

Uncovering Plant Virus Species Forming Novel Provisional Taxonomic Units Related to the Family *Benyviridae*

Andrey G. Solovyev^{1,2,3}  and Sergey Y. Morozov^{1,2,*} ¹ A. N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia² Department of Virology, Biological Faculty, Moscow State University, 119234 Moscow, Russia³ All Russia Research Institute of Agricultural Biotechnology, 127550 Moscow, Russia

* Correspondence: morozov@genebee.msu.ru; Tel.: +7-(495)-9393198

Abstract: Based on analyses of recent open-source data, this paper describes novel horizons in the diversity and taxonomy of beny-like viruses infecting hosts of the plant kingdom (Plantae or Archaeplastida). First, our data expand the known host range of the family *Benyviridae* to include red algae. Second, our phylogenetic analysis suggests that the evolution of this virus family may have involved cross-kingdom host change events and gene recombination/exchanges between distant taxa. Third, the identification of gene blocks encoding known movement proteins in beny-like RNA viruses infecting non-vascular plants confirms other evidence that plant virus genomic RNAs may have acquired movement proteins simultaneously or even prior to the evolutionary emergence of the plant vascular system. Fourth, novel data on plant virus diversity highlight that molecular evolution gave rise to numerous provisional species of land-plant-infecting viruses, which encode no known potential movement genetic systems.

Keywords: RNA virus; benyviruses; movement gene modules; RNA polymerase; evolution; RNA helicase; phylogeny; land plants; algae



Citation: Solovyev, A.G.;

Morozov, S.Y. Uncovering Plant Virus Species Forming Novel Provisional Taxonomic Units Related to the Family *Benyviridae*. *Viruses* **2022**, *14*, 2680. <https://doi.org/10.3390/v14122680>

Academic Editor: Alexander Gorbalenya

Received: 20 September 2022

Accepted: 27 November 2022

Published: 29 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Benyvirus is the only genus officially recognized by ICTV in the family *Benyviridae* (order *Hepelivirales*), with the species *Beet necrotic yellow vein virus* (BNYVV) being the type benyvirus [1,2]. The genus *Benyvirus* includes several viruses that have multipartite genomes and rod-shaped virions infecting plants and are usually transmitted by root-infecting vectors. Their single-stranded, positive-sense RNA genome has four segments containing a 5'-terminal “cap” structure and a 3' poly(A) tract [1]. However, RNAs 1 and 2 may cause productive infection. RNA1 encodes a replicative protein with several typical domains, whereas RNA2 includes the 5'-terminal capsid protein (CP) ORF1 extended by a read-through ORF2, a Triple Gene Block (TGB) movement module and a cysteine-rich protein-encoding ORF6 (Figure 1) [1,3–7].

Our recent papers based on published metagenomic studies have revealed several novel plant virus RNAs encoding benyvirus-related replicative proteins and movement gene modules that are close to a TGB in structural, functional and phylogenetic aspects [8–13]. Metagenomic studies and phylogenetic analyses have also shown that many benyvirus-related sequences seem to correspond to novel non-classified phylogenetic *Benyviridae* clusters comprising viruses infecting a wide range of non-plant hosts, namely, insects and fungi [14–18]. In particular, such insect viruses include *Bemisia tabaci beny-like virus 6* (NCBI accession MW256699), *Bemisia tabaci beny-like virus 4* (MW256697), *Hubei Beny-like virus 1* (MK231108) and *Diabrotica undecimpunctata virus 2* (MN646771), and fungal viruses are represented by *Erysiphe necator associated beny-like virus 1* (MN617775), *Rhizoctonia solani beny-like virus 1* (MK507778) and *Sclerotium rolfsii beny-like virus 1* (MH766487) [15,19–22].

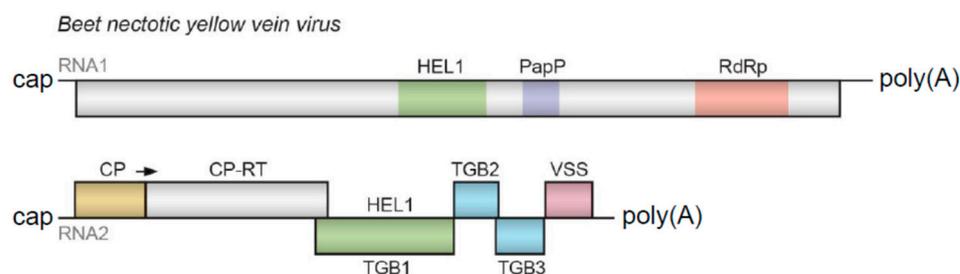


Figure 1. Gene organization of principal RNA genomic components of *Beet necrotic yellow vein virus*. Genes are shown as boxes with the names of the encoded proteins. Domains of RNA replicase ORF encoded by RNA1 include helicase superfamily 1 (HEL1), papain protease (PapP) and RNA-dependent RNA polymerase core domain (RdRp). Genes of proteins potentially involved in cell-to-cell movement (TGB) are shown in green and blue. VSS—virus silencing suppressor; CP-RT—capsid protein read-through protein.

Sets of non-replicative proteins encoded by beny-like plant viruses, on the one hand, and insect/fungal viruses, on the other hand, are generally different from non-plant viruses containing no recognizable movement genes [15,20–22]. In this study, we comprehensively analyzed the gene organization of NCBI-annotated virus genomes and virus-like RNA assemblies (VLRAs), which were identified in transcriptome databases and found to encode replicative proteins related to those in viruses of the genus *Benyvirus*, and showed that the molecular evolution of these replicating RNAs gave rise to numerous provisional virus species, which could infect hosts of the kingdom Plantae and encode no TGB-like gene modules or other known movement genetic systems.

2. Materials and Methods

Benyvirus-like nucleotide sequences were collected from the NCBI plant transcriptome database and 1KP transcriptome database (<https://db.cngb.org/onekp/>, accessed on 26 June 2022). The software used for the viral taxonomic classification is provided by NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy/?term=viruses/>, accessed on 26 June 2022). Our searches were also extended to the most recent plant virus sequence database (<https://riboviria.org/>, accessed on 26 June 2022) [23]. Sequence comparisons were carried out using the BLAST algorithm (BLASTn and BLASTp) at the National Center for Biotechnology Information (NCBI). Open reading frames (ORFs) were identified using ORF Finder programs (http://www.bioinformatics.org/sms2/orf_find.html and <https://www.ncbi.nlm.nih.gov/orffinder/>, accessed on 26 July 2022). Conserved motif searches were conducted in CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>, accessed on 26 July 2022) databases. Hydrophobic protein regions were predicted using Novoprolabs software (<https://www.novoprolabs.com/tools/protein-hydropathy/>, accessed on 26 July 2022). To assemble the 3'-untranslated regions (up to 100–150 nucleotides) of some nearly full-length plant VLRAs, SRA experiments linked to the TSA projects were downloaded using the fastq-dump tool of the NCBI SRA Toolkit (<http://ncbi.github.io/sra-tools/>, accessed on 26 July 2022). Read quality was checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, accessed on 26 July 2022). De novo assembly of the 3'-untranslated regions of some VLRAs was carried out using SPAdes (<http://cab.spbu.ru/software/spades/>, accessed on 26 July 2022) [24] in “RNA mode”. Phylogenetic analysis was performed with “Phylogeny.fr” (a free, simple-to-use web service dedicated to the reconstruction and analysis of phylogenetic relationships between molecular sequences) by constructing maximum likelihood phylogenetic trees (http://www.phylogeny.fr/simple_phylogeny.cgi, accessed on 26 July 2022). Bootstrap percentages obtained with 1000 replications were used. Genome sequences of different benyviruses and beny-like viruses were downloaded from the GenBank database (see text below).

The search for benyvirus-like RNA genomes was initially performed by NCBI TBLASTN analysis with the RNA-dependent RNA polymerase (RdRp) domain of the BNYVV replicative polypeptide as a query. This analysis across the breadth of Excavata (Discoba), Diaphoretickes (Cryptophyta, Rhizaria, Alveolata, Stramenopiles and Archaeplastida) and Amorphea (Fungi, Metazoa and Amoebozoa) [25,26] revealed hundreds of nucleotide contigs encoding proteins with more than 25% identity to the BNYVV RdRp domain (data not shown). RNA contigs under 500 nucleotides in length were excluded from analysis, and thus, the average sizes of the analyzed sequences were about 4500–5000 nucleotides. Further filtering based on pairwise protein sequence similarity NCBI BLASTp searches against the virus protein database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABASE=n/a, accessed on 26 July 2022) for beny-like virus detection was necessarily performed to discriminate highly diverged Benyviridae-like viruses from the representatives of other virus families in the order *Hepelivirales* (*Hepeviridae*, *Matonaviridae*, etc.). For a description of the latter orders, see <https://ictv.global/report/chapter/hepeviridae/hepeviridae> and <https://ictv.global/report/chapter/matonaviridae/matonaviridae> (accessed on 26 July 2022).

The nucleotide sequences of contigs from the NCBI TSA database, which were first identified as VLRA in the present work, can be found using the following accession numbers: red algae and plants—Agarophyton vermiculophyllum VLRA (GILD01050008 and GILD01008494, total 5120 nts), Lithophyllum crustose DN227823 VLRA (GHIV01061204, 5537 nts), Lithophyllum crustose DN218665 VLRA (GHIV01051861, 2575 nts), Gypsophila paniculata VLRA (GILV01257737, 5090 nts), Cenostigma pyramidale VLRA (GIYP01578425, 4555 nts), Salvia miltiorrhiza VLRA (GJJN01058736, 3972 nts), Rhyncholacis cf. penicillata Rhyc27837 VLRA (ICSC01056734, 8281 nts), Rhyncholacis cf. penicillata Rhyc16 VLRA (ICSC01000014, 7893 nts), Astragalus canadensis VLRA (GGNK01006219, 3313 nts), Melampyrum roseum VLRA (IADV01103213, 2174 nts), Striga hermonthica VLRA (ICPL01009187, 8031 nts), Coriandrum sativum VLRA (GGPN01001998, 8098 nts), Camellia reticulata VLRA (GEER01003429, 6431 nts), Viscum album VaGs28290 VLRA (GJLG01028288, 2781 nts), Daiswa yunnanensis VLRA (GFOY01013898, 2707 nts), Ophrys fusca VLRA (GHXI01129489, 7998 nts), Gymnadenia rhellicani VLRA (GHXH01324014, 7925 nts), Ophrys sphegodes VLRA (GHXJ01414654, 8005 nts) and Silene dioica VLRA (GFCC01071918, 9583 nts); arthropods—Xestocephalus desertorum VLRA (GELC01037548, 7329 nts), Pemp-samarca sp. AD-2015 VLRA (GDNN01032725, 4921 nts), Eudarcia simulatricella VLRA (GEOF01029671, 2107 nts), Hydraena subimpressa VLRA (GDPN01031386, 5116 nts), Hypocaccus fitchi VLRA (GDLI01018707, 5112 nts), Porrostoma sp. AD-2015 VLRA (GDLM01020977, 5227 nts), Arrhenodes minutus VLRA (GDPV01017150, 5135 nts), Galerucella pusilla VLRA (HAMG01051171, 7316 nts) and Holacanthella duospinosa VLRA (GFPE01052446 7680 nts).

To generate a phylogenetic tree, we aligned some of the revealed RdRp domains translated from the large number of found Archaeplastida VLRA and selected VLRA of other origins with selected known virus RdRp protein sequences taken from NCBI/GenBank. In total, we included 97 RdRp-encoding contigs (including the outgroup species *Beet yellows closterovirus*) that represent annotated and yet-uncharacterized virus species. VLRA transcripts and annotated virus genomes presented in the tree were taken from 44 Archaeplastida species, while additional included RdRp regions were from viruses with non-plant hosts (mainly fungi and arthropods) and VLRA with unknown hosts (taken from <https://riboviria.org/> (accessed on 26 July 2022)). Among Archaeplastida sequences, we included three sequences from red algae, four from Chara algae and five from non-flowering land plants. The resulting tree is presented in Figure 2.

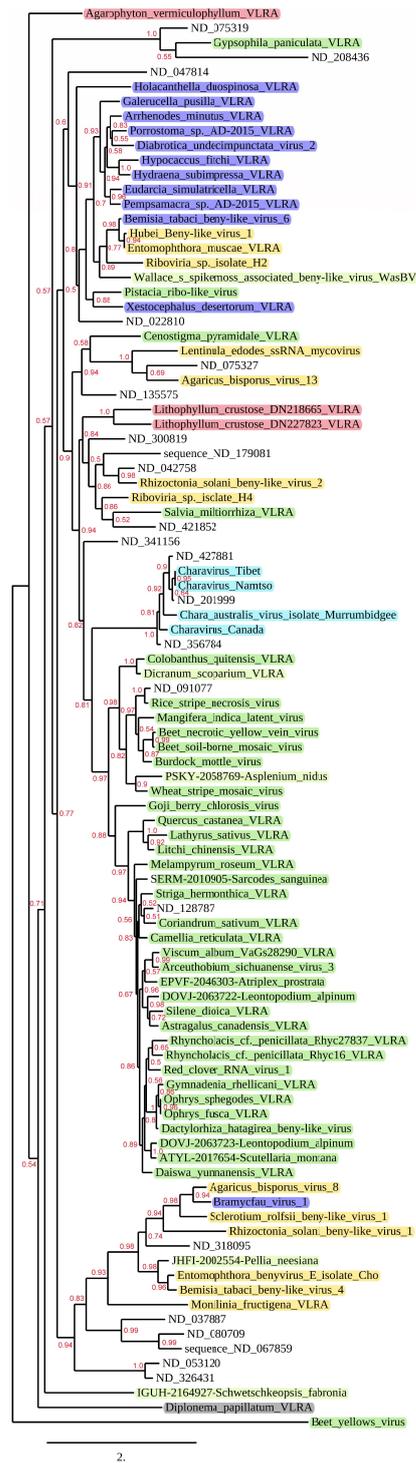


Figure 2. Phylogenetic analysis of the RdRp domains derived from the aligned deduced amino acid sequences of some benyvirus, benyvirus-like VLRA and the selected beny-like viruses. The rooted phylogenetic tree (with the closterovirus species *Beet yellows virus* as the outgroup) was constructed using the maximum likelihood method based on amino acid sequence alignments. The bootstrap values obtained with 1000 replicates are indicated on the branches, and branch lengths correspond to the branch lines’ genetic distances. The genetic distance is shown by the scale bar at the lower left. Host specificity of the selected viruses is indicated by colors: green—flowering plants; light green—non-flowering plants; light blue—charophyte algae; pink—red algae; blue—arthropods; yellow—fungi; shading—euglenozoan protist.

3. Results

The two most diverged benyvirus-like VLRA were revealed in marine organisms. One of these was from the Rhodophyta-specific VLRA of the marine algae *Agarophyton vermiculophyllum* (Figure 2). Another one represented by the VLRA sequence of the marine euglenozoan protist *Diplonema papillatum* also formed no clusters with other sequences. This species belongs to Discoba, a lineage hypothetically proximal to the eukaryotic root, while red algae (Rhodophyta) are the most ancient marine organisms in the kingdom Plantae (Archaeplastida) (<https://www.algaebase.org/browse/taxonomy/>, accessed on 26 July 2022) [26,27]. Therefore, these observations indicate the emergence of beny-like virus replicases before the diversification of eukaryotes or at least the kingdom Plantae. The third basal branch of the tree is again formed by a single representative, the moss *Schwetschkeopsis fabronia* VLRA (Figure 2). Very recently, this VLRA from the 1KP transcriptome project was annotated as the species *Leucodon julaceus associated beny-like virus* (LjBV) [16]. Other well-separated tree branches include more than one sequence representative.

Taking plant-specific virus and VLRA sequences as references, we could discriminate several mixed and well-separated tree clusters: (i) the beny-like RdRp encoded by the VLRA from the non-vascular plant *Pellia neesiana* representing the class Jungermanniopsida is included in the mixed branch together with many fungal viruses and some arthropod-specific viruses; (ii) the Pistacia ribo-like virus RdRp forms a common cluster with many fungal and arthropod-specific viruses; (iii) the flowering plant *Cenostigma pyramidale* VLRA RdRp is included in a common cluster with some fungal viruses; (iv) the flowering plant *Salvia miltiorrhiza* VLRA RdRp forms a common cluster with some fungal viruses and red algae VLRA (Figure 2). Because the phylogenetic placement of the above-mentioned plant viruses is rather close to viruses infecting very distant hosts, from red algae to arthropods, there is a possibility that plant transcriptomes have been contaminated by RNA molecules from other non-plant species as a result of improper sample collection, symbiosis, parasitism or other factors. However, recent estimations of potential contaminations in the 1KP plant transcriptome database suggest that host cross-species transmission in plant and non-plant viruses prevails over different cases of transcriptome artifact contamination [16].

The largest cluster in the presented phylogenetic tree is composed of RdRp sequences of Viridiplantae viruses only. It contains two main sub-branches including Chara-virus-related sequences and land plant virus sequences (Figure 2). The land plant sub-branch includes RdRp sequences of viruses from the genus *Benyvirus* and some non-classified viruses from *Benyviridae*. Moreover, it contains VLRA-derived sequences that may represent virus species forming novel putative taxa in the family *Benyviridae* (see below).

3.1. Species Forming the Most Basal Branch of the Phylogenetic Tree

The sequence forming the most basal branch of the tree is from the Rhodophyta transcriptome of the marine algae *A. vermiculophyllum* (NCBI accessions GILD01050008 and GILD01008494) (Figure 2). This VLRA contains four ORFs (Figure 3). The incomplete ORF1 includes three conserved domains: a viral helicase 1 domain (HEL, pfam01443, E-value 1.09e-20), FtsJ-like methyltransferase (AdoMet methyltransferase superfamily) (FtsJ, pfam01728, E-value 1.69e-12) and an RdRp core motif (pfam00978, E-value 4.33e-09), but it lacks a viral methyltransferase domain (MTR, pfam01660), probably because of sequence truncation (Figure 3).

Pairwise sequence comparisons (Table S1) show that the Rhodophyta-specific viral HEL SF-1 domain is most similar to benyvirus-like species from plants and somewhat less similar to another distantly related Rhodophyta-specific VLRA (see below) as well as arthropod-specific viruses. In general, these data confirm the rather distant relation of the *A. vermiculophyllum* VLRA to the rest of the beny-like viruses with different hosts. Indeed, pairwise sequence comparisons show that the BNYVV core RdRp domain has 28% identity to the type members of families *Hepeviridae* and *Matonaviridae* (species *Human hepatitis E virus* and *Rubella virus*), whereas *A. vermiculophyllum* VLRA RdRp has 33–36% identity to BNYVV and other approved members of the genus *Benyvirus*. It can be hypothesized that

this potential Rhodophyta-specific virus can be regarded as a founding member of a new potential taxon (for example, a subfamily of *Benyviridae*) in the order *Hepelivirales*.

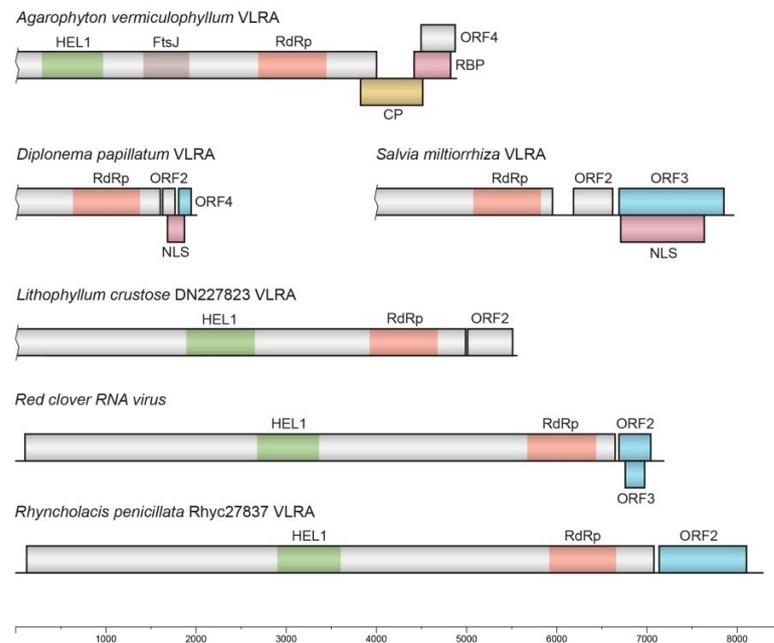


Figure 3. Gene organization of the selected plant viruses and VLRA. The replicase protein domains from all VLRA are shown (see text for details). CP—capsid protein; RBP—RNA-binding protein; NLS—protein with NLS signal. Proteins with putative membrane-bound segments are shown in blue.

The occurrence of the FtsJ-like methyltransferase domain in the *A. vermiculophyllum* VLRA marks its difference in the replicase domain arrangement from benyviruses, which encode no such protein domain [1]. However, this domain has been revealed, particularly in replicase proteins of some members of the family *Kitaviridae* (order *Martellivirales*) [28] and related negeviruses [29]. Amazingly, although FtsJ of the *A. vermiculophyllum* VLRA has only marginal amino acid sequence similarity to this domain in *Martellivirales*, it shows significant similarity to insect DNA-virus-encoded FtsJ proteins, particularly those of *Choristoneura rosaceana nucleopolyhedrovirus* (genus *Alphabaculovirus*; family *Baculoviridae*) (YP_008378434, 40% identity, E-value 2e-07), *Anticarsia gemmatalis nucleopolyhedrovirus* (YP_803463, 37% identity, E-value 2e-06) and *Neophasia sp. alphabaculovirus* (QBC76063, 38% identity, E-value 3e-06). It should be noted that although the RdRp domain of the Rhodophyta-specific VLRA shows obvious similarity to benyvirus replicases, its replicase seems to contain no papain-like proteinase domain, which is conserved in benyvirus replicases [1] (Figure 1).

ORF2 of the *A. vermiculophyllum* VLRA encodes a capsid-like protein (CPL) (Figure 3) of the tobamo-like group that is rather common for rod-shaped viruses [30]. Although the CPL of this VLRA (225 aa in length) has only marginal amino acid sequence identity to CPs of the genus *Benyvirus*, it shows significant similarity to tobamovirus CPs, particularly the cucumber fruit mottle mosaic virus CP (genus *Tobamovirus*; family *Virgaviridae*) (QSM07163, 29% identity, E-value 4e-06), and even higher identity to insect beny-like viruses, particularly the Guiyang benyvirus 1 CPL (family *Benyviridae*) (UHK03085, 40% identity, E-value 7e-21). However, additional TBLAST analysis of the NCBI TSA database shows that the most similar relatives of the *Agarophyton vermiculophyllum* VLRA CPL are encoded by truncated VLRA of another red algal species, *Rhodonematella subimmersa* (GFTI01074942, 44% identity, E-value 3e-12), and the flowering plant *Cenostigma pyramidale* (GIYP01218754, 48% identity, E-value 1e-11). The predicted ORF3 in the *A. vermiculophyllum* VLRA overlaps ORF2 (Figure 3) and encodes a positively charged protein of 135 residues with

$pI = 11.37$, having an uncharged N-terminal region (Figure 4). ORF4 codes for a small protein of 128 residues with no characterized domains.

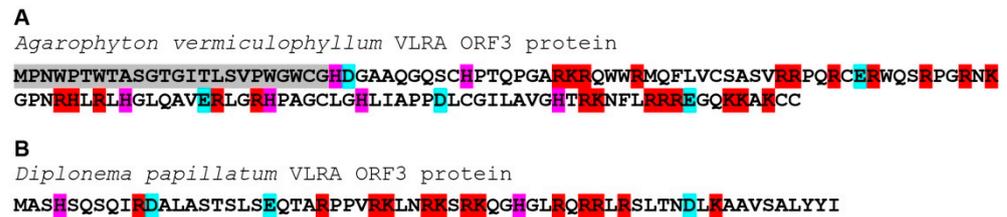


Figure 4. Amino acid sequence of the ORF3 protein encoded by *Agarophyton vermiculophyllum* (A) and *Diplonema papillatum* (B) VLRA. Positively charged residues are shown in red, negatively charged residues are in blue, and histidine residues are in pink; uncharged N-terminal residues are shaded. The putative NLS sequence is underlined.

Another basal-branch-forming VLRA sequence found for the marine euglenozoan protist *D. papillatum* (NCBI accession GJNJ01037323) contains four ORFs (Figure 3). The incomplete ORF1 contains the conserved RdRp domain with a GDD signature. The short ORF2 encodes a small protein of 47 residues containing no characteristic protein signatures. ORF3 overlaps ORF2 (Figure 3) and codes for a positively charged protein of 63 residues with $pI = 11.73$, having a nuclear localization signal (NLS according to NLStradamus software) (Figure 4). These properties allow us to speculate that the ORF3 protein is involved in silencing suppression, as has been shown for some plant viruses [31,32]. ORF4, overlapping with ORF3 (Figure 3), encodes a small protein (44 amino acids in length) with a C-terminal hydrophobic segment (Supplementary Figure S1). It should be noted that the beny-like *D. papillatum* VLRA may represent the second example of a positive-stranded RNA virus infecting Discoba, a lineage hypothetically proximal to the eukaryotic root [24]. To date, only members of the family *Narnaviridae* have been revealed in these hosts [33]. As we found no ORFs coding for capsid-like proteins in the *Diplonema papillatum* VLRA, it can be hypothesized that this potential virus may exist as a “naked” virus, such as narnaviruses [34]. Additionally, assuming a rather low identity level between core RdRp domains of the *D. papillatum* VLRA and BNYVV (27%), it can be speculated that this potential euglenozoan protist virus may represent the founder of a new family in order *Hepelivirales*.

3.2. Cluster of the Phylogenetic Tree including Rhodophyta-Specific VLRA

A cluster of nine RdRp sequences includes two additional Rhodophyta-specific VLRA found for *Lithophyllum crustose*, as well as sequences of fungi-specific viruses and a higher-plant-specific VLRA from *Salvia miltiorrhiza* (family *Lamiaceae*) (Figure 2). Among two Rhodophyta-specific VLRA, the *L. crustose* DN227823 VLRA (GHIV01061204) is longer (likely close in length to the full-size genome) and includes 5537 bases. This VLRA encodes two proteins. The incomplete ORF1 includes two conserved domains: a viral helicase 1 domain (HEL, pfam01443, E-value 5.81e-18) and an RdRp core motif (pfam00978, E-value 7.85e-12) (Figure 3). Unlike the *Agarophyton* VLRA (see above), this replicase encodes no FtsJ-like methyltransferase domain. Additionally, in contrast to the *Agarophyton* VLRA HEL domain, the *Lithophyllum* VLRA HEL domain is closer to replicases of arthropod-specific viruses than to those of plant-specific ones (Table S2). ORF2 of the *Lithophyllum crustose* DN227823 VLRA encodes a negatively charged protein (pI 5.6; 165 aa in length) with no detectable sequence similarity to known virus proteins.

A land-plant-infecting virus in this cluster is represented by the *Salvia miltiorrhiza* VLRA, which is closer, according to the RdRp sequence, to fungal viruses than to Rhodophyta-specific VLRA (Figure 2). An RdRp domain with a GDD signature is found in the C-terminal portion of the incomplete ORF1 (Figure 3). The short ORF2 encodes a small positively charged protein ($pI = 10.4$) of 74 residues containing no characteristic protein signatures. ORF3 codes for a highly hydrophobic, potentially membrane-binding protein of 118 residues in length (Supplementary Figure S2). This ORF is almost completely over-

lapped by ORF4 (Figure 3), which encodes a positively charged protein ($pI = 10.7$; 307 aa in length). This protein contains an NLS sequence motif (RGRGRGRGGFRGRGR, score 0.84) overlapping with the typical N-terminal RGG/RG motif, which is known to participate in protein–RNA interactions of many eukaryotic cell proteins [35,36]. Interestingly, RGG/RG motifs were found to take part in genomic RNA binding by the betacoronavirus nucleocapsid protein [37].

Assuming the percentage of identity between the RdRp core domain of BNYVV and those of the *Lithophyllum crustose* DN227823 VLRA (41%) and the *Salvia miltiorrhiza* VLRA (39%), it can be speculated that the cluster described above may contain viruses of novel subfamilies in the family *Benyviridae*. These considerations are indirectly supported by the fact that some NCBI-approved unclassified members of this family (species *Entomophthora benyvirus E* isolate Cho, *Leucodon julaceus beny-like virus*, *Bemisia tabaci beny-like virus 4*, *Wallace's spikemoss associated beny-like virus*, *Bemisia tabaci beny-like virus 4* and *Rhizoctonia solani beny-like virus 2*) (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1513233> (accessed on 26 July 2022)) express similar or somewhat higher RdRp sequence identities to BNYVV (31–46%).

3.3. Large Cluster of RdRp Domains Encoded by Viridiplantae-Specific Viruses and VLRA s

The largest cluster of the phylogenetic tree includes members of the genus *Benyvirus* along with numerous other plant-infecting viruses (Figure 2). Two viruses exhibiting high sequence identity and similarity in genome organization to *Benyvirus* members are represented by the bipartite species *Wheat stripe mosaic virus* [38,39] and the fern *Asplenium nidus* VLRA [11]. The latter VLRA has been recently annotated as the species *Fern benyvirus* belonging to the family *Benyviridae* [16]. In this cluster, the viruses most distantly related to *Benyvirus* genus members are closely positioned to the monopartite species *Chara australis virus*, which can be tentatively attributed to the Charavirid group (see below). Generally, in this large cluster, we provisionally named three additional groups of putatively monopartite and bipartite plant viruses: Tecimovirids (Tetra-cistron movement block-containing viruses), Binamovirids (Binary movement block-containing viruses) and Reclovirids (Red clover RNA virus 1) (Figure 5A).

3.3.1. Charavirids

Chara australis virus (CAV) [40] has an RNA genome of around 9000 nucleotides (accession JF824737) and encodes a large replicative protein, which shows a relationship with RNA polymerases of benyviruses. Charophyte algae are considered the ancestors of land plants, and *Chara* viruses may be evolutionarily related to ancestor virus species that infected the first plants colonizing terrestrial habitats [40]. Interestingly, beny-like *Chara* viruses are distributed around the globe, since close species were found in Australia and North America (*Charavirus-Canada*) [41].

Previously, we revealed closely related viral RNA metagenomic sequences in the NCBI Sequence Read Archive (SRX8007769), BioProject accession PRJNA615325 (*Charavirus-Tibet*) (Figures 2 and 5A). These data have been derived from samples of fish gills (*Gymnocypris namensis*) collected in Qinghai Lake in the Tibet Autonomous Region of China. A very close virus isolated from the fish body has been recently annotated as *Charavirus-Namtso* (MW483685). The largest encoded protein of these monopartite viruses shows a relationship with RNA polymerases of benyviruses (Figure 2), while the capsid protein is distantly related to the tobamovirus CP [40,41]. Two additional open reading frames (ORFs) code for an RNA helicase and a protein of unknown function. Importantly, this non-replicative “accessory” helicase is related to helicases of the SF-2 superfamily, in contrast to “accessory” TGB1 helicases belonging to SF-1 (Figure 5B) [3,41]. Apparently, the replicase and the “accessory” helicase belong to different lineages of positive-stranded viruses in the orders *Hepelivirales* and *Amarillovirales*, respectively, whereas the “tobamo-like” CPL probably emerged in a common ancestor of the *Hepelivirales* and *Martellivirales* orders [23,41]. It should be noted that VLRA s of unknown origin (ND 427881, ND 201999 and ND 356784

(<https://ribiviria.org> (accessed on 26 July 2022)), see Figure 2) have a genome organization resembling that of *Chara australis virus*. The Charavirid RdRp core domain has 34–36% identity to that of BNYVV, and this cluster can be speculated to include viruses of a new genus/subfamily in the family *Benyviridae*.

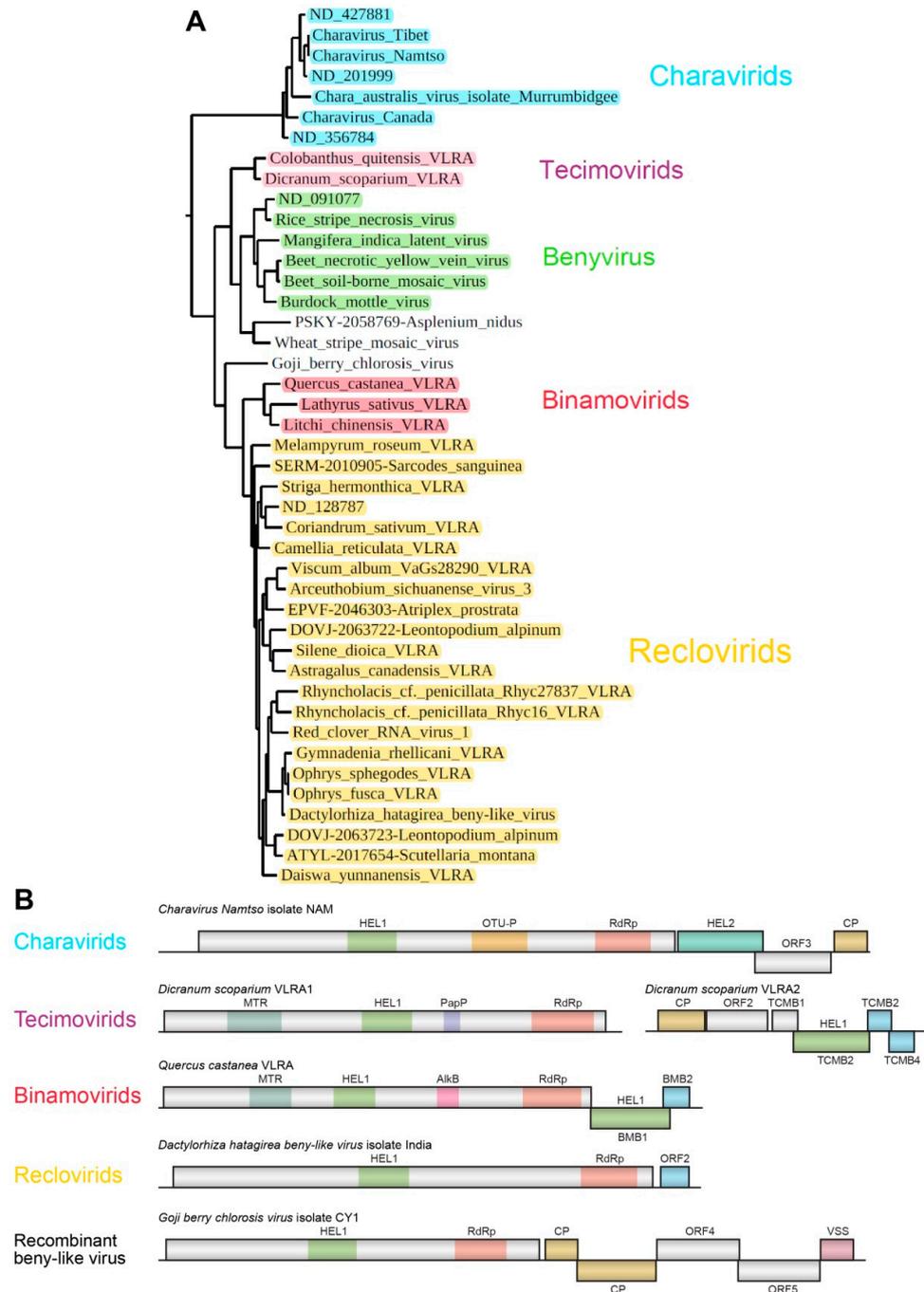


Figure 5. (A) Phylogenetic subtree of the whole tree of RdRp domains from Figure 2 showing a large cluster of Viridiplantae-specific viruses and derived from the aligned amino acid sequences of some benyviruses, benyvirus-like VLRAs and selected beny-like viruses (see text for details). (B) Gene organization of plant viruses and VLRAs that are representatives of Charavirids, Tecimovirids, Binamovirids and Reclovirids. The replicase protein domains from all VLRAs are shown (papain protease—PapP; OTU protease—OTU-P; other domains as in Figure 1). CP—capsid protein; VSS—virus silencing suppressor. Proteins with putative membrane-bound segments are shown in blue.

3.3.2. Goji Berry Chlorosis Virus

Goji berry chlorosis virus (GBCV) is positioned rather close to members of the genus *Benyvirus* in the phylogenetic tree (Figures 2 and 5A). The GBCV RNA genome encodes six polypeptides [42]. The GBCV replicase (ORF1 protein) includes three conserved domains of RNA polymerases, namely, a viral helicase 1 domain (HEL, pfam01443, E-value 1.01e-25) and the RdRp core motif (pfam00978, E-value 1.07e-10), but lacks a viral methyltransferase domain (MTR, pfam01660) (Figure 5B). This domain organization is quite similar to the replicase of the beny-like Rhodophyta-specific VLRA of the marine algae *A. vermiculophyllum* (see above). The GBCV RdRp core domain has 52% identity to that of BNYVV. The GBCV CP (ORF2 protein) represents the “tobamo-like” CPL [42] with the closest sequence similarity to CPLs of arthropod-infecting members of the order *Martellivirales*, particularly the *Abisko virus* CPL (YP_009408587, 29% identity, E-value 2e-06), *Xiangshan martelli-like virus 2* CPL (UDL14010, 28% identity, E-value 1e-04) and *Bemisia tabaci virga-like virus 2* CPL (QWC36454, 25% identity, E-value 1e-06). Strikingly, the CPL-encoding gene in the GBCV genome is duplicated, and its second copy is represented by ORF3, a protein product of which is most closely related to the *Tobacco rattle virus* CP (genus *Tobravirus*) (AAC02063, 28% identity, E-value 3e-04). The ORF4 and ORF5 proteins have no significant similarity to known virus proteins and are marginally related to plant myosin-11 (ORF4 protein) and plant kinesin calmodulin-binding protein (ORF5 protein) (data not shown). ORF6 has been experimentally shown to encode a silencing suppressor protein [42]. This genome organization suggests that GBCV may represent a recombinant between viruses from the families *Benyviridae* and *Virgaviridae* (more specifically, the order *Martellivirales*) [42].

3.3.3. Tecimovirids

Recently, we predicted a novel putative movement gene module, called the tetracistrion movement block (TCMB), in the VLRA of the moss *Dicranum scoparium* and the Antarctic flowering plant *Colobanthus quitensis* [8]. These putative bipartite viruses contain RNA components with a gene organization related to genomes of the genus *Benyvirus*. Replicases of tecimovirids contain RdRp domains with an obviously close relationship to those of viruses in the genus *Benyvirus* (Figures 2 and 5). In particular, the RdRp core domain has 63% identity to that of BNYVV. Like RNA2 of the genus *Benyvirus*, tecimovirid-encoding RNA2 has a 5'-terminal CPL gene, and its RNA helicase gene is included in the 3'-terminal region and is followed by two small overlapping cistrons encoding hydrophobic proteins with distant sequence similarity to the TGB2 and TGB3 proteins. However, unlike the TGB, the TCMB also includes a fourth 5'-terminal gene preceding the helicase gene and coding for a protein showing similarity to the double-stranded RNA-binding proteins of the DSRM AtDRB-like superfamily [8].

3.3.4. Binamovirids

This group of monopartite capsidless viruses is composed of three plant VLRA (Figures 2 and 5), which were initially revealed upon TBLASTN searches of NCBI and 1KP databases with *Hibiscus green spot virus* (HGSV) “accessory” SF-1 helicase as a query [9]. HGSV belongs to the family *Kitaviridae* (order *Martellivirales*) [28]. The HGSV RNA2-encoded “accessory” helicase is the first gene of the binary movement block [12,13] and shows the most pronounced similarity to three long VLRA of the plants *Lathyrus sativus* (NCBI accession GBSS01016353, 7970 nucleotides), *Quercus castanea* (NCBI accession GHJU01198988, 7776 nucleotides) and *Litchi chinensis* (1KP accession WAXR-2010981, 7388 nucleotides). The RdRp core domain of these viruses has 50–52% identity to that of BNYVV. The BMB helicase sequence is somewhat closer to the tecimovirid “accessory” helicase than to the TGB1 helicase [8]. The replicase of binamovirids includes three conserved domains: a viral methyltransferase domain (MTR, pfam01660), a viral helicase 1 domain (HEL, pfam01443) and an RdRp core motif (pfam00978) (see Figure 1 from [9]).

Recently, an additional plant VLRA encoding a binary movement block was revealed in the transcriptome of *Shorea curtisii* (accession GJM01032282). The unusual organization of

this VLRA showed plus and minus chains of the RNA genome assembled in a long hairpin (data not shown). However, despite this probable artifact of sequencing data assembly, the VLRA code for ORF1 replicase with high similarity to the binamovirids *Lathyrus sativus*, *Quercus castanea* and *Litchi chinensis* (RdRp core domain identity 68–75%) and ORF2/3 proteins, which are similar to the BMB1 (Table S3) and BMB2 proteins, respectively.

Despite the fact that binamovirids are quite close to members of the genus *Benyvirus* and tecimovirids, their genomes encode no coat protein. To check the possibility of the presence of additional RNA components in the genomes of binamovirids, we performed a BLAST analysis of transcriptome databases using the 3'-terminal regions of their replicase-encoding RNA segments as queries. However, this analysis revealed no additional genomic segments. Instead, this search showed the nucleotide sequence conservation of the extreme 3'-terminal areas between binamovirids, which confirms their close evolutionary relatedness (Figure 6).

```

Quercus castanea ACTTTTAGTTTGAAGTTTACTGTAGGTTTCTGATTGAGAGGT-TGCTTGGTAGTAGGGTCTGTAGATAGATCCTTT-TGTTTGTGTTAAA
Litchi chinensis ACTCGGGCTTTAC---TTTGGTAGGCGCACTCGATTGAGGAGTGTAGTTGTTCCATGGGTCTGTAGATAGATCCTTTATGTTTTTGTAAA

```

Figure 6. Pairwise nucleotide sequence alignment of the 3'-terminal regions preceding poly(A) in the predicted VLRA RNAs of binamovirids. Conserved nucleotides and RNA blocks are highlighted by yellow background. AAA at the 3'-terminus means poly(A) tract.

3.3.5. Reclovirids

The largest subcluster of the phylogenetic tree showing benyvirus and beny-like RdRps is represented by numerous plant-specific VLRA along with three annotated virus sequences, namely, *Red clover RNA virus 1* (MG596242), *Arceuthobium sichuanense virus 3* (BK059270 [43]) and *Dactylorhiza hatagirea beny-like virus* (BK013327) [44]) (Figures 2 and 5). Strikingly, none of these 22 viruses code for genes of the CP or potential movement proteins. Therefore, we checked the possibility of the presence of additional RNA components in the genomes of reclovirids. As in the case of binamovirids, BLAST analyses of transcriptome databases using the 3'-terminal regions of their replicase-encoding RNA segments as queries have revealed no additional genomic segments. Strikingly, our analyses have demonstrated the nucleotide sequence conservation of the extreme 3'-terminal areas between a dozen reclovirids (Figure 7). Moreover, virus genomes with conserved 3'-non-coding regions are scattered throughout almost the whole reclovirid cluster (Figure 8).

```

Red clover virus 1 TATTGAGTGGTATGGCATTGTAGACCTCTGTAGATAGAGGTA (n)
Arceuthobium sichuanense virus 3 AATTGAGTAGTGGGTCATCGTAATCCTCTGTAGATAGAGGTA (n)
Vicia faba TATTGAGTAGTTTTCTATGTAGGCCTCTGTAGATAGAGGTA (n)
Leontopodium_alpinum TCTTGAGTGGTTTTATCTTGTAGGCCTCTGTAGATAGAGGTA (n)
Rhyncholacis penicillata Rhyc27837 AATTGAG-GGTTTTCCGTGATAGTCCTCTGTAGATAGAGGTA (n)
Gymnadenia rhellican TATTGAGTGGTTACCTAGGGTAGACCTCTGTAGATAGAGGTA (n)
Rhyncholacis penicillata Rhyc16 AATTGAGTGGTTTTACATCGTGGACCTCTGTAGATAGAGGTA (n)
Striga hermonthic AATTGAGTAGTTTACTGAGGTGGACCTCTGTGATAGAGGTA (n)
Cistanche tubulo AATTGAGTAGTTTTCTGTGTTAGGCCTCTGTAGATAGAGGTA (n)
Coriandrum sativum Unigene_31026 GATTGAGTAGTTTTCTGTGGGGACCTCTGTAGATAGAGGTA (n)
Coriandrum sativum Unigene_50954 GATTGAGTAGTTTTCTGTGGTGGACCTCTGTAGATAGAGGTA (n)
Viscum album GATTGAGTAGTCCGTCATCGTAATCCTCTGTAGATAGAGGTA (n)

```

Figure 7. Multiple nucleotide sequence alignment of the 3'-terminal regions preceding poly(A) in the predicted VLRA RNAs of reclovirids. Conserved nucleotides and RNA blocks are highlighted by yellow background. Residues identical to *Arceuthobium sichuanense virus 3* are shown in green. A(n) at the 3'-terminus means poly(A) tract. Truncated VLRA that are not shown in phylogenetic subtree include *Vicia faba* VLRA (GISP01006645) and *Cistanche tubulosa* VLRA (GJRS01079843).



Figure 8. Phylogenetic subtree of the whole tree of RdRp domains from Figure 2 showing Reclovirid cluster of Viridiplantae-specific viruses. Viruses with conserved genomic 3'-termini are marked by 3'-SEQ (see text for details).

To delve deeper into the monopartite genomic organization of previously poorly characterized reclovirids, we initially compared ORF maps within three groups including annotated virus sequences and the most closely related (according to the phylogenetic tree) VLRA, namely, *Red clover RNA virus 1* with *Rhyncholacis penicillata* Rhyc27837 VLRA (ICSC01056734) and *Rhyncholacis penicillata* Rhyc16 VLRA (ICSC01000014); *Dactylorhiza hatagirea beny-like virus* with *Gymnadenia rhellicani* VLRA (GHXH01324014), *Ophrys fusca* VLRA (GHXI01129489) and *Ophrys sphegodes* VLRA (GHXJ01414654); and *Arceuthobium sichuanense virus 3* with *Viscum album* VLRA (GJLG01028288), *Atriplex prostrata* VLRA (EPVF-2046303), *Leontopodium alpinum* VLRA (DOVJ-2063722), *Silene dioica* VLRA (GFCG01071918) and *Astragalus canadensis* VLRA (GGNK01006218) (Figures 2 and 8).

Red clover RNA virus 1 RNA (7190 bases in length) encodes three proteins (Figure 3). The 5'-terminal ORF1 codes for RNA polymerase (2178 aa) with two characteristic domains: a viral helicase 1 domain (HEL, pfam01443, E-value 1.13e-18) and an RdRp core motif (pfam00978, E-value 4.10e-17). ORF2 and its nested ORF3 encode two hydrophobic proteins of 112 and 66 residues, respectively (Figure 3). We presume that both proteins represent “orphan” hydrophobic viral proteins [11] and have membrane-binding properties.

The *Rhyncholacis penicillata* Rhyc27837 VLRA (8274 bases in length), despite the close amino acid sequence similarity and domain organization of the encoded RNA polymerase (ORF1 protein) to *Red clover RNA virus 1* RNA polymerase (identity 50% for the full-length proteins), codes for two proteins only (Figure 3). The ORF2 protein of the *R. penicillata* Rhyc27837 VLRA (313 residues in length) exhibits no similarity to the *Red clover RNA virus 1* ORF2 and 3 proteins. The *R. penicillata* Rhyc16 VLRA (7871 bases in length) also encodes two proteins, and its ORF1-encoded RNA polymerase has 58% identity of the full-length protein to the *R. penicillata* Rhyc27837 ORF1 protein. The ORF2 protein has no sequence similarity to *R. penicillata* Rhyc27837 VLRA-encoded ORF2; nevertheless, both

proteins possess potential membrane-bound segments in the N- and C-terminal parts (Supplementary Figures S3 and S4).

Dactylorhiza hatagirea beny-like virus RNA (7749 bases in length), which represents an isolated group of reclovirids (Figures 2 and 5), encodes two proteins. ORF1-encoded RNA polymerase (2283 aa) has two characteristic domains, a viral helicase 1 domain (HEL, pfam01443, E-value 3.06e-18) and an RdRp core motif (pfam00978, E-value 6.81e-13). The ORF2-encoded protein is predicted to have hydrophobic regions (137aa) (Supplementary Figure S5). ORF1 proteins of the *Gymnadenia rhellicani* VLRA, *Ophrys fusca* VLRA and *Ophrys sphegodes* VLRA belonging to the same group of reclovirids (Figure 5) have similar domain organizations (Figure 5A). The three above-mentioned reclovirid viruses show lower sequence identity (50–53%) in pairwise sequence comparisons with the BNYVV RdRp core domain in comparison with viruses belonging to the genus *Benyvirus* (73–95%) and can tentatively be regarded as members of another genus of *Benyviridae*.

Interestingly, we have found that the N-terminal 60% of the ORF2 protein (excluding its highly hydrophobic C-terminal segment (Supplementary Figure S5)) shows obvious sequence similarity to the ORF2 proteins encoded by VLRA identified for orchids *Gymnadenia rhellicani*, *Ophrys fusca* and *Ophrys sphegodes*. Moreover, similar proteins are encoded by three additional truncated VLRA lacking the RdRp-encoding region. All of these proteins contain the characteristic CX(3)CX(10)CX(3)C motif in the N-terminal region (Figure 9) and hydrophobic segments in the C-terminal half (Supplementary Figures S6 and S7). These cysteine-rich motifs resemble Cys-His fingers, which may be involved in nucleic acid binding and protein–protein interactions [45].

DHBLV	MANVNANDVVEILKYTTICVDN C MHRRELEYQND C AFK C LHVLNNTTVGFAVRQ-TCATEL-VPSNPQLGNILYIIFSGFLMLFRW	[82aa]
GFHX01045452	MAGVDANDIREVLKYTTICVD T CLGRRLEAYRND C AFK C LHLLNNTTVAFSLQK-TCATEM-AVVNPPLGNIMYIIFSGFLMLFNW	(73%)
GHXH01128483	MADIDAKDIAEVLKFTVNCVD T CLGRRLEHRND C AFK C LHLLNNTTVGFTHK-NCAGQL-MIMNPALSDILYLLFSGFLVMLFKW	(59%)
GHXJ01414654	MTKATAEHIKLLHLTVDCVEVCLKLPKASQSD C SFK C FGVLTNSVNQTMCD-PVSTAM-VATNPSSLNIFMILSGFIMLFKW	(41%)
GHXI01129489	MTKATAEHIKLLHLTVDCV E LCLKLPKASQSD C SFK C FGVLTNSVNQTMCD-PLSTGM-VATNPSSLNSILFMIISGLIMLFKW	(39%)
GHXH01324014	MDRINGPLELDTQDVMQILRFTLKC V DECLDK E PERQNS C HT C LRLLNNTVNYVGSNFQ C STDVAIAAPP T FAICMYI I YSGY M LLKW	(37%)
GHXE01021908	-----SRD V LEV L K F T I NC V K V C M Q Y K I V A Q E CT W NC L H I LN N T I GV T T Q N-VCATEL-VPTD F SLNS I LL M LS A LL M LFKW	(47%)

Figure 9. Multiple nucleotide sequence alignment of the N-terminal regions of ORF2 proteins encoded in the 3'-terminal regions of *Dactylorhiza hatagirea beny-like virus* (DHBLV) and predicted closely related VLRA of reclovirids: *Dactylorhiza fuchsii* VLRA (GFHX01045452); *Gymnadenia rhellicani* VLRA TR40174 (GHXH01128483); *Ophrys sphegodes* VLRA (GHXJ01414654); *Ophrys fusca* VLRA (GHXI01129489); *Gymnadenia rhellicani* VLRA TR101479 (GHXH01324014); *Gymnadenia x densiflora* VLRA (GHXE01021908). Amino acid residues identical to those in DHBLV ORF2 protein are shaded. Conserved cysteines in the N-terminal sequences are in yellow. The hydrophobic sequences are underlined. The length of the presented DHBLV sequence is shown in square brackets, and identity to DHBLV protein is indicated on the right.

Arceuthobium sichuanense virus 3 RNA (7341 bases in length) (Figures 2 and 5) encodes two proteins: ORF1-encoded RNA polymerase (2071 aa) with two typical domains, a viral helicase domain (HEL, pfam01443, E-value 3.06e-18) and an RdRp core motif (pfam00978, E-value 6.81e-13), and an ORF2-encoded hydrophobic protein (137aa). Among related subcluster groups (Figures 2 and 5), the *Arceuthobium sichuanense virus 3* ORF2 protein has the only counterpart with obvious similarity (identity 38%) found in the *Viscum album* VLRA (hydrophobic ORF2 protein) (Supplementary Figures S8 and S9). Four other viruses close to *Arceuthobium sichuanense virus 3* encode either one ORF2 hydrophobic protein, similar to the *Leontopodium alpinum* VLRA (DOVJ-2063722), or two ORF2 and ORF3 hydrophobic proteins, with no sequence similarity to the *Arceuthobium sichuanense virus 3* ORF2 protein, as in the case of the *Atriplex prostrata* VLRA (data not shown).

Finally, it should be noted that among other Reclovirids included in Figures 2 and 5, a rather common genome organization toward the 3'-terminus from ORF1 (the RNA polymerase gene) is revealed. All of these representatives encode one or two orphaned proteins usually including potential membrane-bound hydrophobic segments (data not shown).

4. Discussion

The data presented in this paper expand the known host range of the multipartite viruses in the family *Benyviridae* to include algae hosts. First, the phylogenetic analysis confirms the placement of monopartite charophyte algae-infecting viruses as a separate branch in the phylogenetic tree of beny-like viruses. These algae viruses are tentatively named “Charavirids”. Second, the presented analyses indicate that benyvirus-like replicases started their evolutionary radiation in red algae (Rhodophyta), which are the most ancient organisms in the kingdom Plantae (Archaeplastida) (<https://www.algaebase.org/browse/taxonomy/>, accessed on 26 July 2022). Rhodophyta-infecting potentially monopartite beny-like viruses are revealed in two clusters of the phylogenetic tree based on the RdRp core domain. We have found two red algae species carrying potential beny-like viruses, namely, *Lithophyllum crustose* and *Agarophyton vermiculophyllum*. The latter virus is positioned in the most basal part of the RdRp phylogenetic tree, and the corresponding VLRA also encodes a tobamo-like CPL, as is known for several members of the genus *Benyvirus* and some beny-like plant and insect viruses. In particular, the tobamo-like CPL is also encoded by the genomes of Charavirids (see above). Generally, the observed wide distribution of tobamo-like CPL genes [30] in monopartite and multipartite viruses of the family *Benyviridae* (order *Hepelivirales*) and family *Virgaviridae* (order *Martellivirales*) strongly suggests that horizontal gene transfer has played a significant role in the evolution of these virus taxa [15]. The divergence of the Charavirid CP from that of tobamoviruses is estimated to have occurred 212 million years ago (mya) [41]. On the other hand, the time of divergence between charavirus/benyvirus RdRp (order *Hepelivirales*) and virga-like RdRp genes (order *Martellivirales*) is estimated to be ~900 mya [41]. Therefore, benyvirus-like replicating RNAs likely started their evolutionary history in the late Precambrian, i.e., 850–1100 mya [27]. It should be noted that the evolutionary emergence of marine red algae, ranging from single-celled species to multicellular “plant-like” organisms, is estimated to have occurred approximately 1000–1600 mya [27,46,47]. Therefore, the divergence of replication proteins in viruses of the orders *Hepelivirales* and *Martellivirales* could have occurred in marine red algae species. In support of the proposed role of Rhodophyta as a host for common ancestors of *Hepelivirales* and *Martellivirales*, both types of viruses are found in present-day red algae hosts [48].

The analysis presented in this paper also highlights important clues to the evolution of plant cell-to-cell movement systems. First, the available data suggest that MP evolution started at least in lower land plants, as in these hosts, we have found typical movement gene blocks, such as the TCMB in Bryophyta (VLRA of the moss *Dicranum scoparium*) and the TGB in Polypodiopsida (*Fern benyvirus*) [8,11].

Novel data on plant virus diversity suggest that molecular evolution may have given rise to numerous potential plant-infecting virus species, which encode no known movement genetic systems. Previously, this fact has been a point of discussion. In fact, it is known that persistent plant viruses lack cell-to-cell movement systems and do not cause visible symptoms; accordingly, they are transmitted only vertically via gametes. Persistent plant viruses have been found in a few virus families, such as *Endornaviridae* and *Partitiviridae* [49–52]. Further, three large flavi-like replicating RNAs have been isolated from symptomatic plants; these RNAs, encoding single polymerase ORFs and no detectable CP and MP genes, have been termed *Snake River alfalfa virus*, *Carrot flavi-like virus 1* and *Gentian Kobu-sho-associated virus* [53–56]. It is proposed that these viruses could be arthropod-transmitted and spread inside the plant body with the help of other coinfecting viruses.

Similar to many persistent plant viruses, all Reclovirid replicating RNAs have been identified by high-throughput sequencing [43,44]. Therefore, these RNAs can be envisaged as spreading throughout plant bodies with the help of other viruses, as proposed for known persistent viruses. On the other hand, potentially monopartite Reclovirids can be proposed to use novel, yet-undescribed movement protein systems. Indeed, their polymerases are clearly related to benyviruses and beny-like viruses exploiting movement gene blocks, TGB, BMB or TCMB, which encode one, two or three small proteins with hydrophobic membrane-

bound segments. As we have recently demonstrated, such proteins may account for the induction of ER tubule constrictions and increase the PD permeability, thereby facilitating intercellular viral spread, as has been found for the TGB2 and BMB2 proteins [13,57]. Thus, we hypothesize that similar functions can be performed by a broad range of Reclovirid hydrophobic orphan proteins [10]. Apparently, further detailed experimental studies are required to verify whether these proteins can serve as virus MPs.

5. Conclusions

In this work, we compared the genome sequences (mostly partial) of some novel plus-RNA viruses from red algae, a marine euglenozoan protist and land plants. Despite the concerns related to the potential risks of contig misassembly and contamination and the existence of unknown genome segments, in general, the reported sequence analyses of beny-like genomes confirm previously reported evolutionary relationships between viruses of non-flowering plants, flowering plants, arthropods and fungi, as well as suggest novel evolutionary links between viruses of red algae and land plants. We proposed several viruses as founding members of new potential taxa (subfamilies or genera) in the virus family Benyviridae. Additionally, the identification of gene blocks coding for movement proteins in viruses of non-vascular plants indicates that beny-like viruses may have acquired movement proteins prior to the emergence of the plant vascular system.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v14122680/s1>. Figure S1: Prediction of potential hydrophobic membrane-bound regions in ORF4 protein of *Diplonema papillatum* VLRA; Figure S2: Prediction of potential hydrophobic membrane-bound regions in *Salvia miltiorrhiza* VLRA ORF3 protein; Figure S3: Prediction of potential hydrophobic membrane-bound regions in *Red clover RNA virus 1* VLRA ORF2 protein; Figure S4: Prediction of potential hydrophobic membrane-bound regions in *Red clover RNA virus 1* VLRA ORF3 protein; Figure S5: Prediction of potential hydrophobic membrane-bound regions in *Rhyncholacis penicillata* Rhyc27837 VLRA ORF2 protein; Figure S6: Prediction of potential hydrophobic membrane-bound regions in *Rhyncholacis penicillata* Rhyc16 VLRA ORF2 protein; Figure S7: Prediction of potential hydrophobic membrane-bound regions in ORF2 protein of *Dactylorhiza hatagirea beny-like virus*; Figure S8: Prediction of potential hydrophobic membrane-bound regions in ORF2 protein of *Arceuthobium sichuanense virus 3*; Figure S9: Prediction of potential hydrophobic membrane-bound regions in ORF2 proteins of *Viscum album* VLRA; Table S1: Pairwise sequence comparisons of *Agarophyton vermiculophyllum* VLRA-encoded HEL domain with selected viral helicase domains of other viruses and VLRA; Table S2: Pairwise sequence comparisons of *Lithophyllum crustose* coralline algae DN227823 VLRA-encoded HEL domain with selected viral helicase domains of other viruses and VLRA; Table S3: Pairwise sequence comparisons of *Shorea curtisii* VLRA-encoded HEL domain of ORF2 (BMB1) protein with selected viral helicase domains of other viruses and VLRA.

Author Contributions: Conceptualization, S.Y.M. and A.G.S.; methodology, S.Y.M.; software, A.G.S.; validation, S.Y.M. and A.G.S.; formal analysis, A.G.S.; data curation, A.G.S.; writing—original draft preparation, S.Y.M.; writing—review and editing, A.G.S.; supervision, A.G.S.; project administration, A.G.S.; funding acquisition, A.G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Russian Science Foundation (grant 22-14-00063).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gilmer, D.; Ratti, C. ICTV Virus Taxonomy Profile: Benyviridae. *J. Gen. Virol.* **2017**, *98*, 1571–1572. [[CrossRef](#)] [[PubMed](#)]
2. Koonin, E.V.; Dolja, V.V.; Krupovic, M.; Varsani, A.; Wolf, Y.I.; Yutin, N.; Zerbini, F.M.; Kuhn, J.H. Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **2020**, *84*, e00061. [[CrossRef](#)] [[PubMed](#)]
3. Morozov, S.Y.; Solovyev, A.G. Triple gene block: Modular design of a multi-functional machine for plant virus movement. *J. Gen. Virol.* **2003**, *84*, 1351–1366. [[CrossRef](#)] [[PubMed](#)]
4. Verchot-Lubicz, J.; Torrance, L.; Solovyev, A.G.; Morozov, S.Y.; Jackson, A.O.; Gilmer, D. Varied movement strategies employed by triple gene block-encoding viruses. *Mol. Plant-Microbe Interact.* **2010**, *23*, 1231–1247. [[CrossRef](#)]
5. Chiba, S.; Hleibieh, K.; Delbianco, A.; Klein, E.; Ratti, C.; Ziegler-Graff, V.; Bouzoubaa, S.; Gilmer, D. The benyvirus RNA silencing suppressor is essential for long-distance movement, requires both zinc-finger and NoLS basic residues but not a nucleolar localization for its silencing-suppression activity. *Mol. Plant Microbe Interact.* **2013**, *26*, 168–181. [[CrossRef](#)]
6. Pakdel, A.; Mounier, C.; Klein, E.; Hleibieh, K.; Monsion, B.; Mutterer, J.; Erhardt, M.; Bouzoubaa, S.; Ratti, C.; Gilmer, D. On the interaction and localization of the beet necrotic yellow vein virus replicase. *Virus Res.* **2015**, *196*, 94–104. [[CrossRef](#)]
7. Saito, M.; Kiguchi, T.; Kusume, T.; Tamada, T. Complete nucleotide sequence of the Japanese isolate S of beet necrotic yellow vein virus RNA and comparison with European isolates. *Arch. Virol.* **1996**, *141*, 2163–2175. [[CrossRef](#)]
8. Morozov, S.Y.; Solovyev, A.G. Novel Genetic Module Related to Triple Gene and Binary Movement Blocks of Plant Viruses: Tetra-cistron Movement Block. *Biomolecules* **2022**, *12*, 861. [[CrossRef](#)]
9. Morozov, S.Y.; Solovyev, A.G. Phylogenetic relationship of some ‘accessory’ helicases of plant positive-stranded RNA viruses: Toward understanding the evolution of triple gene block. *Front. Microbiol.* **2015**, *6*, 508. [[CrossRef](#)]
10. Morozov, S.Y.; Solovyev, A.G. Small hydrophobic viral proteins involved in intercellular movement of diverse plant virus genomes. *AIMS Microbiol.* **2020**, *6*, 305–329. [[CrossRef](#)]
11. Solovyev, A.G.; Morozov, S.Y. Non-replicative integral membrane proteins encoded by plant alpha-like viruses: Emergence of diverse orphan ORFs and movement protein genes. *Front. Plant. Sci.* **2017**, *8*, 1820. [[CrossRef](#)] [[PubMed](#)]
12. Lazareva, E.A.; Lezzhov, A.A.; Komarova, T.V.; Morozov, S.Y.; Heinlein, M.; Solovyev, A.G. A novel block of plant virus movement genes. *Mol. Plant Pathol.* **2017**, *18*, 611–624. [[CrossRef](#)]
13. Lazareva, E.A.; Lezzhov, A.A.; Chergintsev, D.A.; Golyshev, S.A.; Dolja, V.V.; Morozov, S.Y.; Heinlein, M.; Solovyev, A.G. Reticulon-like properties of a plant virus-encoded movement protein. *New Phytol.* **2021**, *229*, 1052–1066. [[CrossRef](#)] [[PubMed](#)]
14. Kondo, H.; Hirano, S.; Chiba, S.; Andika, I.B.; Hirai, M.; Maeda, T.; Tamada, T. Characterization of burdock mottle virus, a novel member of the genus Benyvirus, and the identification of benyvirus-related sequences in the plant and insect genomes. *Virus Res.* **2013**, *177*, 75–86. [[CrossRef](#)] [[PubMed](#)]
15. 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.; et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539–543. [[CrossRef](#)]
16. Mifsud, J.C.O.; Gallagher, R.V.; Holmes, E.C.; Geoghegan, J.L. Transcriptome Mining Expands Knowledge of RNA Viruses across the Plant Kingdom. *J. Virol.* **2022**, e0026022, (Ahead of print). [[CrossRef](#)]
17. Etebari, K.; Lenancker, P.; Powell, K.S.; Furlong, M.J. Transcriptomics Reveal Several Novel Viruses from Canegrubs (Coleoptera: Scarabaeidae) in Central Queensland, Australia. *Viruses* **2022**, *14*, 649. [[CrossRef](#)]
18. De Miccolis Angelini, R.M.; Landi, L.; Raguseo, C.; Pollastro, S.; Faretra, F.; Romanazzi, G. Tracking of Diversity and Evolution in the Brown Rot Fungi *Monilinia fructicola*, *Monilinia fructigena*, and *Monilinia laxa*. *Front. Microbiol.* **2022**, *13*, 854852. [[CrossRef](#)]
19. Zhu, J.Z.; Zhu, H.J.; Gao, B.D.; Zhou, Q.; Zhong, J. Diverse, Novel Mycoviruses From the Virome of a Hypovirulent *Sclerotium rolfsii* Strain. *Front. Plant Sci.* **2018**, *9*, 1738. [[CrossRef](#)]
20. Picarelli, M.A.S.C.; Forgia, M.; Rivas, E.B.; Nerva, L.; Chiapello, M.; Turina, M.; Colariccio, A. Extreme Diversity of Mycoviruses Present in Isolates of *Rhizoctonia solani* AG2-2 LP From *Zoysia japonica* From Brazil. *Front. Cell Infect. Microbiol.* **2019**, *9*, 244. [[CrossRef](#)]
21. Gilbert, K.B.; Holcomb, E.E.; Allscheid, R.L.; Carrington, J.C. Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PLoS ONE* **2019**, *14*, e0219207. [[CrossRef](#)] [[PubMed](#)]
22. Liu, S.; Valencia-Jiménez, A.; Darlington, M.; Vélez, A.M.; Bonning, B.C. Diabrotica undecimpunctata virus 2, a Novel Small RNA Virus Discovered from Southern Corn Rootworm, *Diabrotica undecimpunctata howardi* Barber (Coleoptera: Chrysomelidae). *Microbiol. Resour. Announc.* **2020**, *9*, e00380-20. [[CrossRef](#)] [[PubMed](#)]
23. Neri, U.; Wolf, Y.I.; Roux, S.; Camargo, A.P.; Lee, B.; Kazlauskas, D.; Chen, I.M.; Ivanova, N.; Allen, L.Z.; Paez-Espino, D.; et al. A five-fold expansion of the global RNA virome reveals multiple new clades of RNA bacteriophages. *bioRxiv* **2022**. [[CrossRef](#)]
24. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
25. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016147. [[CrossRef](#)]
26. Burki, F.; Roger, A.J.; Brown, M.W.; Simpson, A.G.B. The new tree of eukaryotes. *Trends Ecol. Evol.* **2020**, *35*, 43–55. [[CrossRef](#)] [[PubMed](#)]
27. Strassert, J.F.H.; Irisarri, I.; Williams, T.A.; Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **2021**, *12*, 1879. [[CrossRef](#)]

28. Olmedo-Velarde, A.; Hu, J.; Melzer, M.J. A Virus Infecting *Hibiscus rosa-sinensis* Represents an Evolutionary Link Between Cileviruses and Higreviruses. *Front. Microbiol.* **2021**, *12*, 660237. [[CrossRef](#)] [[PubMed](#)]
29. Nunes, M.R.T.; Contreras-Gutierrez, M.A.; Guzman, H.; Martins, L.C.; Feitoza Barbirato, M.; Savit, C.; Balta, V.; Uribe, S.; Vivero, R.; Suaza, J.D.; et al. Genetic characterization, molecular epidemiology and phylogenetic relationships of insect-specific viruses in the taxon Negevirus. *Virology* **2017**, *504*, 152–167. [[CrossRef](#)]
30. Dolja, V.V.; Boyko, V.P.; Agranovsky, A.A.; Koonin, E.V. Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: Two families with distinct patterns of sequence and probably structure conservation. *Virology* **1991**, *184*, 79–86. [[CrossRef](#)]
31. Du, Z.; Chen, A.; Chen, W.; Liao, Q.; Zhang, H.; Bao, Y.; Roossinck, M.J.; Carr, J.P. Nuclear-cytoplasmic partitioning of cucumber mosaic virus protein 2b determines the balance between its roles as a virulence determinant and an RNA-silencing suppressor. *J. Virol.* **2014**, *88*, 5228–5241. [[CrossRef](#)] [[PubMed](#)]
32. Lu, Q.Y.; Yang, L.; Huang, J.; Zheng, L.; Sun, X. Identification and subcellular location of an RNA silencing suppressor encoded by mulberry crinkle leaf virus. *Virology* **2019**, *526*, 45–51. [[CrossRef](#)]
33. Grybchuk, D.; Kostygov, A.Y.; Macedo, D.H.; Votýpka, J.; Lukeš, J.; Yurchenko, V. RNA viruses in *Blechnomonas* (Trypanosomatidae) and evolution of Leishmaniavirus. *mBio* **2018**, *9*, e01932-18. [[CrossRef](#)]
34. Hillman, B.I.; Cai, G. The family *Narnaviridae*: Simplest of RNA viruses. *Adv. Virus Res.* **2013**, *86*, 149–176. [[PubMed](#)]
35. Thandapani, P.; O'Connor, T.R.; Bailey, T.L.; Richard, S. Defining the RGG/RG motif. *Mol. Cell.* **2013**, *50*, 613–623. [[CrossRef](#)] [[PubMed](#)]
36. Chong, P.A.; Vernon, R.M.; Forman-Kay, J.D. RGG/RG motif regions in RNA binding and phase separation. *J. Mol. Biol.* **2018**, *430*, 4650–4665. [[CrossRef](#)] [[PubMed](#)]
37. Cai, T.; Yu, Z.; Wang, Z.; Liang, C.; Richard, S. Arginine methylation of SARS-Cov-2 nucleocapsid protein regulates RNA binding, its ability to suppress stress granule formation, and viral replication. *J. Biol. Chem.* **2021**, *297*, 100821. [[CrossRef](#)]
38. Valente, J.B.; Pereira, F.S.; Stempkowski, L.A.; Farias, M.; Kuhnem, P.; Lau, D.; Fajardo, T.V.M.; Nhani Junior, A.; Casa, R.T.; Bogo, A.; et al. A novel putative member of the family Benyviridae is associated with soil-borne wheat mosaic disease in Brazil. *Plant Pathol.* **2019**, *68*, 588–600. [[CrossRef](#)]
39. Terefe, T.G.; Visser, B.; Botha, W.; Kozana, A.; Roberts, R.; Thompson, G.D.; Prinsloo, G.; Read, D.A. Detection and molecular characterization of Wheat stripe mosaic virus on wheat in South Africa. *Crop Prot.* **2021**, *143*, 105464. [[CrossRef](#)]
40. Gibbs, A.J.; Torronen, M.; Mackenzie, A.M.; Wood, J.T., II; Armstrong, J.S.; Kondo, H.; Tamada, T.; Keese, P.L. The enigmatic genome of *Chara australis* virus. *J. Gen. Virol.* **2011**, *92*, 2679–2690. [[CrossRef](#)]
41. Vlok, M.; Gibbs, A.J.; Suttle, C.A. Metagenomes of a Freshwater Charavirus from British Columbia Provide a Window into Ancient Lineages of Viruses. *Viruses* **2019**, *11*, 299. [[CrossRef](#)] [[PubMed](#)]
42. Kwon, S.-J.; Choi, G.-S.; Choi, B.; Seo, J.-K. Molecular characterization of an unusual new plant RNA virus reveals an evolutionary link between two different virus families. *PLoS ONE* **2018**, *13*, e0206382. [[CrossRef](#)] [[PubMed](#)]
43. Sidharthan, V.K.; Chaturvedi, K.K.; Baranwal, V.K. Diverse RNA viruses in a parasitic flowering plant (spruce dwarf mistletoe) revealed through RNA-seq data mining. *J. Gen. Plant Pathol.* **2022**, *88*, 138–144. [[CrossRef](#)]
44. Sidharthan, V.K.; Kalaivanan, N.S.; Baranwal, V.K. Discovery of putative novel viruses in the transcriptomes of endangered plant species native to India and China. *Gene* **2021**, *786*, 145626. [[CrossRef](#)] [[PubMed](#)]
45. Krishna, S.S.; Majumdar, I.; Grishin, N.V. Structural classification of zinc fingers: Survey and summary. *Nucleic Acids Res.* **2003**, *31*, 532–550. [[CrossRef](#)]
46. Schön, M.E.; Zlatogursky, V.V.; Singh, R.P.; Poirier, C.; Wilken, S.; Mathur, V.; Strasser, J.F.H.; Pinhassi, J.; Worden, A.Z.; Keeling, P.J.; et al. Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat. Commun.* **2021**, *12*, 6651. [[CrossRef](#)]
47. Carlisle, E.M.; Jobbins, M.; Pankhania, V.; Cunningham, J.A.; Donoghue, P.C.J. Experimental taphonomy of organelles and the fossil record of early eukaryote evolution. *Sci. Adv.* **2021**, *7*, eabe9487. [[CrossRef](#)] [[PubMed](#)]
48. Morozov, S.Y.; Solovyev, A.G. Origin of three related movement gene moles in plant viruses: Evolutionary radiations of benyvirus-like replicons. *bioRxiv* **2022**. [[CrossRef](#)]
49. Roossinck, M.J.; Sabanadzovic, S.; Okada, R.; Valverde, R.A. The remarkable evolutionary history of endornaviruses. *J. Gen. Virol.* **2011**, *92*, 2674–2678. [[CrossRef](#)]
50. Fukuhara, T. Endornaviruses: Persistent dsRNA viruses with symbiotic properties in diverse eukaryotes. *Virus Genes* **2019**, *55*, 165–173. [[CrossRef](#)] [[PubMed](#)]
51. Otulak-Kozieł, K.; Kozieł, E.; Escalante, C.; Valverde, R.A. Ultrastructural Analysis of Cells From Bell Pepper (*Capsicum annuum*) Infected With Bell Pepper Endornavirus. *Front. Plant Sci.* **2020**, *11*, 491. [[CrossRef](#)]
52. Roossinck, M.J. Lifestyles of plant viruses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 1899–1905. [[CrossRef](#)] [[PubMed](#)]
53. Atsumi, G.; Tomita, R.; Kobayashi, K.; Sekine, K.-T. Prevalence and genetic diversity of an unusual virus associated with Kobu-sho disease of gentian in Japan. *J. Gen. Virol.* **2013**, *94*, 2360–2365. [[CrossRef](#)] [[PubMed](#)]
54. Shaffer, C.M.; Michener, D.C.; Vlasava, N.B.; Chotkowski, H.; Botermans, M.; Starre, J.; Tzanetakis, I.E. First Report of Gentian Kobu-sho-Associated Virus Infecting Peony in the United States and the Netherlands. *Plant Dis.* **2022**, *106*, 1311. [[CrossRef](#)]
55. Dahan, J.; Wolf, Y.I.; Orellana, G.E.; Wenninger, E.J.; Koonin, E.V.; Karasev, A.V. A Novel Flavi-like Virus in Alfalfa (*Medicago sativa* L.) Crops along the Snake River Valley. *Viruses* **2022**, *14*, 1320. [[CrossRef](#)]

-
56. Schönegger, D.; Marais, A.; Faure, C.; Candresse, T. A new flavi-like virus identified in populations of wild carrots. *Arch. Virol.* **2022**, *167*, 2407–2409. [[CrossRef](#)] [[PubMed](#)]
 57. Solovyev, A.G.; Atabekova, A.K.; Lezzhov, A.A.; Solovieva, A.D.; Chergintsev, D.A.; Morozov, S.Y. Distinct mechanisms of endomembrane reorganization determine dissimilar transport pathways in plant RNA viruses. *Plants* **2022**, *11*, 2403. [[CrossRef](#)] [[PubMed](#)]