Transcriptome Mining Reveals a Spectrum of RNA Viruses in Primitive Plants

Jonathon C.O. Mifsud[a,b#], Rachael V. Gallagher[b.c], Edward C. Holmes[a], Jemma L. Geoghegan[d,e]

[a]Sydney Institute for Infectious Diseases, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

[b]Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109 Australia

[c]Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

[d]Department of Microbiology and Immunology, University of Otago, Dunedin 9016, New Zealand.

[e]Institute of Environmental Science and Research, Wellington 5018, New Zealand.

Running Head: Mining Reveals RNA Viruses in Primitive Plants

#Address correspondence to: Jonathon C.O. Mifsud, jmif9945@uni.sydney.edu.au

Word Counts: Abstract 250, Importance 125, Text 9074.

Keywords: plant virus, virus discovery, algae virus, *Benyviridae*, *Bunyavirales*, *Secoviridae*, evolution

Abstract

Current knowledge of plant viruses stems largely from those affecting economically important plants. Yet, plant species in cultivation represent a small and bias subset of the plant kingdom. Here, we describe virus diversity and abundance from a survey of 1079 transcriptomes from species across the breadth of the plant kingdom (Archaeplastida) by analysing open-source data from the One Thousand Plant Transcriptomes Initiative (1KP). We identified 104 potentially novel viruses, of which 40% comprised single-stranded positive-sense RNA viruses across eight orders, including members of the *Hepelivirales*, *Tymovirales*, *Cryppavirales*, *Martellivirales* and *Picornavirales*. One-third of the newly described viruses comprised double-stranded RNA viruses from the orders *Durnavirales* and *Ghabrivirales*. The remaining were negative-sense RNA viruses from the *Rhabdoviridae*, *Aspiviridae, Yueviridae, Phenuiviridae* and the newly proposed *Viridisbunyaviridae.* Our analysis considerably expands the known host range of 13 virus families to include lower plants (e.g., *Benyviridae* and *Secoviridae*) and four virus families to include algae hosts (e.g., *Tymoviridae* and *Chrysoviridae)*. The discovery of the first 30 kDa movement protein in a non-vascular plant, suggests that the acquisition of plant virus movement proteins occurred prior to the emergence of the plant vascular system. More broadly, however, a co-phylogeny analysis revealed that the evolutionary history of these families is largely driven by cross-species transmission events. Together, these data highlight that numerous RNA virus families are associated with older evolutionary plant lineages than previously thought and that the scarcity of RNA viruses found in lower plants to date likely reflects a lack of investigation rather than their absence.

Importance

Our knowledge of plant viruses is mainly limited to those infecting economically important host species. In particular, we know little about those viruses infecting primitive plant lineages such as the ferns, lycophytes, bryophytes and charophytes. To expand this understanding, we conducted a broad-scale viral survey of species across the breadth of the plant kingdom. We find that primitive plants harbour a wide diversity of RNA viruses including some that are sufficiently divergent to comprise a new virus family. The primitive plant virome we reveal offers key insights into the evolutionary history of core plant virus gene modules and genome segments. More broadly, this work emphasises that the scarcity of viruses found in these species to date likely reflects the absence of research in this area.

1. Introduction

Viruses are responsible for almost 50% of all emerging plant disease (1). Historically, virus identification and characterisation have focused on pathogenic viruses that infect species of economic importance with 69% of the current phytovirosphere — the total assemblage of viruses across the plant kingdom — discovered in cultivated plant species even though they represent less than 0.17% of all known plant diversity (2, 3). Importantly, the advent of metagenomic sequencing technology enables the comprehensive screening of plant tissues for novel and known viruses (4). Despite this, virus diversity in the vast majority of plants remains unquantified (5).

Our ability to infer the origins and diversification of the phytovirosphere from genomic data requires adequate sampling of the viruses across the plant kingdom. Several key plant groups are severely underrepresented or absent in previous studies of the phytovirosphere, including green algae (excluding the Chlorophytes), lower plants, gymnosperms and several angiosperm orders (5, 6). Improving knowledge across these groups will undoubtedly help uncover the evolutionary history of plant virus lineages. For instance, an analysis of the evolutionary history of viruses from algal ancestors might reveal deep associations that shaped the trajectory of plant evolution, including how the key evolutionary transitions of plants – such as terrestrialisation – have shaped the contemporary land plant virome (5). Similarly, through broad sampling across the plant kingdom, we can gain a stronger understanding of the acquisition of viruses through cross-species transmission from plant-associated organisms such as invertebrates, fungi, or protists (5).

The majority (68%) of the currently documented genera of plant viruses have positive-sense single-stranded RNA (+ssRNA) genomes and the majority of virus diversity is known only from angiosperms (7) (Figure 1). Currently, 16 viruses belonging to 12 virus families have been found in gymnosperms (8-12). Outside of several viruses found in ferns, we know little of the diversity of viruses in the lycophytes, bryophytes and charophytes that together encompass ~27,000 species (13-16) (Figure 1). A partial analysis of published transcriptome data detected homologs of the canonical RNA virus RNA-dependent RNA polymerase (RdRp) in algae, several lower plants and gymnosperms (17). However, it is yet to be determined whether viruses that infect freshwater algae – that include the *Zygnematophyceae* ancestors of land plants – resemble those infecting angiosperms or that of the green algae (chlorophytes) which are dominated by double-stranded DNA (dsDNA) viruses particularly from the *Phycodnaviridae* (18). To date, two +ssRNA viruses related to the benyvirids have been identified in freshwater algae (19, 20). Unlike the Chlorophyta, the Charophyta characteristically contain plasmodesmata and homologs of the key components of the land plant innate immune system, both of which have been speculated to explain the absence of double-strand (ds) DNA viruses in land plants (5, 21, 22). An understanding of the viruses infecting the Charophyta and other lower plants is required to effectively test these ideas.

Transcriptome mining has become an inexpensive and efficient method of virus discovery that leverages previous investment (23-29). To this end, we mined the transcriptome data generated by the One Thousand Plant Transcriptomes Initiative (1KP) using sequence homology searches of known plant viruses. The 1KP project provides a major untapped source of polyA-selected transcriptome data for virus discovery drawn from species across the breadth of the plants in a broad sense including green plants (Viridiplantae), glaucophytes (Glaucophyta), red algae (Rhodophyta) (30, 31). Our broad aim was to revise our understanding of the phytovirosphere using data across the plant kingdom and undertake phylogenetic analyses of plant viruses to provide insights into their origins and diversification.

2. Methods

2.1 Transcriptome data generation

The 1KP generated RNA sequencing libraries from 1,143 species across the breadth of the plant kingdom (30). In addition, 30 Chromista and red alga species were also included. Due to the diversity of species examined, samples were obtained from multiple sources including field collections, greenhouses, culture collections and laboratory specimens (32). For the majority of species, young leaves or shoots were collected, although occasionally a mix of vegetative and reproductive tissues was used. To avoid RNA degradation, RNA extraction was performed immediately after tissue collection or tissue was frozen in liquid nitrogen and stored in a -80˚C until extraction (32). Several extraction protocols were used including CTAB and TRIzol (see (32) for complete details). All sequencing was conducted at BGI-Shenzhen, China, using a combination of in-house protocols or TruSeq chemistry (32). All libraries were prepared from polyA RNA. Paired-end sequencing was initially completed using Illumina GAII machines (11% of libraries) with a ~72bp read length but later the HiSeq platform was used (89% of libraries) with a 90 bp read length (32).

2.2 Surveying for viruses in the 1KP

Raw transcriptomes (n = 1079, belonging to 960 plants species) from the 1KP major release were downloaded from the NCBI Short Read Archive (SRA) database (BioProject accession PRJEB21674) and converted to FASTQ format using the SRA Toolkit program fastq-dump in combination with the parallel-fastq-dump wrapper (https://github.com/rvalieris/parallel-fastq-dump) (33). One hundred transcriptomes within the BioProject were not publicly available (released 22/08/2019) at the commencement of this study and thus not analysed. Transcriptomes from the 1KP pilot study (BioProject accession PRJEB4921) and secondary project (BioProject accession PRJEB8056) were similarly not analysed. To reduce the downstream computing resources needed, raw sequences were mapped to their respective host genome scaffold using bowtie2 (34). Genome scaffolds were assembled as part of a previous study (30). Where genome scaffolds were not available (n = 2) all reads were assembled *de novo*. Trinity RNA-seq (v2.1.1) was used to quality trim and assemble *de novo* the unaligned reads captured from mapping (35). The assembled contigs were then assigned to known virus families and annotated through similarity searches against the NCBI nucleotide database (nt), the non-redundant protein database (nr) and a custom viral RdRp database using BLASTN and Diamond (BLASTX) (36, 37). To filter out weak BLAST sequence matches an e-value cut-off of 1×10^{-10} was employed. To identify potential false positives, putative viral contigs were manually compared across the three BLAST searches (nt, nr and RdRp) to ensure matches to virus-associated sequences were consistent.

2.3 Virus filtering and abundance calculations

For all analyses, we focused on virus families known to infect plants or algae. As our analyses rely on sequence-based similarity searches for virus detection it is necessarily biased towards viruses that exhibit to existing virus families. Together, the Virus-Host database (38) and the International Committee on Taxonomy of Viruses (39) were used to develop a list of plant virus families and genera to filter out virus-like contigs associated with vertebrate, invertebrate or fungi hosts based upon their top BLASTx and BLASTn matches. Packages within the Tidyverse collection (v1.3.0) in RStudio were used to complete these tasks (40-42). Where the host was ambiguous (e.g., belonged to a family or genera known to infect both plant and fungal species) the contig was inspected manually.

The relative abundance of each transcript within the host transcriptome was calculated using RNA-Seq by Expectation-Maximization (v1.2.28) (43). To account for variation in the number of unaligned reads between libraries after mapping, contig abundance was standardised by the total number of unaligned paired reads. Contigs under 200 nucleotides in length were excluded from further analysis.

2.4 Genome extension and annotation

Where a novel virus-like contig was discovered, we re-assembled the complete library – without removing host reads – in an attempt to recover a complete virus genome. For all re-assembled libraries, we recalculated abundance measurements to account for both host and non-host reads. The recalculated abundance measurements are shown in Supplementary Table 4. We further re-assembled all libraries belonging to non-flowering plants (n = 402). Reads were mapped onto virus-like contigs using Bbmap and heterogeneous coverage and potential misassemblies were manually resolved using Geneious (v11.0.9) (44, 45).

To determine whether a virus was novel, we followed the criteria as specified by The International Committee on Taxonomy of Viruses (39) (http://www.ictvonline.org/). Novel viruses were named using a combination of the host common name - if documented – and the associated virus taxonomic group (e.g., *Interrupted club-moss deltapartitivirus*). In cases where host assignment proved difficult the suffix “associated” was added to the host name to signify this (e.g., *Calypogeia fissa associated deltaflexivirus*). Where the taxonomic position of a virus was ambiguous the suffix “-like” was used (e.g., *Goldenrod fern qin-like virus*). Virus acronyms were created using a combination of the first and/or second letters of the host common name - if documented – and virus taxonomic group (e.g., *Leucodon julaceus beny-like virus* (LjBV)). Where multiple related viruses were found in the same host, we assigned each a number (e.g., *Odontoschisma prostratum bunyavirus 3* (OdprBV3)).

The percentage identity among virus sequences was calculated via multiple sequence alignments using Clustal Omega (v1.2.3) (46). The RdRp protein coding domain was used for all sequence alignments. Percentage identity matrices were converted to heat map plots using a custom R script provided by (28).

To characterise functional domains, predicted protein sequences along with their closest viral relatives were subjected to a domain-based search using the Conserved Domain Database (v3.18) (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and cross-referenced with the PFAM (v34.0) and Uniclust30 (v2018_08) databases available within the MMseqs2 webserver (47). To recover additional annotations, we used HHpred within the MPI Bioinformatics Toolkit webserver to query the PDB_mmCIF70 (v.12_Oct), SCOPe70 (v2.08), UniProt-SwissProt-viral70 (v3_Nov_2021) and TIGRFAMs (v15.0) databases (48). Virus genome diagrams were produced using the program littlegenomes (49). Where available NCBI/GenBank CDS information was used to annotate reference virus sequences (50).

2.5 Detection of endogenous virus elements

All genome scaffolds produced by the 1KP were used as a database in which we queried using the protein translations of the viruses discovered in this study. Endogenous viral elements (i.e., EVEs) were detected using the tblastn algorithm (51). The search threshold was limited to 100 amino acids in length with an e-value cut off of 1×10^{-20}. Where multiple hits across several plant scaffolds were observed we manually examined the sequence. Suspected endogenous virus sequences were queried against a subset whole-genome shotgun contig database which included green plants (taxid: 33090) and red algae (taxid: 2763). In addition, the virus-like sequences discovered in this study were checked for host gene contamination using the contamination function implemented in CheckV (v0.8.1) (52). All potential endogenous sequences were removed from further analyses.

2.6 Assessing library contamination by eukaryotes, bacteria, and protozoa

For libraries in which a novel virus was discovered we investigated whether reads belonging to other eukaryotes were also present in the sequencing libraries. To achieve this, we obtained taxonomic identification for raw reads in each library – without the removal of host reads – by aligning them to the NCBI nt database using the KMA aligner and the CCMetagen program (53, 54). Sequence abundance was calculated by counting the number of nucleotides matching the reference sequence with an additional correction for template length (the default parameter in KMA). Krona charts generated by CCMetagen were edited were further edited in Adobe Illustrator (https://www.adobe.com) (55). Library contamination was also assessed by the 1KP and used to inform our host-virus assignments (31).

2.7 Phylogenetic analysis of plant viruses

Phylogenetic trees of the plant-associated viruses discovered here were inferred using a maximum likelihood approach. We combined our translated virus contigs with known virus protein sequences from each respective virus family taken from NCBI/GenBank (50). Sequences were then aligned with the program Clustal Omega (v1.2.3) with default parameters (46). Sites of ambiguity were removed using trimAl (v1.2) (56). To estimate phylogenetic trees, selection of the best-fit model of amino acid substitution was determined using the Akaike information criterion, corrected AIC, and the Bayesian information criterion with the ModelFinder function (-m MFP) in IQ-TREE (57, 58). All phylogenetic trees were created using IQ-TREE with 1000 bootstrap replicates. Phylogenetic trees were annotated with FigTree (v1.4.4) (59) and further edited in Adobe Illustrator (https://www.adobe.com).

To visualise the occurrence of cross-species transmission and virus-host co-divergence across plant virus families, we reconciled the co-phylogenetic relationship between viruses and their hosts. For each select plant virus family, a vascular plant host cladogram was constructed using trees from (60) and (61), using the R package V.PhyloMaker (v0.1.0) (62). As lower plants and non-plant species are not present in the V.PhyloMaker megatree, these hosts were added to the cladogram using the software phyloT, a phylogenetic tree generator based on NCBI taxonomy (http://phylot.biobyte.de/) as well as topologies available in the appropriate literature. The host information was obtained from the NCBI Virus database (accessed 14/12/2021) and available literature (63) A tanglegram that graphically represents the correspondence between host and virus trees was created using the R packages phytools (v0.7-80) and APE (v5.5) (64, 65). Virus sequences from each family were obtained through a broad survey of all virus genomic data available on GenBank. The virus phylogenies used in the co-phylogenies were constructed as detailed above. To quantify the relative frequencies of cross-species transmission versus virus-host co-divergence we reconciled the co-phylogenetic relationship between viruses and their hosts using the Jane co-phylogenetic software package (66). Jane employs a maximum parsimony approach to determine the best 'map' of the virus phylogeny onto the host phylogeny. The cost of duplication, host-jump and extinction event types were set to one, while host-virus co-divergence was set to zero as it was considered the likely null event. Following the parsimony principle, the reconciliation proceeds by minimising the total event cost. The number of generations and the population size was both set to 100. Jane was chosen over its successor eMPRess as it allows for a virus to be associated with multiple host species and handle polytomies (67). For a multi-host virus, we represented each association as a polytomy on the virus phylogeny.

2.8 Assigning plant host clades

Each plant host was assigned to each clade in a previous study based upon their phylogenetic positioning and lineage information (30). To improve clarity when colouring the phylogenies (although not the tanglegrams) we reduced the number of clades from 25 to ten (core eudicots, basal eudicots, monocots, basalmost angiosperms, gymnosperms, fern and fern allies, non-vascular, green algae, red algae and lastly Chromista) by combining those that were closely related or potentially overlapping to increase the number of species in each group (SI Table 1).

2.9 Data availability

The raw One Thousand Plant Transcriptomes Initiative sequence reads are available at BioProject PRJEB21674. All viral genomes and corresponding sequences assembled in this study have been deposited in NCBI GenBank and assigned accession numbers xxxx-yyyy.

3. Results

We characterised the viruses found in the transcriptomes of 960 plant species within the 1KP major release. The transcriptomes represented a broad taxonomic sampling across the Archaeplastida (green plants, glaucophytes and red algae). Sequencing libraries had a median of 25,187,714 paired reads (range 10,156,464–46,650,336). A median of 82% of reads (range 1%-96%) in these libraries mapped to host genome scaffolds and were subsequently removed. *De novo* assembly of the sequencing reads resulted in a median of 36,015 contigs (range 1,396–146,217) per library, with a total of 41,256,176 contigs generated (SI Table 2).

3.1 Diversity and abundance of plant viruses

In total, virus-like transcripts were found for 603 plant species; 69% of these were plant-associated while numerous identified sequences shared high similarity to non-plant associated viruses including those known to infect fungi, invertebrate and vertebrate hosts. Among the non-plant-associated virus transcripts, 34% were unclassified (10% of total virus-like transcripts) such that they were most closely related to a virus sequence with little to no taxonomic information (i.e., a virus sequence classified as only belonging to the *Riboviria*). If an RdRp-like region was detected in an unclassified virus-like transcript we further assessed whether it could be plant-associated (see Phylogenetic analysis of identified viruses). The remaining non-plant-associated virus transcripts were largely classified within the *Orthomyxoviridae* (vertebrate associated) (25%), *Rhabdoviridae* (invertebrate associated) (17%), *Partitiviridae* (fungus associated) (10%), *Mimiviridae* (amoeboid associated) (10%) and *Adenoviridae* (vertebrate associated) (7%) and excluded from the remainder of this study. These sequences are discussed in more detail in the section on "Presence of contaminants in sequencing libraries" below. Although some of these viruses could represent plant infection it remains challenging to discern and we, therefore, made the conservative decision to remove them from the analysis.

We detected transcripts closely associated with viruses containing single and double-stranded DNA and RNA genomes. The majority of virus-like sequences belonged to families with +ssRNA genomes (61%) or reverse-transcribing dsDNA viruses (22%) (Figure 1). The +ssRNA virus transcripts were predominately classified within the *Betaflexiviridae* (30%), *Potyviridae* (19%), *Secoviridae* (16%) and *Alphaflexiviridae* (10%) (SI Table 3). Negative-sense single-stranded RNA (-ssRNA) virus transcripts were classified within the *Aspiviridae* (0.04%), *Rhabdoviridae* (6%) and *Tospoviridae* (3%) (*Phenuiviridae* and *Yueviridae* transcripts were later detected in the unclassified virus-like transcripts) (SI Table 3). dsDNA virus transcripts with sequence similarities to the *Phycodnaviridae* were detected across the algae samples. These phycodna-like virus transcripts frequently encoded the chitinase and DNA ligase genes which are homologous to those in distantly related host organisms including fungi and bacteria. Due to the difficulties discerning whether these transcripts represent *Phycodnaviridae* sequences or contamination, we excluded all phycodnavirus-related sequences. All remaining dsDNA viruses were exclusively reverse-transcribing viruses from the *Caulimoviridae*. We failed to detect any sequences that shared homology with several plant virus families including *Reoviridae*, *Nanoviridae* and *Fimoviridae* (although see the Discussion for caveats).

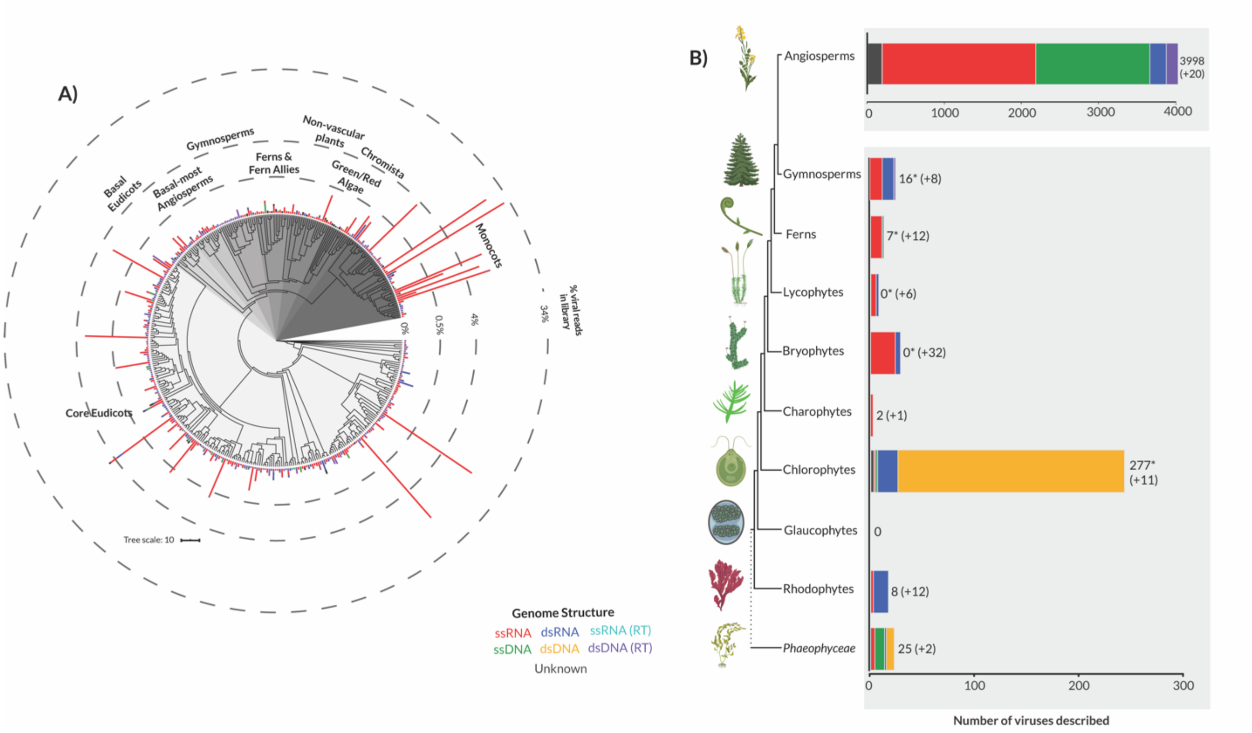


Figure 1. (A) Phylogram of virus composition across the One Thousand Plant Transcriptomes Initiative (1KP) samples. Plant-associated virus abundance was summarised for each plant species and normalised using a Box-Cox transformation. The height of each bar represents the percentage of virus reads detected in each plant species (after the removal of host reads). Plant clades are labelled and differentiated by shades of grey. The 1KP ASTRAL tree was used as the basis for this tree (30). Clade and abundance annotations were added using the Interactive Tree of Life (iTOL) web-based tool (109). (B) The phytovirosphere across the Plantae and *Phaeophyceae*. A schematic tree of the evolution of major plant groups. Each bar represents the number of total viruses formally or likely associated with each host group and is coloured by virus genome composition. The total number of viruses for each plant group plus those found in this study is also shown at the end of each bar. The Virus-Host (38) and NCBI virus databases (110) combined with literature searches were used to obtain virus counts. Lineage branches are not drawn to scale. To our knowledge, no viruses have been found in the Glaucophytes. Plant and algae images were obtained from BioRender.com or drawn in Adobe Illustrator

(https://www.adobe.com). *Transcriptome scaffolds from libraries belonging to these host groups shared homology to virus RdRps and were partially analysed but not assembled or deposited to GenBank (17).

There was a large range of total viral abundance in each library (5×10^{-6}% – 31% reads after host-associated reads were removed). Viruses with +ssRNA genomes accounted for the vast majority (99.8%) of virus abundance detected (Figure 1, SI Table 3). As expected, virus discovery was concentrated in the flowering plants (angiosperms), which have the highest number of previously classified viruses. For instance, plant virus-like sequences were frequently discovered in the core eudicots and monocots (i.e., 73% of libraries in which plant virus transcripts were found). The detection rate of plant viruses was highest in the most basal angiosperms (57%) and monocots (50%). No significant difference in virus abundance was observed between sequencing platforms (Genome Analyzer II and Illumina HiSeq 2000; p=0.327).

3.2 Presence of contaminants in sequencing libraries

The bacterial, fungal and insect species that live in or on plant tissues are commonly sampled within plant sequencing libraries (31), although contamination from other plants is also a possibility during sample preparation or sequencing. To quantify the extent of library contamination we used the KMA and CCMetagen tools (Figure 2). Among the libraries analysed (n = 95), bacteria were consistently detected representing a median of 1.5% of total abundance (range 0.01%-33%). A median of 2% of library abundance was associated with fungi sequences (range 0%-53%). Arthropods and chordates were also commonly detected across libraries (found in 87 and 89 libraries, respectively) but at lower abundance (median 0.15%, range 0%-11.4%). The presence of chordate associated reads is likely attributed to various routes of sample contamination (e.g., faeces) or during sample processing and sequencing.

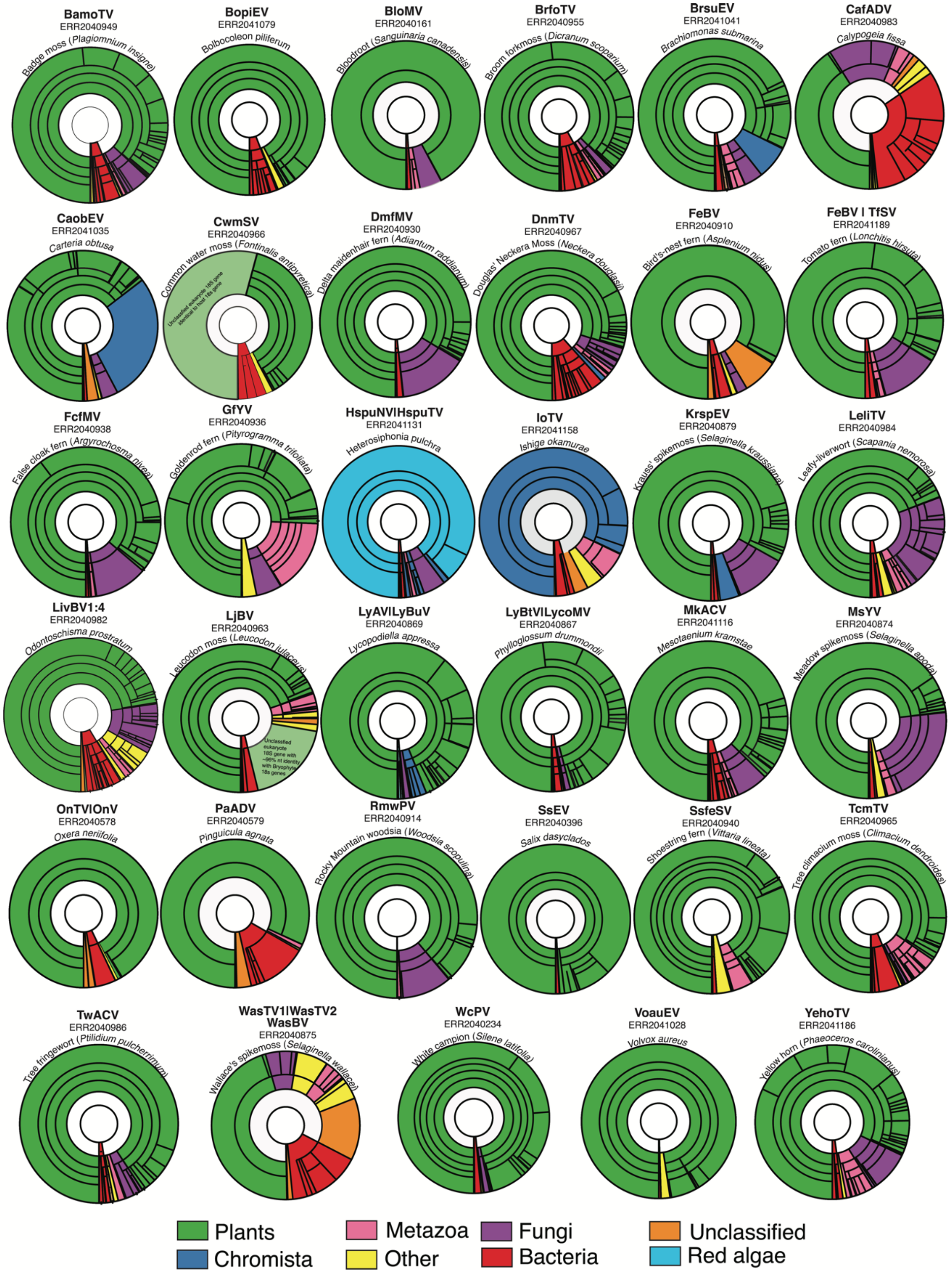


Figure 2. Taxonomic assignments of reads in select One Thousand Plant Transcriptomes Initiative (1KP) libraries. Each Krona graph illustrates the relative abundance of taxa in a

metatranscriptome at varying taxonomic levels. For clarity, a maximum depth of five taxonomic levels was chosen for each graph. The library Sequence Read Archive accession number, host species, and the corresponding virus of interest are annotated above each graph. Segments are highlighted based upon the species taxonomic grouping (plants = green, Chromista = blue, unclassified = orange, bacteria = red, metazoa = pink, fungi = purple, red algae = light blue, other = yellow). Here "plants" encompasses the Viridiplantae. Reads without any match in the nt database are not shown.

The detection of four vertebrate associated viruses across several libraries provided further evidence of library contamination. Sequences belonging to these viruses - *Influenza A virus* (16 libraries), *Human mastadenovirus C* (30 libraries), *Human immunodeficiency virus* (15 libraries) and *Parainfluenza virus 5* (3 libraries) – were present at low abundance and showed little genetic variation between libraries. Notably, chordate-associated reads were only present in 66% of libraries in which these viruses were found. The failure to consistently detect potential hosts for these viruses suggests contamination during sequencing. The four vertebrate associated viruses were largely absent in libraries in which novel plant-associated viruses were discovered, except for the *Larix speciosa*, *Brachiomonas submarina, Climacium dendroides*, *Silene latifolia and Oxera neriifolia* transcriptomes.

In addition, the 1KP compared all assembled sequences to a reference set of nuclear 18S ribosomal RNA sequences from the SILVA small subunit rRNA database using BLASTn (31, 68). Where a sample had several alignments to any other plant sequences outside of the expected source family the sample was described as having "worrisome contamination" (31). This applied to eleven plant libraries in which novel viruses were identified. Below, we discuss library contaminates from viewpoint of virus-host associations.

3.3 Phylogenetic analysis of identified viruses

To infer phylogenetic relationships between identified viruses, order and family-level phylogenetic trees were estimated using the highly conserved viral region that comprises the RdRp. In total, we assembled 104 RdRp contigs that likely represent novel virus species, of which 41 were considered as unclassified or non-plant associated due to their similarities to virus groups known to infect non-plant hosts (SI Table 4). Further analysis of these contigs revealed that they are likely plant-associated.

3.3.1 Positive-sense single-stranded RNA ((+)ssRNA) viruses

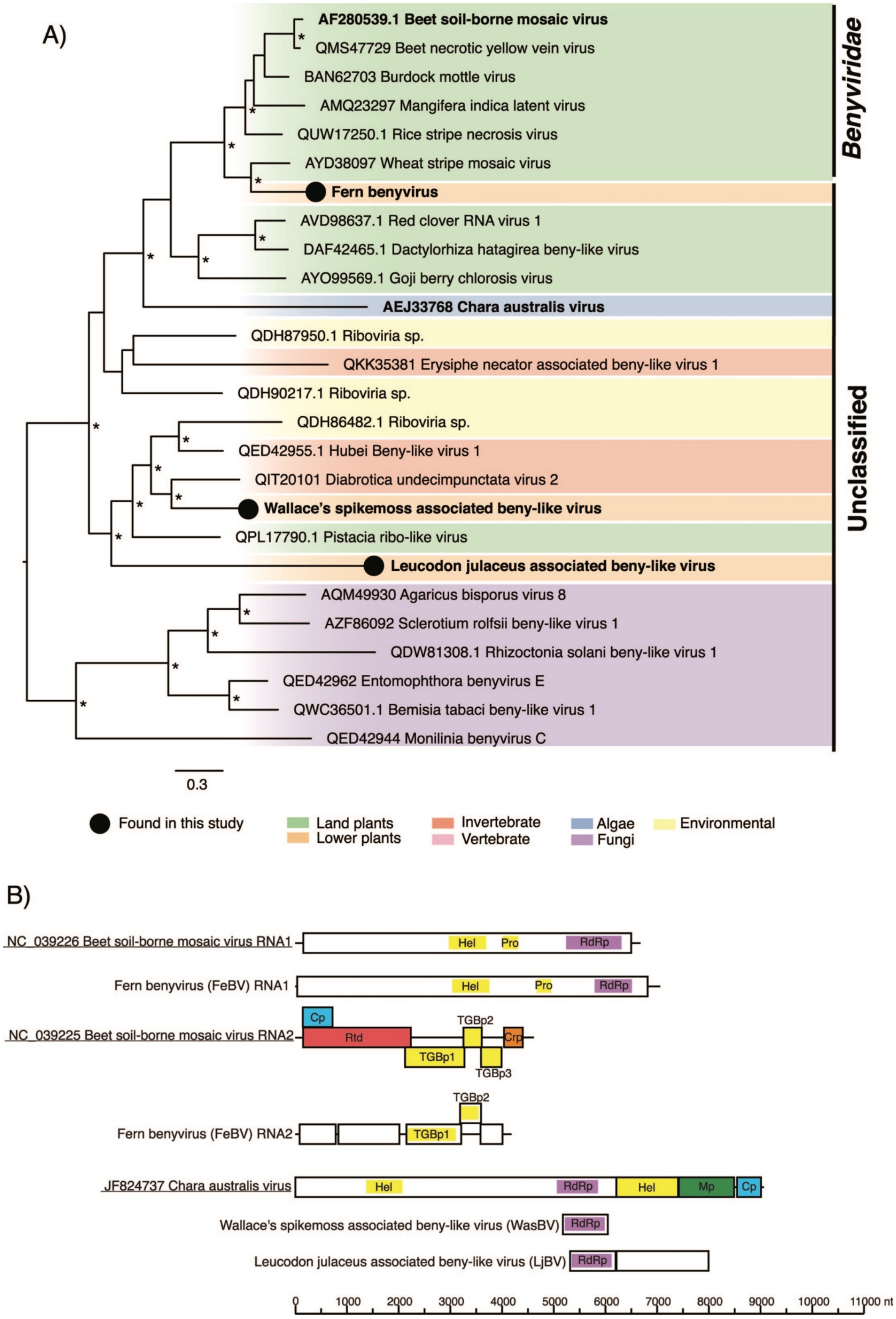
Hepelivirales

Benyviridae. We identified three beny-like sequences that to our knowledge represent the first benyvirid found in lower plants. The first sequence, tentatively named *Fern benyvirus* (FeBV), was found in both the bird's-nest fern (*Asplenium nidus*) and tomato fern (*Lonchitis hirsuta*). Together with *Wheat stripe mosaic virus*, FeBV represents a well-supported clade separate from the remaining plant benyviruses (Figure 3).

The triple gene block (TGB) is a hallmark gene module of the *Benyviridae* among several other virus families in the class *Alsuviricetes* (69)). In both fern libraries, proteins resembling the TGB were assembled (Figure 3). The TGB proteins shared ~34% amino acid identity with the TGB protein of other benyvirids. To our knowledge, this is the first TGB protein found outside of flowering plants. Phylogenetic analysis placed the TGB1 protein of FeBV basal to the *Benyviridae* (SI figure 1).

Two additional beny-like viruses, named here *Leucodon julaceus associated beny-like virus* (LjBV) and *Wallace's spikemoss associated beny-like virus* (WasBV) were assembled. LjBV and WasBV cluster with unclassified algae, invertebrates, fungi and soil-derived viruses forming a group basal to all plant benyvirids and potentially constitute a novel virus group (Figure 3). LjBV contains a second open reading frame (ORF) with no detectable homology to known sequences (Figure 3).

Due to the phylogenetic placement of LjBV and WasBV close to viruses infecting distant hosts (e.g., invertebrates and fungi), we investigated the potential of contamination from other eukaryotes as the source of these viruses. Of note, the Wallace's spikemoss metatranscriptome contained reads that matched various fungi orders (7% of all reads) as well as those matching the plant-parasitic oomycete *Albugo laibachii* (7%) which makes inferring virus-host relationships challenging (Figure 2). Reads belonging to various fungi species accounted for 10% of the bird's-nest fern transcriptome and 12% of the tomato fern transcriptome (Figure 2). Despite the presence of fungi-associated reads, the phylogenetic position of FeBV suggests that FeBV is likely plant-associated (Figure 3). No concerning contaminants were detected in the *Leucodon julaceus* transcriptome.


A)
AF280539.1 Beet soil-borne mosaic virus
QMS47729 Beet necrotic yellow vein virus
BAN62703 Burdock mottle virus
AMQ23297 Mangifera indica latent virus
QUW17250.1 Rice stripe necrosis virus
AYD38097 Wheat stripe mosaic virus
Fern benyvirus
AVD98637.1 Red clover RNA virus 1
DAF42465.1 Dactylorhiza hatagirea beny-like virus
AYO99569.1 Goji berry chlorosis virus
AEJ33768 Chara australis virus
QDH87950.1 Riboviria sp.
QKK35381 Erysiphe necator associated beny-like virus 1
QDH90217.1 Riboviria sp.
QDH86482.1 Riboviria sp.
QED42955.1 Hubei Beny-like virus 1
QIT20101 Diabrotica undecimpunctata virus 2
Wallace's spikemoss associated beny-like virus
QPL17790.1 Pistacia ribo-like virus
Leucodon julaceus associated beny-like virus
AQM49930 Agaricus bisporus virus 8
AZF86092 Sclerotium rolfsii beny-like virus 1
QDW81308.1 Rhizoctonia solani beny-like virus 1
QED42962 Entomophthora benyvirus E
QWC36501.1 Bemisia tabaci beny-like virus 1
QED42944 Monilinia benyvirus C
Benyviridae
Unclassified
0.3
● Found in this study
Land plants
Lower plants
Invertebrate
Vertebrate
Algae
Fungi
Environmental
B)
NC_039226 Beet soil-borne mosaic virus RNA1
Hel
Pro
RdRp
Fern benyvirus (FeBV) RNA1
Hel
Pro
RdRp
NC_039225 Beet soil-borne mosaic virus RNA2
Cp
Rtd
TGBp1
TGBp2
TGBp3
Crp
Fern benyvirus (FeBV) RNA2
TGBp1
TGBp2
JF824737 Chara australis virus
Hel
RdRp
Hel
Mp
Cp
Wallace's spikemoss associated beny-like virus (WasBV)
RdRp
Leucodon julaceus associated beny-like virus (LjBV)
RdRp
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000 nt

387

Figure 3. (A) Phylogenetic relationships of the beny-like viruses identified in this study. ML phylogenetic tree based on the RNA-1 replicase protein shows the topological position of virus-like sequences discovered in this study (black circles) in the context of their closest relatives. Branches are highlighted to represent host clade (land plants = green, lower plants = orange, invertebrate = red, vertebrate = pink, algae = blue, fungi = purple, yellow = environmental, Chromista = light blue, red algae = dark green). Here "Land plants" encompasses both angiosperms and gymnosperms while "Lower plants" includes the bryophytes, lycophytes, and ferns. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. (B) Genomic organization of the beny-like virus sequences identified in this study and representative species used in the phylogeny. Beet soil-borne mosaic virus RNA three and four are not pictured here. The data underlying this figure and definitions of acronyms used are presented in SI Table 5.

Tymovirales

Betaflexiviridae. We identified 18 virus sequences that fell within the order *Tymovirales*. Four virus transcripts were associated with the *Betaflexiviridae*. The first, named *Sea beet betaflexivirus* (SbBV) clusters with *Agapanthus virus A*, an unclassified betaflexivirus (Figure 4). The remaining sequences denoted *Iranian poppy betaflexivirus* (IpBV), *Linum macraei betaflexivirus* (LimBV) and *Lycopod associated betaflexivirus* (LyBtV) resemble capilloviruses. Notably, LyBtV may extend the known host range of the *Betaflexiviridae* from angiosperms to lower plants. All sequences phylogenetically cluster with known capilloviruses and potentially represent novel virus species (Figure 4). The *Phylloglossum drummondii* library in which LyBtV was assembled had contamination from lycopod and dicot species (Figure 2). As the majority of plant-associated reads were assigned to lycophytes (50%), LyBtV has been tentatively assigned to this group.

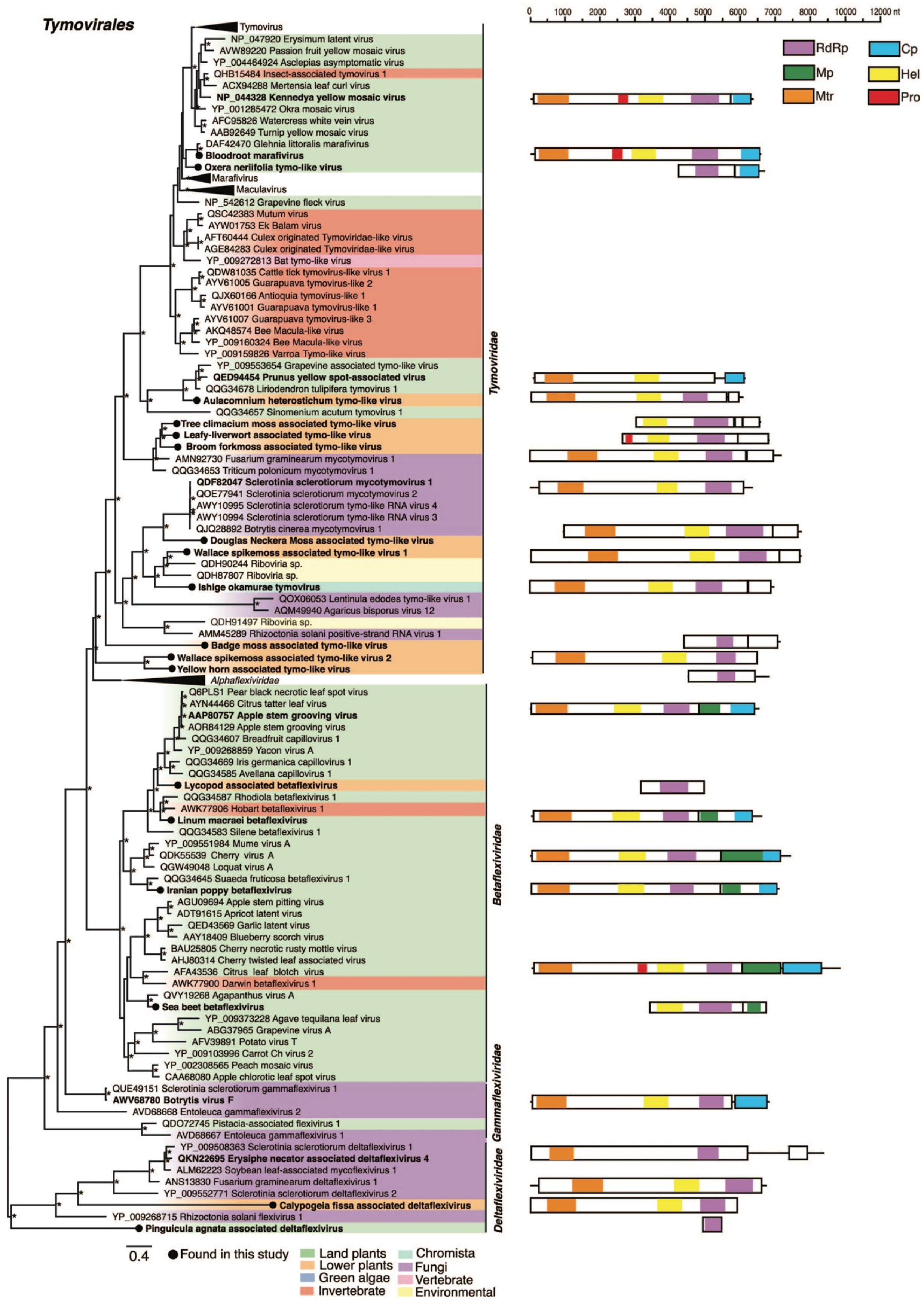

Tymovirales
Tymovirus
NP_047920 Erysimum latent virus
AVW89220 Passion fruit yellow mosaic virus
YP_004464924 Asclepias asymptomatic virus
QHB15484 Insect-associated tymovirus 1
ACX94288 Mertensia leaf curl virus
NP_044328 Kennedya yellow mosaic virus
YP_001285472 Okra mosaic virus
AFC95826 Watercress white vein virus
AAB92649 Turnip yellow mosaic virus
DAF42470 Glehnia littoralis marafivirus
Bloodroot marafivirus
Oxera neriifolia tymo-like virus
Marafivirus
Maculavirus
NP_542612 Grapevine fleck virus
QSC42383 Mutum virus
AYW01753 Ek Balam virus
AFT60444 Culex originated Tymoviridae-like virus
AGE84283 Culex originated Tymoviridae-like virus
YP_009272813 Bat tymo-like virus
QDW81035 Cattle tick tymovirus-like virus 1
AYV61005 Guarapuava tymovirus-like 2
QJX60166 Antioquia tymovirus-like 1
AYV61001 Guarapuava tymovirus-like 1
AYV61007 Guarapuava tymovirus-like 3
AKQ48574 Bee Macula-like virus
YP_009160324 Bee Macula-like virus
YP_009159826 Varroa Tymo-like virus
YP_009553654 Grapevine associated tymo-like virus
QED94454 Prunus yellow spot-associated virus
QQG34678 Liriodendron tulipifera tymovirus 1
Aulacomnium heterostichum tymo-like virus
QQG34657 Sinomenium acutum tymovirus 1
Tree climacium moss associated tymo-like virus
Leafy-liverwort associated tymo-like virus
Broom forkmoss associated tymo-like virus
AMN92730 Fusarium graminearum mycotymovirus 1
QQG34653 Triticum polonicum mycotymovirus 1
QDF82047 Sclerotinia sclerotiorum mycotymovirus 1
QOE77941 Sclerotinia sclerotiorum mycotymovirus 2
AWY10995 Sclerotinia sclerotiorum tymo-like RNA virus 4
AWY10994 Sclerotinia sclerotiorum tymo-like RNA virus 3
QJQ28892 Botrytis cinerea mycotymovirus 1
Douglas Neckera Moss associated tymo-like virus
Wallace spikemoss associated tymo-like virus 1
QDH90244 Riboviria sp.
QDH87807 Riboviria sp.
Ishige okamurae tymovirus
QOX06053 Lentinula edodes tymo-like virus 1
AQM49940 Agaricus bisporus virus 12
QDH91497 Riboviria sp.
AMM45289 Rhizoctonia solani positive-strand RNA virus 1
Badge moss associated tymo-like virus
Wallace spikemoss associated tymo-like virus 2
Yellow horn associated tymo-like virus
Tymoviridae
Alphaflexiviridae
Q6PLS1 Pear black necrotic leaf spot virus
AYN44466 Citrus tatter leaf virus
AAP80757 Apple stem grooving virus
AOR84129 Apple stem grooving virus
QQG34607 Breadfruit capillovirus 1
YP_009268859 Yacon virus A
QQG34669 Iris germanica capillovirus 1
QQG34585 Avellana capillovirus 1
Lycopod associated betaflexivirus
QQG34587 Rhodiola betaflexivirus 1
AWK77906 Hobart betaflexivirus 1
Linum macraei betaflexivirus
QQG34583 Silene betaflexivirus 1
YP_009551984 Mume virus A
QDK55539 Cherry virus A
QGW49048 Loquat virus A
QQG34645 Suaeda fruticosa betaflexivirus 1
Iranian poppy betaflexivirus
AGU09694 Apple stem pitting virus
ADT91615 Apricot latent virus
QED43569 Garlic latent virus
AAY18409 Blueberry scorch virus
BAU25805 Cherry necrotic rusty mottle virus
AHJ80314 Cherry twisted leaf associated virus
AFA43536 Citrus leaf blotch virus
AWK77900 Darwin betaflexivirus 1
QVY19268 Agapanthus virus A
Sea beet betaflexivirus
YP_009373228 Agave tequilana leaf virus
ABG37965 Grapevine virus A
AFV39891 Potato virus T
YP_009103996 Carrot Ch virus 2
YP_002308565 Peach mosaic virus
CAA68080 Apple chlorotic leaf spot virus
Betaflexiviridae
QUE49151 Sclerotinia sclerotiorum gammaflexivirus 1
AWV68780 Botrytis virus F
AVD68668 Entoleuca gammaflexivirus 2
QDO72745 Pistacia-associated flexivirus 1
AVD68667 Entoleuca gammaflexivirus 1
Gammaflexiviridae
YP_009508363 Sclerotinia sclerotiorum deltaflexivirus 1
QKN22695 Erysiphe necator associated deltaflexivirus 4
ALM62223 Soybean leaf-associated mycoflexivirus 1
ANS13830 Fusarium graminearum deltaflexivirus 1
YP_009552771 Sclerotinia sclerotiorum deltaflexivirus 2
Calypogeia fissa associated deltaflexivirus
YP_009268715 Rhizoctonia solani flexivirus 1
Pinguicula agnata associated deltaflexivirus
Deltaflexiviridae
*
0.4
● Found in this study
🟩 Land plants
🟧 Lower plants
🟦 Green algae
🟥 Invertebrate
🟩 Chromista
🟪 Fungi
🟥 Vertebrate
🟨 Environmental
0
1000
2000
3000
4000
5000
6000
7000
8000
10000
12000 nt
🟪 RdRp
🟦 Cp
🟩 Mp
🟨 Hel
🟧 Mtr
🟥 Pro

Figure 4. Left: Phylogenetic relationships of the viruses within the order *Tymovirales*. ML phylogenetic tree based on the replication protein shows the topological position of virus-like sequences discovered in this study (black circles) in the context of their closest relatives. See Figure 3 for the colour scheme. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny.

Tymoviridae. We identified 12 virus-like sequences that clustered within the *Tymoviridae* and related viruses. *Ishige okamurae associated tymo-like virus* (IoTV) was detected in the brown alga *Ishige okamurae* and likely represents the first virus in the order *Tymovirales* from brown algae. IoTV, along with ten sequences assembled from hornworts, liverworts and bryophytes grouped with tymo-like viruses from fungus and environmental samples (Figure 4). It is uncertain whether the true hosts of the novel tymo-like viruses discovered here are plants. Fungi contaminates were detected across these libraries but varied in abundance (range 1%-21%, mean = 6%). Despite their clustering with mycotymoviruses, *Broom forkmoss associated tymo-like virus* (BrfoTV) and *Tree climacium moss associated tymo-like virus* (TcmTV) were assembled from libraries with ~1% fungal reads, highlighting the inherent difficulties in host-virus assignment. Importantly, <1% of reads in *Ishige okamurae* transcriptome belonged to species of fungi (Figure 2).

We assembled two tymo-like virus sequences denoted *Oxera neriifolia tymo-like virus* (OnTV) and *Bloodroot marafivirus* (BloMV). BloMV and OnTV grouped with the unclassified *Glehnia littoralis marafivirus* (Figure 4). Marafiviruses and tymoviruses are commonly distinguished from each other based upon a highly conserved 16 nucleotide (nt) sequence known as the “tymobox”

[GAGUCUGAAUUGCUUC] in tymoviruses and the “marafibox” [CA(G/A)GGUGAAUUGCUUC] in marafiviruses (70, 71). While these two novel viruses cluster together phylogenetically, they differ in terms of genome structure and motifs. A “marafibox” like sequence appears to be present in BloMV (CAACGCGAAUUGCUUU) (5606-5621 nt) albeit differing by several residues. This finding, combined with the BloMV genome likely consisting of a single large ORF, supports the assignment of BloMV as a *Marafivirus*. OnTV, like members of the *Tymovirus* genera, contains both a second ORF – likely encoding a coat protein (CP) – and a tymobox (1493-1508 nt) (Figure 4). Phylogenetic analysis of the coat protein sequence places OnTV and BloMV in a clade with macula- and marafi-like viruses (SI Figure 1).

Deltaflexiviridae. We assembled two sequences that share similarities to members of the mycotymovirus family, *Deltaflexiviridae*. The first sequence was detected in the liverwort *Calypogeia fissa*, tentatively named *Calypogeia fissa associated deltaflexivirus* (CafADV) and appeared distantly related to delta- and gammaflexiviruses. A second related partial sequence, named here *Pinguicula agnata virus* (PaV), shared 32% amino acid identity with mycoflexivirus, *Botrytis virus F*. In a phylogenetic analysis with members of the *Tymovirales,* CafADV and PaAGV are placed with the deltaflexivirids (Figure 4).

It is unclear whether the source of these virus sequences is from plants or contamination from other eukaryotes. The *C. fissa* library contained numerous contaminates including algae, fungi and bacteria representing 1%, 15% and 33% of total reads respectively, which make discerning the host association for CafAV challenging (Figure 2). Interestingly, no fungi-associated reads were found in the *P. agnata* library suggesting a potential plant origin (Figure 2).

Picornavirales

Secoviridae. We identified four sequences that shared similarities to members of the *Secoviridae* denoted *Common water moss secovirus* (CwmSV), *Salix dasyclados secovirus* (SadSV), *Tomato fern secovirus* (TfSV) and *Shostring fern secovirus* (SfSV). CwmSV, TfSV and SfSV cluster within the nepoviruses and likely represented the first seco-like virus detected in the bryophytes and ferns (Figure 5).

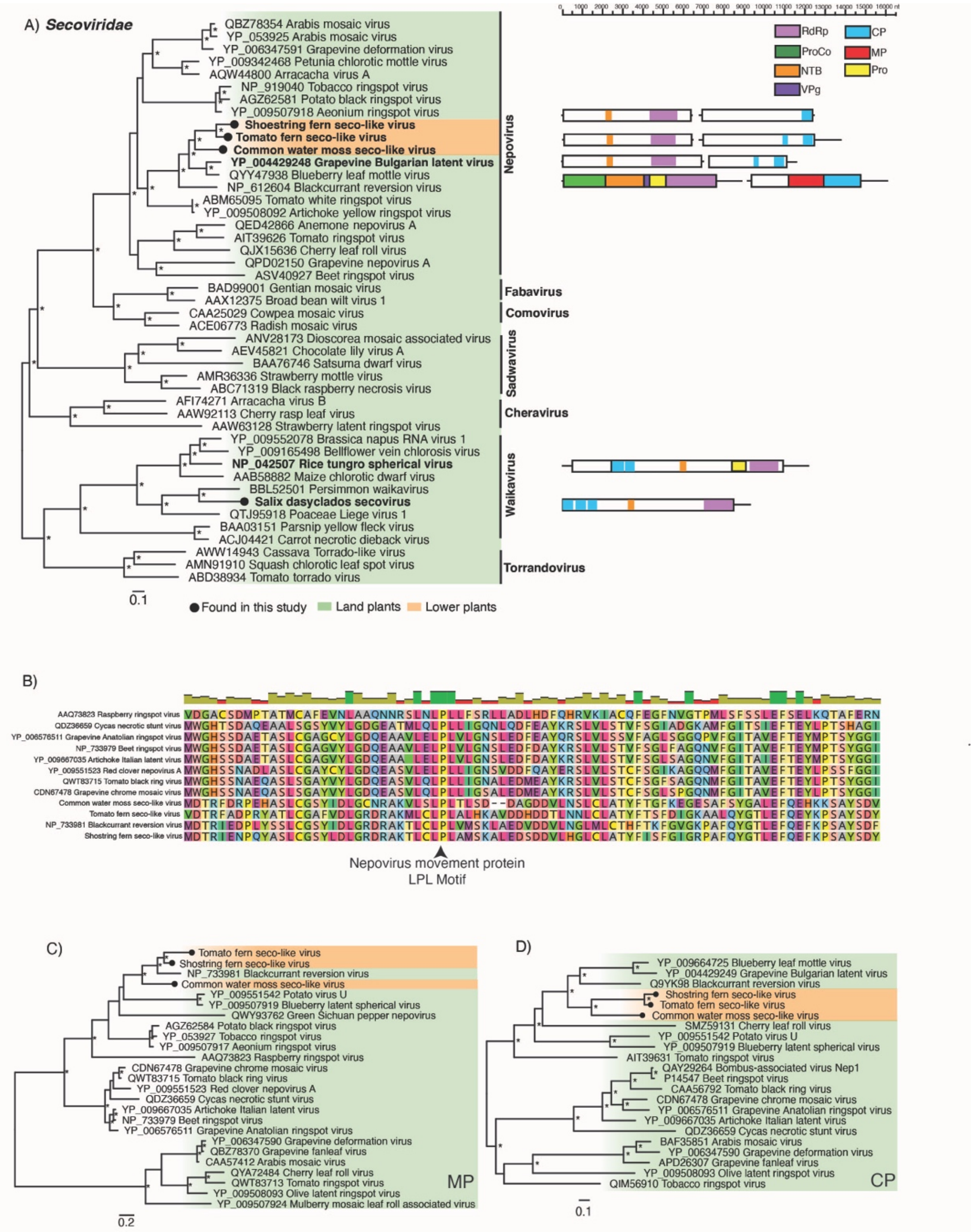

A) Secoviridae
QBZ78354 Arabis mosaic virus
YP_053925 Arabis mosaic virus
YP_006347591 Grapevine deformation virus
YP_009342468 Petunia chlorotic mottle virus
AQW44800 Arracacha virus A
NP_919040 Tobacco ringspot virus
AGZ62581 Potato black ringspot virus
YP_009507918 Aeonium ringspot virus
Shoestring fern seco-like virus
Tomato fern seco-like virus
Common water moss seco-like virus
YP_004429248 Grapevine Bulgarian latent virus
QYY47938 Blueberry leaf mottle virus
NP_612604 Blackcurrant reversion virus
ABM65095 Tomato white ringspot virus
YP_009508092 Artichoke yellow ringspot virus
QED42866 Anemone nepovirus A
AIT39626 Tomato ringspot virus
QJX15636 Cherry leaf roll virus
QPD02150 Grapevine nepovirus A
ASV40927 Beet ringspot virus
BAD99001 Gentian mosaic virus
AAX12375 Broad bean wilt virus 1
CAA25029 Cowpea mosaic virus
ACE06773 Radish mosaic virus
ANV28173 Dioscorea mosaic associated virus
AEV45821 Chocolate lily virus A
BAA76746 Satsuma dwarf virus
AMR36336 Strawberry mottle virus
ABC71319 Black raspberry necrosis virus
AFI74271 Arracacha virus B
AAW92113 Cherry rasp leaf virus
AAW63128 Strawberry latent ringspot virus
YP_009552078 Brassica napus RNA virus 1
YP_009165498 Bellflower vein chlorosis virus
NP_042507 Rice tungro spherical virus
AAB58882 Maize chlorotic dwarf virus
BBL52501 Persimmon waikavirus
Salix dasyclados secovirus
QTJ95918 Poaceae Liege virus 1
BAA03151 Parsnip yellow fleck virus
ACJ04421 Carrot necrotic dieback virus
AWW14943 Cassava Torrado-like virus
AMN91910 Squash chlorotic leaf spot virus
ABD38934 Tomato torrado virus
Nepovirus
Fabavirus
Comovirus
Sadwavirus
Cheravirus
Waikavirus
Torrandovirus
0.1
● Found in this study
🟩 Land plants
🟧 Lower plants
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 14000 15000 16000 nt
RdRp
CP
ProCo
MP
NTB
Pro
VPg
B)
AAQ73823 Raspberry ringspot virus VDGACSDMPTATMCAFEVNLAAQNNRSLNLPLLFSRLLADLHDFQHRVKIACQFEGFNVGTPMLSFSSLEFSELKQTAFERN
QDZ36659 Cycas necrotic stunt virus MWGHTSDAQEAALSGSYVYLGDGEATMLQLPLLIGQNLQDFEAYKRSLVLSTVFSGIADGKAMFGITSIEFTEYLPTSHAGI
YP_006576511 Grapevine Anatolian ringspot virus MWGHSSDAETASLCGAGCYLGDQEAAVLELPLVLGNSLEDFEAYQRSLVLSSVFAGLSGGQPVFGITAVEFTEYMPTSYGGI
NP_733979 Beet ringspot virus MWGHSSDAETASLCGAGVYLGDQEAAVLELPLVLGNSLEDFDAYKRSLVLSTVFSGLFAGQNVFGITAVEFTEYMPTSYGGI
YP_009667035 Artichoke Italian latent virus MWGHSSDAETASLCGAGVYLGDQEAA LELPLVLGNSLEDFDAYKRSLVLSTVFSGLSAGQNVFGITAIEFTEYMPTSYGGI
YP_009551523 Red clover nepovirus A MWGHSSNADLASLCGAYCYLGDQEASVLELPLLIGNSVDDFQAYERSLVLSTCFSGIKAGQQMFGITAVEFTEYLPSSFGGI
QWT83715 Tomato black ring virus MWGHSSNAEQASLSGAYVYLGDQEASVLQLPLLIGNALEDMEAYKRSLVLSTCFSGFSAGQNMFGITAVEFTEYLPTSYGGI
CDN67478 Grapevine chrome mosaic virus MWGHSSNAEQASLSGAYVYLGDQEASVLQLPLLIGSALEDMEAYKRSLVLSTCFSGLSPGQNMFGITAVEFTEYLPTSYGGI
Common water moss seco-like virus MDTRFDRPEHASLCGSYIDLGCNRAKVLSLPLTLSD--DAGDDVLNSLCLATYFTGFKEGESAFSYGALEFQEHKKSAYSDV
Tomato fern seco-like virus VDTRFADPRYATLCGAFVDLGRDRAKMLCLPLALHKAVDDHDDTLNNLCLATYFTSFDIGKAALQYGTLEFQEYKPSAYSDY
NP_733981 Blackcurrant reversion virus MDTRIEDPLYSSLCGSYIDLGRDRAKTLCLPLVMSKLAEDVDDVLNGLMLCTHFTKFGVGKPAFQYGTLEFQEFKPSAYSDF
Shostring fern seco-like virus MDTRIENPQYASLCGSYLDLGRDRAKTLCLPLAMSKALEDSDDVLHGLCLATYFISFGIGRPAFQYGTLEFQEFKPSAYSDY
Nepovirus movement protein
LPL Motif
C)
Tomato fern seco-like virus
Shostring fern seco-like virus
NP_733981 Blackcurrant reversion virus
Common water moss seco-like virus
YP_009551542 Potato virus U
YP_009507919 Blueberry latent spherical virus
QWY93762 Green Sichuan pepper nepovirus
AGZ62584 Potato black ringspot virus
YP_053927 Tobacco ringspot virus
YP_009507917 Aeonium ringspot virus
AAQ73823 Raspberry ringspot virus
CDN67478 Grapevine chrome mosaic virus
QWT83715 Tomato black ring virus
YP_009551523 Red clover nepovirus A
QDZ36659 Cycas necrotic stunt virus
YP_009667035 Artichoke Italian latent virus
NP_733979 Beet ringspot virus
YP_006576511 Grapevine Anatolian ringspot virus
YP_006347590 Grapevine deformation virus
QBZ78370 Grapevine fanleaf virus
CAA57412 Arabis mosaic virus
QYA72484 Cherry leaf roll virus
QWT83713 Tomato ringspot virus
YP_009508093 Olive latent ringspot virus
YP_009507924 Mulberry mosaic leaf roll associated virus
MP
0.2
D)
YP_009664725 Blueberry leaf mottle virus
YP_004429249 Grapevine Bulgarian latent virus
Q9YK98 Blackcurrant reversion virus
Shostring fern seco-like virus
Tomato fern seco-like virus
Common water moss seco-like virus
SMZ59131 Cherry leaf roll virus
YP_009551542 Potato virus U
YP_009507919 Blueberry latent spherical virus
AIT39631 Tomato ringspot virus
QAY29264 Bombus-associated virus Nep1
P14547 Beet ringspot virus
CAA56792 Tomato black ring virus
CDN67478 Grapevine chrome mosaic virus
YP_006576511 Grapevine Anatolian ringspot virus
YP_009667035 Artichoke Italian latent virus
QDZ36659 Cycas necrotic stunt virus
BAF35851 Arabis mosaic virus
YP_006347590 Grapevine deformation virus
APD26307 Grapevine fanleaf virus
YP_009508093 Olive latent ringspot virus
QIM56910 Tobacco ringspot virus
CP
0.1

Figure 5. (A) Left: Phylogenetic relationships of the viruses identified within the virus family *Secoviridae*. ML phylogenetic trees based on the Pro-pol region show the topological position of virus-like sequences discovered in this study (black circles) in the context of their closest relatives. Right: Genomic organization of the seco-like sequences identified in this study and representative species used in the phylogeny. (B) Multiple amino acid sequence alignment of the 30K movement protein "LPL" motifs which are highly conserved throughout the nepoviruses. (C) Phylogenetic relationships of the Nepovirus 30K movement proteins. D) Phylogenetic relationships of the Nepovirus coat proteins. For all trees, branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. See Figure 3 for the colour scheme. Viruses discovered in this study are signified using a black circle on the tree tip.

A putative RNA2 ORF was assembled for the three nepovirus-like sequences each containing a complete CP (Figure 5). The CPs fall within the nepovirus subgroup C (Figure 5D). While a movement protein (MP) domain was not formally detected, we predict that the region upstream of the CP contains a putative movement-like protein. For CwmSV, this region (amino acid position 312-883) displayed sequence homology to the MP of *Blackcurrant reversion virus* (E-value: 5.42e-86, amino acid identity: 46%). Both TfSV and SfSV displayed similar levels of homology in this region. We detected the LPL motif which is commonly found in nepovirus MPs in all three viruses (Figure 5B). Phylogenetic analysis of the putative MPs placed these viruses with *Blackcurrant reversion virus* in the genera Nepovirus (Figure 5C).

We found little evidence that these viruses were detected due to contamination by land plants or other eukaryotes. The *F. antipyretica* transcriptome was composed of reads closely related to a feather moss belonging to the order Hypnales to which *F. antipyretica* is also found. Furthermore, a large proportion of reads were assigned to an uncultured eukaryote 18S rRNA gene (54%) (HG421124.1) that was identical to the *F. antipyretica* 18S rRNA (AF023714.1) among other bryophyte 18S rRNA genes in a blastn search (e-value = 2e-102, nucleotide identity = 100%) (Figure 2). Fungi represented 12% of reads in the *L. hirsute* transcriptome.

Despite this, it is unlikely that TfSV is fungi-associated as no fungal contamination was detected in the *Vittaria lineata* transcriptome in which the closely related SfSV sequence (amino acid identity: 78%) was assembled (Figure 2).

Lenarviricota

Mitoviridae. We identified six virus sequences that cluster within the *Mitoviridae* - denoted *Chinese swamp cypress mitovirus* (CscMV)*, Asian bayberry mitovirus* (AsbaMV)*, False cloak ferns mitovirus* (FcfMV)*, Delta maidenhair fern mitovirus* (DmfMV) and *Lycopod associated mitovirus* (LycoMV). The fern (FcfMV and DmfMV) and lycophyte (LycoMV) associated sequences cluster with the fern *Azolla filiculoides mitovirus 1* and form a sister group to the plant mitoviruses and non-retroviral endogenous RNA viral elements (NERVEs) (Figure 6) (16). The gymnosperm associated sequences form a sister all the plant-associated mitoviruses and NERVEs. KpcMV extends the known host range of plant mitoviruses from ferns to lycophytes. Another mito-virus sequence was detected in the green alga *Bolbocoleon piliferum*, denoted *Bolbocoleon piliferum mito-like virus* (BopiMV). BopiMV falls basal to the mitoviruses, distinct from various unclassified mito-like viruses including the green algae associated *mito-like picolinusvirus* (QOW97241) (Figure 6). All novel sequences show strong conservation of the motifs characteristic of mitovirus RdRps (SI Figure 2) (72).

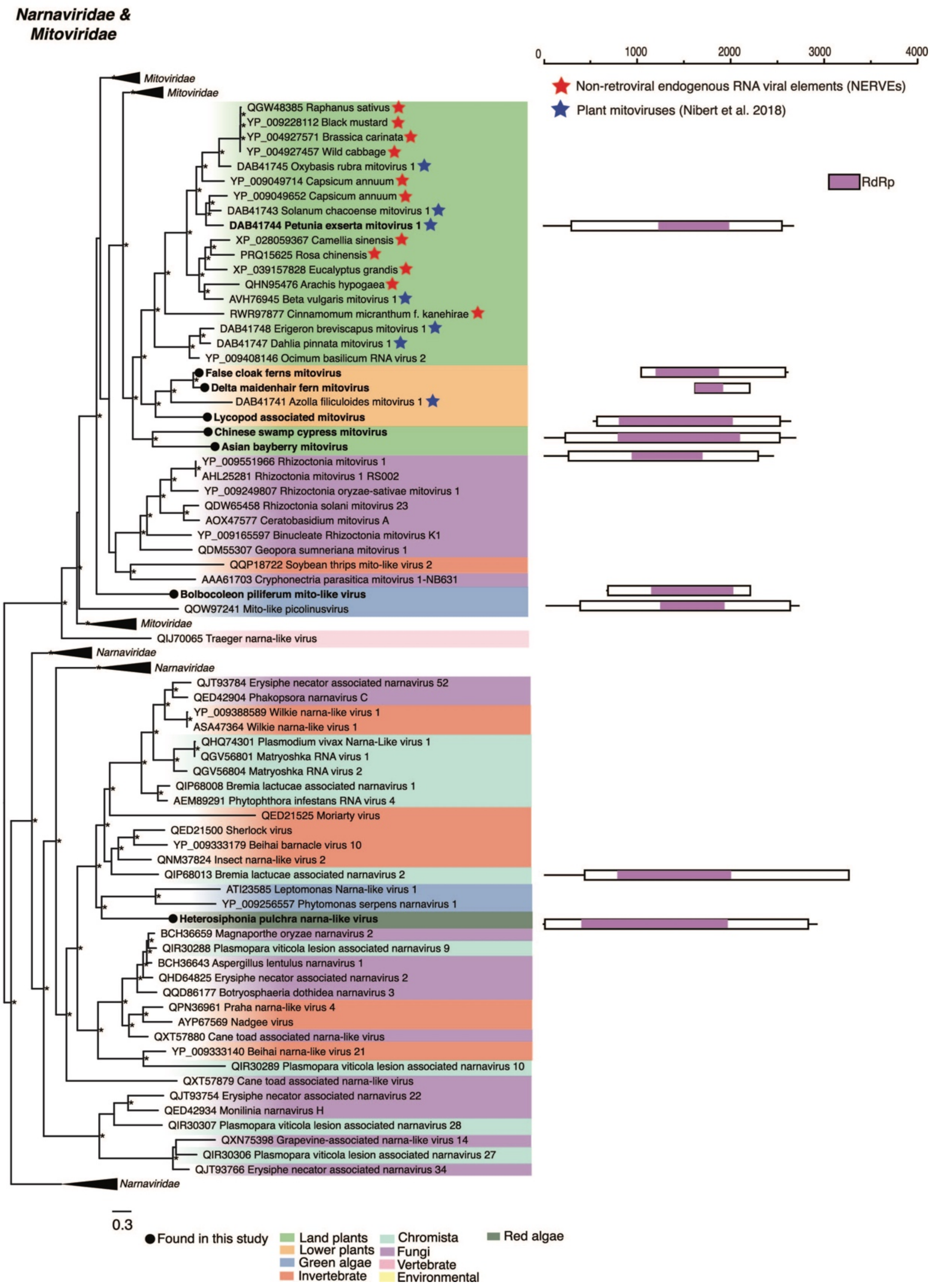

Narnaviridae & Mitoviridae
0
1000
2000
3000
4000
★ Non-retroviral endogenous RNA viral elements (NERVEs)
★ Plant mitoviruses (Nibert et al. 2018)
RdRp
Mitoviridae
Mitoviridae
QGW48385 Raphanus sativus
YP_009228112 Black mustard
YP_004927571 Brassica carinata
YP_004927457 Wild cabbage
DAB41745 Oxybasis rubra mitovirus 1
YP_009049714 Capsicum annuum
YP_009049652 Capsicum annuum
DAB41743 Solanum chacoense mitovirus 1
DAB41744 Petunia exserta mitovirus 1
XP_028059367 Camellia sinensis
PRQ15625 Rosa chinensis
XP_039157828 Eucalyptus grandis
QHN95476 Arachis hypogaea
AVH76945 Beta vulgaris mitovirus 1
RWR97877 Cinnamomum micranthum f. kanehirae
DAB41748 Erigeron breviscapus mitovirus 1
DAB41747 Dahlia pinnata mitovirus 1
YP_009408146 Ocimum basilicum RNA virus 2
False cloak ferns mitovirus
Delta maidenhair fern mitovirus
DAB41741 Azolla filiculoides mitovirus 1
Lycopod associated mitovirus
Chinese swamp cypress mitovirus
Asian bayberry mitovirus
YP_009551966 Rhizoctonia mitovirus 1
AHL25281 Rhizoctonia mitovirus 1 RS002
YP_009249807 Rhizoctonia oryzae-sativae mitovirus 1
QDW65458 Rhizoctonia solani mitovirus 23
AOX47577 Ceratobasidium mitovirus A
YP_009165597 Binucleate Rhizoctonia mitovirus K1
QDM55307 Geopora sumneriana mitovirus 1
QQP18722 Soybean thrips mito-like virus 2
AAA61703 Cryphonectria parasitica mitovirus 1-NB631
Bolbocoleon piliferum mito-like virus
QOW97241 Mito-like picolinusvirus
Mitoviridae
QIJ70065 Traeger narna-like virus
Narnaviridae
Narnaviridae
QJT93784 Erysiphe necator associated narnavirus 52
QED42904 Phakopsora narnavirus C
YP_009388589 Wilkie narna-like virus 1
ASA47364 Wilkie narna-like virus 1
QHQ74301 Plasmodium vivax Narna-Like virus 1
QGV56801 Matryoshka RNA virus 1
QGV56804 Matryoshka RNA virus 2
QIP68008 Bremia lactucae associated narnavirus 1
AEM89291 Phytophthora infestans RNA virus 4
QED21525 Moriarty virus
QED21500 Sherlock virus
YP_009333179 Beihai barnacle virus 10
QNM37824 Insect narna-like virus 2
QIP68013 Bremia lactucae associated narnavirus 2
ATI23585 Leptomonas Narna-like virus 1
YP_009256557 Phytomonas serpens narnavirus 1
Heterosiphonia pulchra narna-like virus
BCH36659 Magnaporthe oryzae narnavirus 2
QIR30288 Plasmopara viticola lesion associated narnavirus 9
BCH36643 Aspergillus lentulus narnavirus 1
QHD64825 Erysiphe necator associated narnavirus 2
QQD86177 Botryosphaeria dothidea narnavirus 3
QPN36961 Praha narna-like virus 4
AYP67569 Nadgee virus
QXT57880 Cane toad associated narna-like virus
YP_009333140 Beihai narna-like virus 21
QIR30289 Plasmopara viticola lesion associated narnavirus 10
QXT57879 Cane toad associated narna-like virus
QJT93754 Erysiphe necator associated narnavirus 22
QED42934 Monilinia narnavirus H
QIR30307 Plasmopara viticola lesion associated narnavirus 28
QXN75398 Grapevine-associated narna-like virus 14
QIR30306 Plasmopara viticola lesion associated narnavirus 27
QJT93766 Erysiphe necator associated narnavirus 34
Narnaviridae
0.3
● Found in this study
Land plants
Lower plants
Green algae
Invertebrate
Chromista
Fungi
Vertebrate
Environmental
Red algae

Figure 6. Left: Phylogenetic relationships of the viruses within the virus families *Narnaviridae* and *Mitoviridae*. ML phylogenetic trees based on the replication protein show the topological position of virus-like sequences discovered in this study (black circles) in the context of their

closest relatives. See Figure 3 for the colour scheme. Blue stars signify mitovirus sequences identified in (16). Red stars signify non-retroviral endogenous RNA viral elements (NERVEs). All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny.

There is little evidence to suggest that these sequences are derived from a non-plant organism. While the FcfMV and DmfMV libraries were contaminated with fungi, (12% and 15% of reads respectively) fungi-associated reads were absent in the libraries of all other mitoviruses. As the codon UGA encodes tryptophan (Trp) in fungal mitochondria this codon assignment is also present in fungal mitoviruses (73-75). In contrast, the UGA codon in plant mitochondria is a stop codon and hence absent from plant mitovirus sequences except as a stop codon (16). The absence of internal UGA codons in these sequences is further evidence that these sequences are plant-derived (16, 76). Although additional analyses are required, we found no evidence through searches of the 1KP genome scaffolds and the WGS shotgun database that these sequences are mitochondrial or nuclear NERVEs. Furthermore, CscMV, AsbaMV and LycoMV contain complete RdRps and their UTRs share similarities in length and identity with plant mitoviruses.

Narnaviridae A partial narna-like virus sequence was identified in the red alga *Heterosiphonia pulchra* denoted Heterosiphonia pulchra narna-like virus (HspuNV). HspuNV clusters with unclassified trypanosomatid associated viruses. While ~5% of reads in this library were associated with fungi the phylogenetic position of this virus suggests that it is not derived from fungi (Figure 2, Figure 6).

Tolivirales

Tombusviridae. An alphacarmo-like virus tentatively named *Ihi tombusvirus* (IhiTV) was identified in an Ihi (*Portulaca molokiniensis*) sample. IhiTV is phylogenetically positioned within the alphacarmoviruses (SI Figure 3).

Patatavirales

Potyviridae. We identified three virus-like sequences that clustered with plant viruses in the family *Potyviridae* – *Traubia modesta potyvirus* (TramPV), *Common milkweed potyvirus* (ComPV) and *Salt wort potyvirus* (SawPV). TramPV and ComPV shared 87% amino acid identity and may therefore represent a single virus species. The potyvirus-like sequences discovered all group with known potyviruses in a phylogenetic analysis of the Nib gene (SI Figure 3).

Martellivirales

Endornaviridae. Six alphaendorna-like virus sequences were detected in the four green algae species and one lycophyte*.* The green algae and lycophyte associated alphaendorna-like viruses termed *Bolbocoleon piliferum endorna-like virus* (BopiEV), *Volvox aureus endorna-like virus* (VoauEV), *Carteria obtusa associated endorna-like virus* (CaobEV), *Brachiomonas submarina associated endorna-like virus* (BrsuEV), *Staurastrum sebaldi endornavirus* (SsEV) and *Krauss' spikemoss associated endorna-like virus* (KrspEV) fall across the alphaendornavirus phylogeny and predominately cluster with algae and fungi associated viruses (Figure 7). There was little evidence of algae (non-host) or fungi contamination in the *S. sebaldi*, *B. piliferum* and *V. aureus* transcriptomes with <1% of all reads associated with these groups (Figure 2). Non-green algae contaminants were present in the *C. obtus* (28%)*, B. submarina* (7%) and *S. kraussiana* (4%) transcriptomes where fungi also appeared as a notable contaminate representing 11% of all reads (Figure 2). To our knowledge, these sequences represent the first endornavirus associated with charophytes, chlorophytes and lycophytes although further work is needed to confirm the virus-host associations.

Unclassified. We identified a virus-like sequence in an *Oxera neriifolia* library, termed *Oxera neriifolia associated virus*. The sequence, 10,214 nt in length contained four ORFs. The first ORF (7,536 nt) comprised of a viral methyltransferase, helicase, and RNA polymerase while the third ORF (513 nt) most closely resembled a CP. ORF one and ORF three shared the greatest sequence similarity with *Culex pipiens associated Tunisia virus* (32% amino acid identity). The second and fourth ORF share no homology to known viruses. The genome organization of OnV is distinct from the other related plant virus families (Figure 7). OnV forms a distinct and well-supported outgroup to the *Closterviridae, Bromoviridae* and *Mayoviridae* families. As such, OnV may potentially constitute a new virus family (Figure 7). We found little evidence that OnV was detected due to contamination by other eukaryotes (Figure 2).

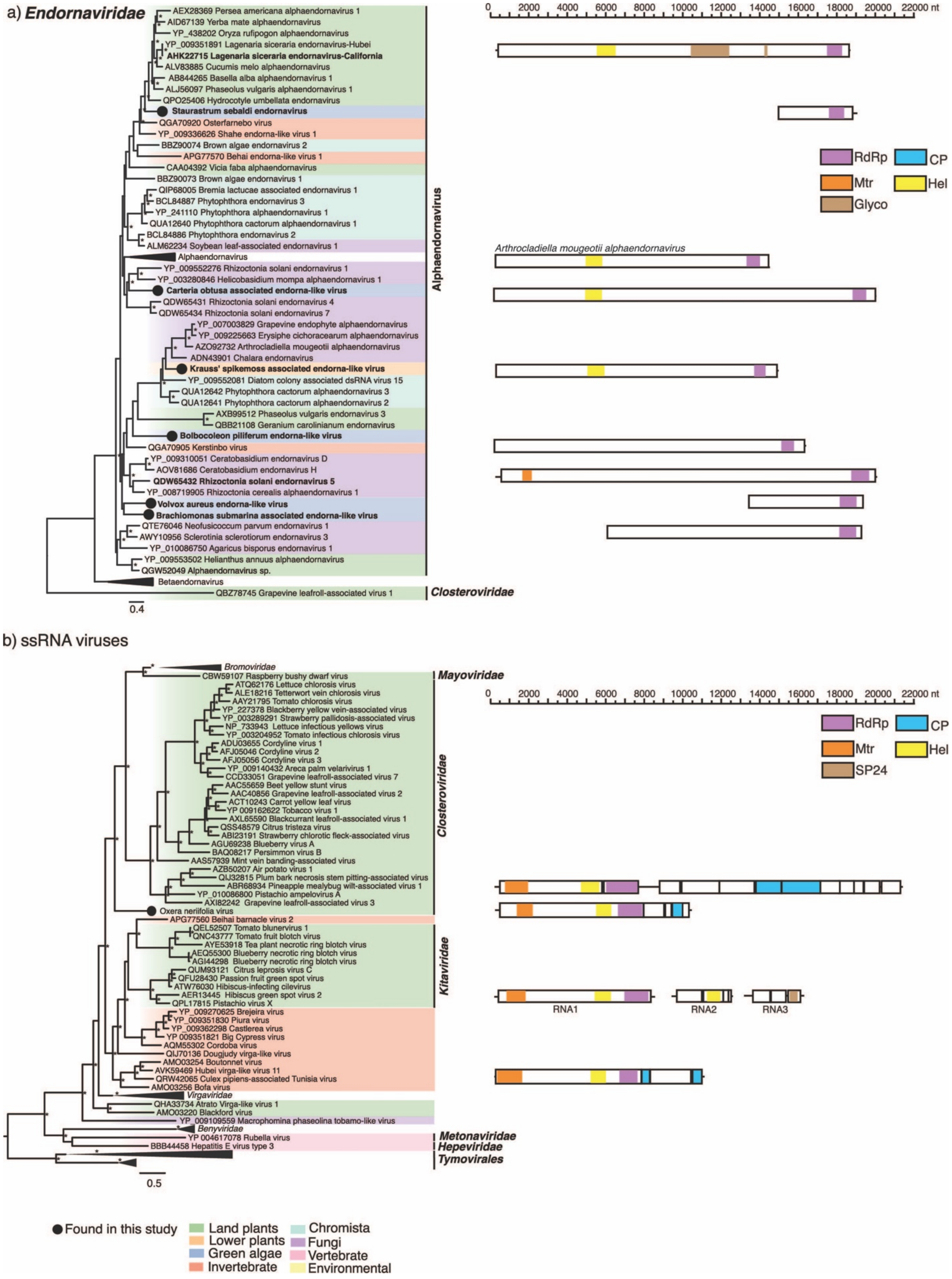

a) Endornaviridae
AEX28369 Persea americana alphaendornavirus 1
AID67139 Yerba mate alphaendornavirus
YP_438202 Oryza rufipogon alphaendornavirus
YP_009351891 Lagenaria siceraria endornavirus-Hubei
AHK22715 Lagenaria siceraria endornavirus-California
ALV83885 Cucumis melo alphaendornavirus
AB844265 Basella alba alphaendornavirus 1
ALJ56097 Phaseolus vulgaris alphaendornavirus 1
QPO25406 Hydrocotyle umbellata endornavirus
Staurastrum sebaldi endornavirus
QGA70920 Osterfarnebo virus
YP_009336626 Shahe endorna-like virus 1
BBZ90074 Brown algae endornavirus 2
APG77570 Behai endorna-like virus 1
CAA04392 Vicia faba alphaendornavirus
BBZ90073 Brown algae endornavirus 1
QIP68005 Bremia lactucae associated endornavirus 1
BCL84887 Phytophthora endornavirus 3
YP_241110 Phytophthora alphaendornavirus 1
QUA12640 Phytophthora cactorum alphaendornavirus 1
BCL84886 Phytophthora endornavirus 2
ALM62234 Soybean leaf-associated endornavirus 1
Alphaendornavirus
YP_009552276 Rhizoctonia solani endornavirus 1
YP_003280846 Helicobasidium mompa alphaendornavirus 1
Carteria obtusa associated endorna-like virus
QDW65431 Rhizoctonia solani endornavirus 4
QDW65434 Rhizoctonia solani endornavirus 7
YP_007003829 Grapevine endophyte alphaendornavirus
YP_009225663 Erysiphe cichoracearum alphaendornavirus
AZO92732 Arthrocladiella mougeotii alphaendornavirus
ADN43901 Chalara endornavirus
Krauss' spikemoss associated endorna-like virus
YP_009552081 Diatom colony associated dsRNA virus 15
QUA12642 Phytophthora cactorum alphaendornavirus 3
QUA12641 Phytophthora cactorum alphaendornavirus 2
AXB99512 Phaseolus vulgaris endornavirus 3
QBB21108 Geranium carolinianum endornavirus
Bolbocoleon piliferum endorna-like virus
QGA70905 Kerstinbo virus
YP_009310051 Ceratobasidium endornavirus D
AOV81686 Ceratobasidium endornavirus H
QDW65432 Rhizoctonia solani endornavirus 5
YP_008719905 Rhizoctonia cerealis alphaendornavirus 1
Volvox aureus endorna-like virus
Brachiomonas submarina associated endorna-like virus
QTE76046 Neofusicoccum parvum endornavirus 1
AWY10956 Sclerotinia sclerotiorum endornavirus 3
YP_010086750 Agaricus bisporus endornavirus 1
YP_009553502 Helianthus annuus alphaendornavirus
QGW52049 Alphaendornavirus sp.
Betaendornavirus
QBZ78745 Grapevine leafroll-associated virus 1
Alphaendornavirus
Closteroviridae
0.4
0 2000 4000 6000 8000 10000 12000 14000 16000 18000 20000 22000 nt
RdRp
CP
Mtr
Hel
Glyco
Arthrocladiella mougeotii alphaendornavirus
b) ssRNA viruses
Bromoviridae
CBW59107 Raspberry bushy dwarf virus
ATQ62176 Lettuce chlorosis virus
ALE18216 Tetterwort vein chlorosis virus
AAY21795 Tomato chlorosis virus
YP_227378 Blackberry yellow vein-associated virus
YP_003289291 Strawberry pallidosis-associated virus
NP_733943 Lettuce infectious yellows virus
YP_003204952 Tomato infectious chlorosis virus
ADU03655 Cordyline virus 1
AFJ05046 Cordyline virus 2
AFJ05056 Cordyline virus 3
YP_009140432 Areca palm velarivirus 1
CCD33051 Grapevine leafroll-associated virus 7
AAC55659 Beet yellow stunt virus
AAC40856 Grapevine leafroll-associated virus 2
ACT10243 Carrot yellow leaf virus
YP 009162622 Tobacco virus 1
AXL65590 Blackcurrant leafroll-associated virus 1
QSS48579 Citrus tristeza virus
ABI23191 Strawberry chlorotic fleck-associated virus
AGU69238 Blueberry virus A
BAQ08217 Persimmon virus B
AAS57939 Mint vein banding-associated virus
AZB50207 Air potato virus 1
QIJ32815 Plum bark necrosis stem pitting-associated virus
ABR68934 Pineapple mealybug wilt-associated virus 1
YP_010086800 Pistachio ampelovirus A
AXI82242 Grapevine leafroll-associated virus 3
Oxera neriifolia virus
APG77560 Beihai barnacle virus 2
QEL52507 Tomato blunervirus 1
QNC43777 Tomato fruit blotch virus
AYE53918 Tea plant necrotic ring blotch virus
AEQ55300 Blueberry necrotic ring blotch virus
AGI44298 Blueberry necrotic ring blotch virus
QUM93121 Citrus leprosis virus C
QFU28430 Passion fruit green spot virus
ATW76030 Hibiscus-infecting cilevirus
AER13445 Hibiscus green spot virus 2
QPL17815 Pistachio virus X
YP_009270625 Brejeira virus
YP_009351830 Piura virus
YP_009362298 Castlerea virus
YP 009351821 Big Cypress virus
AQM55302 Cordoba virus
QIJ70136 Dougjudy virga-like virus
AMO03254 Boutonnet virus
AVK59469 Hubei virga-like virus 11
QRW42065 Culex pipiens-associated Tunisia virus
AMO03256 Bofa virus
Virgaviridae
QHA33734 Atrato Virga-like virus 1
AMO03220 Blackford virus
YP_009109559 Macrophomina phaseolina tobamo-like virus
Benyviridae
YP 004617078 Rubella virus
BBB44458 Hepatitis E virus type 3
Mayoviridae
Closteroviridae
Kitaviridae
Metonaviridae
Hepeviridae
Tymovirales
0.5
0 2000 4000 6000 8000 10000 12000 14000 16000 18000 20000 22000 nt
RdRp
CP
Mtr
Hel
SP24
RNA1
RNA2
RNA3
● Found in this study
Land plants
Lower plants
Green algae
Invertebrate
Chromista
Fungi
Vertebrate
Environmental

Figure 7. Left: Phylogenetic relationships of the (A) endorna-like and (B) unclassified (+)ssRNA virus identified in this study. ML phylogenetic trees based on the replication protein show the topological position of virus-like sequences discovered in this study in the context of those obtained previously. Right: Genomic organization of the (A) endorna-like and (B) unclassified ssRNA virus sequence identified in this study and representative species used in the phylogeny. For all trees, branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. See Figure 3 for the colour scheme. Viruses discovered in this study are signified using a black circle on the tree tip.

3.3.2 Negative-sense single-stranded RNA ((-)ssRNA) viruses

Bunyavirales

Phenuiviridae. A phenui-like virus sequence termed *Brown algae phenui-like virus* (BralPV) was recovered from a *Sargassum thunbergii* transcriptome. The partial L segment clusters with the unclassified plant and fungi viruses (Figure 8). No additional phenui-like virus segments were recovered. There were no concerning contaminants were detected in the *S. thunbergii* transcriptome (Figure 2).

***Viridisbunyaviridae*.** We identified 16 bunya-like virus sequences from eight liverwort, moss and lycophyte libraries. Three libraries contained multiple distinct putative complete and partial viruses. The overall pairwise nucleotide identity was <70% between each sequence (Figure 8). As such we consider each a different bunya-like viruses. These sequences group together to form a novel clade of unclassified bunya-like viruses distantly related to oomycete, fungi, and invertebrate viruses (Figure 8). Bunyaviruses typically comprise three segments (L, M, and S), although only the L segment was recovered for these sequences. These sequences represent the first plant-associated viruses that cluster near the unofficially named *Deltamycobunyaviridae* (77) (Figure 8). As the complete coding sequences of the viruses discovered share <30% amino acid identity to the nearest relatives in the *Deltamycobunyaviridae,* they may constitute a new virus family. We tentatively name this virus family the *Viridisbunyaviridae*, *(Viridis* meaning green, while bunya is derived from the virus *order Bunyavirales* in which this clade falls within*).* There was no evidence suggesting that these sequences originated from non-plant contaminants. Host assignment was unclear for *Lycopod associated bunyavirus* and *Liverwort associated bunyavirus 1:4* as reads belonging to several lycophyte and liverwort species, respectively, were found in the source transcriptomes (Figure 2).

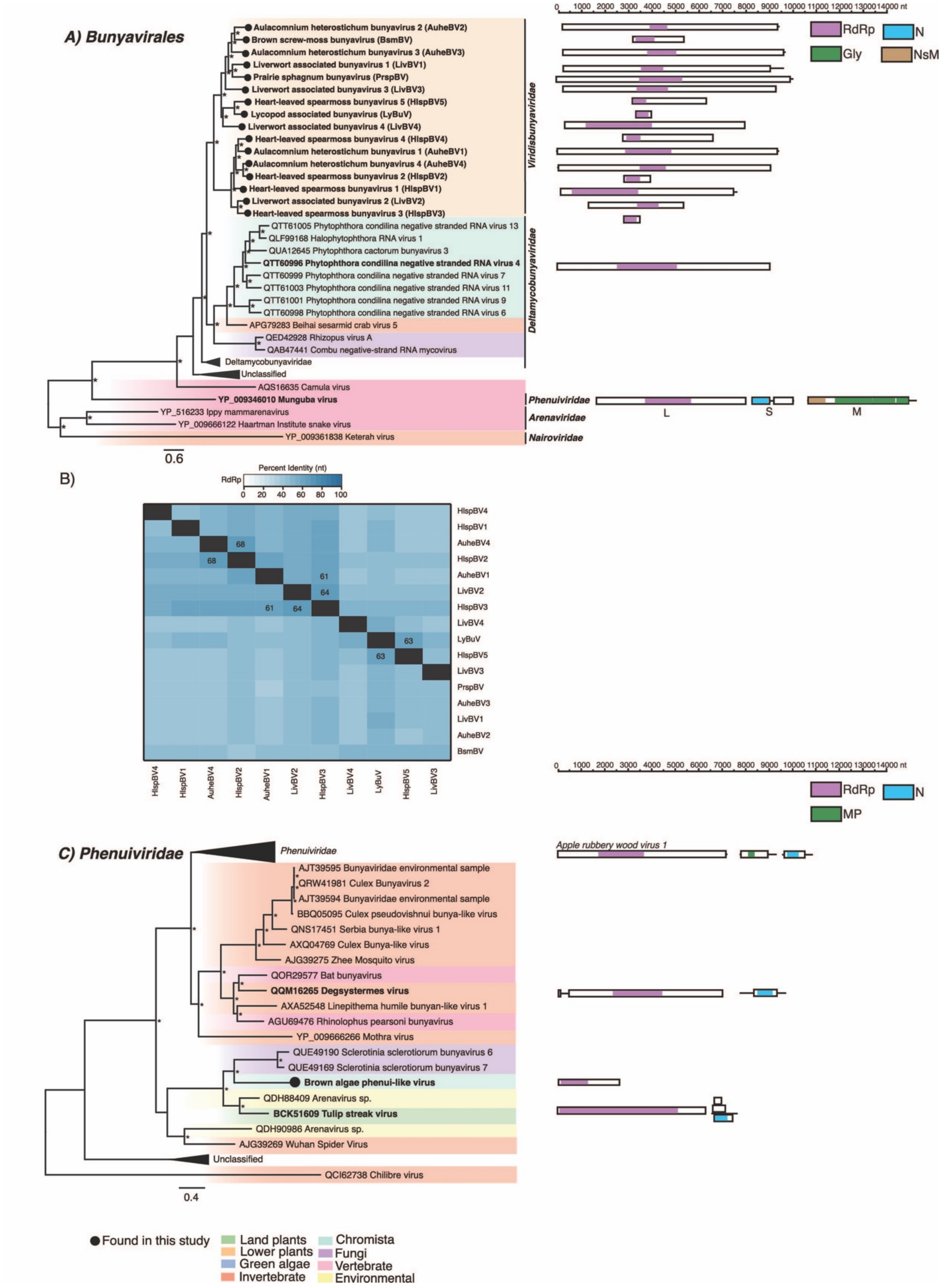

A) Bunyavirales
Aulacomnium heterostichum bunyavirus 2 (AuheBV2)
Brown screw-moss bunyavirus (BsmBV)
Aulacomnium heterostichum bunyavirus 3 (AuheBV3)
Liverwort associated bunyavirus 1 (LivBV1)
Prairie sphagnum bunyavirus (PrspBV)
Liverwort associated bunyavirus 3 (LivBV3)
Heart-leaved spearmoss bunyavirus 5 (HlspBV5)
Lycopod associated bunyavirus (LyBuV)
Liverwort associated bunyavirus 4 (LivBV4)
Heart-leaved spearmoss bunyavirus 4 (HlspBV4)
Aulacomnium heterostichum bunyavirus 1 (AuheBV1)
Aulacomnium heterostichum bunyavirus 4 (AuheBV4)
Heart-leaved spearmoss bunyavirus 2 (HlspBV2)
Heart-leaved spearmoss bunyavirus 1 (HlspBV1)
Liverwort associated bunyavirus 2 (LivBV2)
Heart-leaved spearmoss bunyavirus 3 (HlspBV3)
QTT61005 Phytophthora condilina negative stranded RNA virus 13
QLF99168 Halophytophthora RNA virus 1
QUA12645 Phytophthora cactorum bunyavirus 3
QTT60996 Phytophthora condilina negative stranded RNA virus 4
QTT60999 Phytophthora condilina negative stranded RNA virus 7
QTT61003 Phytophthora condilina negative stranded RNA virus 11
QTT61001 Phytophthora condilina negative stranded RNA virus 9
QTT60998 Phytophthora condilina negative stranded RNA virus 6
APG79283 Beihai sesarmid crab virus 5
QED42928 Rhizopus virus A
QAB47441 Combu negative-strand RNA mycovirus
Deltamycobunyaviridae
Unclassified
AQS16635 Camula virus
YP_009346010 Munguba virus
YP_516233 Ippy mammarenavirus
YP_009666122 Haartman Institute snake virus
YP_009361838 Keterah virus
Viridisbunyaviridae
Deltamycobunyaviridae
Phenuiviridae
Arenaviridae
Nairoviridae
0.6
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 14000 nt
RdRp
N
Gly
NsM
L
S
M
B)
RdRp
Percent Identity (nt)
0 20 40 60 80 100
68
68
61
64
61
64
63
63
HlspBV4
HlspBV1
AuheBV4
HlspBV2
AuheBV1
LivBV2
HlspBV3
LivBV4
LyBuV
HlspBV5
LivBV3
PrspBV
AuheBV3
LivBV1
AuheBV2
BsmBV
HlspBV4
HlspBV1
AuheBV4
HlspBV2
AuheBV1
LivBV2
HlspBV3
LivBV4
LyBuV
HlspBV5
LivBV3
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 14000 nt
RdRp
N
MP
C) Phenuiviridae
Phenuiviridae
Apple rubbery wood virus 1
AJT39595 Bunyaviridae environmental sample
QRW41981 Culex Bunyavirus 2
AJT39594 Bunyaviridae environmental sample
BBQ05095 Culex pseudovishnui bunya-like virus
QNS17451 Serbia bunya-like virus 1
AXQ04769 Culex Bunya-like virus
AJG39275 Zhee Mosquito virus
QOR29577 Bat bunyavirus
QQM16265 Degsystermes virus
AXA52548 Linepithema humile bunyan-like virus 1
AGU69476 Rhinolophus pearsoni bunyavirus
YP_009666266 Mothra virus
QUE49190 Sclerotinia sclerotiorum bunyavirus 6
QUE49169 Sclerotinia sclerotiorum bunyavirus 7
Brown algae phenui-like virus
QDH88409 Arenavirus sp.
BCK51609 Tulip streak virus
QDH90986 Arenavirus sp.
AJG39269 Wuhan Spider Virus
Unclassified
QCI62738 Chilibre virus
0.4
● Found in this study
Land plants
Lower plants
Green algae
Invertebrate
Chromista
Fungi
Vertebrate
Environmental

Figure 8. Phylogenetic relationships of the viruses (A) Left: A phylogeny depicting a novel clade of viruses related to the *Deltamycobunyaviridae* in the context of the *Bunyavirales*. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny (B) Percent identity matrix of the novel bunya-like viruses. Identity scores are calculated from an alignment of the RdRp protein coding sequence. For clarity, the 100% identity along the diagonal has been removed. Where sequence identity is >= 60% the value is shown. (C) Left: A phylogeny depicting the phenui-like virus identified in this study in the context of the *Phenuiviridae*. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny. For all trees, branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. See Figure 3 for the colour scheme. Viruses discovered in this study are signified using a black circle on the tree tip.

Mononegavirales

Rhabdoviridae. We identified seven sequences that clustered with plant viruses in the family *Rhabdoviridae* denoted *Canadian violet rhabdovirus 1* (CvRV1), Canadian violet rhabdovirus 2 (CvRV2), *Common ivy rhabdovirus* (CoiRV) and *Indian pipe rhabdovirus* (InpRV), *Tree fern varicosa-like virus* (TfVV), *Monoclea gottschei varicosa-like virus* (MgVV) and *Bug moss associated rhabdo-like virus* (BmRV). Notably, TfVV and MgVV expand the host range of the rhabdoviruses from angiosperms and gymnosperm to ferns and liverworts. RNA2 segments were recovered for both viruses, TfVV RNA2 contained five genes while MgVV contained four (Figure 9C). Two partial segments sharing similarities to the nucleocapsid (N) of *Black grass varicosavirus-like virus* (YP_009130620.1) were found in the Indian pipe library and share 50%

amino acid identity. All sequences likely represent novel species within known plant infecting genera (Figure 9C).

BmRV is a partial sequence (693 nt) most closely related to the unclassified *Hubei rhabdo-like virus 2* (44% amino acid identity). Further evidence is needed to confirm BmRV as the first moss rhabdovirus, but the relatively low proportion of contaminates in this library (3% algae and 3% fungi) suggests that this virus is plant-associated (Figure 2). While 53% of reads in the MgVV library were fungi associated the phylogenetic position of MgVV suggests it is derived from plants (Figure 2, Figure 9C).

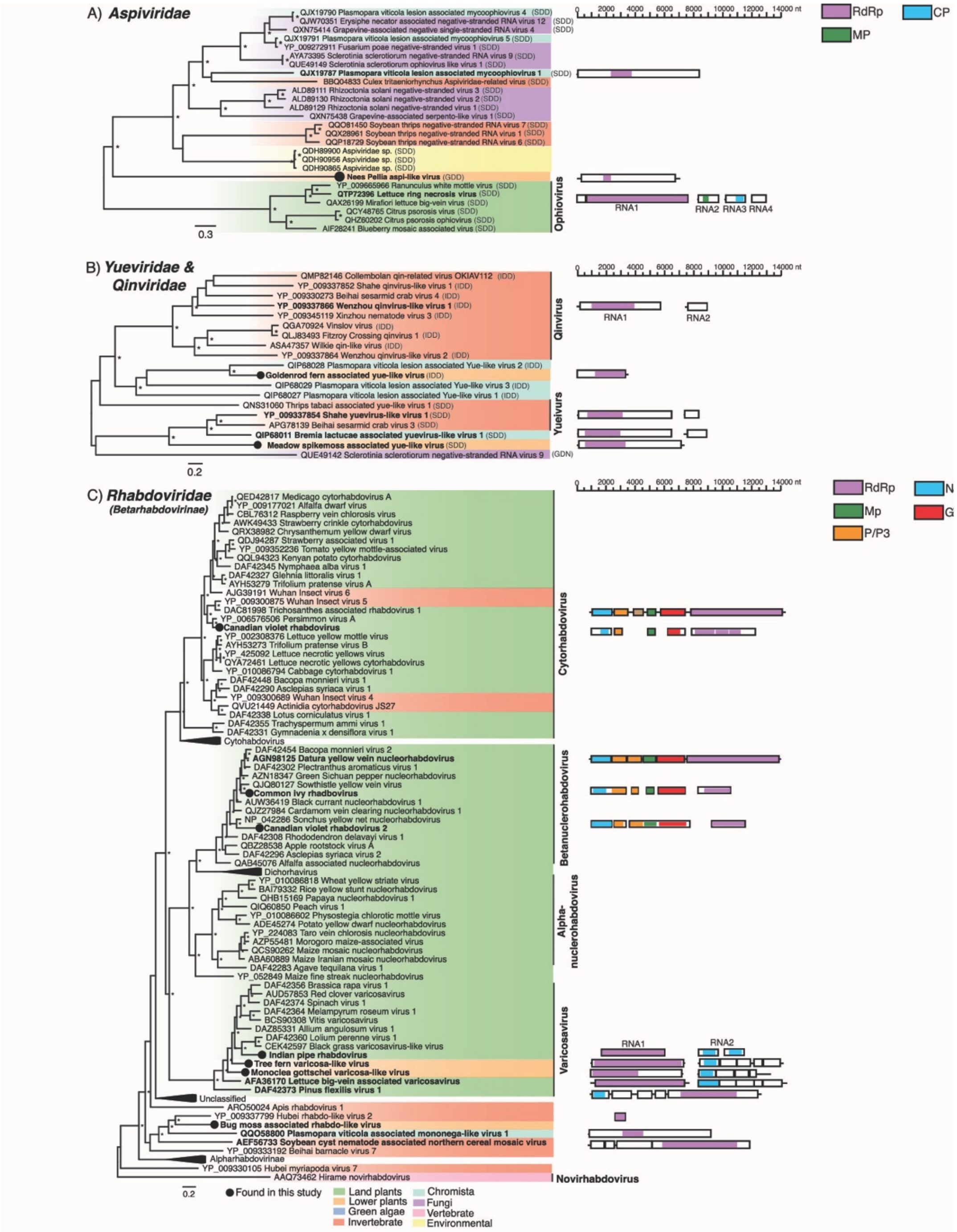

A) Aspiviridae
QJX19790 Plasmopara viticola lesion associated mycoophiovirus 4 (SDD)
QJW70351 Erysiphe necator associated negative-stranded RNA virus 12 (SDD)
QXN75414 Grapevine-associated negative single-stranded RNA virus 4 (SDD)
QJX19791 Plasmopara viticola lesion associated mycoophiovirus 5 (SDD)
YP_009272911 Fusarium poae negative-stranded virus 1 (SDD)
AYA73395 Sclerotinia sclerotiorum negative-stranded RNA virus 9 (SDD)
QUE49149 Sclerotinia sclerotiorum ophiovirus like virus 1 (SDD)
QJX19787 Plasmopara viticola lesion associated mycoophiovirus 1 (SDD)
BBQ04833 Culex tritaeniorhynchus Aspiviridae-related virus (SDD)
ALD89111 Rhizoctonia solani negative-stranded virus 3 (SDD)
ALD89130 Rhizoctonia solani negative-stranded virus 2 (SDD)
ALD89129 Rhizoctonia solani negative-stranded virus 1 (SDD)
QXN75438 Grapevine-associated serpento-like virus 1 (SDD)
QQO81450 Soybean thrips negative-stranded RNA virus 7 (SDD)
QQX28961 Soybean thrips negative-stranded RNA virus 1 (SDD)
QQP18729 Soybean thrips negative-stranded RNA virus 6 (SDD)
QDH89900 Aspiviridae sp. (SDD)
QDH90956 Aspiviridae sp. (SDD)
QDH90865 Aspiviridae sp. (SDD)
Nees Pellia aspi-like virus (GDD)
YP_009665966 Ranunculus white mottle virus (SDD)
QTP72396 Lettuce ring necrosis virus (SDD)
QAX26199 Mirafiori lettuce big-vein virus (SDD)
QCY48765 Citrus psorosis virus (SDD)
QHZ60202 Citrus psorosis ophiovirus (SDD)
AIF28241 Blueberry mosaic associated virus (SDD)
Ophiovirus
0.3
0 2000 4000 6000 8000 10000 12000 14000 nt
RdRp
CP
MP
RNA1
RNA2 RNA3 RNA4
B) Yueviridae & Qinviridae
QMP82146 Collembolan qin-related virus OKIAV112 (IDD)
YP_009337852 Shahe qinvirus-like virus 1 (IDD)
YP_009330273 Beihai sesarmid crab virus 4 (IDD)
YP_009337866 Wenzhou qinvirus-like virus 1 (IDD)
YP_009345119 Xinzhou nematode virus 3 (IDD)
QGA70924 Vinslov virus (IDD)
QLJ83493 Fitzroy Crossing qinvirus 1 (IDD)
ASA47357 Wilkie qin-like virus (IDD)
YP_009337864 Wenzhou qinvirus-like virus 2 (IDD)
QIP68028 Plasmopara viticola lesion associated Yue-like virus 2 (IDD)
Goldenrod fern associated yue-like virus (IDD)
QIP68029 Plasmopara viticola lesion associated Yue-like virus 3 (IDD)
QIP68027 Plasmopara viticola lesion associated Yue-like virus 1 (IDD)
QNS31060 Thrips tabaci associated yue-like virus 1 (SDD)
YP_009337854 Shahe yuevirus-like virus 1 (SDD)
APG78139 Beihai sesarmid crab virus 3 (SDD)
QIP68011 Bremia lactucae associated yuevirus-like virus 1 (SDD)
Meadow spikemoss associated yue-like virus (SDD)
QUE49142 Sclerotinia sclerotiorum negative-stranded RNA virus 9 (GDN)
Qinvirus
Yueivurs
0.2
0 2000 4000 6000 8000 10000 12000 14000 nt
RNA1
RNA2
C) Rhabdoviridae (Betarhabdovirinae)
QED42817 Medicago cytorhabdovirus A
YP_009177021 Alfalfa dwarf virus
CBL76312 Raspberry vein chlorosis virus
AWK49433 Strawberry crinkle cytorhabdovirus
QRX38982 Chrysanthemum yellow dwarf virus
QDJ94287 Strawberry associated virus 1
YP_009352236 Tomato yellow mottle-associated virus
QQL94323 Kenyan potato cytorhabdovirus
DAF42345 Nymphaea alba virus 1
DAF42327 Glehnia littoralis virus 1
AYH53279 Trifolium pratense virus A
AJG39191 Wuhan Insect virus 6
YP_009300875 Wuhan Insect virus 5
DAC81998 Trichosanthes associated rhabdovirus 1
YP_006576506 Persimmon virus A
Canadian violet rhabdovirus
YP_002308376 Lettuce yellow mottle virus
AYH53273 Trifolium pratense virus B
YP_425092 Lettuce necrotic yellows virus
QYA72461 Lettuce necrotic yellows cytorhabdovirus
YP_010086794 Cabbage cytorhabdovirus 1
DAF42448 Bacopa monnieri virus 1
DAF42290 Asclepias syriaca virus 1
YP_009300689 Wuhan Insect virus 4
QVU21449 Actinidia cytorhabdovirus JS27
DAF42338 Lotus corniculatus virus 1
DAF42355 Trachyspermum ammi virus 1
DAF42331 Gymnadenia x densiflora virus 1
Cytohabdovirus
DAF42454 Bacopa monnieri virus 2
AGN98125 Datura yellow vein nucleorhabdovirus
DAF42302 Plectranthus aromaticus virus 1
AZN18347 Green Sichuan pepper nucleorhabdovirus
QJQ80127 Sowthistle yellow vein virus
Common ivy rhadbovirus
AUW36419 Black currant nucleorhabdovirus 1
QJZ27984 Cardamom vein clearing nucleorhabdovirus 1
NP_042286 Sonchus yellow net nucleorhabdovirus
Canadian violet rhabdovirus 2
DAF42308 Rhododendron delavayi virus 1
QBZ28538 Apple rootstock virus A
DAF42296 Asclepias syriaca virus 2
QAB45076 Alfalfa associated nucleorhabdovirus
Dichorhavirus
YP_010086818 Wheat yellow striate virus
BAI79332 Rice yellow stunt nucleorhabdovirus
QHB15169 Papaya nucleorhabdovirus 1
QIQ60850 Peach virus 1
YP_010086602 Physostegia chlorotic mottle virus
ADE45274 Potato yellow dwarf nucleorhabdovirus
YP_224083 Taro vein chlorosis nucleorhabdovirus
AZP55481 Morogoro maize-associated virus
QCS90262 Maize mosaic nucleorhabdovirus
ABA60889 Maize Iranian mosaic nucleorhabdovirus
DAF42283 Agave tequilana virus 1
YP_052849 Maize fine streak nucleorhabdovirus
DAF42356 Brassica rapa virus 1
AUD57853 Red clover varicosavirus
DAF42374 Spinach virus 1
DAF42364 Melampyrum roseum virus 1
BCS90308 Vitis varicosavirus
DAZ85331 Allium angulosum virus 1
DAF42360 Lolium perenne virus 1
CEK42597 Black grass varicosavirus-like virus
Indian pipe rhabdovirus
Tree fern varicosa-like virus
Monoclea gottschei varicosa-like virus
AFA36170 Lettuce big-vein associated varicosavirus
DAF42373 Pinus flexilis virus 1
Unclassified
ARO50024 Apis rhabdovirus 1
YP_009337799 Hubei rhabdo-like virus 2
Bug moss associated rhabdo-like virus
QQO58800 Plasmopara viticola associated mononega-like virus 1
AEF56733 Soybean cyst nematode associated northern cereal mosaic virus
YP_009333192 Beihai barnacle virus 7
Alpharhabdovirinae
YP_009330105 Hubei myriapoda virus 7
AAQ73462 Hirame novirhabdovirus
Cytorhabdovirus
Betanucleorhabdovirus
Alpha-nucleorhabdovirus
Varicosavirus
Novirhabdovirus
0.2
0 2000 4000 6000 8000 10000 12000 14000 nt
RdRp
N
Mp
G
P/P3
RNA1
RNA2
● Found in this study
Land plants
Lower plants
Green algae
Invertebrate
Chromista
Fungi
Vertebrate
Environmental

Figure 9. Left: Phylogenetic relationships of the viruses within the families, (A) *Aspiviridae*, (B) *Yue*- and *Qinviridae* and (C) *Rhabdoviridae.* Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny. For all trees, branches are scaled to the number of amino acid substitutions per site and trees were

mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. See Figure 3 for the colour scheme. Viruses discovered in this study are signified using a black circle on the tree tip. For trees (A) and (B), the RdRp motif C trimer of each sequence is shown in brackets at the end of the tip label.

Serpentovirales

Aspiviridae. We identified an aspi-like sequence termed *Nees' Pellia aspi-like virus* (NpAV). A complete RNA1 segment (6989 nt) was assembled, although no other segments were recovered (Figure 9A). NpAV most closely resembles *Rhizoctonia solani negative-stranded virus 3* (amino acid identity: 22%) and falls basal to all the unclassified aspi-like viruses including those found in fungi, invertebrates, and oomycetes. NpAV is the first aspi-like virus identified in plants outside of the angiosperms and may constitute a novel virus group (Figure 9A). Notably, unlike the other aspiviruses that possess a SDD sequence in motif C of the RdRp – a known signature for segmented negative-stranded RNA viruses – NpAV has a GDD sequence (Figure 9A).

Goujianvirales

Yueviridae. An yue-like virus sequence termed *Meadow spikemoss associated yue-like virus* (MsYV) was found in the lycophyte *Selaginella apoda* and most closely resembles algae associated *Bremia lactucae associated yuevirus-like virus 1* (amino acid identity: 26%). Phylogenetic analysis supports the assignment of MsYV as the first plant yuevirus (Figure 9B)

A second partial yue-like virus sequence was detected in a *Pityrogramma trifoliata* library and termed *Goldenrod fern associated yue-like virus* (GfYV). GfYV falls with a group of oomycete associated viruses. Consistent with the qin-like viruses, GfYV has an IDD (Ile-Asp-Asp) sequence motif instead of the common GDD (Gly-Asp-Asp) in the catalytic core of its RdRp, while MsYV contains SDD (Ser-Asp-Asp) in the same manner as many yue-like viruses (Figure 9B). The libraries from which GfYV and MsYV were assembled are contaminated with fungal reads (5% and 21%, respectively) as such host assignment is made with caution (Figure 2). Reads belonging to oomycetes were not found in either library.

3.3.3 Double-stranded RNA (dsRNA) viruses

Durnavirales

Amalgaviridae. We detected five sequences that cluster with amalga-like viruses. *Lycopod associated amalgavirus* (LycoAV) is a partial RdRp containing a sequence that falls basal to the *Amalgaviridae* and represents the first amalga-like virus in the lycophytes (Figure 10). Three amalga-like sequences were discovered in green and red algae transcriptomes and cluster with *Diatom colony associated dsRNA virus 2* (Figure 10). As noted in the case of *Bryopsis mitochondria-associated dsRNA virus* and several green algae associated viruses (78) when translated into amino acids using the protozoan mitochondrial code, two overlapping ORFs are present: the first, encoding a hypothetical protein, while the second, a replicase through a –1 ribosomal frameshift (79). For two of the amalga-like sequences identified in this study – *Nucleotaenium eifelense virus* (NueiV) and *Rhodella violacea virus* (RhviV) – a similar structure was observed but we were unable to identify any ribosomal frameshift motifs in either sequence (Figure 10). Further work is needed to confirm if these sequences should be translated through the mitochondrial genetic code.

A contig containing what appears to be a complete coding sequence (3259 nt) and RdRp motifs was assembled in the *Woodsia scopulina* transcriptome and tentatively named *Rocky Mountain woodsia associated virus* (RmwPV). The predicted RdRp region (918-1921 nt) of RmvPV shares similarity to both partiti-like viruses (e.g., *Ustilaginoidea virens nonsegmented virus 2*, 26% aa identity) and the unclassified *Phytophthora infestans RNA virus 1* (42% aa identity) which has been shown to likely constitutes a novel virus family (80). The resemblance RmwPV shares with two seemingly distantly related virus groups suggest its position within the *Durnavirales* should be treated with caution (Figure 10).

The transcriptome in which RmwPV was discovered is contaminated with fungal reads (10% (Figure 2). If RmwPV was derived from fungal contaminates this could potentially explain the phylogenetic placement of RmwPV (Figure 10). The *Lycopodiella appressa* transcriptome in which LycoAV was discovered is contaminated by reads belonging to species across various land plant groups. Reads belonging to land plants comprised 35% of plant-associated reads while lycopod associated reads comprised 65% (Figure 2).

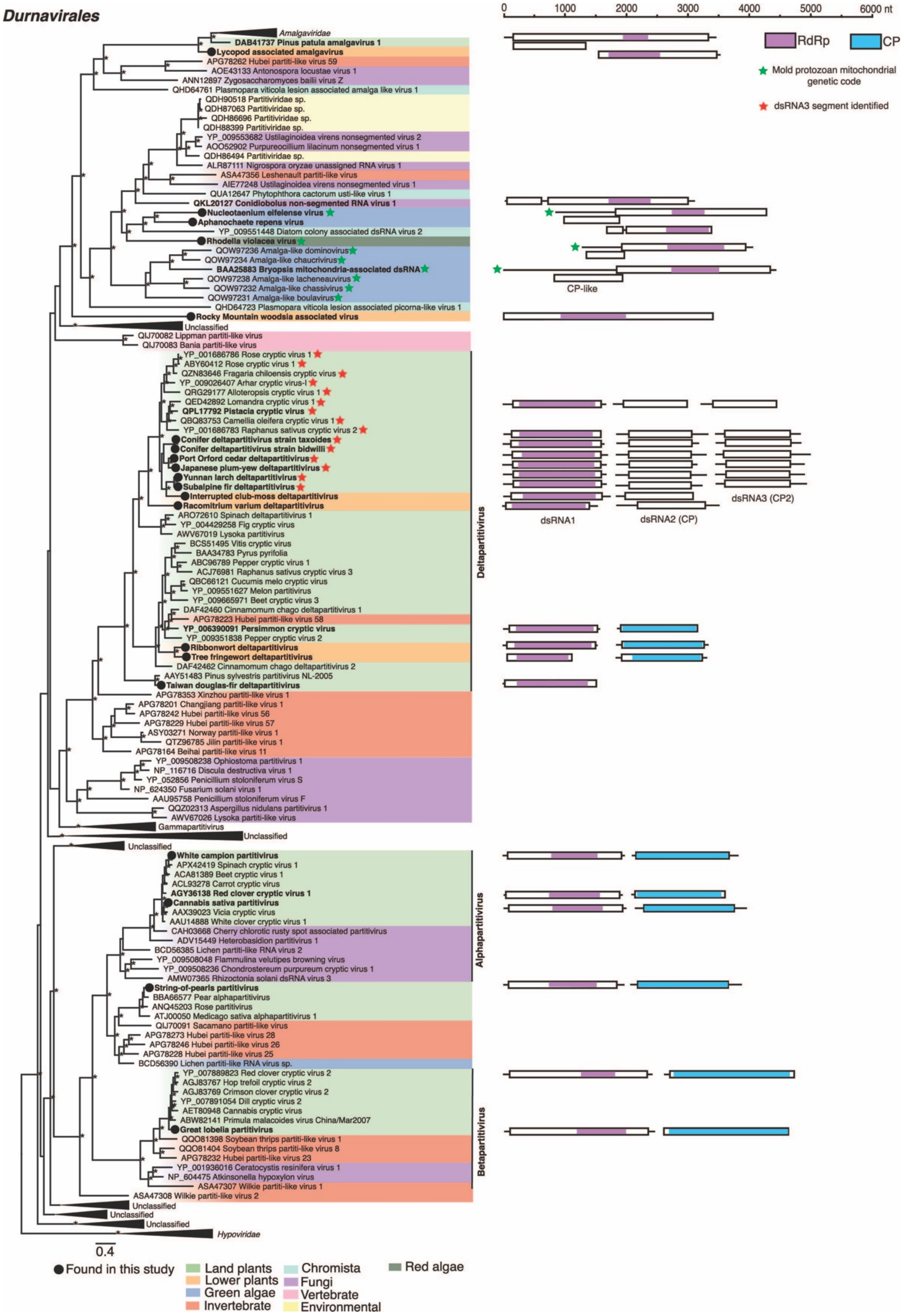

Durnavirales
0
1000
2000
3000
4000
5000
6000 nt
RdRp
CP
★ Mold protozoan mitochondrial genetic code
★ dsRNA3 segment identified
Amalgaviridae
DAB41737 Pinus patula amalgavirus 1
● Lycopod associated amalgavirus
APG78262 Hubei partiti-like virus 59
AOE43133 Antonospora locustae virus 1
ANN12897 Zygosaccharomyces bailii virus Z
QHD64761 Plasmopara viticola lesion associated amalga like virus 1
QDH90518 Partitiviridae sp.
QDH87063 Partitiviridae sp.
QDH86696 Partitiviridae sp.
QDH88399 Partitiviridae sp.
YP_009553682 Ustilaginoidea virens nonsegmented virus 2
AOO52902 Purpureocillium lilacinum nonsegmented virus 1
QDH86494 Partitiviridae sp.
ALR87111 Nigrospora oryzae unassigned RNA virus 1
ASA47356 Leshenault partiti-like virus
AIE77248 Ustilaginoidea virens nonsegmented virus 1
QUA12647 Phytophthora cactorum usti-like virus 1
QKL20127 Conidiobolus non-segmented RNA virus 1
● Nucleotaenium eifelense virus ★
● Aphanochaete repens virus
YP_009551448 Diatom colony associated dsRNA virus 2
● Rhodella violacea virus ★
QOW97236 Amalga-like dominovirus ★
QOW97234 Amalga-like chaucrivirus ★
BAA25883 Bryopsis mitochondria-associated dsRNA ★
QOW97238 Amalga-like lacheneauvirus ★
QOW97232 Amalga-like chassivirus ★
QOW97231 Amalga-like boulavirus ★
QHD64723 Plasmopara viticola lesion associated picorna-like virus 1
● Rocky Mountain woodsia associated virus
Unclassified
CP-like
QIJ70082 Lippman partiti-like virus
QIJ70083 Bania partiti-like virus
YP_001686786 Rose cryptic virus 1 ★
ABY60412 Rose cryptic virus 1 ★
QZN83646 Fragaria chiloensis cryptic virus ★
YP_009026407 Arhar cryptic virus-I ★
QRG29177 Alloteropsis cryptic virus 1 ★
QED42892 Lomandra cryptic virus 1 ★
QPL17792 Pistacia cryptic virus ★
QBQ83753 Camellia oleifera cryptic virus 1 ★
YP_001686783 Raphanus sativus cryptic virus 2 ★
● Conifer deltapartitivirus strain taxoides ★
● Conifer deltapartitivirus strain bidwillii ★
● Port Orford cedar deltapartitivirus ★
● Japanese plum-yew deltapartitivirus ★
● Yunnan larch deltapartitivirus ★
● Subalpine fir deltapartitivirus ★
● Interrupted club-moss deltapartitivirus
● Racomitrium varium deltapartitivirus
ARO72610 Spinach deltapartitivirus 1
YP_004429258 Fig cryptic virus
AWV67019 Lysoka partitivirus
BCS51495 Vitis cryptic virus
BAA34783 Pyrus pyrifolia
ABC96789 Pepper cryptic virus 1
ACJ76981 Raphanus sativus cryptic virus 3
QBC66121 Cucumis melo cryptic virus
YP_009551627 Melon partitivirus
YP_009665971 Beet cryptic virus 3
DAF42460 Cinnamomum chago deltapartitivirus 1
APG78223 Hubei partiti-like virus 58
YP_006390091 Persimmon cryptic virus
YP_009351838 Pepper cryptic virus 2
● Ribbonwort deltapartitivirus
● Tree fringewort deltapartitivirus
DAF42462 Cinnamomum chago deltapartitivirus 2
AAY51483 Pinus sylvestris partitivirus NL-2005
● Taiwan douglas-fir deltapartitivirus
Deltapartitivirus
dsRNA1
dsRNA2 (CP)
dsRNA3 (CP2)
APG78353 Xinzhou partiti-like virus 1
APG78201 Changjiang partiti-like virus 1
APG78242 Hubei partiti-like virus 56
APG78229 Hubei partiti-like virus 57
ASY03271 Norway partiti-like virus 1
QTZ96785 Jilin partiti-like virus 1
APG78164 Beihai partiti-like virus 11
YP_009508238 Ophiostoma partitivirus 1
NP_116716 Discula destructiva virus 1
YP_052856 Penicillium stoloniferum virus S
NP_624350 Fusarium solani virus 1
AAU95758 Penicillium stoloniferum virus F
QQZ02313 Aspergillus nidulans partitivirus 1
AWV67026 Lysoka partiti-like virus
Gammapartitivirus
Unclassified
Unclassified
● White campion partitivirus
APX42419 Spinach cryptic virus 1
ACA81389 Beet cryptic virus 1
ACL93278 Carrot cryptic virus
AGY36138 Red clover cryptic virus 1
● Cannabis sativa partitivirus
AAX39023 Vicia cryptic virus
AAU14888 White clover cryptic virus 1
CAH03668 Cherry chlorotic rusty spot associated partitivirus
ADV15449 Heterobasidion partitivirus 1
BCD56385 Lichen partiti-like RNA virus 2
YP_009508048 Flammulina velutipes browning virus
YP_009508236 Chondrostereum purpureum cryptic virus 1
AMW07365 Rhizoctonia solani dsRNA virus 3
Alphapartitivirus
● String-of-pearls partitivirus
BBA66577 Pear alphapartitivirus
ANQ45203 Rose partitivirus
ATJ00050 Medicago sativa alphapartitivirus 1
QIJ70091 Sacamano partiti-like virus
APG78273 Hubei partiti-like virus 28
APG78246 Hubei partiti-like virus 26
APG78228 Hubei partiti-like virus 25
BCD56390 Lichen partiti-like RNA virus sp.
YP_007889823 Red clover cryptic virus 2
AGJ83767 Hop trefoil cryptic virus 2
AGJ83769 Crimson clover cryptic virus 2
YP_007891054 Dill cryptic virus 2
AET80948 Cannabis cryptic virus
ABW82141 Primula malacoides virus China/Mar2007
● Great lobelia partitivirus
QQO81398 Soybean thrips partiti-like virus 1
QQO81404 Soybean thrips partiti-like virus 8
APG78232 Hubei partiti-like virus 23
YP_001936016 Ceratocystis resinifera virus 1
NP_604475 Atkinsonella hypoxylon virus
ASA47307 Wilkie partiti-like virus 1
ASA47308 Wilkie partiti-like virus 2
Betapartitivirus
Unclassified
Unclassified
Unclassified
Hypoviridae
0.4
● Found in this study
🟩 Land plants
🟧 Lower plants
🟦 Green algae
🟥 Invertebrate
▨ Chromista
🟪 Fungi
▨ Vertebrate
🟨 Environmental
▨ Red algae

Figure 10. Left: Phylogenetic relationships of the viruses within the order *Durnavirales*. ML phylogenetic trees based on the replication protein show the topological position of the virus-like sequences discovered in this study (black circles) in the context of their closest relatives. See Figure 3 for the colour scheme. Green stars are used to signify sequences that have been translated using the protozoan mitochondrial genetic code. Red stars are used to signify sequences for which a dsRNA3 coat protein-like segment has been described. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Tip labels are bolded when the genome structure is shown on the right. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny.

Partitiviridae. We detected 14 sequences that share a resemblance with members of the *Partitiviridae*. For each of these sequences, complete dsRNA1 and dsRNA2 segments were recovered. Ten sequences were found in non-flowering plants and cluster within the deltapartitiviruses. A clade within the deltpartitiviruses is known to encode a third segment comprising of a divergent dsRNA2 full-length capsid protein with unknown function (Figure 10). We identified dsRNA3 segments in related conifer associated sequences but not in those found in moss and lycophyte libraries (Figure 10). Phylogenies estimated on the coding sequences of dsRNA2 and dsRNA3 reveal essentially the same grouping which is largely consistent with the host phylogeny (SI Figure 4). We extend the known host range of the deltapartitiviruses to include liverworts, mosses, and lycophytes. The remaining sequences were found in eudicots and cluster with known plant partitiviruses (Figure 10). The white campion was judged to have contamination from ginseng and chickweed (31). However, the relatively low proportion of the library these contaminates compose (<1%) suggests that it is unlikely these species are the host of WcPV (Figure 2). There is no evidence that the other partiti-like sequences discovered are derived from contaminates.

Ghabrivirales

Chrysoviridae. We identified two partial sequences that share resemblance with members of the alphachrysoviruses denoted *Mesotaenium kramstae alphachyrso-like virus* (MkACV) and *Tree fringewort alphachyrso-like virus* (TwACV) (Figure 11). A complete RNA2 segment was recovered for MkACV which shared similarity with the p98 of various chrysoviruses (Figure 11). The MkACV RNA2 segment did not contain the “PGDGXCXXHX” motif commonly found in this protein (81). To our knowledge, these sequences represent the first chyrsoviruses in liverworts and algae. While reads belonging to fungi were found in the libraries MkACV and TwACV were assembled from, the phylogenetic positioning of the viruses suggest that they are plant-derived (Figure 2, Figure 11).

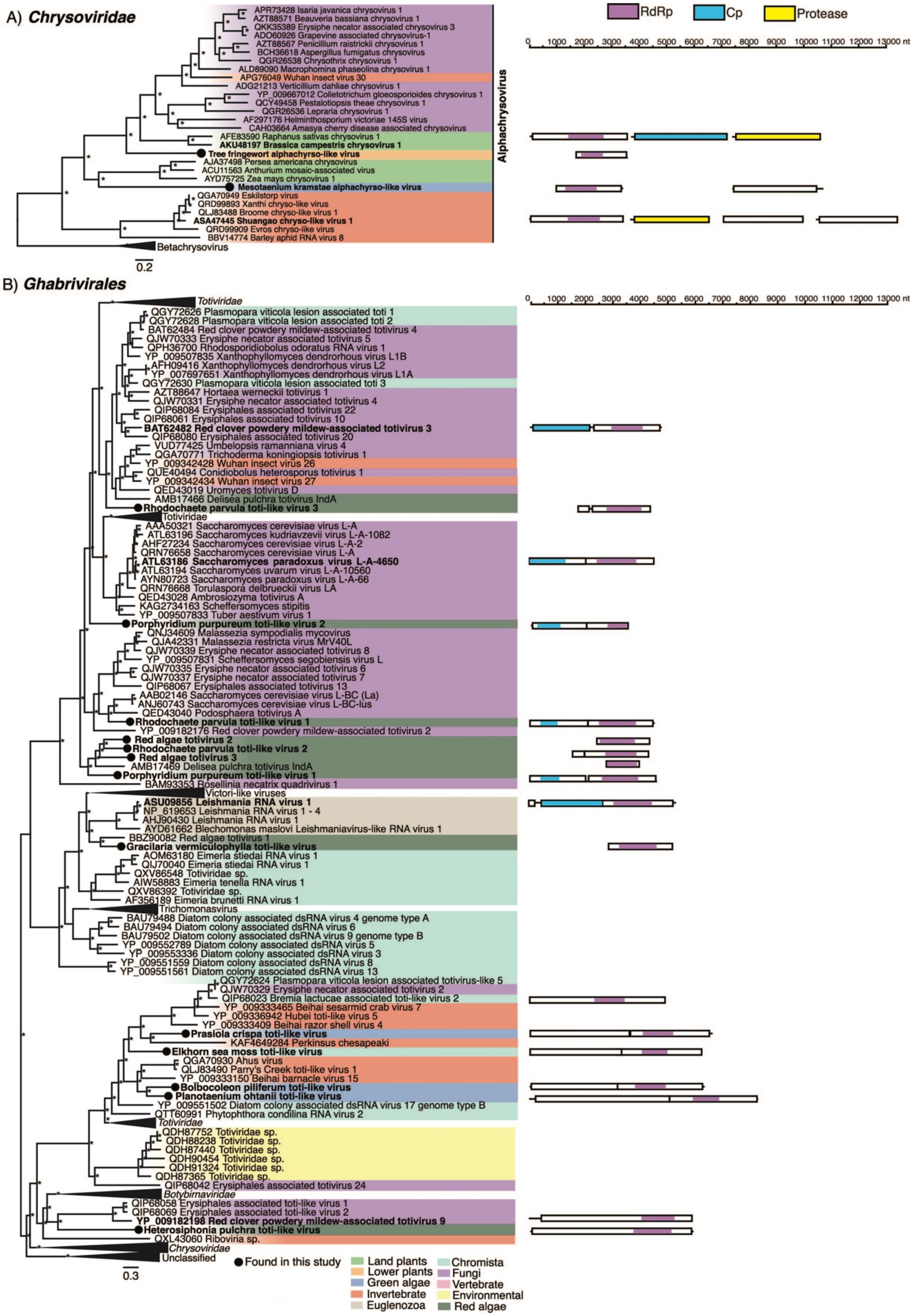

A) Chrysoviridae
RdRp
Cp
Protease
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 nt
APR73428 Isaria javanica chrysovirus 1
AZT88571 Beauveria bassiana chrysovirus 1
QKK35389 Erysiphe necator associated chrysovirus 3
ADO60926 Grapevine associated chrysovirus-1
AZT88567 Penicillium raistrickii chrysovirus 1
BCH36618 Aspergillus fumigatus chrysovirus
QGR26538 Chrysothrix chrysovirus 1
ALD89090 Macrophomina phaseolina chrysovirus 1
APG76049 Wuhan insect virus 30
ADG21213 Verticillium dahliae chrysovirus 1
YP_009667012 Colletotrichum gloeosporioides chrysovirus 1
QCY49458 Pestalotiopsis theae chrysovirus 1
QGR26536 Lepraria chrysovirus 1
AF297176 Helminthosporium victoriae 145S virus
CAH03664 Amasya cherry disease associated chrysovirus
AFE83590 Raphanus sativas chrysovirus 1
AKU48197 Brassica campestris chrysovirus 1
Tree fringewort alphachyrso-like virus
AJA37498 Persea americana chrysovirus
ACU11563 Anthurium mosaic-associated virus
AYD75725 Zea mays chrysovirus 1
Mesotaenium kramstae alphachyrso-like virus
QGA70949 Eskilstorp virus
QRD99893 Xanthi chryso-like virus
QLJ83488 Broome chryso-like virus 1
ASA47445 Shuangao chryso-like virus 1
QRD99909 Evros chryso-like virus
BBV14774 Barley aphid RNA virus 8
Betachrysovirus
Alphachrysovirus
0.2
B) Ghabrivirales
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 nt
Totiviridae
QGY72626 Plasmopara viticola lesion associated toti 1
QGY72628 Plasmopara viticola lesion associated toti 2
BAT62484 Red clover powdery mildew-associated totivirus 4
QJW70333 Erysiphe necator associated totivirus 5
QPH36700 Rhodosporidiobolus odoratus RNA virus 1
YP_009507835 Xanthophyllomyces dendrorhous virus L1B
AFH09416 Xanthophyllomyces dendrorhous virus L2
YP_007697651 Xanthophyllomyces dendrorhous virus L1A
QGY72630 Plasmopara viticola lesion associated toti 3
AZT88647 Hortaea werneckii totivirus 1
QJW70331 Erysiphe necator associated totivirus 4
QIP68084 Erysiphales associated totivirus 22
QIP68061 Erysiphales associated totivirus 10
BAT62482 Red clover powdery mildew-associated totivirus 3
QIP68080 Erysiphales associated totivirus 20
VUD77425 Umbelopsis ramanniana virus 4
QGA70771 Trichoderma koningiopsis totivirus 1
YP_009342428 Wuhan insect virus 26
QUE40494 Conidiobolus heterosporus totivirus 1
YP_009342434 Wuhan insect virus 27
QED43019 Uromyces totivirus D
AMB17466 Delisea pulchra totivirus IndA
Rhodochaete parvula toti-like virus 3
Totiviridae
AAA50321 Saccharomyces cerevisiae virus L-A
ATL63196 Saccharomyces kudriavzevii virus L-A-1082
AHF27234 Saccharomyces cerevisiae virus L-A-2
QRN76658 Saccharomyces cerevisiae virus L-A
ATL63186 Saccharomyces paradoxus virus L-A-4650
ATL63194 Saccharomyces uvarum virus L-A-10560
AYN80723 Saccharomyces paradoxus virus L-A-66
QRN76668 Torulaspora delbrueckii virus LA
QED43028 Ambrosiozyma totivirus A
KAG2734163 Scheffersomyces stipitis
YP_009507833 Tuber aestivum virus 1
Porphyridium purpureum toti-like virus 2
QNJ34609 Malassezia sympodialis mycovirus
QJA42331 Malassezia restricta virus MrV40L
QJW70339 Erysiphe necator associated totivirus 8
YP_009507831 Scheffersomyces segobiensis virus L
QJW70335 Erysiphe necator associated totivirus 6
QJW70337 Erysiphe necator associated totivirus 7
QIP68067 Erysiphales associated totivirus 13
AAB02146 Saccharomyces cerevisiae virus L-BC (La)
ANJ60743 Saccharomyces cerevisiae virus L-BC-lus
QED43040 Podosphaera totivirus A
Rhodochaete parvula toti-like virus 1
YP_009182176 Red clover powdery mildew-associated totivirus 2
Red algae totivirus 2
Rhodochaete parvula toti-like virus 2
Red algae totivirus 3
AMB17469 Delisea pulchra totivirus IndA
Porphyridium purpureum toti-like virus 1
BAM93353 Rosellinia necatrix quadrivirus 1
Victori-like viruses
ASU09856 Leishmania RNA virus 1
NP_619653 Leishmania RNA virus 1 - 4
AHJ90430 Leishmania RNA virus 1
AYD61662 Blechomonas maslovi Leishmaniavirus-like RNA virus 1
BBZ90082 Red algae totivirus 1
Gracilaria vermiculophylla toti-like virus
AOM63180 Eimeria stiedai RNA virus 1
QIJ70040 Eimeria stiedai RNA virus 1
QXV86548 Totiviridae sp.
AIW58883 Eimeria tenella RNA virus 1
QXV86392 Totiviridae sp.
AF356189 Eimeria brunetti RNA virus 1
Trichomonasvirus
BAU79488 Diatom colony associated dsRNA virus 4 genome type A
BAU79494 Diatom colony associated dsRNA virus 6
BAU79502 Diatom colony associated dsRNA virus 9 genome type B
YP_009552789 Diatom colony associated dsRNA virus 5
YP_009553336 Diatom colony associated dsRNA virus 3
YP_009551559 Diatom colony associated dsRNA virus 8
YP_009551561 Diatom colony associated dsRNA virus 13
QGY72624 Plasmopara viticola lesion associated totivirus-like 5
QJW70329 Erysiphe necator associated totivirus 2
QIP68023 Bremia lactucae associated toti-like virus 2
YP_009333465 Beihai sesarmid crab virus 7
YP_009336942 Hubei toti-like virus 5
YP_009333409 Beihai razor shell virus 4
Prasiola crispa toti-like virus
KAF4649284 Perkinsus chesapeaki
Elkhorn sea moss toti-like virus
QGA70930 Ahus virus
QLJ83490 Parry's Creek toti-like virus 1
YP_009333150 Beihai barnacle virus 15
Bolbocoleon piliferum toti-like virus
Planotaenium ohtanii toti-like virus
YP_009551502 Diatom colony associated dsRNA virus 17 genome type B
QTT60991 Phytophthora condilina RNA virus 2
Totiviridae
QDH87752 Totiviridae sp.
QDH88238 Totiviridae sp.
QDH87440 Totiviridae sp.
QDH90454 Totiviridae sp.
QDH91324 Totiviridae sp.
QDH87365 Totiviridae sp.
QIP68042 Erysiphales associated totivirus 24
Botybirnaviridae
QIP68058 Erysiphales associated toti-like virus 1
QIP68069 Erysiphales associated toti-like virus 2
YP_009182198 Red clover powdery mildew-associated totivirus 9
Heterosiphonia pulchra toti-like virus
QXL43060 Riboviria sp.
Chrysoviridae
Unclassified
0.3
● Found in this study
Land plants
Lower plants
Green algae
Invertebrate
Euglenozoa
Chromista
Fungi
Vertebrate
Environmental
Red algae

Figure 11. Left: Phylogenetic relationships of the viruses within the order *Ghabrivirales*. (A) A phylogeny of the *Chrysoviridae,* (B) an order level phylogeny. ML phylogenetic trees based on the replication protein show the topological position of the virus-like sequences discovered in this study (black circles) in the context of their closest relatives. See Figure 3 for the colour scheme. Green stars are used to signify sequences that have been translated using the protozoan mitochondrial genetic code. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support. Virus taxonomic names are labelled to the right. Right: Genomic organization of the virus sequences identified in this study and representative species used in the phylogeny.

Totiviridae. Thirteen sequences sharing similarities to toti-like viruses were discovered in eight red and green algae transcriptomes. All sequences share less than 50% amino acid identity across their coding sequence, as such we consider each a putative toti-like viruses. Among these sequences four cluster with *Delisea pulchra totivirus IndA* (AMB17469.1) to form a red alga associated clade basal to the totiviruses (Figure 11). *Gracilaria vermiculophylla toti-like virus* (GrveTV) along with *Red algae totivirus 1* (BBZ90082) form a sister group to the protozoan infecting leishmaniaviruses (Figure 11). The remaining sequences are phylogenetically positioned across the tree of toti-like viruses, commonly occupying basal positions (Figure 11).

Prasiola crispa is contaminated by reads from the fungi, *Candida albicans*. *Prasiola crispa toti-like virus* (PrcrTV), clusters with unclassified protist, fungi, invertebrate and algae viruses including *Elkhorn sea moss toti-like virus* (EsmTV) (Figure 2, Figure 11). The Kappaphycus alvarezii transcriptome in which EsmTV was found showed no evidence of contamination suggesting that PrcrTV may also be derived from algae (Figure 2). The *Mazzaella japonica* transcriptome in which *Red algae toti-like virus 2:3* (RedTV2/3) were discovered was predominantly composed of reads associated with the red algae genera Chondrus. As >99% of reads in this library belong to red algae species RedTV2 and RedTV3 have been assigned to this group. The *Porphyridium purpureum* transcriptome is highly contaminated by reads belonging to flowering plants and an unidentified cloning vector (M10197.1) (Figure 2). The phylogenetic positioning of the viruses discovered from this transcriptome (*Porphyridium purpureum toti-like virus* 1 & 2) point towards being derived from red algae rather than flowering plants (Figure 11).

3.4 Long-term virus-host evolutionary relationships

To examine the frequency of cross-species transmission and co-divergence among plant viruses, we estimated tanglegrams that depict pairs of rooted phylogenetic trees displaying the evolutionary relationship between a virus family and their hosts. This revealed cross-species transmission as the predominate evolutionary event predicted among all the RNA virus groups analysed (median 65%, range 46%-79%) (Figure 12). Cross-species transmission was most frequent in the *Betaflexiviridae* (79%) and the subfamily *Betarhabdovirinae* (79%). Virus-host co-divergence (median 23%, range 14%-29%) and to a lesser extent duplication (i.e., speciation) (median 4.6%, range 1.4%-24%) and extinction events (median 2.9%, range 0%-11%) were detected across plant virus families (Figure 12). Co-divergence was most frequently predicted in the *Benyviridae* and *Tymoviridae* representing 29% and 26% of events respectively. Importantly, however, the results of our co-phylogenetic analysis are undoubtedly influenced by the sample of plant viruses and will likely change as the number of plant viruses identified increases.

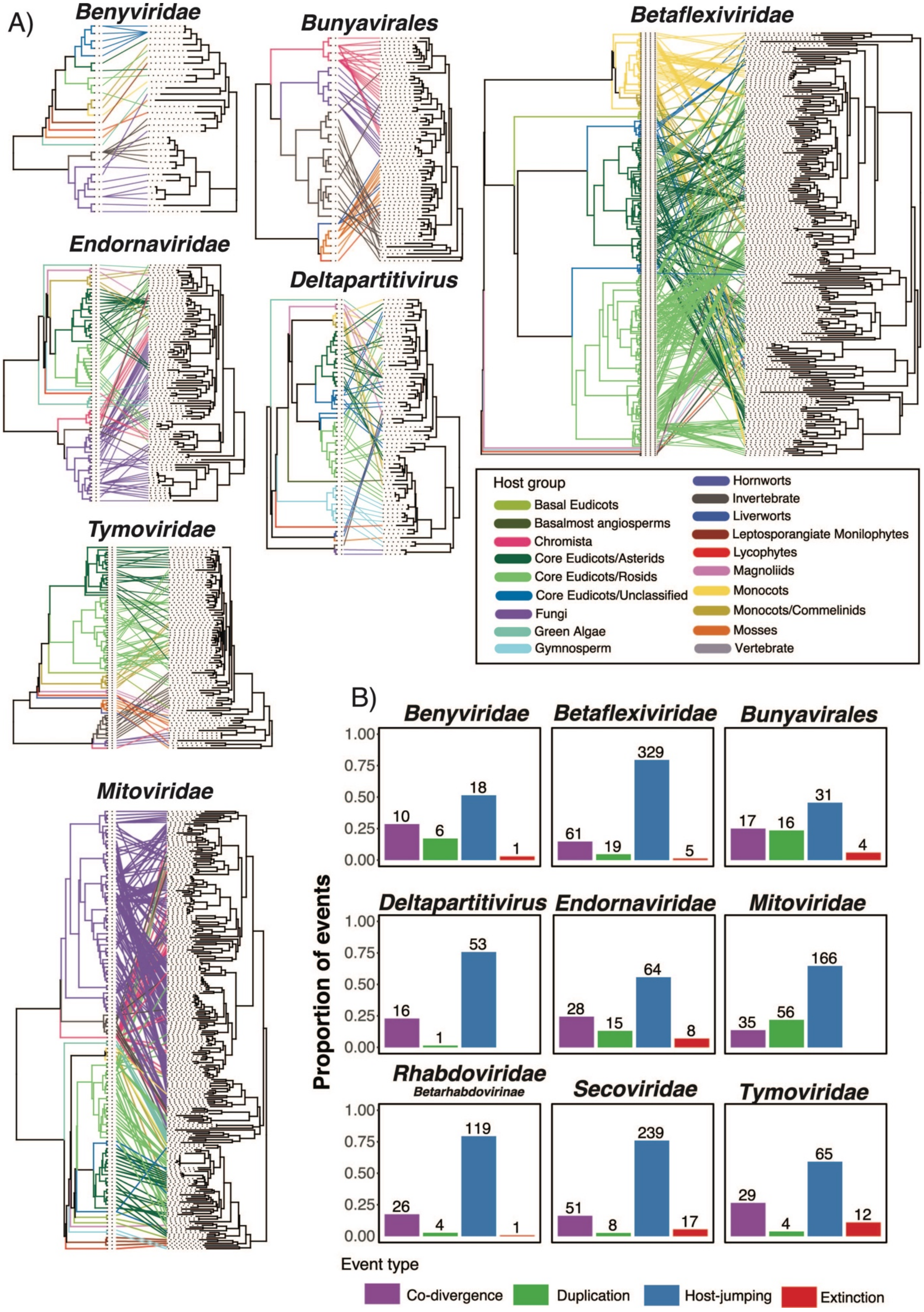

A)
Benyviridae
Bunyavirales
Betaflexiviridae
Endornaviridae
Deltapartitivirus
Tymoviridae
Mitoviridae
Host group
Basal Eudicots
Basalmost angiosperms
Chromista
Core Eudicots/Asterids
Core Eudicots/Rosids
Core Eudicots/Unclassified
Fungi
Green Algae
Gymnosperm
Hornworts
Invertebrate
Liverworts
Leptosporangiate Monilophytes
Lycophytes
Magnoliids
Monocots
Monocots/Commelinids
Mosses
Vertebrate
B)
Proportion of events
1.00
0.75
0.50
0.25
0.00
Benyviridae
10
6
18
1
Betaflexiviridae
61
19
329
5
Bunyavirales
17
16
31
4
Deltapartitivirus
16
1
53
Endornaviridae
28
15
64
8
Mitoviridae
35
56
166
Rhabdoviridae
Betarhabdovirinae
26
4
119
1
Secoviridae
51
8
239
17
Tymoviridae
29
4
65
12
Event type
Co-divergence
Duplication
Host-jumping
Extinction

Figure 12. (A) Tanglegram of rooted phylogenetic trees for select virus groups and their hosts. Lines and branches are coloured to represent host clade. The cophylo function implemented in phytools (v0.7-80) was used to maximise the congruence between the host (left) and virus (right) phylogenies. Supplementary Figure 5 provides the names of the hosts and viruses along with additional tanglegrams for the *Secoviridae* and *Rhabdoviridae*. (B). Reconciliation analysis of select virus groups. Barplots illustrate the range of the proportion of possible events and are coloured by event type.

4. Discussion

Our ability to reconstruct the evolutionary history of plant viruses and understand the drivers of their emergence has been constrained by inadequate sampling across the enormous, extant diversity of plant species. Here, we provide a large-scale virus discovery project based on mining transcriptomes from across the entire breadth of the plant kingdom. In doing so we have identified 104 potentially novel virus species. We considerably expand upon the known host range of 13 virus families to now include lower plants and expand a further four virus families to include host associations with algae. We also find the first evidence of a movement protein with a predicted molecular weight of ~30 kDa (herein referred to as a “30K MP”) in a virus of non-vascular plants. Collectively, this new knowledge advances our understanding of RNA virus diversity across the Archaeplastida.

4.1 RNA viruses are widespread across lower plant lineages

To date, viral surveys in basal plant lineages (namely ferns, bryophytes and algae) have revealed only the minimal occurrence of (+)RNA viruses (5, 17, 20, 78, 82, 83), supporting the idea that RNA viromes in angiosperms evolved as they diversified during the Cretaceous (84).

However, our results potentially challenge this paradigm as we detected the first evidence of sets of (+)ssRNA viruses in lower plants and algae, implying that these groups are associated with older lineages of plants. Several of these viruses are deep branching and sit basal to angiosperm infecting viruses (e.g., LycoAV and LyBuV) in phylogenetic trees. Other viruses discovered here occupy ambiguous positions between established plant virus families (e.g., OnV) or cluster in large numbers to form novel plant-associated clades (e.g., the *Viridisbunyaviridae* in the *Bunyavirales*). Benyviruses are typically transmitted by the root-infecting plasmodiophorids *Polymyxa betae* and *Polymyxa graminis* (85, 86). The Phytomyxids (plasmodiophorids and phagomyxids) are parasites of plants, diatoms, oomycetes and brown algae and have been shown to demonstrate cross-kingdom host shifts (e.g., between angiosperms and oomycetes) (87). As such the plasmodiophorids may be a vehicle for cross-species transmission between aquatic protists and land plants (5). FeBV, a beny-like virus identified in this study, formed a clade along with *Wheat stripe mosaic virus* distinct from members of the genus *Benyvirus*. Deciphering the evolutionary history and mode of transmission for the lower plant beny-like viruses will require further studies with particular emphasis on these taxa. Interestingly, no plasmodiophorid-associated reads were detected in any of the libraries from which we assembled a beny-like virus. LjBV and WasBV appear distantly related to the benyviruses. These viruses group with a suite of unclassified viruses assembled from a soil metatranscriptome study suggesting that, like the benyviruses, this larger group of unclassified viruses may involve soil-borne parasites like the plasmodiophorids (88). Our detection of tymovirid-like sequences in the lycophytes, bryophytes and brown algae dramatically expands the known host range of the *Tymovirales*. Several of these viruses were similar to unclassified Riboviria species assembled from a recent survey of common wild oat soil rhizosphere and detritosphere (88) (Figure 4). The metatranscriptome of the sequenced soil samples from the common wild oat study was largely composed of Viridiplantae, fungi, Amoebozoa, protists, nematodes, and other eukaryotes. As such, using phylogenetic clusters to infer host associations of our viruses remains challenging. Indeed, these viruses may result from contamination from other eukaryotes (e.g., fungi or invertebrates) although we found no consistent evidence among these viruses (Figure 2). Assuming these viruses are plant-associated, their phylogenetic pattern suggests that they may have resulted from cross-kingdom transmission events that frequent the Alsuviricetes.

The partial deltaflexi-like virus we detected in *P. agnata* (PaADV) is particularly noteworthy. The deltaflexiviruses are only known to infect fungi, although no fungi associated reads were found in the *P. agnata* metatranscriptome (Figure 2). The mycovirus families *Delta*- and *Gammaflexiviridae* are thought to have been derived from the plant alpha- and betaflexivirids through cross-species transmission (5, 89)). As such PaAGV could potentially represent an intermediate between the plant and fungi flexiviruses or perhaps a more recent fungus to plant transmission. As only a fragment of the polymerase gene was assembled for this virus future work should confirm the presence of PaAGV and its phylogenetic position.

4.2 The extension of the *Mitoviridae* to a lycophyte host

Through the analysis of mitoviruses-like, non-retroviral endogenous RNA viral elements (NERVEs), it was argued that the origin of plant mitovirus NERVEs was a single horizontal transfer from a fungal mitovirus before the origin of vascular plants in the early Silurian, ~400 MYA (90). Evidence of contemporary mitoviruses in flowering plants and a fern have challenged this view, suggesting that a lineage of plant rather than fungal mitoviruses are the immediate ancestors of plant mitovirus NERVEs (16). Indeed, plant-to-fungus transmission would eliminate code conflicts between fungi and plant mitochondrial genetic codes (76). Herein, we demonstrate the existence of a lower plant-associated sister clade to the angiosperm mitoviruses and NERVEs. This clade includes a clubmoss associated mitovirus, the most primitive plant mitovirus sequence to date. This finding aligns with the estimation of the origin of plant mitovirus NERVEs occurring as early as the evolution of the clubmoss (90). The recent finding of mitoviruses in green algae – including BopiMV in this study – highlight the broad host range of mitoviruses (78, 83). The phylogenetic position of these viruses and the absence of NERVEs from these groups suggest that they are not the ancestors of land plant mitoviruses and NERVEs.

4.4 Establishment of a new virus family in the *Bunyavirales*: *Viridisbunyaviridae*

We identified 16 bunya-like viruses assembled from six non-vascular plant libraries including liverwort, moss, and lycophyte species. These viruses form a novel clade within the *Bunyavirales* and represent the first viruses in this order to be associated with lower plants. This clade likely represents a novel virus family which we have tentatively named the *Viridisbunyaviridae.* Several libraries contained up to five distinct viruses (each sharing <70% nucleotide identity). Virus co-infections are frequently observed in plants and have been reported in the closest relatives of these viruses, the *Deltamycobunyaviridae* (91, 92). As with previous studies we were only able to recover the bunyavirus L segment (92, 93). Further studies are needed to recover the missing small and medium-sized segments and to confirm the presence of mixed infections in plants.

4.5 Discovery of the first 30 kDa movement protein in non-vascular plants

Through the discovery of lower plant-associated viruses, we have gained insights into how the genome structure and composition of contemporary flowering plants viruses have evolved. The detection of secovirid-like sequences in bryophytes and ferns represents the first occurrence of plant secoviruses outside of angiosperms and the first evidence of a 30K MP homolog in non-vascular plants. These proteins aid the cell-to-cell movement of viruses in plants. For example, the MP of *Cucumber mosaic virus* increases the size exclusion limit of plasmodesmata allowing virus particles to pass through cell walls (94). To date, homologs of 30K MP have only been detected in plant viruses infecting angiosperms to the lycophytes (17, 95). Further work is needed to confirm the presence and function of 30K MPs in viruses infecting the bryophytes and other lower plants.

4.6 Detection of Deltaparitivirus dsRNA3 segments in gymnosperms but not in non-vascular plants

Our discovery of six tri-segmented deltapartitivirus species provides insights into the evolution of the deltapartitivirus dsRNA3 segment. dsRNA3 segments have been found in several alpha- and deltapartitiviruses infecting flowering plants (96-99). These segments typically encode seemingly full-length capsid protein or in the case of alphapartitivirus *Rosellinia necatrix partitivirus 2*, a truncated version of the RdRp which may serve as an interfering RNA (100). There is some debate as to the source of dsRNA3 segments, particularly whether they are satellite viruses that co-opt the RdRp of the co-infecting helper viruses or that the additional segment is a result of coinfection of two different plant partitiviruses and the second RdRp-encoding segment is lost after the initial infection (101). For the first time, we find dsRNA3 segments in conifer associated viruses but not in those found in lower plants including bryophytes and lycophytes. The absence of dsRNA3 in non-vascular plants means that it is possible that this segment evolved after the divergence of vascular and non-vascular plants in the Silurian period (102). It is possible that dsRNA3 segments exist for the non-vascular plant infecting deltapartitiviruses but was not detected due to the large degree of divergence between this segment and reference sequences (including those found in this study). However, the dsRNA1 and dsRNA2 segments of the putative lower plant deltapartitiviruses shared >50% aa identity with the tri-segmented deltapartitiviruses - well above the detection limit for tools such as Diamond BlastX (37). dsRNA3 segments typically appear no more divergent than dsRNA2 segments therefore it is unlikely that we would be able to detect both the dsRNA1 and dsRNA2 segments without detecting dsRNA3. Further work is needed to confirm the presence of deltapartitivirus dsRNA3 segments.

4.7 Discovery of an unsegmented varicosavirus-like viruses in ferns and liverworts

Finally, the recently discovered gymnosperm varicose-like *Pinus flexilis virus 1* in the family *Rhabdoviridae* contains an unsegmented genome organisation that differs from the typical bi-segmented structure of the varicosaviruses (25, 103). We find the bi-segmented structure in varicosavirus-like viruses for the first time in ferns and liverworts (TfVV and MgVV) which predate the gymnosperms.

4.8 Caveats

Importantly, the data generated under the 1KP were not explicitly created for virus discovery, such that there are important caveats associated with the methods and metatranscriptomic data mined for virus contigs. For instance, as axenic cultures are not a viable option in most instances, the 1KP samples are commonly contaminated by nucleic acids belonging to bacterial, fungal, and insect species. We addressed this by using a combination of host/virus abundance measurements and phylogenetic analyses to improve the accuracy of virus-host assignments. For most of the viruses described, phylogenetic placement within plant infecting virus families strongly supports their association with plants. However, several of the viruses found in algae and lower plants were associated with lineages known to infect invertebrates and fungi or unclassified viruses recovered from environmental samples. The association between the viruses of lower plants and algae with that of fungi and invertebrate viruses may reflect the absence of algal and lower plant viruses in reference sequence databases. Experimental confirmation is needed to formally assign the viruses discovered in this study to their hosts.

The average sequencing depth of the 1KP libraries was 1.99 gigabases of sequence per sample (range 1.3-3.0), lower than many other virus discovery studies (6, 104, 105). Sequencing depth has been shown to correlate with the ability to detect viruses present at low abundance (106, 107). Further, a large proportion of the virus transcripts detected were from viruses whose full-length genomic or subgenomic mRNAs were polyadenylated at the 3′ end (SI Table 4, Figure 1). Although this was anticipated (i.e. the libraries generated by the 1KP initiative were prepared from polyA+ RNA), it limited the detection of non-polyadenylated viruses (e.g., dsRNA, dsDNA) and may have contributed to the lack of phycodnavirus sequences detected in algae (107).

To reduce the computational burden of assembly, we attempted to remove host-associated reads before contig assembly by mapping them to the host scaffolds provided by the 1KP initiative. While this step reduces the occurrence of false-positive virus detection it also risks removing virus reads, particularly reverse-transcribing plant viruses (108). While we frequently detected transcripts associated with the reverse-transcribing family *Caulimoviridae,* no members of the *Metaviridae* or *Pseudoviridae* were detected.

Acknowledgements

We thank Richard Miller for computational support and Justine Charon for advice on alignments. This work would not have been possible without the data generously provided by the One Thousand Plant Transcriptomes Initiative.

Funding

J.C.O.M. was supported by the William Macleay Microbiological Research Fund from the Linnean Society of NSW. R.V.G. was funded by a Discovery Early Career Researcher Award (DECRA) (DE170100208) and E.C.H. was funded by an ARC Australian Laureate Fellowship (FL170100022) from the Australian Research Council. J.L.G. was funded by a New Zealand Royal Society Rutherford Discovery Fellowship (RDF-20-UOO-007).

References

1. Anderson JT. 2016. Plant fitness in a rapidly changing world. New Phytologist 210:81-87.
2. Wren JD, Roossinck MJ, Nelson RS, Scheets K, Palmer MW, Melcher U. 2006. Plant virus biodiversity and ecology. PLoS Biology 4:80.
3. Mifsud JCO. 2020. Explorations of the plant virosphere.
4. Roossinck MJ, Martin DP, Roumagnac P. 2015. Plant virus metagenomics: advances in virus discovery. Phytopathology 105:716-727.
5. Dolja VV, Krupovic M, Koonin EV. 2020. Deep roots and splendid boughs of the global plant virome. Annual review of phytopathology 58:23-53.
6. Shates TM, Sun P, Malmstrom CM, Dominguez C, Mauck KE. 2019. Addressing Research Needs in the Field of Plant Virus Ecology by Defining Knowledge Gaps and Developing Wild Dicot Study Systems. Frontiers in Microbiology 9.
7. Dolja VV, Koonin EV. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Research 244:36-52.
8. Veliceasa D, Enünlü N, Kós PB, Köster S, Beuther E, Morgun B, Deshmukh SD, Lukács N. 2006. Searching for a new putative cryptic virus in Pinus sylvestris L. Virus Genes 32:177-186.
9. Sidharthan VK, Kalaivanan NS, Baranwal VK. 2021. Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. Gene 786:145626.
10. Han S, Karasev A, Ieki H, Iwanami T. 2002. Nucleotide sequence and taxonomy of Cycas necrotic stunt virus. Archives of virology 147:2207-2214.

11. Yang S, Shan T, Wang Y, Yang J, Chen X, Xiao Y, You Z, He Y, Zhao M, Lu J. 2020. Virome of riverside phytocommunity ecosystem of an ancient canal.

12. Nibert ML, Pyle JD, Firth AE. 2016. A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. Virology 498:201-208.

13. Dawes C. 2016. Chapter 4 - Macroalgae Systematics, p 107-148. *In* Fleurence J, Levine I (ed), Seaweed in Health and Disease Prevention doi:https://doi.org/10.1016/B978-0-12-802772-1.00004-X. Academic Press, San Diego.

14. Christenhusz MJ, Byng JW. 2016. The number of known plants species in the world and its annual increase. Phytotaxa 261:201-217.

15. Valverde RA, Sabanadzovic S. 2009. A novel plant virus with unique properties infecting Japanese holly fern. Journal of General Virology 90:2542-2549.

16. Nibert ML, Vong M, Fugate KK, Debat HJ. 2018. Evidence for contemporary plant mitoviruses. Virology 518:14-24.

17. Mushegian A, Shipunov A, Elena SF. 2016. Changes in the composition of the RNA virome mark evolutionary transitions in green plants. BMC biology 14:68.

18. Short SM, Staniewski MA, Chaban YV, Long AM, Wang DL. 2020. Diversity of Viruses Infecting Eukaryotic Algae. Current Issues in Molecular Biology 39:29-61.

19. Gibbs AJ, Torronen M, Mackenzie AM, Wood JT, Armstrong JS, Kondo H, Tamada T, Keese PL. 2011. The enigmatic genome of Chara australis virus. Journal of General Virology 92:2679-2690.

20. Vlok M, Gibbs AJ, Suttle CA. 2019. Metagenomes of a freshwater charavirus from British Columbia provide a window into ancient lineages of viruses. Viruses 11.

21. Han G-Z. 2019. Origin and evolution of the plant immune system. New Phytologist 222:70-83.

22. Brunkard JO, Zambryski PC. 2017. Plasmodesmata enable multicellularity: new insights into their evolution, biogenesis, and functions in development and immunity. Current Opinion in Plant Biology 35:76-83.

23. Greninger AL. 2018. A decade of RNA virus metagenomics is (not) enough. Virus Research 244:218-229.

24. Miller AK, Mifsud JCO, Costa VA, Grimwood RM, Kitson J, Baker C, Brosnahan CL, Pande A, Holmes EC, Gemmell NJ, Geoghegan JL. 2021. Slippery when wet: cross-species transmission of divergent coronaviruses in bony and jawless fish and the evolutionary history of the Coronaviridae. Virus Evolution doi:10.1093/ve/veab050.

25. Bejerman N, Dietzgen RG, Debat H. 2021. Illuminating the Plant Rhabdovirus Landscape through Metatranscriptomics Data. Viruses 13:1304.

26. Parry R, Wille M, Turnbull OMH, Geoghegan JL, Holmes EC. 2020. Divergent Influenza-Like Viruses of Amphibians and Fish Support an Ancient Evolutionary Association. Viruses 12:1042.

27. Grimwood RM, Holmes EC, Geoghegan JL. 2021. A Novel Rubi-Like Virus in the Pacific Electric Ray (Tetronarce californica) Reveals the Complex Evolutionary History of the Matonaviridae. Viruses 13.

28. Gilbert KB, Holcomb EE, Allscheid RL, Carrington JC. 2019. Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. PLoS One 14:e0219207.

29. Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, Börold J, Salzburger W, Kaderali L, Briggs JAG, Bartenschlager R. 2017. Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. Cell host & microbe 22:387-399.e6.

30. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, Porsch M, Quint M, Rensing SA, Soltis DE, Soltis PS, Stevenson DW, Ullrich KK, Wickett NJ, DeGironimo L, Edger PP, Jordon-Thaden IE, Joya S, Liu T, Melkonian B, Miles NW, Pokorny L, Quigley C, Thomas P, Villarreal JC, Augustin MM, Barrett MD, Baucom RS, Beerling DJ, Benstein RM, Biffin E, Brockington SF, Burge DO, Burris JN, Burris KP, Burtet-Sarramegna V, Caicedo AL, Cannon SB, Çebi Z, Chang Y, Chater C, Cheeseman JM, Chen T, Clarke ND, Clayton H, Covshoff S, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574:679-685.

31. Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, Li L, Du W, K. Ullrich K, Wickett NJ, Barkmann TJ, Barker MS, Leebens-Mack JH, Wong GK-S. 2019. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). GigaScience 8.

32. Johnson MT, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 7:e50226.

33. Leinonen R, Sugawara H, Shumway M, Collaboration INSD. 2010. The sequence read archive. Nucleic acids research 39:19-21.

34. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

35. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A.

2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8:1494-1512.

36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

37. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59-60.

38. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H. 2016. Linking Virus Genomes with Host Taxonomy. Viruses 8:66.

39. Gilmer D, Ratti C, Consortium IR. 2017. ICTV Virus taxonomy profile: Benyviridae. The Journal of general virology 98:1571.

40. RStudio T. 2020. RStudio: integrated development for R.

41. Team RC. 2013. R: A language and environment for statistical computing. Vienna, Austria.

42. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J. 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.

43. Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

44. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States),

45. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

46. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology 7:539.

47. Mirdita M, Steinegger M, Söding J. 2019. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics 35:2856-2858.

48. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. Journal of Molecular Biology 430:2237-2243.

49. Lay CL. 2021. biolumber/littlegenomes: First release. doi:10.5281/ZENODO.5081375.

50. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I. 2021. GenBank. Nucleic acids research 49:D92-D96.

51. Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC biology 4:1-14.

52. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature biotechnology 39:578-585.

53. Marcelino VR, Clausen PTLC, Buchmann JP, Wille M, Iredell JR, Meyer W, Lund O, Sorrell TC, Holmes EC. 2020. CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. Genome Biology 21:103.

54. Clausen PT, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC bioinformatics 19:1-8.

55. Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385.

56. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.

57. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution 32:268-274.

58. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods 14:587-589.

59. Rambaut A, Drummond A. 2012. FigTree: Tree figure drawing tool, version 1.4.0. Institute of Evolutionary Biology, University of Edinburgh.

60. Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506:89-92.

61. Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. Am J Bot 105:302-314.

62. Jin Y, Qian H. 2019. V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. Ecography 42:1353-1359.

63. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brister JR. 2017. Virus Variation Resource–improved response to emergent viral outbreaks. Nucleic acids research 45:D482-D490.

64. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528.

65. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217-223.

66. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms for Molecular Biology 5:1-10.

67. Santichaivekin S, Yang Q, Liu J, Mawhorter R, Jiang J, Wesley T, Wu Y-C, Libeskind-Hadas R. 2020. eMPRess: a systematic cophylogeny reconciliation tool. Bioinformatics doi:10.1093/bioinformatics/btaa978.

68. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research 41:D590-D596.

69. Morozov SY, Solovyev AG. 2003. Triple gene block: modular design of a multifunctional machine for plant virus movement. Journal of General Virology 84:1351-1366.

70. Hammond R, Ramirez P. 2001. Molecular characterization of the genome of Maize rayado fino virus, the type member of the genus Marafivirus. Virology 282:338-347.

71. Ding S, Howe J, Keese P, Mackenzie A, Meek D, Osorlo-Keese M, Skotnicki M, Srifah P, Torronen M, Gibbs A. 1990. The tymobox, a sequence shared by most tymoviruses: its use in molecular studies of tymoviruses. Nucleic acids research 18:1181-1187.

72. Xie J, Ghabrial SA. 2012. Molecular characterizations of two mitoviruses co-infecting a hyovirulent isolate of the plant pathogenic fungus Sclerotinia sclerotiorum. Virology 428:77-85.

73. Heinze C. 2012. A novel mycovirus from Clitocybe odora. Archives of virology 157:1831-1834.

74. Nibert ML. 2017. Mitovirus UGA (Trp) codon usage parallels that of host mitochondria. Virology 507:96-100.

75. Zhang T, Li W, Chen H, Yu H. 2015. Full genome sequence of a putative novel mitovirus isolated from Rhizoctonia cerealis. Archives of virology 160:1815-1818.

76. Shackelton LA, Holmes EC. 2008. The role of alternative genetic codes in viral evolution and emergence. Journal of Theoretical Biology 254:128-134.

77. Nerva L, Turina M, Zanzotto A, Gardiman M, Gaiotti F, Gambino G, Chitarra W. 2019. Isolation, molecular characterization and virome analysis of culturable wood fungal endophytes in esca symptomatic and asymptomatic grapevine plants. Environmental microbiology 21:2886-2904.

78. Charon J, Marcelino VR, Wetherbee R, Verbruggen H, Holmes EC. 2020. Metatranscriptomic identification of diverse and divergent RNA viruses in green and Chlorarachniophyte algae cultures. Viruses 12.

79. Koga R, Horiuchi H, Fukuhara T. 2003. Double-stranded RNA replicons associated with chloroplasts of a green alga, Bryopsis cinicola. Plant molecular biology 51:991-999.

80. Cai G, Myers K, Hillman BI, Fry WE. 2009. A novel virus of the late blight pathogen, Phytophthora infestans, with two RNA segments and a supergroup 1 RNA-dependent RNA polymerase. Virology 392:52-61.

81. Covelli L, Coutts RH, Di Serio F, Citir A, Açıkgöz S, Hernandez C, Ragozzino A, Flores R. 2004. Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus Chrysovirus. Journal of General Virology 85:3389-3397.

82. Rousvoal S, Bouyer B, López-Cristoffanini C, Boyen C, Collén J. 2016. Mutant swarms of a totivirus-like entities are present in the red macroalga Chondrus crispus and have been partially transferred to the nuclear genome. Journal of phycology 52:493-504.

83. Charon J, Murray S, Holmes EC. 2021. Revealing RNA virus diversity and evolution in unicellular algae transcriptomes. Virus Evolution 7.

84. Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. Nature 389:33-39.

85. Valente JB, Pereira FS, Stempkowski LA, Farias M, Kuhnem P, Lau D, Fajardo TVM, Nhani Junior A, Casa RT, Bogo A, da Silva FN. 2019. A novel putative member of the family Benyviridae is associated with soilborne wheat mosaic disease in Brazil. Plant Pathology 68:588-600.

86. Tamada T, Schmitt C, Saito M, Guilley H, Richards K, Jonard G. 1996. High resolution analysis of the readthrough domain of beet necrotic yellow vein virus readthrough protein: a KTER motif is important for efficient transmission of the virus by Polymyxa betae. Journal of General Virology 77:1359-1367.

87. Neuhauser S, Kirchmair M, Bulman S, Bass D. 2014. Cross-kingdom host shifts of phytomyxid parasites. BMC Evolutionary Biology 14:33.

88. Starr EP, Nuccio EE, Pett-Ridge J, Banfield JF, Firestone MK. 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. Proceedings of the National Academy of Sciences 116:25900.

89. Ghabrial SA, Caston JR, Jiang D, Nibert ML, Suzuki N. 2015. 50-plus years of fungal viruses. Virology 479-480:356-68.

90. Bruenn JA, Warner BE, Yerramsetty P. 2015. Widespread mitovirus sequences in plant genomes. Peerj 3.

91. Moreno Goncalves AB, Lopez-Moya JJ. 2019. When viruses play team sports: mixed infections in plants. Phytopathology 110:29-48.

92. Botella L, Jung T. 2021. Multiple Viral Infections Detected in Phytophthora condilina by Total and Small RNA Sequencing. Viruses 13:620.

93. Botella L, Janoušek J, Maia C, Jung MH, Raco M, Jung T. 2020. Marine Oomycetes of the Genus Halophytophthora Harbor Viruses Related to Bunyaviruses. Frontiers in Microbiology 11.

94. Su S, Liu Z, Chen C, Zhang Y, Wang X, Zhu L, Miao L, Wang X-C, Yuan M. 2010. Cucumber Mosaic Virus Movement Protein Severs Actin Filaments to Increase the Plasmodesmal Size Exclusion Limit in Tobacco The Plant Cell 22:1373-1387.

95. Mushegian AR, Elena SF. 2015. Evolution of plant virus movement proteins from the 30K superfamily and of their homologs integrated in plant genomes. Virology 476:304-315.

96. Kumar S, Subbarao BL, Kumari R, Hallan V. 2017. Molecular characterization of a novel cryptic virus infecting pigeonpea plants. PloS one 12:e0181829.

97. Sabanadzovic S, Ghanem-Sabanadzovic NA. 2008. Molecular characterization and detection of a tripartite cryptic virus from rose. Journal of Plant Pathology:287-293.

98. Chen L, Chen J, Liu L, Yu X, Yu S, Fu T, Liu W. 2006. Complete nucleotide sequences and genome characterization of double-stranded RNA 1 and RNA 2 in the Raphanus sativus-root cv. Yipinghong. Archives of virology 151:849-859.

99. Wu LP, Du YM, Xiao H, Peng L, Li R. 2020. Complete genomic sequence of tea-oil camellia deltapartitivirus 1, a novel virus from Camellia oleifera. Archives of Virology 165:227-231.

100. Chiba S, Lin YH, Kondo H, Kanematsu S, Suzuki N. 2013. Effects of Defective Interfering RNA on Symptom Induction by, and Replication of, a Novel Partitivirus from a Phytopathogenic Fungus, Rosellinia necatrix. Journal of Virology 87:2330-2341.

101. Nibert ML, Ghabrial SA, Maiss E, Lesker T, Vainio EJ, Jiang D, Suzuki N. 2014. Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. Virus Research 188:128-141.

102. Harrison CJ, Morris JL. 2018. The origin and early evolution of vascular plant shoots and leaves. Philosophical Transactions of the Royal Society B: Biological Sciences 373:20160496.

103. Walker PJ, Blasdell KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM, Tesh RB, Tordo N. 2018. ICTV virus taxonomy profile: Rhabdoviridae. Journal of General Virology 99:447-448.

104. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S, Buchmann J, Wang W, Xu J, Holmes EC, Zhang Y-Z. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539.

105. Hao X, Zhang W, Zhao F, Liu Y, Qian W, Wang Y, Wang L, Zeng J, Yang Y, Wang X. 2018. Discovery of plant viruses from tea plant (Camellia sinensis (L.) O. Kuntze) by metagenomic sequencing. Frontiers in Microbiology 9:2175.

106. Maclot F, Candresse T, Filloux D, Malmstrom CM, Roumagnac P, van der Vlugt R, Massart S. 2020. Illuminating an ecological blackbox: using high throughput Sequencing to characterize the plant virome across scales. Frontiers in Microbiology 11:2575.

107. Visser M, Bester R, Burger JT, Maree HJ. 2016. Next-generation sequencing for virus detection: covering all the bases. Virology Journal 13:85.

108. Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biology Direct 4:41.

109. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic acids research 47:W256-W259.

110. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. Nucleic Acids Res 43:D571-7.

Supplementary Information

Supplementary Figure 1. (A) Phylogram of the triple gene block (TGB) protein 1. ML phylogenetic trees show the topological position of the newly discovered TGB sequence in the tomato fern (black circle) in the context of the closest relatives. (B) Phylogram of the Tymoviridae virus coat proteins (CP). ML phylogenetic trees show the topological position of the newly discovered CP sequences in (black circle) in the context of the closest relatives. Branches are highlighted to represent virus taxonomy (Maculavirus = green, Marafivirus = orange, Tymovirus = red and unclassified = grey). For each colour, a lighter shade signifies that this virus is related to but has not formally been assigned to this genus. (C) Phylogram of the *Oxera neriifolia associated virus* coat protein (CP). ML phylogenetic trees show the topological position of the newly discovered CP sequence (black circle) in the context of the closest relatives. For all trees, branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. Numbers at the nodes indicate bootstrap support over 70% (1000 replicates).

Supplementary Figure 2. Multiple sequence alignment conserved amino acid motifs in RNA-dependent RNA polymerase (RdRp) regions of the mitoviruses discovered in this study along with reference mitoviruses. The bar above each residue is green if 100% of residues in that column are identical, green-brown if they are 30%-99%, and red if under 30%. The numbers under each section correspond to regions containing motifs identified in (72).

Supplementary Figure 3. (A) Phylogenetic relationships of the viruses identified within the virus families *Potyviridae and Tombusviridae*. ML phylogenetic trees based upon alignments of the amino acid sequences of the RdRp protein show the topological position of discovered virus-like sequences (black circles) from this study in the context of their closest relatives. See Figure 3 for the colour scheme. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. An asterisk indicates node support of >70% bootstrap support.

Supplementary Figure 4. Phylogram of the deltapartitii-like virus (A) coat protein/RNA2 (CP) and (B) RNA3/coat protein 2. ML phylogenetic trees show the topological position of the newly discovered CP sequences in (black circle) in the context of the closest relatives. All branches are scaled to the number of amino acid substitutions per site and trees were mid-point rooted for clarity only. Numbers at the nodes indicate bootstrap support over 70% (1000 replicates).

Supplementary Figure 5. Tanglegram of rooted phylogenetic trees for select virus families and their hosts. Lines and branches are coloured to represent host clade. The cophylo function implemented in phytools (v0.7-80) was used to maximise the congruence between the host (left) and virus (right) phylogenies.

Supplementary Table 1. Clade assignment for all One Thousand Plant Transcriptomes Initiative (1KP) species for which a virus was detected.

Supplementary Table 2. Summary information for each One Thousand Plant Transcriptomes Initiative (1KP) libraries analysed.

Supplementary Table 3. Proportion of transcripts and abundance assigned to each plant virus family.

Supplementary Table 4. Summary table of the viruses discovered in this study

Supplementary Table 5. Genome annotation information underlying the annotation graphs

Supplementary References

1. Xie J, Ghabrial SA. 2012. Molecular characterizations of two mitoviruses co-infecting a hyovirulent isolate of the plant pathogenic fungus Sclerotinia sclerotiorum. Virology 428:77-85.