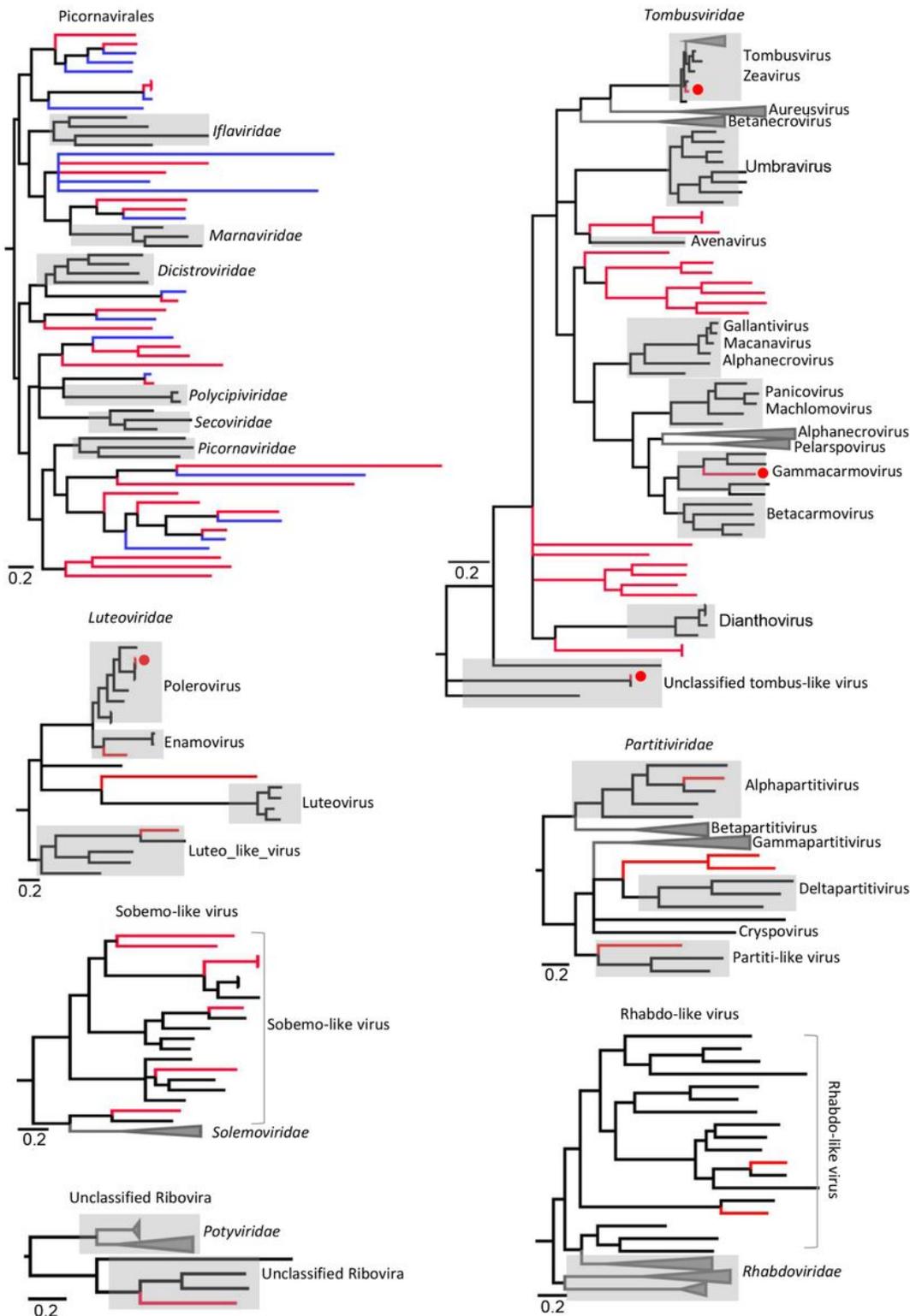


Figure 3

The phylogenies of potentially new viruses. Seven Bayesian inference trees were constructed using MrBayes v3.2 based on virus RdRp domains, within each tree, the viruses found in this study are marked with red line. In the phylogenetic tree of Picornavirales the best matched virus based BLASTp searching

using RdRp sequence of each novel virus are labeled with blue color. The name of the virus family or genus is shown on the right side of each cluster. Each scale bar indicates 0.2 amino acid substitutions per site.

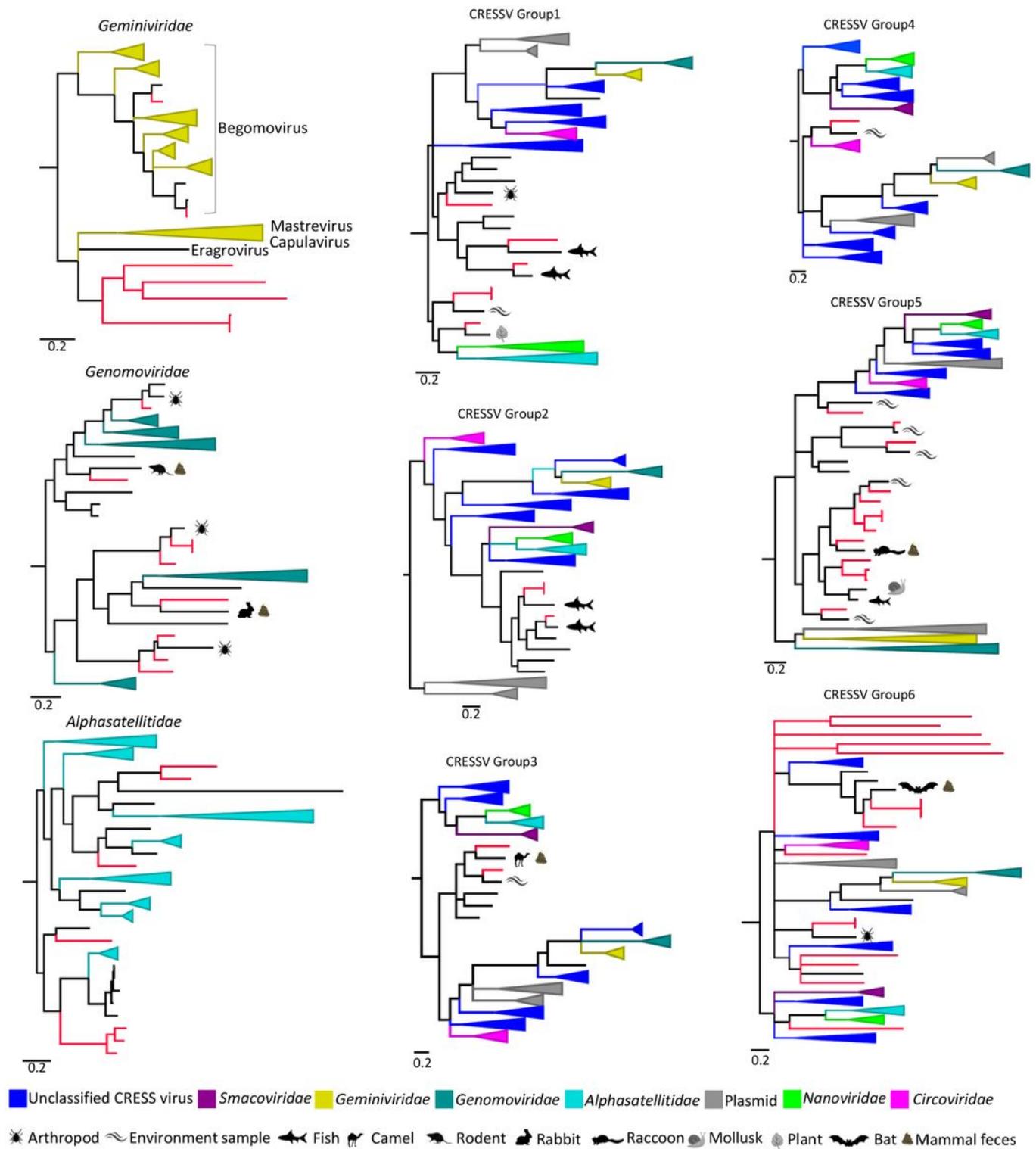


Figure 4

The phylogenies of potentially new CRESS DNA viruses. Nine Bayesian inference trees were established using MrBayes v3.2 based on REP proteins, within each tree, the viruses found in this study are marked

with red line. Hosts are indicated with different silhouette of mammal, arthropod, plant leaf, or waves standing for virus environmental source. The name of the virus family or genus is shown on the right of each cluster. Each scale bar indicates 0.2 amino acid substitutions per site.

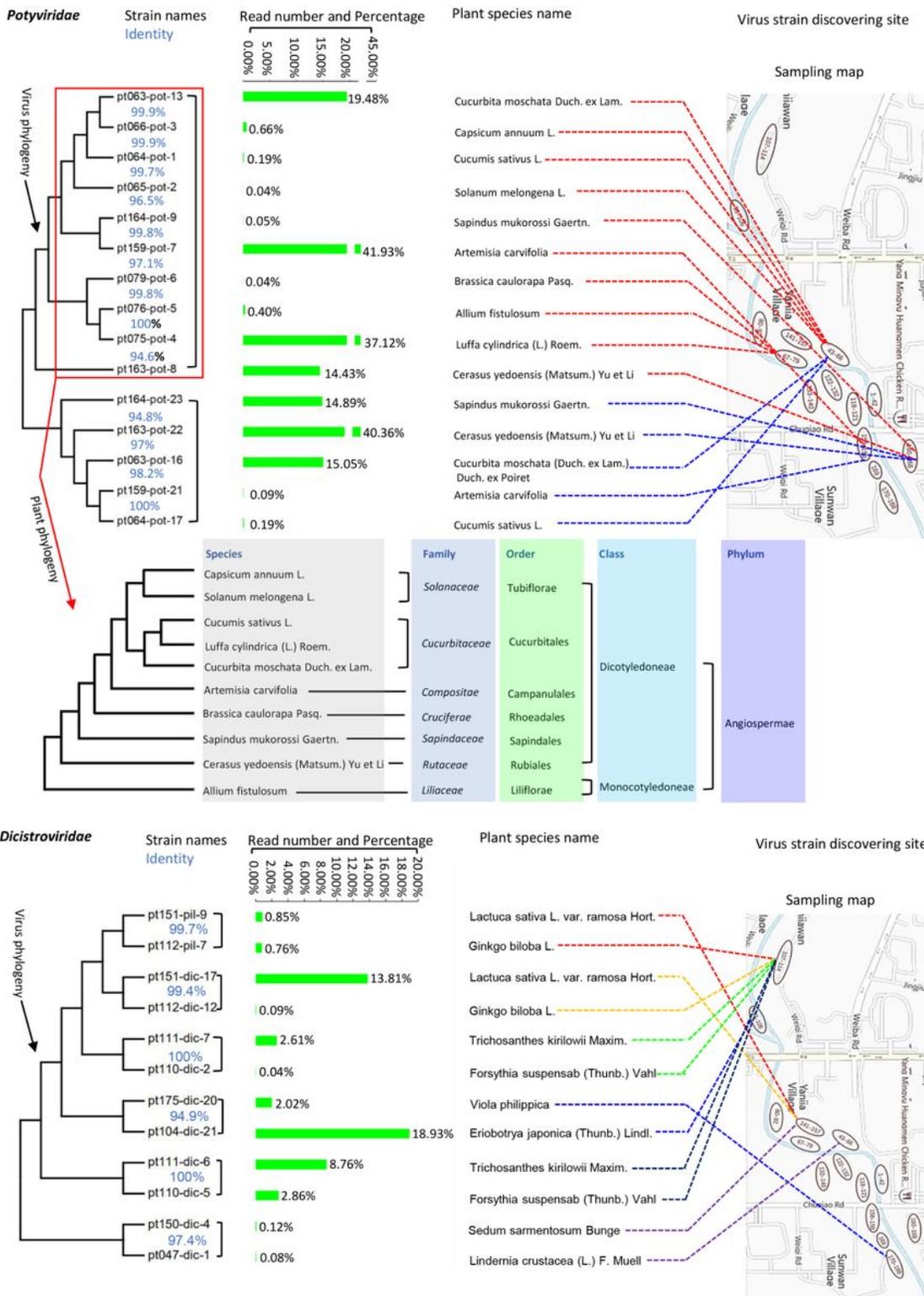


Figure 5

Potential cross-species infection of viruses among different species of plants. Neighbor-joining trees based on nucleotide sequence of complete genomes of viruses belonging in the family Potyviridae or

Dicistroviridae are shown here. Virus strain names are labeled on each branch followed by percentage of viral reads number against total unique reads number in corresponding library. Virus host plant species name are shown, dotted lines are used here to indicate the sampling sites of different species of plant. The collecting plants and its geographical location are connected by dotted lines of different colors. In the same phylogenetic cluster, the color of dotted lines behind virus hosts are the same. Sequence identity based on complete genome between neighboring sequences in the same cluster are labeled between branches in phylogenetic tree. The phylogeny and classification of host species of dicistrovirus strains in the larger cluster are shown below the dicistrovirus tree to indicated the levels over which the cross-species infection occurred.

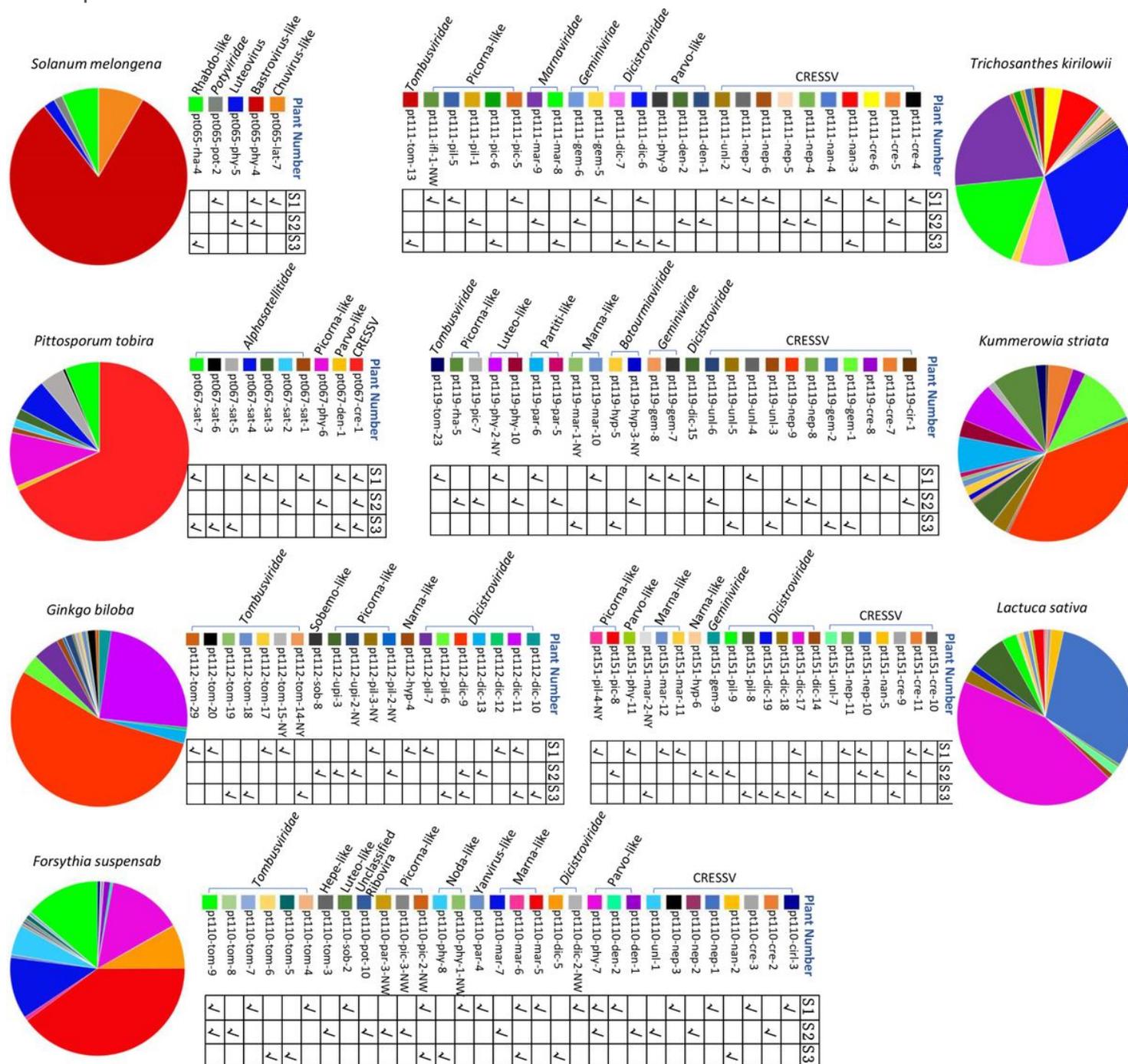


Figure 6

Co-infecting viruses in plants. Pie charts describe the viruses with complete genome in those libraries containing more than 3 different viruses. Sector area in each pie chart represents the proportion of the number of reads mapped to the complete viral genome in the library. Three individual plant in the same plant species are marked with S1, S2 and S3, respectively. Check marks below virus strains names show positive of virus in (RT-)PCR screening.

Supplementary Files

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

- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable1.xlsx](#)
- [Supplementaryinformation.pdf](#)
- [Supplementaryinformation.pdf](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable1.xlsx](#)