microorganisms

MDPI

Article

Identification and Full Genome Analysis of the First Putative Virus of Sea Buckthorn (*Hippophae rhamnoides* L.)

Ina Balke [1,*], Vilija Zeltina [2], Nikita Zrelovs [3], Ieva Kalnciema [2], Gunta Resevica [2], Rebeka Ludviga [1], Juris Jansons [4], Inga Moročko-Bičevska [5], Dalija Segliņa [6] and Andris Zeltins [2,*]

1 Plant Virus Protein Research Group, Latvian Biomedical Research and Study Centre, Ratsupites Street 1, k-1, LV-1067 Riga, Latvia
2 Plant Virology Group, Latvian Biomedical Research and Study Centre, Ratsupites Street 1, k-1, LV-1067 Riga, Latvia
3 Bioinformatics Core Facility, Latvian Biomedical Research and Study Centre, Ratsupites Street 1, k-1, LV-1067 Riga, Latvia
4 Cell Biology and Microscopy Core Facility, Latvian Biomedical Research and Study Centre, Ratsupites Street 1, k-1, LV-1067 Riga, Latvia
5 Unit of Plant Pathology and Entomology, Institute of Horticulture, Latvia University of Life Sciences and Technologies, Graudu Street 1, LV-3701 Dobele, Latvia
6 Unit of Processing and Biochemistry, Institute of Horticulture, Latvia University of Life Sciences and Technologies, Graudu Street 1, LV-3701 Dobele, Latvia
* Correspondence: inab@biomed.lu.lv (I.B.); anze@biomed.lu.lv (A.Z.)

Abstract: The agricultural importance of sea buckthorn (SBT; *Hippophae rhamnoides* L.) is rapidly increasing. Several bacterial and fungal pathogens infecting SBT have been identified and characterized; however, the viral pathogens are not yet known. In this study, we identified, isolated, and sequenced a virus from a wild plantation of SBT for the first time. Sequence analysis of the obtained viral genome revealed high similarity with several viruses belonging to the genus *Marafivirus*. The genome of the new virus is 6989 nucleotides (nt) in length according to 5′, 3′ RACE (without polyA-tail), with 5′ and 3′ 133 and 109 nt long untranslated regions, respectively. The viral genome encoded two open reading frames (ORFs). ORF1 encoded a polyprotein of 1954 amino acids with the characteristic marafivirus non-structural protein domains—methyltransferase, Salyut domain, papain-like cysteine protease, helicase, and RNA-dependent RNA polymerase. ORF1 was separated from ORF2 by 6 nt, encoding the coat protein (CP) with typical signatures of minor and major forms. Both CP forms were cloned and expressed in a bacterial expression system. Only the major CP was able to self-assemble into 30 nm virus-like particles that resembled the native virus, thus demonstrating that minor CP is not essential for virion assembly.

Keywords: sea buckthorn; *Hippophae*; *Marafivirus*; RNA-seq; marafibox; plant virus

Citation: Balke, I.; Zeltina, V.; Zrelovs, N.; Kalnciema, I.; Resevica, G.; Ludviga, R.; Jansons, J.; Moročko-Bičevska, I.; Segliņa, D.; Zeltins, A. Identification and Full Genome Analysis of the First Putative Virus of Sea Buckthorn (*Hippophae rhamnoides* L.). *Microorganisms* **2022**, *10*, 1933. https://doi.org/10.3390/microorganisms10101933

Academic Editor: Xiao-Wei Wang

Received: 24 August 2022
Accepted: 22 September 2022
Published: 28 September 2022

1. Introduction

Sea buckthorn (SBT, *Hippophae rhamnoides* L., genus *Hippophae* L., family *Elaeagnaceae*) is an exceptionally valuable plant that is currently domesticated and cultivated in orchards, particularly in Europe, Canada, and the USA [1]. SBT is a spinescent, deciduous, and dioecious berry-producing shrub or small tree [2]. In addition, it is a pioneer species that is highly adaptable to extreme climatic and soil conditions and is tolerant to drought and extreme temperatures ranging from −43 °C to 55 °C [3,4]. It is an ideal plant for soil erosion control, land recovery, wildlife habitat enhancement, and farmstead protection. SBT has attracted considerable attention from researchers, producers, and various industries because it has a high nutritional and medicinal value for humans [5]. SBT is not a native plant to Latvia, although fossil pollen records indicate its presence in postglacial raw soils in this region. The deliberate cultivation of SBT can be traced back to more than 100 years [6]. The first recorded trials with introduced SBT in Latvia were recorded in the 1970s when

the seedlings ('Bogatirskaja' and 'Malutka') originating from Russia (Kaliningrad region or Barnaul) were planted along the roadsides as windbreakers, as well as for recultivation of the exhausted dolomite, sand, and gravel pits [7,8]. However, it was not until 1980 that SBT was grown as a plant of agricultural importance in Latvia [9].

To date, limited research related to disease and pest control has been reported in SBT due to the comparatively short time of domestication and cultivation in large orchards [10,11]. Diseases and pests, which can alter the morphology and developmental in almost all stages of SBT, directly affect the cultivation of SBT. To date, several pests and diseases of SBT have been reported, and more are likely to be identified with the increase in the number of SBT plantations [11,12]. The major fungal diseases reported in SBT include verticillium wilt, Fusarium wilt, stem canker, damping off, brown rot, scab, and dried-shrink disease. The common pathogenic fungi reported in SBT include species of the genera *Fusarium*, *Verticillium*, *Diaporthe*, *Hymenopleella*, *Alternaria*, *Pythium*, *Monilia*, *Valsa*, and *Stigmina* [11,13]. Forty-seven fungal species affecting SBT have been reported in Russia, with *Fusarium sporotrichiella* reported to cause the maximum damage [14]. In Finland, the genus *Stigmina* was reported to be the cause of stem canker, which can kill the entire shrub of susceptible cultivars [15]. The incidence of powdery mildew in SBT has been recorded in Himachal Pradesh [16]. Fungal endophytes (*Aspergillus niger*, *Mortierella minutissima*, and sterile mycelia of *Basidiomycotina*) and spores of four species of vesicular–arbuscular mycorrhiza (*Glomus albidum*, *G. fasciculatum*, *G. macrocarpum*, and *Gigaspora margariata*) have been isolated from different SBT plant parts and associated soil samples [17]. *Emericella quadrilineata* was isolated from the leaves of *H. salicifolia* in India [18]. Moreover, root rot caused by *Rhizoctonia solani* is a major problem in greenhouses in Uttarakhand [14]. In China, species of *Fusarium* and *Phomopsis* were found to be associated with dried-shrink disease in SBT [19]. Several fungi and bacteria (*Verticillium*, *Fusarium*, *Diaporthe*, *Eutypa*, *Hymenopleella*, and *Pseudomonas* spp.) were identified as possible causes of SBT decline in Latvia [20,21]. Although many fungal species have been reported as pathogens of SBT, disease symptoms using bioassays and fulfilment of Koch's postulates have been verified only for a few pathogens, including *V. dahliae*, *F. sporotrichioides*, *F. acuminatum*, *F. oxysporum*, *F. camptoceras*, *Stigmina* sp., and *Lepteutypa* sp. [11]. However, surprisingly, we did not find any reports of viruses infecting *H. rhamnoides*.

Since 2009, various RNA sequencing methods have been developed to identify phytoviruses, including sequencing of total RNA, ribosomal-RNA-depleted total RNA, double-stranded RNA, virus-derived small interfering RNA, RNA from purified or partially purified viral particles, polyadenylated RNA, and RNA after subtractive hybridization with healthy plant RNA [22]. High-throughput sequencing (HTS) technologies have quickly become a go-to "gold standard" method for novel virus identification and monitoring in various sample types. HTS eliminates the need for prior knowledge of the expected viral genomic sequences, thus providing a substantial advantage over traditional methods such as PCR amplification or microarray hybridization, which are dependent on target-specific primers [23]. HTS allows the identification of phytoviruses from various different material types at a potentially unlimited resolution. HTS has been successfully applied for such purposes in mixed infections [24], wastewaters [25], symptomless plants [26], herbarium [27], large field surveys [28], and human feces [29]. The plant virus genus *Marafivirus* is one of the many examples in which the introduction of HTS technology has resulted in the identification of multiple novel species. Consequently, the number of recognized *Marafivirus* species has increased from 4 (*Bermuda grass etched-line virus*, *Citrus sudden death-associated virus*, *Maize rayado fino virus*, and *Oat blue dwarf virus*) in 2009 to 11 in 2020 (*Alfalfa virus F*, *Blackberry virus S*, *Grapevine-asteroid-mosaic-associated virus*, *Grapevine Syrah virus 1* (also called *Grapevine virus Q*), *Nectarine marafivirus M*, *Olive latent virus 3*, and *Peach marafivirus D*), and not counting the tentative species [30]. The *Marafivirus* genus name was derived from the reference strain of the ***Ma***ize ***ra***yado ***fi***no virus [31].

Viruses belonging to the genus *Marafivirus* within the family *Tymoviridae* are small isometric plant viruses with a monopartite positive-sense single-stranded RNA genome

(gRNA) varying from 6.3 kb to 7.1 kb in length (excluding the poly-A tail). gRNA has a 5′ cap, a 3′-part encoding a tRNA-like structure, and a poly-A tail [32]. Marafiviruses contain the "marafibox" sequence [CA(G/A)GGUGAAUUGCUUC] of 16 nt that is conserved and comparable to the "tymobox" (differs by three or four nucleotides), which has been shown to be a subgenomic promoter in tymoviruses [33]. On the basis of the predicted activity and location of the signature domains encoded within their genomes, marafiviruses are included in the alphavirus-like superfamily [34,35]. However, viruses presently classified as marafiviruses exhibit some diversity in their genome architectures [36–40]. The very narrow host range of marafiviruses makes host susceptibility an important species demarcation criterion. The genomes of marafiviruses have a high cytidine (C) content (36–45%) and usually encode a single large precursor polyprotein containing methyltransferase (MT), Salyut domain (SD), papain-like cysteine protease (PRO), helicase (HEL), RNA-dependent RNA polymerase (RdRp), and coat protein (CP) domains. CP is commonly found in virions in two forms: major (small p21 kDa) and minor (large p23–25 kDa) [41,42]. Interestingly, three CP fractions have been previously identified for oat blue dwarf virus (OBDV) and citrus sudden-death-associated virus (CSDaV) after virus purification [38,43]. Major and minor CPs were found in the virus particles at a molar ratio of approximately 3:1 [36,44]. The minor CP is initially produced as a C-terminal fusion of the replication protein, whereas the major CP is produced from subgenomic RNA (sgRNA) [41]. The in planta experiments using the infectious cDNA of CSDaV revealed that major p21 is indeed a product of direct translation by leaky scanning from the second start codon in sgRNA. The minor CPs, p25 and p23, are produced by direct translation from the first start codon in sgRNA and by trans-proteolytic cleavage processing derived from the p25 precursor, but not as the fusion part of the polyprotein [43]. For OBDV CP, the major CP is translated directly from the sgRNA, while the minor CP is cleaved from both the polyprotein and a minor CP precursor translated from the sgRNA [45]. However, for Maize rayado fino virus (MRFV), minor and major CPs are largely translated from the sgRNA [46]. Unlike other marafiviruses, Alfalfa virus F (AVF) and putative marafivirus Medicago sativa marafivirus 1 (MsMV1) do not have a second initiation codon for the coding region of the major CP1 and only encode methionine (Met) for the minor CP [47,48]. In this case, a possible strategy to produce the two CPs could be direct translation of the sgRNA for the minor CP and posttranslational cleavage of the larger precursor to produce a major CP [36]. The reason for the multiple expression strategies for minor CP is unclear. Edwards and Weiland [44] proposed that the cleavage of the replicase polyprotein does not produce stoichiometrically sufficient amounts of minor CP necessary for virion assembly. They also suggested an evolutionary transition toward CP production solely via sgRNAs and that the readthrough of a larger replicase polyprotein is vestigial. However, the production of a minor CP through a cleavage mechanism provides a regulatory feature with probable functional significance for both replication and encapsidation [44].

Marafiviruses are phloem limited and, thus, are generally not sap transmissible. MRFV has been shown to be transmissible by vascular puncture [49,50]. Marafiviruses are thought to be transmitted by leafhoppers in a persistent and propagative manner. Nevertheless, relatively few plant viruses naturally infect both insects and plant cells. Only the viral families *Rhabdoviridae*, *Reoviridae*, and *Bunyaviridae* and the genus *Marafivirus* are propagatively transmitted [51]. It was speculated that the complexity of CP expression in marafiviruses relative to that in tymoviruses might be related to the infection of both plant and insect hosts by marafiviruses [44]. Minor CP-depleted mutants of MRFV do not retain encapsidation and systemic infectivity, but they retain leafhopper transmissibility, indicating that the 37 amino acid (aa) N-terminal extension of minor CP is a leafhopper transmission signal sequence. Additionally, loss of major CP expression did not result in systemic infection [46].

In this study, we isolated an isometric virus approximately 30 nm in diameter from SBT plants obtained from a local wild plantation and sequenced its complete genome. HTS data analysis revealed that the isolated virus demonstrated genome and functional domain organization consistent with those of phytoviruses belonging to the genus *Marafivirus*. A

close relationship with the recognized and tentative *Marafivirus* representatives was further supported by phylogenetic analyses of the genome and protein sequences of the novel virus. On the basis of the complete genome sequence obtained by combining RNA-seq, RT-PCR, Sanger sequencing, and 5′ and 3′ rapid amplification of cDNA ends (RACE), as well as the genome organization of the isolated virus, we propose that the new virus can be considered a novel viral species within the genus *Marafivirus* and should be named as Sea buckthorn marafivirus (SBuMV) isolate BU1. Additionally, cloning and expression of both *CP* gene product variants (minor and major) of SBuMV in the *E. coli* expression system were investigated.

2. Materials and Methods

2.1. Plant Source

Leaf material of SBT was collected in August 2017 (Figure S1A) and August 2021 (Figure S1B–D) from the same three individual SBT shrubs with leaves bearing necrotic spots. SBT source plants were sampled at a wild plantation grown since at least the 1950s at the Bul̦lupe riverbank to control soil erosion (GPS location:57.018110N; 24.004251E).

2.2. Virus and RNA Purification

The viral fraction was separated from 30 g of leaf material of each SBT sample that was collected in August 2017. Leaves were grinded in the electric mixer ETA 0010 (ETA, Prague, Czech Republic) in 100 mL of homogenization buffer (0.02 M HEPES (pH 6.8), 0.2 M sucrose, 20 mM sodium azide (NaN_3), 10 mM β-mercaptoethanol (β-ME), 1 mM PMSF, 2 mM EDTA). The suspension was centrifuged in a low-speed centrifuge 5804R (Eppendorf, Hamburg, Germany) at 11,000 rpm (15,557× *g*) and 4 °C for 15 min. The resulting supernatant was filtered using a filter paper. TX-100 was added to the clarified supernatant drop by drop on a magnetic stirrer at 4 °C until the 5% saturation was reached. The solution was then loaded on a 20% sucrose cushion in a 0.02 M HEPES buffer (1:1 volume ratio; pH 8.2) and sedimented by ultracentrifugation on Optima-XL (Backman Coulter, Brea, CA, USA) using a Type-70Ti rotor (Backman Coulter, USA) at 50,000 rpm (183,960× *g*) and 4 °C for 1.5 h. The obtained supernatant was discarded, and the pellet was dissolved in 6 mL of buffer (0.1 M sodium borate (NaB; pH 6.8), 5 mM EDTA, overnight (ON)) at 4 °C. The solution was then clarified by centrifugation in a low-speed centrifuge 5804R (Eppendorf, Hamburg, Germany) at 10,000 rpm (12,857× *g*) and room temperature for 5 min. Soluble fraction (6 mL) was further purified on a sucrose gradient containing layered sucrose fractions ranging from 10% to 40% (with intervals of 10%) in 0.1 M NaB (pH 8.2). Sucrose gradient centrifugation was performed as previously described [52]. After sucrose gradient centrifugation, 6.5 mL fractions from the bottom of the tubes were carefully collected. Sucrose fraction analysis was performed using sodium dodecyl sulphate-polyacrylamide gel (12.5%) electrophoresis (SDS-PAGE) followed by Coomassie G250 (Sigma-Aldrich, St. Louis, MO, USA) staining. Two of the fractions (40% and 30%) were dialyzed against 100 volumes of 0.05 M NaB (pH 8.2) ON on a magnetic stirrer. Dialyzed solution was loaded on a 20% sucrose cushion in 0.05 M NaB and 2 mM EDTA and sedimented by ultracentrifugation on Optima-XL (Backman Coulter, Brea, CA, USA) in Type-70Ti rotor at 50,000 rpm (183,960× *g*) and 4 °C for 4 h. The supernatant was removed, and the pellet was dissolved in 0.3 mL solution containing 0.05 M NaB and 2 mM EDTA. The contents of the dissolved pellet solution were further analyzed using transmission electron microscopy (TEM). RNA was extracted from 100 µL of purified virus sample using the innuPREP Virus DNA/RNA Kit (Analytik Jena, Jena, Germany) according to the manufacturer's protocol. The purified RNA was then analyzed using a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) to evaluate the average RNA concentration and assess the quality of the isolated RNA at 260/280 nm. Next, the purified RNA specimen was analyzed using an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with an Agilent RNA 6000 Pico kit (Agilent Technologies, Santa Clara, CA, USA) to evaluate the fragmentation level of extracted RNA and length distribution of the obtained RNA

fragments. Precise total RNA concentration was determined using a Qubit 2.0 (Thermo Fisher Scientific, Waltham, MA, USA) with a Qubit RNA high sensitivity (HS) assay kit (Thermo Fisher Scientific, Waltham, MA, USA).

For leaf samples collected in August 2021, total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Purified RNA was then analyzed using a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) to assess the quality of isolated RNA at 260/280 nm. RNA concentration was determined using a Qubit RNA broad range assay kit (Thermo Fisher Scientific, Waltham, MA, USA) on a Qubit 2.0.

2.3. Transmission Electron Microscopy (TEM)

Purified virus or VLP samples were adsorbed onto carbon-formvar-coated grids (300 mesh Copper/Palladium 3.05 mm; Laborimpex, Forest, Belgium) or carbon-formvar-coated grids (hexagonal 200 mesh nickel 3.05 mm; Laborimpex, Forest, Belgium) for 1 min. The grids were then washed three times with 1 mM EDTA solution by brief immersion in the solution. The droplet was removed using a filter paper. Washed grids with the adsorbed samples were then negatively stained with a 0.5% uranyl acetate aqueous solution for 30 s. The stained droplet was removed using a filter paper. The samples were allowed to dry on filter paper in a Petri dish for 1 h before performing TEM. Grids were examined using a JEM-1230 electron microscope (JEOL, Tokyo, Japan) at an accelerating voltage of 100 kV.

2.4. Sea Buckthorn Marafivirus (SBuMV) gRNA RNA-seq Library Preparation for HTS on the MGI Platform

The HTS library was generated for the viral RNA extracted from the pooled plant samples collected in August 2017 using the MGIEasy RNA Directional Library Prep Set for 16 reactions (MGI, Shenzhen, China) according to the manufacturer's protocol for pair-end reads of 150 bp. Libraries were verified on an Agilent 2100 bioanalyzer with a HS DNA Kit (Agilent Technologies, Santa Clara, CA, USA) for the average fragment length calculation and evaluation of quality and on Qubit 2.0, with Qubit 1X dsDNA HS Assay Kits (Thermo Fisher Scientific, Waltham, MA, USA) for the measurement of library fragment concentration. Library pooling, circularization, and cleaning were performed according to protocol guidelines. The final concentration of the resultant library was measured with a Qubit ssDNA Assay kit (Thermo Fisher Scientific, Waltham, MA, USA), and HTS was performed on a flow cell PE150 (MGI, Shenzhen, China) using a DNBSEQ-G400 (MGI, Shenzhen, China) system.

2.5. SBuMV Genome De Novo Assembly and Preliminary Annotation

Demultiplexed read data were inspected using FastQC v0.11.5 [53]. Transcriptome de novo assembly was performed using rnaSPAdes v3.13.1 [54]. Initial functional annotation of transcripts was performed using the ORF finder [55] and conserved domain search [56] of NCBI within the putative products of the predicted ORFs. On the basis of the conserved domain content of their putative products, transcripts of presumably viral origin were subjected to a BLASTn [57] search against the non-redundant nucleotide database (nr/nt) restricted to sequences of viral origin (taxid:10239) to gain an overview of their placement within the known virus diversity. Pairwise alignment of viral transcripts to those of the respective highest-scoring BLASTn hits was performed using an EMBOSS Needle [58].

2.6. SBuMV gRNA 5′ and 3′ RACE and gRNA RT-PCR Fragment Verification by Sanger Sequencing

Prior to starting the 5′/3′-end RACE, the isolated viral RNA was tested by one-step RT-PCR using the Verso 1-Step RT-PCR Kit (Hot Start) (Thermo Fisher Scientific, Waltham, MA, USA) and nested primers—MarBox1F and MarBox1R. The corresponding PCR product (428 bp) was extracted from the gel after analysis in 0.8% native agarose gel (NAG) using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA) and cloned

into the linearized vector pTZ-57 using an InsTAclon PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA). Conventional sequencing transformations into XL1-Blue Super competent cells (Agilent Technologies, Santa Clara, CA, USA) were used for all ligations. PCR-fragment-containing clones were selected by digestion analysis using NcoI and XhoI restriction enzymes (Thermo Fisher Scientific, Waltham, MA, USA). Three positive clones were sequenced by the Sanger sequencing method using the ABI PRISM BigDye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA) and ABI PRISM 3130xl sequencer (Thermo Fisher Scientific, Waltham, MA, USA) with the corresponding primers M13seq-F and M13seq-R. Sequence assemblies were prepared using SeqMan software. To determine the complete genome sequence of SBuMV up to the last base, the 5′ and 3″ untranslated regions (UTRs) of the virus were determined using a SMARTer®RACE 5′/3′ Kit (Takara Bio, Kusatsu, Japan). Viral RNA isolated in August 2017 was used for UTR elucidation and resequencing. First-strand cDNA was amplified according to the manufacturer's protocol. The 5′ (Figure 1C, track 1) and 3′ (Figure 1C, track 7) ends were amplified with the genome-specific primers MetPro-seq2-R and MarRdRp-seq2-F, respectively, and 10xUPS (SMARTer®RACE 5′/3′ Kit). An additional amplification step was employed because PCR products could not be initially detected in NAG. The second PCR was performed with Green Phusion polymerase (GFpol, Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol using the same template as used for the first PCR, the MarCP-2F (3′ end RACE), and short UPS (SMARTer®RACE 5′/3′ Kit) or the first-strand cDNA, primer MetPro-seq1-R (5′ end RACE), and short UPS. PCR reactions were analyzed in 0.8% NAG, and the corresponding PCR products were extracted from the gel; adenine overlaps were added using Taq polymerase (Thermo Fisher Scientific, Waltham, MA, USA) for cloning into the linearized vector pTZ-57 or directly after the purification into the linearized vector pJET1.2 using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA). Insert-containing clones were selected by restriction testing using EcoRI and HindIII restriction enzymes. At least three clones with the characteristic restriction pattern and direct PCR products were sequenced using the primers M13seq-F and M13seq-R for pTZ57 and pJET1.2-F and pJET1.2-R for pJET1.2. The generated sequences were aligned to the de novo assembled sequences from HTS data. The corresponding replication-associated polyprotein (RP) domains and CP were further amplified and verified by Sanger sequencing (Figure 1). The 5′ UTR-Pro fragment (Figure 1C, track 2; 2023 bp) was amplified using 5′ RACE cDNA with GFpol in GC buffer and the primers Mar-5UTR-RACE-F and MetPro-seq2-R. The PRO domain (Figure 1C, track 3; 1775 bp) was amplified using 5′ RACE cDNA with GFpol in GC buffer and the primers Mar-Pro2-F and Mar-Pro2-R. The Hel-RdRp domain (Figure 1C, track 4; 2861 bp) was amplified using 3′ RACE cDNA with GFpol and primers Mar-Hel-F and Mar-RdRp-R, and the CP-3′ UTR fragment (Figure 1C, track 6; 975 bp) was amplified using 3′ RACE cDNA with GFpol in GC buffer and primers Mar-CPL-F and Mar-3UTR-RACE-R. The RdRp-C-term-CP fragment (Figure 1C, track 5; 1066 bp) was amplified using 3′ RACE cDNA with GFpol in GC buffer and the primers SB-RdRp-CP-seq-F and MarCP2R. All corresponding PCR products were extracted from NAG and cloned into pTZ-57 or pJET1.2 cloning vectors. Clone selection and sequencing were performed as described above. MT pTZ-57 clones were sequenced with M13seq-F and M13seq-R; PRO pTZ-57 clones were sequenced with M13seq-F, M13seq-R, and internal primers—MetPro-seq2-F and MetPro-seq2-R—employing 5% dimethyl sulfoxide (DMSO) and a heating step (5 min at 95 °C) before the main sequencing program [59]. HEL-RdRp pJET1.2 clones were sequenced using the pJET1.2-F, pJET1.2-R, and internal primers—MarHel-seq3F and MarHel-seq4F—with a prior heating step and 5% DMSO, MarRdRp-seq2-F, MarRdRp-seq2-R, MarRdRp-seq3-R, and MarRdRp-seq5-R. CP-3′-UTR pJET1.2 clones were sequenced using pJET1.2-F and pJET1.2-R. The 5′-UTR-Pro pJET1.2 clones were sequenced using pJET1.2-F, pJET1.2-R, and the internal primers—MetPro-seq1-F and MetPro-seq1-R. RdRp-C-term-CP pJET1.2 clones were sequenced using pJET1.2-F and pJET1.2-R. All the primers used are listed in Table S1.

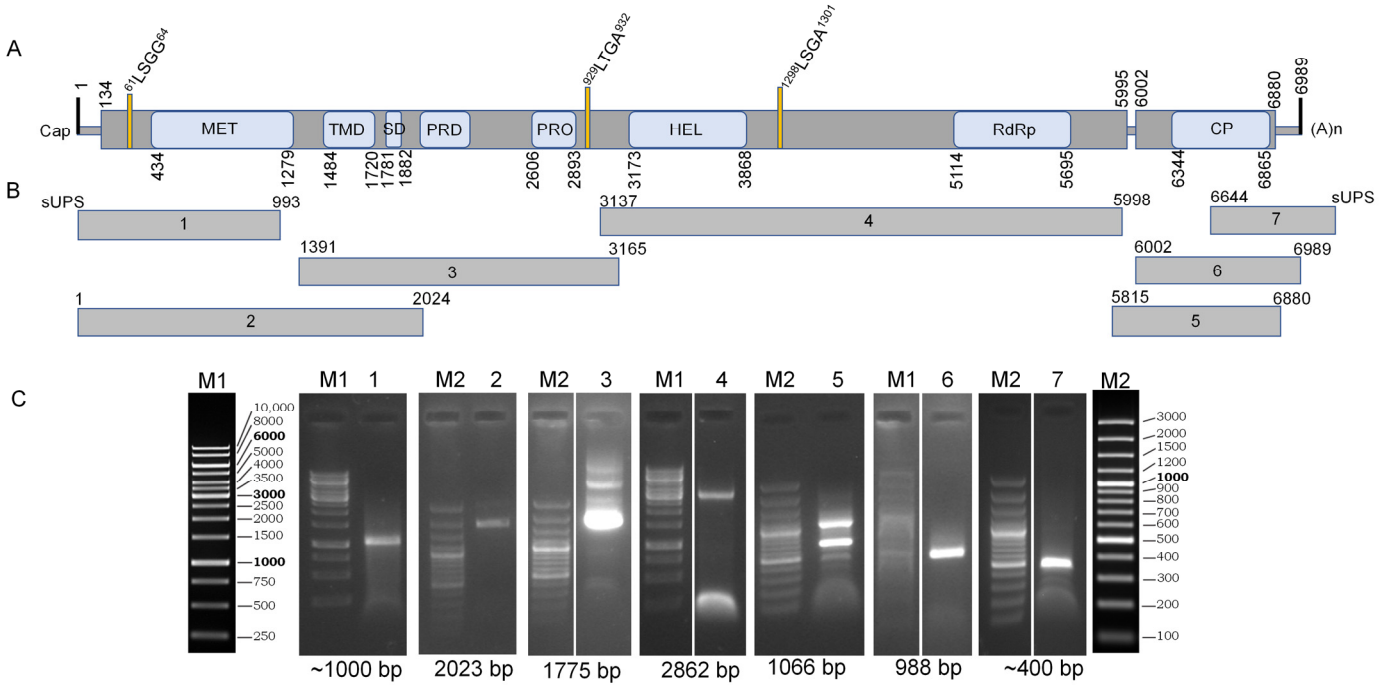

Figure 1. Schematic diagram of the SBuMV genomic sequence verification. MET—methyltransferase; TMD—transmembrane domain; SD—Salyut domain; PRD—proline-rich domain; PRO—papain-like cysteine protease; HEL—helicase, RdRp—RNA-dependent RNA polymerase; CP—coat protein. (**A**) Schematic genome diagram; (**B**) schematic SBuMV genome verification schema; (**C**) RT-PCR fragment analysis in native agarose gel. M1—GeneRuler 1kb DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA); M2—GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA) (see Table S1 for additional details).

2.7. Tymoviridae Sequence Dataset Acquisition

For reconstruction of SBuMV evolutionary relationships with other *Tymoviridae* viruses, a dataset (Table S2) comprising public biological sequence repository accessions was additively built as follows: (1) exemplar isolates of species belonging to the family *Tymoviridae* from the Virus Metadata Repository number 18 of ICTV (19 October 2021; MSL36); (2) *Tymoviridae* entries from RefSeq [60] (including those that do not yet have a standing in the official virus taxonomy as per ICTV Master Species List 2021.v1 (1 April 2022)); (3) *Tymoviridae* entries longer than 6000 bases representing presumably complete or near-complete genomes from the NCBI nucleotide database [61]. Respective nucleotide sequences were retrieved along with their taxonomy and the replicase/polyprotein and/or capsid protein sequences they encode (Table S2). As some recognized and tentative *Marafiviruses* are known to have their CP encoded within the polyprotein gene, polyprotein sequences from entries that did not have an individual ORF encoding for CP were subjected to conserved domain prediction using batch CD-Search under default settings [62]. Entries with a discernible CP domain at the C-terminus of the polyprotein sequence were then partitioned by moving the last 300 aa of the polyproteins into the presumable CP aa sequence entries. Thus, for entries with CP encoded within the C-terminus of replicase polyprotein, truncated polyprotein aa sequence (without the last 300 aa; qualifier "_woCP" added to the original protein sequence accession in the label) was used for replicase phylogenies, and C-terminal sequences of 300 aa were used for CP phylogenies (qualifier "_300last" added to the original protein sequence accession in the label). MAFFT v7.453 [63] was used to generate multiple sequence alignments (MSAs) of the (1) complete or near-complete genome (hereafter referred to as genome MSA and, accordingly, genome tree), (2) polyprotein (without CP domain) aa, and (3) CP (including polyprotein-derived CP

sequences) aa in automatic mode. Thereafter, each MSA was used to generate the respective maximum-likelihood and neighbor-joining trees.

2.8. Evolutionary Relationship Analysis of the Recognized and Tentative Tymoviridae Representatives

Maximum likelihood (ML) trees were constructed using IQ-TREE v. 2.0.3 [64], and ModelFinder [65] was used for the best substitution model selection according to the Bayesian information criterion, allowing for polytomies and using 1000 ultrafast bootstrap (UFBoot; [66]) replicates to determine the branch supports. Neighbor-joining (NJ) trees [67] were constructed using MEGA v 7.0.26 [68], eliminating all MSA positions with less than 90% site coverage, assuming uniform substitution rates, and determining branch supports using 1000 bootstrap [69] replicates. The trees were rooted using the respective outgroup (Botrytis virus F) sequences and visualized using FigTree v 1.4.4 [70]. Respective ML and NJ trees were then placed side-by-side and annotated using Inkscape v 1.0.1 [71], with distal nodes of the well-supported branches ($\geq$95% UFboot for ML trees and $\geq$80% bootstrap for NJ trees) being colored in green. The technical parameters of the MSAs and inferred trees are shown in Table S3.

2.9. Expression, Purification, and Analysis of SBuMV CP Cloned into the Bacterial Expression Vector

Minor (p31) and major (p21.2) *CP* genes, which were obtained from the RNA extracted from the samples collected in August, 2017, were amplified by RT-PCR using the Verso 1-Step RT-PCR Kit with Thermo-Start Taq (Hot Start; Thermo Fisher Scientific, Waltham, MA, USA) using primers Mar-CPL-F and MarCP2R (876 bp) for the *p31* variant, and MarCP1F and MarCP2R (602 bp) for the *p21.2* variant. PCR products were cloned into the pTZ-57R/T vector (Thermo Fisher Scientific, Waltham, MA, USA). Clones containing *CP* inserts were identified by analyzing the pattern generated upon digestion with restriction enzymes NcoI and HindIII and confirmed by Sanger sequencing using M13seq-F and M13seq-R primers. Plasmid-harboring clones containing a *p31* or *p21.2* insert were digested with NcoI and HindIII, and DNA fragments were purified and cloned into the NcoI and HindIII sites of *E. coli* expression vector pRSFDuet1 (Novagen, Madison, WI, USA), resulting in the plasmids pRSFDu-p31 or pRSFDu-21.2, respectively. For p31 and p21.2 co-expression, *p21.2* was amplified with the MARCP-short-Nde-F and MARCP-short-Xho-R primers. The PCR products were cloned into the pTZ-57R/T vector. Clones containing the *p21.2* insert were identified by analyzing the pattern generated upon digestion with restriction enzymes NdeI and XhoI and confirmed by Sanger-based sequencing using M13seq-F and M13seq-R primers. A plasmid-harboring clone containing a *p21.2* insert was digested with NdeI and XhoI, and DNA fragments were purified and cloned into the NdeI and XhoI sites of the *E. coli* expression vector pRSFDu-p31, resulting in the plasmid pRSFDu-p31-p21.2. Plasmid clones without sequence ambiguities were transformed into the C2566 *E. coli* expression strain (New England Biolabs, Ipswich, MA, USA). Cell cultivation and expression of CPs as well as cell disruption conditions were the same as previously described for cocksfoot mottle virus and rice yellow mottle virus [52]. In brief, VLPs containing *E. coli* were disrupted by ultrasound in 1$\times$ PBS, 5 mM β-ME, and 0.5% TX-100 buffer with the ultrasonic lab devices UP200S (Hielscher Ultrasonics, Teltow, Germany) at a period of 0.5 and intensity 70% for 16 min. The cell lysate was clarified by centrifugation at 11,000 rpm (15,557$\times$ *g*) for 10 min. The soluble fraction further was purified on sucrose gradient (sucrose % in fractions 60%, 50%, 40%, 30%, and 20%) by ultracentrifugation on Optima-XL in SW-32 rotor (Backman Coulter, Brea, CA, USA) at 25,000 rpm (106,559$\times$ *g*) and 18 °C for 6 h. SBuMV CP from sucrose fractions (40% and 30%) were collected by ultracentrifugation on Optima-XL in Type-70Ti rotor at 50,000 rpm (183,960$\times$ *g*) and 4 °C for 4 h. The supernatant was removed, and the pellet was dissolved in 3 mL of 1$\times$ PBS. All samples were analyzed by NAG and SDS-PAGE and purified CP samples with TEM.

2.10. SBuMV Detection by RT-PCR

The total RNA isolated from SBT samples collected in August 2021 was used for SBuMV detection. One-step RT-PCR using a SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) was performed using a combination of primers (Mar-CPL-F and MarCP2R (876 bp)) designed specifically for minor *CP* amplification.

3. Results

3.1. Virus Purification from SBT Leaf Samples

The leaf material collected from the SBT shrubs was used for virus purification. Homogenization buffer (100 mL) used for leaf homogenization was concentrated 333 times using several purification steps, including sucrose cushions and sucrose gradients. TEM analysis revealed the presence of icosahedral viral particles in all the samples (red arrows next to that Figure 2A part).

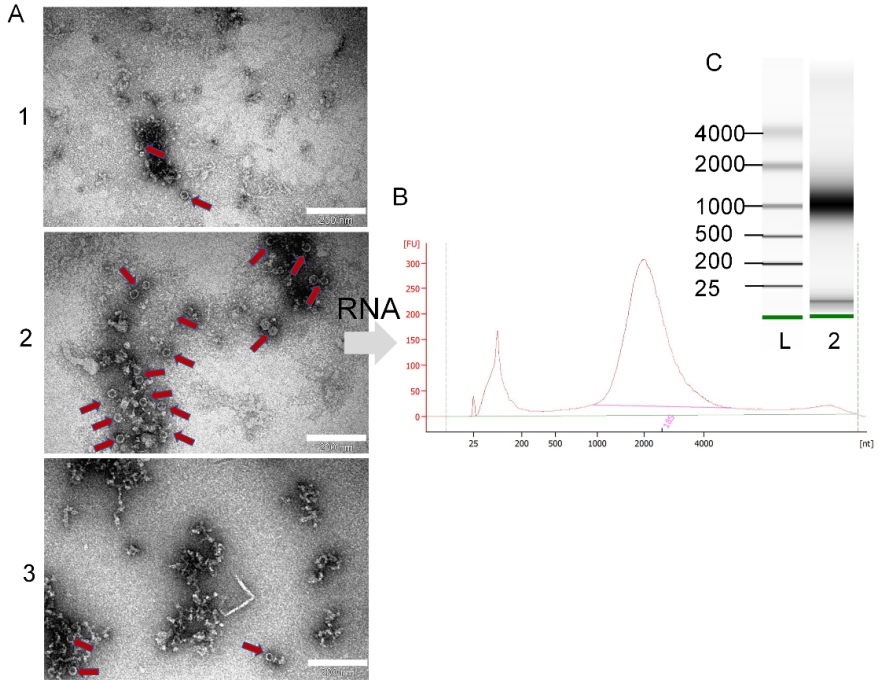

Figure 2. Purification of viruses from SBT and analysis of isolated RNA. (**A**) TEM analysis of purified and concentrated SBT samples, scale 200 nm; (**B**) chromatogram of isolated RNA from corresponding Bioanalyzer assays; (**C**) electropherogram of isolated RNA from corresponding Bioanalyzer assays; 1–3—SBT samples; L—RNA ladder (Agilent Technologies, Santa Clara, CA, USA).

3.2. SBuMV Genome Assembly

Demultiplexed HTS read data inspection using FastQC v0.11.5 [53] revealed that the sequencing run yielded 13,184,928 read (of up to 150 bp in length) pairs that were considered to be of sufficient quality for downstream processing. Transcriptome de novo assembly was performed using rnaSPAdes v3.13.1 [54], generating 147,111 transcripts of up to 47,770 bp long. Four of the de novo assembled transcripts (6895–6959 bp; 123–178 k-mer coverage) were identified to represent the near-complete genome variants of a putative novel *Tymoviridae* representative on the basis of the characteristic presence of the conserved domain in their aa sequence of the putative ORF product. As these transcripts differed by multiple indels and single nucleotide polymorphisms, indicating extensive quasispecies within the host, oligonucleotide primers for both genome termini RACE and validation of the genome were designed on the basis of the sequence of one of these near-complete genome transcripts that was selected randomly. After Sanger-based sequencing and mapping of the obtained reads

onto the selected transcript that validated the sequence and allowed for the extension of the transcript, the genome of this novel putative *Tymoviridae* family representative was determined to constitute 7020 bases (including 3′ poly-A tail) with a 53.8% GC content that had an average HTS read coverage of 819×. ORF prediction and functional annotation were performed using ORF finder tool [55] and conserved domain search of NCBI [56]. BLASTn [57] search against the nr/nt database restricted to the sequences of viral origin (taxid:10239) showed the highest total-scoring hit to the complete genome (7148 bases) of OLV-3 isolate CN/1/1 (Accession: FJ444852.2), which had a query coverage and percent identity (e-value 1e-150) of 60% and 68.81%, respectively. Lower-scoring hits were found to isolates of other marafiviruses (e.g., grapevine-asteroid-mosaic-associated virus (GAMaV), OBDV, Nectarine marafivirus M (NeVM), and CSDaV). Pairwise alignment of this novel virus genome to that of OLV-3 using EMBOSS Needle [58] showed a length of 8065 bases, 4630/8065 (57.41%) identities, and 1962/8065 (24.33%) gaps. The full sequence of SBuMV was submitted to GenBank on 31 March 2022 and was assigned the accession number ON149451.1.

3.3. SBuMV Evolutionary Relationships with Other Viruses

In all three of the tree pairs (genome sequence, CP, and replicase aa sequence), SBuMV shared a well-supported most recent common ancestor (MRCA) with OLV-3 isolate CN1/1 and was grouped in a major *Marafivirus* clade comprising all the taxonomically recognized marafiviruses, with the exception of AVF and some tentative marafiviruses (e.g., MsMV1, Davidia involucrata marafivirus 1 (DiMV1), and Glehnia littoralis marafivirus; Figure 3, Figures S2 and S3).

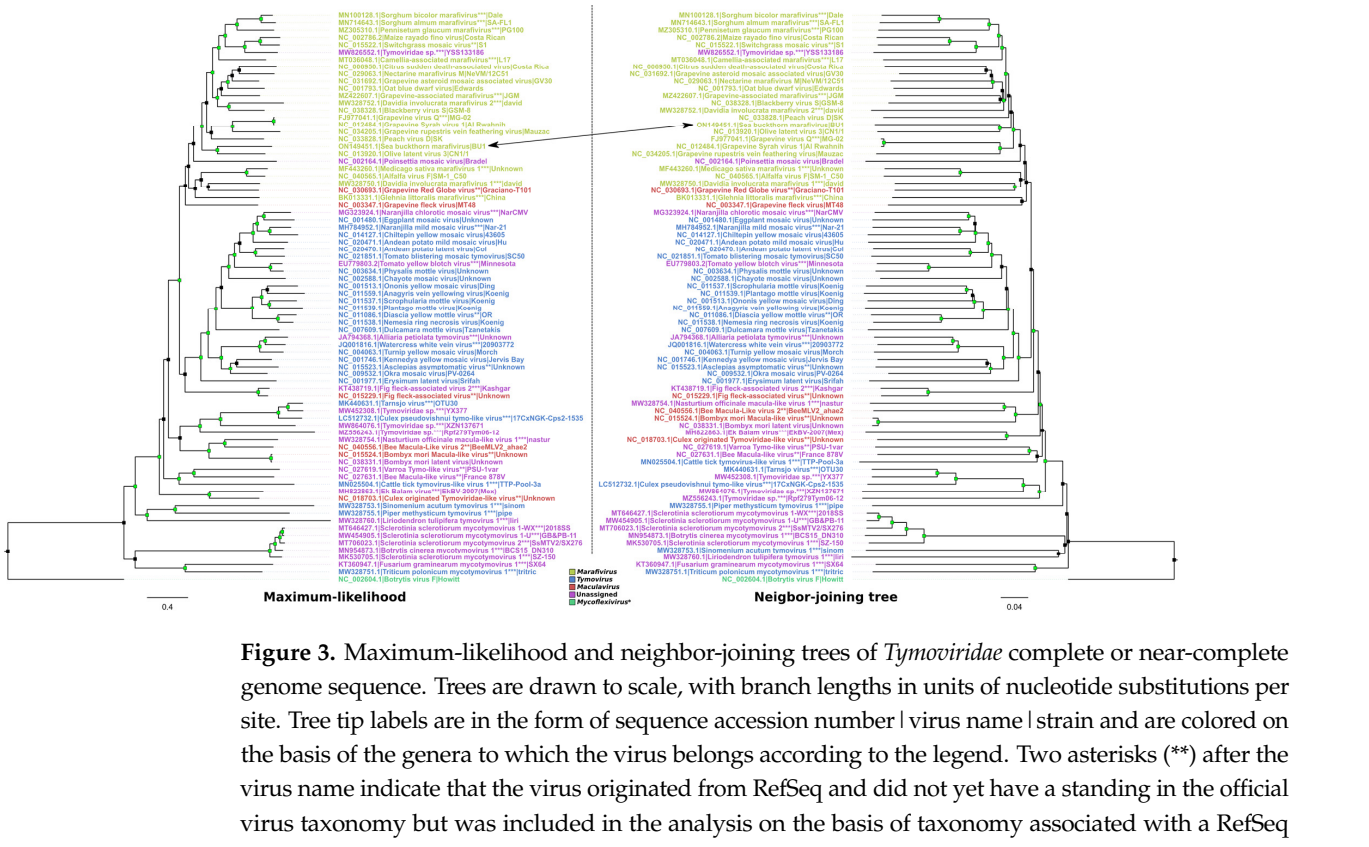

Figure 3. Maximum-likelihood and neighbor-joining trees of *Tymoviridae* complete or near-complete genome sequence. Trees are drawn to scale, with branch lengths in units of nucleotide substitutions per site. Tree tip labels are in the form of sequence accession number | virus name | strain and are colored on the basis of the genera to which the virus belongs according to the legend. Two asterisks (**) after the virus name indicate that the virus originated from RefSeq and did not yet have a standing in the official virus taxonomy but was included in the analysis on the basis of taxonomy associated with a RefSeq entry. Three asterisks (***) after the virus name indicate that the sequence originated from the Nuccore database and did not yet have a standing in the official virus taxonomy but was included in the analysis on the basis of taxonomy associated with a complete or near-complete genome GenBank entry that was longer than 6000 bases in length. Botrytis virus F, which is a member of family *Gammaflexiviridae*, genus *Mycoflexivirus*, serves as an outgroup at which the trees are rooted. The black arrow connects Sea buckthorn marafivirus (SBuMV) leaf in both of the trees.

This aforementioned major clade containing all the officially recognized marafiviruses (except for the AVF) was well supported (≥80% bootstrap support for the NJ trees; ≥95% UFBoot support for ML trees) in all the trees, except for the CP ML tree, and also included sequences from an unassigned *Tymoviridae* representative (isolate YSS133186) that had a well-supported common ancestry with MRFV isolate Costa Rican and some of the tentative marafiviruses. This suggests the possibility of classification of *Tymoviridae* isolate YSS133186 as a representative of *Marafivirus* genus. Interestingly, the AVF isolate SM-1_C50, which is recognized as a marafivirus, was not grouped in the large *Marafivirus* clade with sufficient support in either of the six trees, along with some other yet unrecognized tentative marafiviruses (based on the respective sequence submission-associated taxonomy; e.g., MsMV1, DiMV1 isolate david, and Glehnia littoralis marafivirus isolate China) that tended to cluster together, although they did not have a well-supported MRCA. Additionally, all the ICTV-recognized *Tymovirus* representatives formed a well-supported clade in all the trees, except for the CP NJ tree, with two of the tentative *Tymovirus* representatives forming a distinct clade comprising Sinomenium acutum tymovirus 1 isolate sinom and Piper methysticum tymovirus 1 isolate pipe (only shown in the genome and polyprotein tree as no CP sequence was readily available). Most of the "mycotimoviruses" that were listed as tentative *Tymoviridae* representatives without a genus-level assignment based on the submitted sequence-associated taxonomy [72] also formed a well-supported distinct clade in genome and polyprotein trees and might represent a putative novel genus. In all trees, various arthropod-associated tentative *Tymoviridae* representatives formed several distinct well-supported clades comprising a small number of leaves, and the majority of the sequences representing these leaves did not have a genus assigned to them. However, some of them were provisionally assigned to *Maculavirus* (e.g., Macula-Like virus 2, Bombyx mori Macula-like virus, and Culex-originated *Tymoviridae*-like virus) and *Tymovirus* (e.g., Tarnsjo virus isolate OTU30, cattle tick tymovirus-like virus 1, and Culex pseudovishnui tymo-like virus) genera as per NCBI taxonomy [72]. Overall, the results of our phylogenetic analyses suggest that some of the *Tymoviridae*-related virus sequences, without standing in the official virus taxonomy, have accumulated in the public biological sequence repositories, which, we believe, requires the attention of the ICTV representatives to formalize their place within the scope of the official virus taxonomy, thus encouraging further *Tymoviridae*-like virus diversity studies and clarifying the uncertainty regarding the intrafamily relationships of tentative *Tymoviridae* representatives that have been recently revealed.

3.4. SBuMV Genome Annotation, Resequencing, and 5′ and 3′ End Mapping with RACE

To complete the de novo assembled viral genome, the 5′ and 3′ terminal ends and internal genome fragments were resequenced and validated using Sanger-based sequencing. To sequence both flanking ends of the viral genome, we used the switching mechanism at 5′ end of RNA template (SMART) RACE method. The 5′ or 3′ ends were amplified with custom SBuMV nested primers designed using processed RNA-seq data (de novo assembled transcripts) and SMARTer® RACE 5′/3′ kit primers. Custom primers for nested SBuMV PCR fragment amplification and resequencing were designed on the basis of de novo assembled near-complete SBuMV genome transcripts. At least four successful 5′ RACE clones were sequenced, and according to the obtained sequence alignment onto the de novo assembled transcripts from HTS data, the 5′ UTR was determined to be 133 nt long, starting with guanidine (G). The G at the first position of gRNA could indicate that the genome of SBuMV and other completely sequenced marafiviruses that contain G as the first nt [36–40,48,73–78] of their gRNA is 5′ capped when compared with other members of the family *Tymoviridae*, which had G at the gRNA 5′ end and were capped with m7G [41]. RNAfold web server [79] analysis demonstrated three putative hairpin structures (Figure 4D) that showed similar secondary RNA structure to that of tymoviruses [79,80]. The 3′ UTR was 109 nt long (without poly-A tail), demonstrating a putative tRNA-like secondary structure as determined by the RNAfold web server (Figure 4C). Furthermore,

the complete genome of SBuMV was validated by resequencing several gRNA segments (Figure 2). As revealed by the HTS data mapped onto the assembled transcripts and Sanger-based resequencing, virtually every genomic feature except for the 5′ and 3′ UTRs (which were missing from the initial de novo assembled transcripts) showed a plethora of nt substitutions, some of which were nonsynonymous and resulted in aa changes for ORF products in some of the putative genotypes. However, a high heterogeneity of RNA virus populations termed "viral quasispecies" is not an uncommon phenomenon, which has previously raised questions as to what extent a single RNA virus population-derived sequence, representing major allele frequencies, can describe the mutant clouds present in the sample in reality [81]. Determination of the 5′ and 3′ genome termini showed that the full-length SBuMV genome is a monopartite RNA molecule consisting of 6989 nt (excluding the poly-A tail).

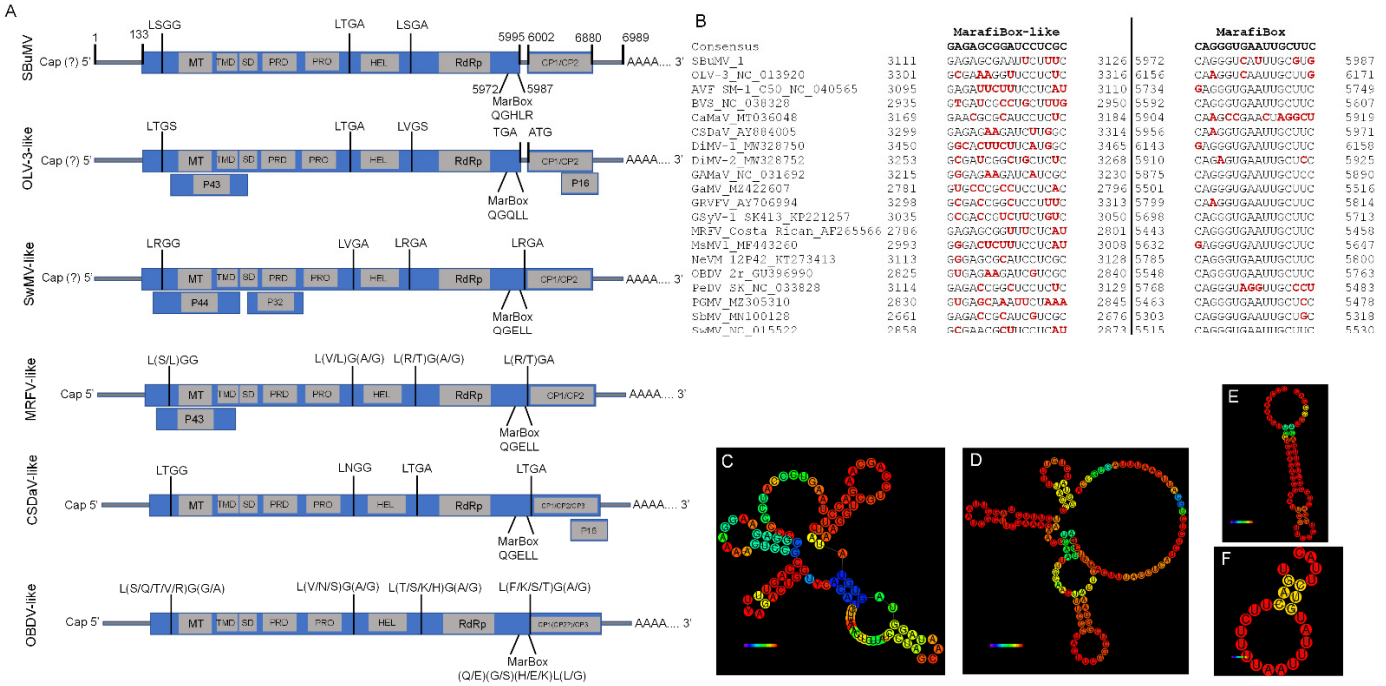

Figure 4. SBuMV genomic features and genome organization. (**A**) Schematic representation of marafivirus genome organizations. MT—methyltransferase; SD—Salyut domain; PRO—protease; HEL—helicase; RdRp—RNA dependent RNA polymerase; MarBox—"marafibox" represented as aa; (**B**) 16 nt "marafibox" and "marafibox-like" sequence alignment; (**C**) predicted RNA structure of SBuMV 3′ UTR; (**D**) predicted RNA structure of SBuMV 5′ UTR; (**E**) predicted secondary RNA structure of SBuMV "marafibox-like" sequence; (**F**) predicted secondary RNA structure of SBuMV "marafibox".

Marafiviruses discovered thus far have demonstrated at least six different genome organizations (Figure 4A). Here, we classified them according to the well-studied marafiviruses: MRFV-like, CSDaV-like, OLV-3-like, OBDV-like, switchgrass mosaic virus (SwMV)-like, and SBuMV-like. The newly identified SBuMV genome organization resembles that of OBDV, but with a separate ORF encoding a CP, which has been observed only in OLV-3 among all *Marafivirus* representatives. The OBDV-like genome organization group consists of six taxonomically recognized members (peach marafivirus D (PeDV), AVF, blackberry virus S (BlVS), NeVM, GAMaV, and grapevine rupestris vein feathering virus (GRVFV)) and several tentative *Marafivirus* members (MsMV1, DiMV1, Davidia involucrata marafivirus 2 (DiMV2), Camellia-associated marafivirus (CaMaV), Sorghum bicolor marafivirus (SbMV), grapevine-associated marafivirus (GaMV), and Pennisetum glaucum marafivirus (PGMV)). The MRFV-like genome organization is also characterized by grapevine Syrah virus 1 (GSyV-1). The CSDaV-like and OLV-3-like genome organization groups only contained their

reference members. A new tentative member of the genus *Marafivirus*, SwMV, possesses a fifth genome organization type (SwMV-like) with the main ORF encoding a polyprotein, as in OBDV-like group, but has two additional ORFs located in close proximity and nested within the polyprotein-encoding ORF (at the 5′ end). This shows that SBuMV genome organization could be viewed as a sixth distinct type of marafivirus genome organization, with SBuMV being the only known marafivirus to possess it. SBuMV ORF1 begins at the 134th nt with methionine and ends at the 5998th nt with a stop codon UAA. The RP was 1954 aa long, with a calculated molecular mass of 216.5 kDa. RP-encoded domains were determined using the conserved domain database (CDD) [82], suggesting that SBuMV ORF1 contains at least five distinct protein domains: viral MT (pfam01660; 101–382 aa), Salyut domain (cl41199; 550–583 aa), tymovirus endopeptidase (cl05113; 825–920 aa), viral (superfamily 1) RNA HEL (pfam01443; 1014–1245 aa), and RdRp (cl03049; 1661–1854 aa). MSA of SBuMV RP and RP of other officially recognized and selected tentative marafiviruses (20 in total) using the PROMALS3D server [83] allowed us to identify many characteristic domain motifs. Hence, all motifs corresponding to the MT domain (motifs I, II, and III) [84] were identified. Additionally, PRO motifs I (CLL) and II (H(F/Y)), which are conserved between tymo-, marafi-, and macula-viruses and contain a catalytic dyad (C^{829} and H^{918}), HEL motifs I to VI, and RdRp motifs I to VIII, could be identified according to the OBDV model (File S1) [37,38]. The GSyV-1 permuted RdRp motif VI with a conserved viral polymerase aa sequence (GDD) was clearly visible (File S1) [85]. Two transmembrane domains were identified within the 422–607 aa region and flanked by the putative PRO catalytic sites ^{422}LVGW^{425} and ^{604}LYGN^{607}. In tymovirus TYMV, an internal sequence (41 aa) of a 140 K protein was previously identified as a chloroplast targeting domain (CTD), similar to SBuMV, which had a C-terminal extension of the MT domain that behaves as an integral membrane protein during infection [86].

ORF2 was in the same reading frame (+2) as ORF1 but was separated by six nt. According to CDD search results, ORF2 encoded a product showing features of a tymovirus-like CP domain (cl03052; 115–288 aa), thus possibly encoding for a CP of SBuMV. The first start codon of ORF2 (minor CP) began at 6002 nt, and the second (major CP) at 6275 nt; however, both ended at 6880 nt with a stop codon UAA. The minor CP was 292 aa long and had a predicted molecular mass of 31 kDa (p31), with proline-(29.21%) and serine-rich (15.73%) N-terminal (89 aa), which is similar to OLV-3 CP [37]. This seems to be the largest minor CP variant among not only all marafiviruses, but also among all representatives of *Tymoviridae* [41]. Major CP variant is translated from a second in-frame methionine, and the resulting product is 201 aa long, with a predicted molecular mass 21.2 kDa (p21.2). The CP gene (ORF2) contains several conserved aa sequences, which are typical for marafivirus CPs—motif I (^{136}PFQW^{138}), motif II (^{169}YRYA^{174}), and motif III (^{222}GGPV^{225}) (File S1). PFQ conserved aa triplet is present in all of the sequenced viruses belonging to the family *Tymoviridae* [37]. According to the MSA of CP (File S1), the majority of the marafiviruses studied thus far have a second methionine (with the exception of AVF and MsMV1) and a putative PRO cleavage site (except for OLV-3 and SBuMV, which share an MRCA). Interestingly, similar to luteoviruses, SBuMV and OLV-3 have a C-rich region 15 nt after a stop codon, which is a readthrough signal that produces a minor CP [87].

MT (61.5%), PRO (44.14%), RdRp (67.2%), and CP (66.03%) aa sequences of SBuMV shared the highest similarity to the corresponding domains of OLV-3 (YP_003475889.1; [37]). Only HEL shared the highest similarity (66.22%) with the CaMaV (QID59002.1) [76] while having a slightly lower identity (63.46%) with its counterpart from OLV-3. The highest whole polyprotein aa sequence similarity with the ICTV assigned *Marafivirus* species also was with OLV-3 (Table 1). These aa sequence identity values are in accordance with the criteria set for the demarcation of novel species within the genus *Marafivirus* [41].

The SBuMV RP N-terminal contains a 70 aa protein domain with a molecular mass of 7.5 kDa, which is separated from PR by a putative papain-like cysteine PRO cleavage site ^{61}LSGG^{64}, which was also identified in other marafiviruses (File S1). According to Protein BLAST and CDD results, no similarities were detected with other proteins or their domains,

although these might potentially be revealed by employing a more sensitive sequence-profile-based search strategy that was not undertaken. The RP N-terminal region is also rich in proline (15.87%), serine (14.29%), and threonine (9.59%) residues. PRO cleavage sites between the PRO and HEL domains (^{929}LTGA^{932}) and between the HEL and RdRp domains (^{1298}LSGA^{1301}) were also present.

Table 1. Similarity of the SBuMV isolate BU1 polyprotein amino acid sequence (UWS64431.1) to polyprotein sequences from the ICTV recognized *Marafivirus* species isolates.

Virus	Abbreviation	Polyprotein Length	Polyprotein Accession	Total Score *	Query Coverage	Identity
Sea buckthorn marafivirus	SBuMV	1954	UWS64431.1	3963	100%	100.00%
Olive latent virus 3	OLV3	2000	YP_003475889.1	2036	90%	61.61%
Nectarine marafivirus M	NeVM	2067	UBZ25923.1	1870	85%	63.67%
Grapevine asteroid mosaic associated virus	GAMaV	2158	UTM04229.1	1840	82%	65.99%
Citrus sudden death-associated virus	CSDaV	2188	YP_224218.1	1839	84%	64.46%
Blackberry virus S	BlVS	2035	YP_009505639.1	1815	82%	67.40%
Peach virus D	PeVD	2055	QCC30253.1	1796	82%	65.32%
Oat blue dwarf virus	OBDV	2067	ADD13602.1	1759	88%	56.76%
Maize rayado fino virus	MRFV	2028	NP_115454.2	1758	83%	65.16%
Grapevine Syrah virus 1	GSV1	2081	YP_002756536.1	1635	89%	50.81%
Alfalfa virus F	AVF	2129	YP_009551972.1	1635	82%	64.97%

* Of note, details of the highest scoring hits are provided, which might originate from isolates other than exemplar isolate of a given virus.

A "marafibox"-related sequence of SBuMV was located at nt positions 5972–5987, with a putative sgRNA transcription start site of CAAC located at the nt positions 6014–6017. However, CAAU and CAAG were present instead of CAAC in other marafiviruses. SBuMV "marafibox"-related sequences had four changes from the sgRNA 16 nt consensus sequence (5′ CAGGGU**C**A**U**UUGC**G**U**G** 3′; QGHLR). Differences in this genomic feature have also been reported for other marafiviruses. For example, the PeVD and OLV-3 "marafiboxes" differed from the consensus "marafibox" by six and three nt, respectively (Figure 4B) [37,73]. The consensus aa sequence encoded by the "marafiboxes" is (Q/E)(A/G/S)(E/Q/K/H)L(L/G/P/R) [76]. Multiple aa sequence alignment of AFV, MsMV1, and DiMV1 "marafiboxes", where glutamic acid is present instead of a more common glutamine (File S1), was also consistent with the phylogenetic signals that delineated them into a distinct clade (Figure 2), suggesting that this could already be a feature of their MRCA. The sequence preceding ORF2 formed a secondary stem RNA structure (Figure 4F). Additionally, a sequence similar to the "marafibox" was identified at nt positions 3111–3126 (5′ GAGAGCGAAUUCUUUC 3′). It was located precisely upstream of the HEL domain and is predicted to form a stem-loop RNA secondary structure (Figure 4E). Before it, a slippery sequence motif "XXXYYYZ" (where X represents any three identical nt, Y represents AAA or UUU, and Z represents A, C, or U) was identified, which is a common signal for a programmed −1 ribosomal frameshift (SBuMV: CCCAAAA; 3097–3103 nt) [62,88]. However, C triplets in such sequences were shown to be the least effective [89].

3.5. Minor and Major CP Expression in Bacterial Expression System

SDS-PAGE analysis of p31 and p21.2 revealed that p31 and p31 + p21.2 expression levels were reduced compared with that of p21.2 (Figure 5, track T,S). All SBuMV CP variants (p31, p21.2, and p31 + p21.2) were purified on a sucrose gradient, and its analysis in SDS-PAGE and NAG showed that the CPs of all variants were located in 40% and 30% gradient fractions (Figure 5, tracks 3,4).

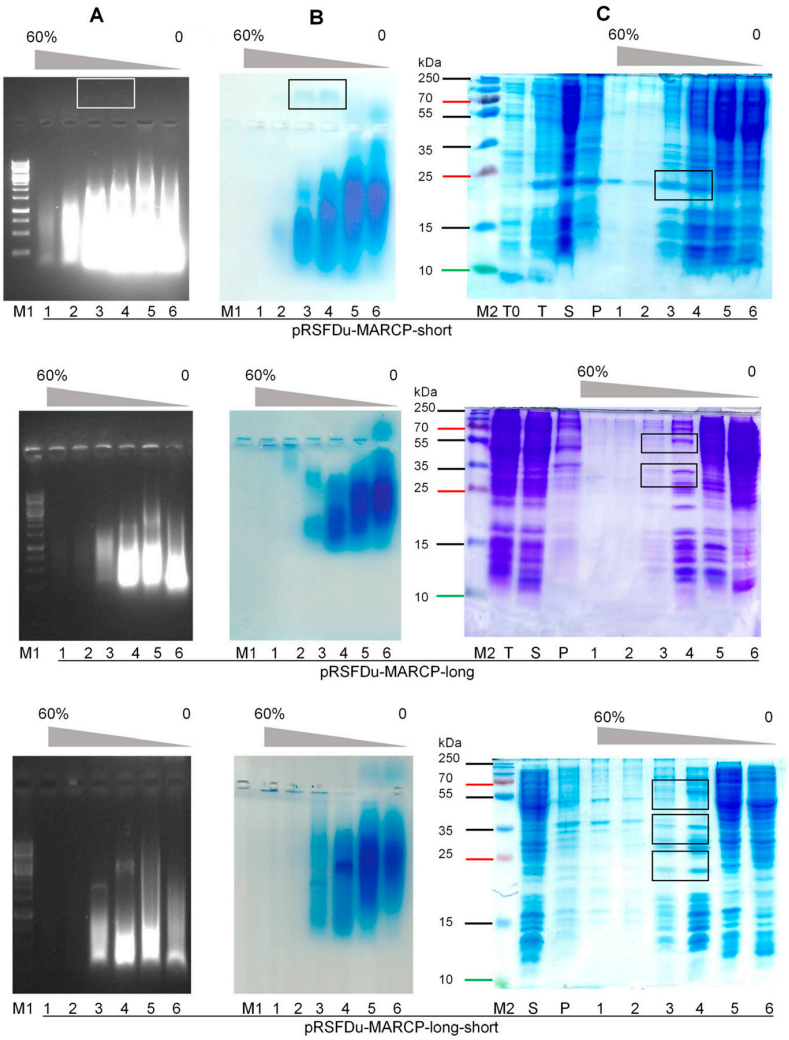

Figure 5. SBuMV native virion and SBuMV VLP purification. (**A**) Native agarose gel stained with ethidium bromide; (**B**) native agarose gel stained with Coomassie G250; (**C**) 12.5% SDS-PAGE stained with Coomassie G250 or R250; 1–6—sucrose gradient fractions starting with 60% (with intervals of 10%) to 20%, overplayed by soluble protein fraction; T0—total cell lysate before induction; T—total cell lysate after expression of CPs; S—supernatant, which contains the soluble protein fraction after cell disruption by ultrasound; P—pellet in the soluble protein fraction after cell disruption by ultrasound; M1—GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA); M2—PageRuler™ Plus Prestained Protein Ladder, 10–250 kDa (Thermo Fisher Scientific, Waltham, MA, USA).

Sucrose fractions containing p21.2, p31, and p31 + p21.2 were pooled and sedimented by ultracentrifugation. Analysis of p21.2, p31, and p31 + p21.2 in ethidium-bromide-stained 0.8% NAG revealed a nucleic acid pattern that migrated towards a negative charge for p21.2, implying that this CP variant has a positive charge (Figure 5, track 1). However, p31 migrated as a negatively charged protein (Figure 5, track 2). After NAG staining with G250, the p21.2 and p31 location signals overlapped with an ethidium-bromide-stained band (Figure 5, track 1,2). Notably, p21.2 migrated in SDS-PAGE as a slightly larger protein (Figure 4, track S) than predicted, whereas p31 showed a migration pattern consistent with a predicted molecular weight of 55 kDa (Figure 5, track 1). However, both p21.2 and p31 formed additional bands.

Analysis of purified CPs using TEM revealed that p21.2 can readily form homogenous self-assembled icosahedral VLPs with a 30 nm diameter (Figure 6B), resembling typical native marafivirus virions. Furthermore, the analysis of p31-containing fraction revealed

only protein aggregates (Figure 6B), while in the case of p31 + p21.2, only some partially assembled VLPs were detected (Figure 6B).

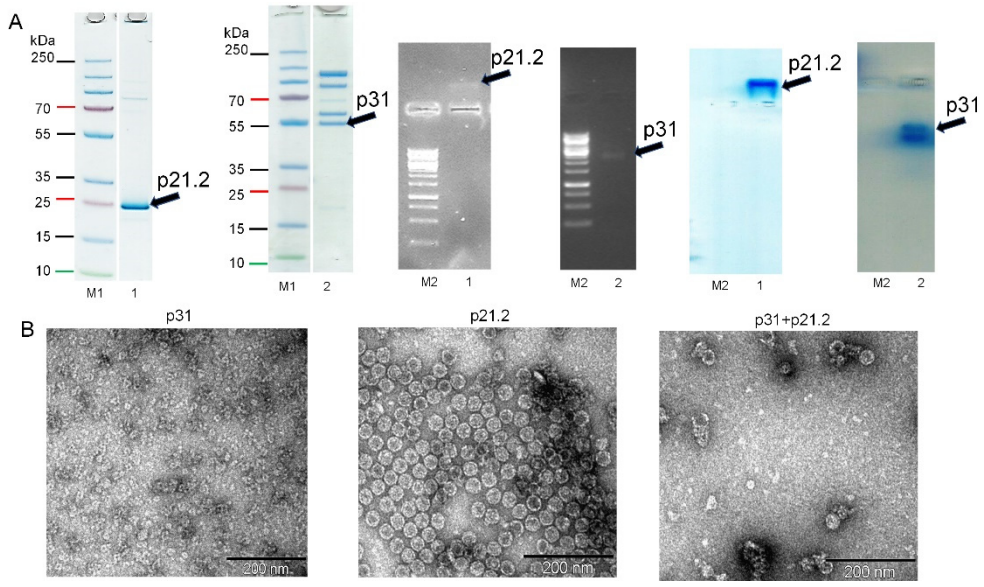

Figure 6. Recombinantly expressed SBuMV CPs analysis. (**A**) SDS-PAGE (stained with G250) and native agarose gel analysis (stained with ethidium bromide and G250); M1—PageRuler™ Plus Prestained Protein Ladder, 10–250 kDa (Thermo Fisher Scientific, Waltham, MA, USA); M2—GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA); 1—purified p21.2; 2—purified p31. (**B**) Transmission electron microscopy analysis, scale 200 nm.

In the case of MRFV, VLPs were obtained by refolding the CP from the inclusion bodies, which could lead to CP proteolysis and possible VLP formation in the case of minor CP. Here, SBuMV VLPs self-assembled directly in the bacterial cells, and the majority of the expressed CPs were soluble. The results of this study suggest that the minor CP is not essential for the assembly of seemingly structurally intact viral particles, meaning that it can have other functions.

3.6. SBuMV Detection in Follow-Up Samples

RT-PCR was used to detect the presence of genetic material for SBuMV minor CP-encoding region. All three of the collected samples were found to be SBuMV positive through RT-PCR analysis in 0.8% NAG (Figure 7).

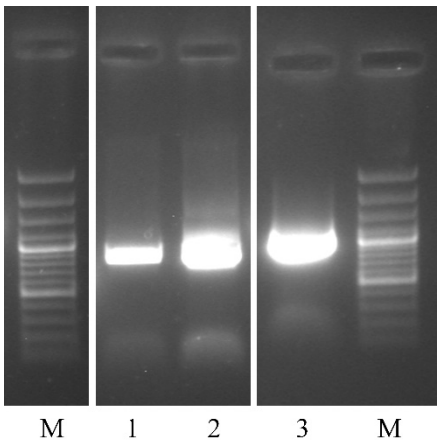

Figure 7. RT-PCR-based SBuMV detection in the follow-up samples. 1–3—SBT samples collected in 2021; M—GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA).

4. Discussion

The discovery of novel viruses provides information that allows for a better understanding of the mechanisms of viral replication, translation, particle assembly, and movement, which are essential for their use as tools for biotechnological applications. Additionally, knowledge of novel possible causative agents of the diseases is invaluable for epidemiology and outbreak containment. Undeniably, the discovery of novel viruses also increases the global understanding of viral diversity, which enables extensive evolutionary studies and further improves the possibilities for conducting comparative-genomics-based studies (such as elucidation of conserved sites within viral proteins and assigning a function to novel viral proteins) that is of utmost importance in the omics era [90].

Moreover, the functional annotation of the SBuMV genome was possible directly due to the fact that many other complete or partial marafivirus genomes have been previously sequenced and are publicly available in the biological sequence repositories. With the growing number of novel viruses being identified using HTS, it can be foreseen that HTS will soon become a conventional method for viral phytopathology and will be applied in diverse studies, complementing classical virological approaches that are slowly becoming obsolete without being complemented by relevant sequence data. A comparison of the marafiviruses uncovered thus far shows that despite having some variation within their genome organization, most of their features are similar. The diversity in the genome organization among the marafiviruses could be due to evolution as a result of adaptation to new host plants; however, sequencing, annotation, and validation errors might also play a role in exaggerating the reality [91,92]. Additionally, recombination between closely related viruses (some of which might not yet be known to science) within the host cells cannot be excluded when trying to deduce the reasons for such a variety in possible genome organizations within a single genus. RNA recombination was found to mediate the rearrangement of viral genes, repair of deleterious mutations, and acquisition of non-self sequences, resulting in ambiguous phylogenetic signals for some viral taxa when the possibilities of recombination were specifically not accounted for. The evidence for recombination not only between closely related viruses but also between distantly related ones and even between the viral and host RNAs suggests that plant viruses unabashedly test the possibility of recombination with any available genetic material [93]. RNA viruses can be regarded as models of efficiency in compressing the maximum amount of information, such as coding and regulatory signals, into the minimum sequence space. They achieve such efficiency by often employing noncanonical translation mechanisms such as overlapping ORFs, some of which can be very short, and arrangement of the ORFs that can additionally regulate their expression. However, small functional ORFs, often lacking conventional initiation sites, can be difficult to detect. Thus, specialized bioinformatics tools are often required to detect the key viral genes [94]. All marafiviruses encode a large RP with five identical functional domains—MT, SD, PRO, HEL, and RdRp—and the documented differences occur only in the presence of an additional ORF at the 5′ or 3′ end of gRNA. For example, MRFV, OLV-3, SwMV, and GsyV-1 genomes have a second overlapping ORF, which shows low sequence identity to the putative movement protein (MP) of tymoviruses and is missing from other marafiviruses [31]. It is a proline-, serine-, and threonine-rich protein, and was shown to be dispensable for the ability to infect maize and leafhopper transmissibility in the MRFV translation mutant experiments. Moreover, unambiguous evidence that the MRFV ORF43 is expressed in vivo is lacking [95]. The absence of the MP gene could be the result of marafivirus phloem limitation [38]. However, in the genomes of CSDaV, OBDV, grapevine fleck virus (genus *Maculavirus*; GFkV), and SBuMV, a degenerate p43-like ORF interrupted by several stop codons was identified. Therefore, it is reasonable to hypothesize that the truncated MPs of these viruses are remnants of an ancestral MP that had degenerated during evolution as these viruses became restricted to phloem cells [40]. The N-terminal domain of RP is proline rich and probably overtakes the function of a missing MP, and the MP can also be expressed from a putative non-canonically translated ORF. This has been observed in members of the family *Luteoviridae*, which are confined

to the phloem (vascular tissue) potentially due to their specialized phloem-specific MP. These proteins are translated from a single viral mRNA, sgRNA1, via initiation at more than a single AUG codon to express overlapping genes and by ribosomal read-through of a stop codon [94]. To maximize coding capacity, RNA viruses often encode overlapping genes and use unusual translational control mechanisms [94]. The absence of convincing overlapping ORFs in the SBuMV genome could stem from the inability to identify a novel genome architecture/genetic material organization and/or lack of publicly available relevant protein domain homologs. It is possible that some putative small ORFs may be facilitated by the use of non-AUG initiation codons. Under certain circumstances, the near-cognate codons CUG, GUG, ACG, AUU, AUA, UUG, and AUC are able to support a significant level of initiation (typically 2–15% of the initiation levels from an AUG codon in a similar context), with CUG being the most efficient non-AUG initiation codon in many systems [96,97]. Initiation by non-AUG codons normally requires a strong initiation context but may also be enhanced in less predictable ways by RNA secondary structures [98]. In several plant viruses, a combination of non-AUG and poor-context AUG initiation codons allows for the production of three or even four functional proteins from a single transcript [99–101].

Readthrough of the *CP* stop codon of *Luteoviridae* family was used for minor CP variant translation. The readthrough domain in the N-terminal region, which is conserved, is required for aphid transmission; however, the C-terminus, which is variable, appears to enhance aphid transmission efficiency and function in the phloem retention of the virus, thus influencing systemic infection, virus accumulation, and symptom development [87]. Probably, the N-terminal region of the marafivirus minor CP has similar functions; however, this is yet to be properly tested. This mechanism seems to be especially feasible for OLV-3 and SBuMV because they have separate ORFs for RP and CP, and the CP N-terminal is proline rich compared with that of other marafiviruses. Although there are no typical PRO cleavage sites before a minor CP, an aa sequence similar to the PRO cleavage site LQGH is present in the "marafibox" of SBuMV, which corresponds to the LQGQ in OLV-3 "marafibox" (File S1). However, the implications of this in silico observation are yet to be verified by functional experiments. Perhaps, this could be a result of adaptation to a different transmission vector, for example, from leafhoppers to aphids. In the case of CSDaV, aphids from CSDaV-positive plants were readily virus positive, but the leafhopper species collected from the same areas were negative for the presence of virus [40]. In addition, Hammond and Ramirez [36] suggested that MRFV p25 (rich in hydrophobic aa and proline) may play a biological role in the transmission and virion packaging of MRFV. This suggestion was based on examples from other viruses, where N- or C-terminal extensions are involved in virus transmission, assembly, replication, and/or spreading [36], and this suggestion has been recently confirmed [46]. On the basis of current knowledge, the N-terminus of a minor CP variant, when co-expressed with a major variant, could facilitate the packaging of viral RNA [102,103]. MRFV and AVF minor *CP* genes were transiently expressed in *Nicotiana benthamiana* [48,104], showing in both cases that VLPs may encapsulate *CP* mRNA and/or host RNA [48,103].

The major CPs of marafiviruses are translated from sgRNA by a leaky scanning mechanism [43]. The indirect evidence of VLP self-assembly without a minor CP, and the absence of VLP during heterologous expression in a bacterial system, could be associated with the fact that major CP is translated from an sgRNA, which holds true for other marafiviruses. The maturation of minor CP from RP is achieved by PRO processing. Marafiviruses (such as tymoviruses) contain a PRO cleavage site. Indirect confirmation for this strategy was observed when AVF minor and MRFV minor and major CP variants were either individually expressed or co-expressed by agroinfiltration of *N. benthamiana* [48,103]. The TEM analysis revealed the presence of isometric MRFV and AVF VLPs of ≈30 nm in diameter (similar to SBuMV VLPs), but it was unclear whether the observed VLPs were composed of a minor CP variant alone or of both proteins, where the minor variant could have been post-translationally cleaved by PVX- or host-encoded proteases from the major

CP variant. Alignment of the marafivirus RP aa sequences revealed that at least three PRO cleavage sites that appear as a consensus motif LXG[G/A] were clearly identifiable in most of them: the first one at the N-terminal region of RP, the second one between PRO and HEL domains, and the third one between the HEL and RdRp domains. For marafiviruses that have a CP fused with the RP, the fourth site can be identified before a minor CP. In vitro translation of MRFV gRNA extracted from virions found in rabbit reticulocyte lysates resulted in the synthesis of polypeptides ranging from 15 to 165 kDa in weight. However, no polypeptides corresponding to the CPs could be detected by immunoprecipitation of the translation products using an antiserum [105]. Translation of the OBDV virion RNA in rabbit reticulocyte lysates resulted in the production of protein domains with molecular masses of 227 kDa (RP with CP), 202 kDa (RP), 133 kDa (MT, PRO, HEL), 94 kDa (RdRp with CP), 70 kDa (RdRp), and 24 kDa (CP), as suggested by the predicted protein sequences [38]. Sequence conservation and similarity to other viruses that utilize papain-like proteases as a means of processing polyproteins and defense against protein degradation via the ubiquitin–proteasome system suggests that similar mechanisms are used by marafiviruses [44]. Comparison of the MRFV PRO 3D structure with known PRO 3D structures revealed turnip yellow mosaic virus (TYMV, family *Tymovirus*) PRO as the closest structural homolog, followed by ovarian-tumor-domain-containing protein 3 (OTUD3) from *Homo sapiens* and OTUD1 from *Saccharomyces cerevisiae*. Additionally, deubiquitinating activity has been shown for BlVS, CSDaV, GsyV1, MRFV, OBDV, and OLV3 [106]. On the basis of their overall core fold, MRFV PRO and TYMV PRO seem to be more related to a viral out-specific ubiquitin hydrolase (DUB) than to papain-like proteases [107,108]. The tymovirus and marafivirus PRO can be classified as OTU-like cysteine proteases [109,110]. The MRFV cleavage site LVGA between PRO and HEL domains has been verified in the past, suggesting that this particular site in this location is highly likely to be active for other marafiviruses as well (File S1) [106]. PRO cleavage sites for several marafiviruses, such as PGMV (INGG), SbMV (VNGG), DiMV2 (LTGS), and DiMV1 (LNGS), differ in either the first or last position (File S1). Different aa in these positions have also been documented in other phytoviruses. TYMV PRO/DUB recognizes the consensus peptide substrate (K/R)LXG(G/A/S) [111]. Other plant viruses, such as representatives of the *Potyviridae* family, use PRO cleavage that resembles a consensus sequence (Y/F/G)xG(A, N, S) [110].

For many ssRNA(+) viruses, viral gRNAs are capped, allowing for efficient translation in eukaryotic cells. Cellular mRNA capping enzymes of the host are located in the nucleus; thus, many viruses that replicate in the cytoplasm encode their own capping enzymes [112]. Multiple aa sequence alignment of the MT domain demonstrated a characteristic consensus sequence in all the analyzed marafivirus sequences (n = 21; File S1), and CDD analysis revealed an SD domain [82]. This evidence suggests that all marafiviruses may have a cap structure at the 5′ ends of their genomes. This type of MT is strictly associated with supergroup 3 RdRp and superfamily 1 HEL and can be classified as a tymo-like MT [84].

All positive-sense RNA viruses with a genome size of over 6 kb encode a putative RNA HEL, which is thought to be involved in the unwinding of a duplex during the viral RNA replication and, perhaps, translation as well [113]. Seven conserved superfamily 1 HEL motifs (File S1) can be clearly identified from the corresponding marafivirus aa sequence alignment, thus confirming that their HEL belongs to superfamily 1 [114]. MT, HEL, and RdRp coding domains are easier to identify because of the presence of more evolutionarily conserved motifs compared with the PRO domain. Conservation of viral papain-like proteases is observed almost exclusively around the catalytic cysteine (C) residue. However, identification of a catalytic histidine (H) is difficult in some of these PRO domains, and no other conserved motifs are detectable in such domains [84]. Although marafiviruses have two clearly recognizable PRO motifs, CLL and GHF (File S1) [37,38], there are additional small, conserved motifs, such as L(W/R), GL, H(F/L), and L(A/C).

These RNA genomes perform the traditional role of being the blueprint for all viral proteins; however, they also contain *cis*-acting RNA elements (REs) that have been shown

to impact many essential viral processes, including protein translation, genome replication, and transcription of sgRNAs. For many viruses, REs are usually located in the 5′ and 3′ UTRs and/or internally within inter-cistronic regions. However, one drawback of this approach is that the size and/or location of the non-coding regions can have some limitations. To avoid such limitations, many plus-strand RNA viruses have evolved to position their REs within their coding regions [115].

One potential drawback of having a RE in a coding region is that the same RNA sequence region physically couples two or more distinct activities; therefore, one or both of the corresponding sequence functions may be compromised compared with a hypothetical free-standing localization. Additionally, in some cases, the relative location of REs within the genome of a given virus may not be optimal for their activity; therefore, compensatory measures may be required. One strategy used by many RNA viruses to deal with sub-optimally positioned REs is to eventually reorganize their relative location within the genome via intramolecular long-range RNA–RNA interactions to mediate translational initiation and viral RNA genome replication and form functional secondary RNA structures [115].

The SBuMV genome contains several putative structural elements in both the 5′ and 3′ UTRs as well as in the coding regions. RNA structure predictions suggested that the 3′ UTR could form a tRNA-like structure similar to that of other viruses. For example, the barley stripe mosaic virus (BSMV, genus: *Hordeivirus*, family: *Virgaviridae*) has a poly-A region followed by a 3′ terminal sequence capable of folding into a tRNA-like structure that can be aminoacylated [116]. The gRNA of tymoviruses contains a 3′-UTR secondary RNA structure that is functionally and structurally related to tRNAs. This peculiar structure allows for tRNA-like domain interaction with tRNA-specific proteins, such as RNase P, tRNA nucleotidyl-transferase, aminoacyl-tRNA synthetases, and elongation factors, thus allowing it to play an important biological role in the viral life cycle [117]. The RNA secondary structure prediction of 5′ UTR suggested the presence of three putative hairpin structures (Figure 4D) that also resemble their counterparts from the tymoviruses, indicating a similar function of the UTR in both.

The internal putative structural element—an additional "marafibox" motif at the beginning of a HEL domain in SBuMV—could be an extra signal sequence. A consensus sequence (GnGAnCGnnUCCUCUC) (Figure 4B) was identified for this putative element from the MSA of the corresponding SBuMV regions to its counterparts from other marafiviruses. An extra "marafibox" motif was also identified in the comparative analysis of GSyV-1, which boasted a secondary motif with an 11 base consensus sequence (CUnnCACUCnC), located at a variable distance from the primary "marafibox". The conserved RNA sequence of the primary "marafibox" motif is generally found to be foldable into a stem structure topped with a UUCA loop. Relative to the overlying protein-coding frame, the second conserved motif is located at variable distances and in a reading frame that is different from the first motif [118]. Al Rwahnih et al. [118] proposed that the second RNA sequence motif in the "marafibox" is not conserved, as it is not an aa coding sequence, but rather a sequence encoding RNA structural information, which is also presumed to be function of the first motif. Before "marafibox," a slippery sequence motif CCCAAAA corresponding to the common XXXYYYZ (X represents any three identical nt; Y represents AAA or UUU; and Z represents A, C, or U) ribosomal frameshift signal can be identified [62,88]. Recently, an unusual aspect of encephalomyocarditis virus (EMCV; genus: *Cardiovirus*, family: *Picornaviridae*) translation by a previously undetected -1 PRF site in an internal region of the polyprotein was discovered, despite EMCV serving as a model virus for more than 50 years [119]. PRF is *trans*-activated by viral protein 2A. Consequently, the frameshifting efficiency increases from 0 to 70% (one of the highest known in a mammalian system) over the course of infection, temporally regulating the expression levels of the viral structural and enzymatic proteins [62]. However, separate experiments are required to prove the PRF activity due to the C triplets being the least effective in the slippery sequence and the absence of a "spacer" between the slippery sequence and the PRF [89]. The second

explanation for the presence of this motif could be that the "marafibox", similar to the "tymobox", is a subgenomic promoter [33]. Several other phytoviruses (e.g., members of *Luteoviridae* family) have three sgRNAs [94].

Pseudoknots, secondary structures resulting from an interaction of stems and loops, represent a structurally diverse group of functional regulatory elements. They play several diverse biological roles, such as forming a catalytic core of various ribozymes, self-splicing introns, and telomerase. Moreover, they play a critical role in altering gene expression by inducing ribosomal frameshifting in many viruses [120].

To identify the symptoms of a disease potentially caused by SBuMV, SBT orchids have to be independently tested, and the results should be carefully analyzed considering the possible presence of other potential causative agents so that all of Koch's postulates are fulfilled. Analysis of the SBuMV genome and protein sequences has raised more questions than it has answered. For example, is the 5′ end capped? Does proteolysis occur at the identified cleavage sites? Which is the vector for SBuMV? All these and many more questions would be answered in the near future, thus paving the way for exciting discoveries and novel knowledge on phytoviruses.

5. Conclusions

In conclusion, here, we, for the first time, present a virus putatively infecting SBT. We characterized its full genome, and cloned and expressed minor and major forms of CP either individually or co-expressed. We also showed that major CP-derived virus-like particles self-assembled directly in the bacterial cells, and the majority of the expressed CPs were soluble. Our study suggests that the minor CP is not essential for the assembly of seemingly structurally intact viral particles, meaning that it can have other functions.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/microorganisms10101933/s1, Figure S1: Sea buckthorn leaf samples collected in August 2017 and August 2021. Figure S2: Maximum-likelihood and neighbor-joining trees generated using *Tymoviridae* capsid protein sequence. Figure S3: *Tymoviridae* polyprotein aa sequence with maximum-likelihood and neighbor-joining trees. Table S1: Primers used for SBuMV *CPs* and genome fragment amplification and sequencing. Table S2: The dataset used for evolutionary relationship analysis. Table S3: Descriptions of the generated MSA and features of the resulting tree. File S1: Alignment of marafivirus RP and CP with coded protein domain motifs using PROMALS3D.

Author Contributions: Conceptualization, I.B., N.Z. and A.Z.; methodology, I.B., V.Z., N.Z., I.K., G.R., R.L. and J.J.; software, N.Z.; validation, I.B. and N.Z.; formal analysis, I.B., V.Z., N.Z., I.K., G.R., R.L. and J.J.; investigation, I.B., V.Z., N.Z., I.K., G.R., R.L. and J.J.; resources, I.B. and A.Z.; data curation, I.B. and N.Z.; writing—original draft preparation, I.B. and N.Z.; writing—review and editing, I.B., N.Z., I.M.-B., D.S. and A.Z.; visualization, I.B. and N.Z.; supervision, I.B. and A.Z.; project administration, I.B. and A.Z.; funding acquisition, A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Latvian Council of Science, grant no. lzp-2019/1-0131.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors declare that the data supporting the findings of this study are available within the article and the datasets generated and analyzed for this study and can be available from the corresponding authors on reasonable request. The dataset used for evolutionary relationship analysis (accessions of the selected publicly available *Tymoviridae* entries and the related sequence metadata) can be seen in Table S2. Generated MSA descriptions and the resulting tree features can be seen in Table S3. SBuMV complete genome sequence identified in this study was deposited in the GenBank database under the accession number ON149451.

Acknowledgments: We are grateful and would like to thank Monta Brīvība, Līga Birzniece, and Kaspars Megnis from the Latvian Biomedical Research and Study Centre's Genome Centre's genotyping and sequencing unit for technical and methodological support during the NGS library preparation

and sequencing. We would like to also thank Davids Fridmanis, and Ivars Silamikelis from the Latvian Biomedical Research and Study Centre's Bioinformatics core facility for continuous support and advice in the NGS data analysis methodology.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Yang, B.R.; Kallio, H. Effects of harvesting time on triacylglycerols and glycerophospholipids of sea buckthorn (*Hippophae rhamnoides* L.) berries of different origins. *J. Food Compos. Anal.* **2002**, *15*, 143–157. [CrossRef]
2. Vescan, A.; Pamfil, D.; Bele, C.; Matea, C.; Sisea, C.R. Several lipophilic components of five elite genotypes of romanian seabuckthorn (*Hipppohae rhamnoides* subs. carpatica). *Not. Bot. Horti Agrobot.* **2010**, *38*, 114–122.
3. Rongsen, A. Sea buckthorn a multipurpose plant for fragile mountains. In *Icimod Occasional Paper No. 20*; International Centre for Integrated Mountain Development (ICIMOD): Kathmandu, Nepal, 1992; pp. 6–7, 18–20.
4. Kondrashov, V.T.; Sokolova, E.P. New wilt-resistant forms of *Hippophaë rhamnoides*. *Byulleten Mosk. Obs. Ispyt. Prir. Biol.* **1990**, *96*, 146–153.
5. Li, T.S.C.; Schroeder, W.R. Sea buckthorn (*Hippophae rhamnoides* L.): A multipurpose plant. *HortTechnology* **1996**, *6*, 370. [CrossRef]
6. Mauriņš, A.; Zvirgzds, A. *[Dendrology]*; LU Akadēmiskais apgāds: Riga, Latvia, 2006.
7. Bruvelis, A. Sea buckthorn cultivation in baltic states. In Proceedings of the 1st Congress of the International Seabuckthorn Association, Berlin, Germany, 14–18 September 2003; pp. 64–66.
8. Segliņa, D. *Sea Buckthorn Fruits and Their Processing Products*; Latvia Univ. of Agriculture: Jelgava, Latvia, 2007.
9. Brūvelis, A. Experiences about sea buckthorn cultivation and harvesting in Latvia. In *Proceedings of the 3rd European Workshop on Sea Buckthorn EuroWorkS2014*; Kauppinen, S., Petruneva, E., Eds.; Association of Latvian Fruit Growers: Riga, Latvia, 2015; pp. 36–41.
10. Lamo, K.; Solanki, S.P.S. Sea buckthorn a boon for trans-himalayan region of Ladakh: A review. *Agric. Rev.* **2019**, *40*, 289–295. [CrossRef]
11. Drevinska, K.; Moročko-Bičevska, I. Sea buckthorn diseases caused by pathogenic fungi: A review. *Proc. Latv. Acad. Sci.* **2022**, *in press*.
12. Kalia, R.K.; Singh, R.; Rai, M.K.; Mishra, G.P.; Singh, S.R.; Dhawan, A.K. Biotechnological interventions in sea buckthorn (*Hippophae* L.): Current status and future prospects. *Trees Struct. Funct.* **2011**, *25*, 559–575. [CrossRef]
13. Li, T.S.C. Sea buckthorn (*Hippophae rhamnoides* L.): Production and utilization. In *Taxonomy, Natural Distribution and Botany*; Li, T.S.C., Beveridge, T.H.J., Eds.; PRC Research Press: Ottawa, ON, Canada, 2003; pp. 7–12.
14. Singh, K.P.; Prasad; Yadav, V.K. The first report of *Rhizoctonia solani* Kuch on Seabuckthorn (*Hippophae salicifolia* d.Don) in Uttaranchal Himalayas. *J. Mycol. Plant Pathol.* **2007**, *37*, 126–127.
15. Parrika, P.; Karhu, S. *Stem Canker of Sea Buckthorn (Hippophae rhamnoides L.) in Finland, 7th International Congress of Plant Pathology*; International Society for Plant Pathology: Edinburg, Scotland, 1998; p. 3.7.51.
16. Bharat, N.K. Occurrence of powdery mildew on seabuckthorn in Himachal Pradesh. *Indian For.* **2006**, *132*, 517.
17. Kumar, S.; Sagar, A. Microbial associates of *Hippophaë rhamnoides* (Seabuckthorn). *Plant Pathol. J.* **2007**, *6*, 299–305. [CrossRef]
18. Saxena, S.; Malik, N.; Guleri, S. First report on Occurrence of *Emericella quadrilineata* on leaves of *Hippophae salcifolia* D. Don from India. *Int. J. Curr. Microbiol. Appl. Sci.* **2015**, *4*, 1006–1009.
19. Ruan, C.J.; Li, H.; Mopper, S. Characterization and identification of ISSR markers associated with resistance to dried-shrink disease in sea buckthorn. *Mol. Breed.* **2009**, *24*, 255–268. [CrossRef]
20. Konavko, D.; Malchev, S.; Pothier, J.F.; Jundzis, M.; Morocko-Bicevska, I.; Rezzonico, F. Diversity and host range of *Pseudomonas* in fruit tree species in Latvia. *Acta Hortic.* **2016**, *1149*, 25–29. [CrossRef]
21. Moročko-Bičevska, I.; Sokolova, O.; Konavko, D.; Vēvere, K.; Jundzis, M.; Fatehi, J. Survey on diseases and fungal pathogens associated with cankers and decline of sea buckthorn. *IOBC WPRS Bull.* **2019**, *144*, 56–61.
22. Pecman, A.; Kutnjak, D.; Gutierrez-Aguirre, I.; Adams, I.; Fox, A.; Boonham, N.; Ravnikar, M. Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Front. Microbiol.* **2017**, *8*, 1998. [CrossRef]
23. Mackay, I.M.; Arden, K.E.; Nitsche, A. Real-time PCR in virology. *Nucleic Acids Res.* **2002**, *30*, 1292–1305. [CrossRef]
24. Pecman, A.; Kutnjak, D.; Mehle, N.; Znidaric, M.T.; Gutierrez-Aguirre, I.; Pirnat, P.; Adams, I.; Boonham, N.; Ravnikar, M. High-throughput sequencing facilitates characterization of a "forgotten" plant virus: The case of a Henbane mosaic virus infecting tomato. *Front. Microbiol.* **2018**, *9*, 2739. [CrossRef]
25. Bacnik, K.; Kutnjak, D.; Pecman, A.; Mehle, N.; Tusek Znidaric, M.; Gutierrez Aguirre, I.; Ravnikar, M. Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Water Res.* **2020**, *177*, 115628. [CrossRef]
26. Zrelovs, N.; Resevica, G.; Kalnciema, I.; Niedra, H.; Lacis, G.; Bartulsons, T.; Morocko-Bicevska, I.; Stalazs, A.; Drevinska, K.; Zeltins, A.; et al. First report of black currant-associated rhabdovirus in blackcurrants in Latvia. *Plant Dis.* **2022**, *106*, 1078. [CrossRef]

27. Hartung, J.S.; Roy, A.; Fu, S.; Shao, J.; Schneider, W.L.; Brlansky, R.H. History and diversity of citrus leprosis virus recorded in herbarium specimens. *Phytopathology* **2015**, *105*, 1277–1284. [CrossRef]
28. Vucurovic, A.; Kutnjak, D.; Mehle, N.; Stankovic, I.; Pecman, A.; Bulajic, A.; Krstic, B.; Ravnikar, M. Detection of four new tomato viruses in Serbia using post-hoc high-throughput sequencing analysis of samples from a large-scale field survey. *Plant Dis.* **2021**, *105*, 2325–2332. [CrossRef] [PubMed]
29. Zhang, T.; Breitbart, M.; Lee, W.H.; Run, J.Q.; Wei, C.L.; Soh, S.W.L.; Hibberd, M.L.; Liu, E.T.; Rohwer, F.; Ruan, Y.J. RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biol.* **2006**, *4*, 108–118. [CrossRef] [PubMed]
30. ICTV. Available online: https://talk.Ictvonline.Org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/245/tymoviridae (accessed on 17 May 2022).
31. Martelli, G.P.; Sabanadzovic, S.; Sabanadzovic, N.A.G.; Edwards, M.C.; Dreher, T. The family tymoviridde. *Arch. Virol.* **2002**, *147*, 1837–1846. [CrossRef] [PubMed]
32. Dreher, T.W.; Edwards, M.C.; Gibbs, A.J.; Haenni, A.-L.; Hammond, R.W.; Jupin, I.; Koenig, R.; Sabanadzovic, S.; Abou GhanemSabanadzovic, N.; Martelli, G.P. *Family Tymoviridae*; Elsevier: San Diego, CA, USA, 2005; pp. 1067–1076.
33. Ding, S.W.; Howe, J.; Keese, P.; Mackenzie, A.; Meek, D.; Osorio-Keese, M.; Skotnicki, M.; Srifah, P.; Torronen, M.; Gibbs, A. The tymobox, a sequence shared by most tymoviruses: Its use in molecular studies of tymoviruses. *Nucleic Acids Res.* **1990**, *18*, 1181–1187. [CrossRef] [PubMed]
34. Goldbach, R.; Le Gall, O.O.; Wellink, J. Alpha-like viruses in plants. *Semin. Virol.* **1991**, *2*, 19–25.
35. Rozanov, M.N.; Koonin, E.V.; Gorbalenya, A.E. Conservation of the putative methyltransferase domain-a hallmark of the sindbis-like supergroup of positive-strand rna viruses. *J. Gen. Virol.* **1992**, *73*, 2129–2134. [CrossRef]
36. Hammond, R.W.; Ramirez, P. Molecular characterization of the genome of maize rayado fino virus, the type member of the genus marafivirus. *Virology* **2001**, *282*, 338–347. [CrossRef]
37. Alabdullah, A.; Minafra, A.; Elbeaino, T.; Saponari, M.; Savino, V.; Martelli, G.P. Complete nucleotide sequence and genome organization of Olive latent virus 3, a new putative member of the family Tymoviridae. *Virus Res.* **2010**, *152*, 10–18. [CrossRef]
38. Edwards, M.C.; Zhang, Z.; Weiland, J.J. Oat blue dwarf marafivirus resembles the tymoviruses in sequence, genome organization, and expression strategy. *Virology* **1997**, *232*, 217–229. [CrossRef]
39. Agindotan, B.O.; Gray, M.E.; Hammond, R.W.; Bradley, C.A. Complete genome sequence of switchgrass mosaic virus, a member of a proposed new species in the genus Marafivirus. *Arch. Virol.* **2012**, *157*, 1825–1830. [CrossRef]
40. Maccheroni, W.; Alegria, M.C.; Greggio, C.C.; Piazza, J.P.; Kamla, R.F.; Zacharias, P.R.; Bar-Joseph, M.; Kitajima, E.W.; Assumpção, L.C.; Camarotte, G.; et al. Identification and genomic characterization of a new virus (Tymoviridae family) associated with citrus sudden death disease. *J. Virol.* **2005**, *79*, 3028–3037. [CrossRef] [PubMed]
41. Dreher, T.W.; Edwards, M.C.; Gibbs, A.J.; Haenni, A.-L.; Hammond, R.W.; Jupin, I.; Koenig, R.; Sabanadzovic, S.; Martelli, G.P. Family-Tymoviridae. In *Virus Taxonomy*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier: San Diego, CA, USA, 2012; pp. 944–952.
42. Ahola, T.; Karlin, D.G. Sequence analysis reveals a conserved extension in the capping enzyme of the alphavirus supergroup, and a homologous domain in nodaviruses. *Biol. Direct* **2015**, *10*, 16. [CrossRef] [PubMed]
43. Matsumura, E.E.; Coletta-Filho, H.D.; Machado, M.A.; Nouri, S.; Falk, B.W. Rescue of Citrus sudden death-associated virus in *Nicotiana benthamiana* plants from cloned cDNA: Insights into mechanisms of expression of the three capsid proteins. *Mol. Plant Pathol.* **2019**, *20*, 611–625. [CrossRef] [PubMed]
44. Edwards, M.C.; Weiland, J.J. Coat protein expression strategy of oat blue dwarf virus. *Virology* **2014**, *450–451*, 290–296. [CrossRef]
45. Edwards, M.C.; Weiland, J.J. First infectious clone of the propagatively transmitted oat blue dwarf virus. *Arch. Virol.* **2010**, *155*, 463–470. [CrossRef] [PubMed]
46. Mlotshwa, S.; Khatri, N.; Willie, K.; Xu, J.; Todd, J.; Tran, H.H.; Stewart, L.R. Coat protein expression strategy of maize rayado fino virus and evidence for requirement of cp1 for leafhopper transmission. *Virology* **2022**, *570*, 96–106. [CrossRef]
47. Kim, H.; Park, D.; Hahn, Y. Identification of novel rna viruses in alfalfa (Medicago sativa): An Alphapartitivirus, a Deltapartitivirus, and a Marafivirus. *Gene* **2018**, *638*, 7–12. [CrossRef]
48. Nemchinov, L.G.; François, S.; Roumagnac, P.; Ogliastro, M.; Hammond, R.W.; Mollov, D.S.; Filloux, D. Characterization of alfalfa virus F, a new member of the genus Marafivirus. *PLoS ONE* **2018**, *13*, e0203477.
49. Louie, R. Vascular puncture of maize kernels for the mechanical transmission of maize white line mosaic-virus and other viruses of maize. *Phytopathology* **1995**, *85*, 139–143. [CrossRef]
50. Madriz-Ordeñana, K.; Rojas-Montero, R.; Lundsgaard, T.; Ramírez, P.; Thordal-Christensen, H.; Collinge, D.B. Mechanical transmission of maize rayado fino marafivirus (MRFV) to maize and barley by means of the vascular puncture technique. *Plant Pathol.* **2000**, *49*, 302–307. [CrossRef]
51. Blanc, S. Vector transmission of plant viruses. In *Encyclopedia of Virology*, 3rd ed.; Mahy, B.W.J., Van Regenmortel, M.H.V., Eds.; Academic Press: Oxford, UK, 2008; pp. 274–282.
52. Balke, I.; Resevica, G.; Zeltins, A. Isolation and characterization of two distinct types of unmodified spherical plant sobemovirus-like particles for diagnostic and technical uses. *Methods Mol. Biol.* **2018**, *1776*, 19–34. [PubMed]
53. Andrews, S. Fastqc: A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed on 17 November 2021).

54. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A.D. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* **2019**, *8*, giz100. [CrossRef] [PubMed]
55. Wheeler, D.L.; Church, D.M.; Federhen, S.; Lash, A.E.; Madden, T.L.; Pontius, J.U.; Schuler, G.D.; Schriml, L.M.; Sequeira, E.; Tatusova, T.A.; et al. Database resources of the national center for biotechnology. *Nucleic Acids Res.* **2003**, *31*, 28–33. [CrossRef]
56. Marchler-Bauer, A.; Bryant, S.H. Cd-search: Protein domain annotations on the fly. *Nucleic Acids Res.* **2004**, *32*, W327–W331. [CrossRef] [PubMed]
57. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
58. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef]
59. Kieleczawa, J. Fundamentals of sequencing of difficult templates—An overview. *J. Biomol. Tech.* **2006**, *17*, 207–217.
60. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2007**, *35*, D61–D65. [CrossRef]
61. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [CrossRef]
62. Napthine, S.; Ling, R.; Finch, L.K.; Jones, J.D.; Bell, S.; Brierley, I.; Firth, A.E. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat. Commun.* **2017**, *8*, 15582. [CrossRef]
63. Katoh, K.; Standley, D.M. Mafft multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
64. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
65. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. Modelfinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
66. Minh, B.Q.; Nguyen, M.A.; von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef] [PubMed]
67. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
68. Kumar, S.; Stecher, G.; Tamura, K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
69. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985**, *39*, 783–791. [CrossRef]
70. Rambaut, A. Figtree v. 1.4.4. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 10 May 2021).
71. Inkscape. Project Inkscape. Available online: https://inkscape.org (accessed on 15 September 2020).
72. Schoch, C.L.; Ciufo, S.; Domrachev, M.; Hotton, C.L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; McVeigh, R.; O'Neill, K.; Robbertse, B.; et al. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* **2020**, *2020*, baaa062. [CrossRef]
73. Igori, D.; Lim, S.; Baek, D.; Kim, S.Y.; Seo, E.; Cho, I.S.; Choi, G.S.; Lim, H.S.; Moon, J.S. Complete nucleotide sequence and genome organization of peach virus D, a putative new member of the genus Marafivirus. *Arch. Virol.* **2017**, *162*, 1769–1772. [CrossRef]
74. Glasa, M.; Predajňa, L.; Šoltys, K.; Sabanadzovic, S.; Olmos, A. Detection and molecular characterisation of Grapevine Syrah virus-1 isolates from Central Europe. *Virus Genes* **2015**, *51*, 112–121. [CrossRef]
75. Villamor, D.E.V.; Mekuria, T.A.; Pillai, S.S.; Eastwell, K.C. High-throughput sequencing identifies novel viruses in nectarine: Insights to the etiology of stem-pitting disease. *Phytopathology* **2016**, *106*, 519–527. [CrossRef] [PubMed]
76. Zhang, S.; Yang, L.; Ma, L.; Tian, X.; Li, R.; Zhou, C.; Cao, M. Virome of Camellia japonica: Discovery of and molecular characterization of new viruses of different taxa in camellias. *Front. Microbiol.* **2020**, *11*, 945. [CrossRef] [PubMed]
77. Wu, Q.; Kehoe, M.; Kinoti, W.M.; Wang, C.; Rinaldo, A.; Tyerman, S.; Habili, N.; Constable, F.E. First report of grapevine rupestris vein feathering virus in grapevine in Australia. *Plant Dis.* **2020**, *105*, 515. [CrossRef] [PubMed]
78. Candresse, T.; Faure, C.; Theil, S.; Beuve, M.; Lemaire, O.; Spilmont, A.S.; Marais, A. First report of Grapevine asteroid mosaic-associated virus infecting Grapevine (*Vitis vinifera*) in France. *Plant Dis.* **2017**, *101*, 1061. [CrossRef]
79. Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neuböck, R.; Hofacker, I.L. The vienna RNA websuite. *Nucleic Acids Res.* **2008**, *36*, W70–W74. [CrossRef] [PubMed]
80. Hellendoorn, K.; Michiels, P.J.; Buitenhuis, R.; Pleij, C.W. Protonatable hairpins are conserved in the 5′-untranslated region of tymovirus RNAs. *Nucleic Acids Res.* **1996**, *24*, 4910–4917. [CrossRef] [PubMed]
81. Froissart, R.; Roze, D.; Uzest, M.; Galibert, L.; Blanc, S.; Michalakis, Y. Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol.* **2005**, *3*, e89. [CrossRef]
82. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. Cdd/sparcle: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [CrossRef]
83. Pei, J.; Kim, B.H. Grishin NV. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300. [CrossRef]
84. Koonin, E.V.; Dolja, V.V.; Morris, T.J. Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 375–430. [CrossRef]

85. Sabanadzovic, S.; Ghanem-Sabanadzovic, N.A.; Gorbalenya, A.E. Permutation of the active site of putative RNA-dependent RNA polymerase in a newly identified species of plant alpha-like virus. *Virology* **2009**, *394*, 1–7. [CrossRef] [PubMed]
86. Moriceau, L.; Jomat, L.; Bressanelli, S.; Alcaide-Loridan, C.; Jupin, I. Identification and molecular characterization of the chloroplast targeting domain of turnip yellow mosaic virus replication proteins. *Front. Plant Sci.* **2017**, *8*, 2138. [CrossRef] [PubMed]
87. Xu, Y.; Ju, H.J.; DeBlasio, S.; Carino, E.J.; Johnson, R.; MacCoss, M.J.; Heck, M.; Miller, W.A.; Gray, S.M. A stem-loop structure in potato leafroll virus open reading frame 5 (ORF5) is essential for readthrough translation of the coat protein ORF stop codon 700 bases upstream. *J. Virol.* **2018**, *92*, e01544-17. [CrossRef]
88. Jacks, T.; Madhani, H.D.; Masiarz, F.R.; Varmus, H.E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **1988**, *55*, 447–458. [CrossRef]
89. Brierley, I.; Jenner, A.J.; Inglis, S.C. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **1992**, *227*, 463–479. [CrossRef]
90. Wren, J.D.; Roossinck, M.J.; Nelson, R.S.; Scheets, K.; Palmer, M.W.; Melcher, U. Plant virus biodiversity and ecology. *PLoS Biol.* **2006**, *4*, 314–315. [CrossRef]
91. Balke, I.; Resevica, G.; Zeltins, A. The ryegrass mottle virus genome codes for a sobemovirus 3c-like serine protease and RNA-dependent RNA polymerase translated via -1 ribosomal frameshifting. *Virus Genes* **2007**, *35*, 395–398. [CrossRef]
92. Somera, M.; Sarmiento, C.; Truve, E. Overview on sobemoviruses and a proposal for the creation of the family sobemoviridae. *Viruses* **2015**, *7*, 3076–3115. [CrossRef]
93. Sztuba-Solińska, J.; Urbanowicz, A.; Figlerowicz, M.; Bujarski, J.J. RNA-RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* **2011**, *49*, 415–443. [CrossRef]
94. Smirnova, E.; Firth, A.E.; Miller, W.A.; Scheidecker, D.; Brault, V.; Reinbold, C.; Rakotondrafara, A.M.; Chung, B.Y.W.; Ziegler-Graff, V. Discovery of a small non-aug-initiated orf in poleroviruses and luteoviruses that is required for long-distance movement. *PLoS Pathog.* **2015**, *11*, e1004868. [CrossRef]
95. Edwards, M.C.; Weiland, J.J.; Todd, J.; Stewart, L.R.; Lu, S. ORF43 of maize rayado fino virus is dispensable for systemic infection of maize and transmission by leafhoppers. *Virus Genes* **2016**, *52*, 303–307. [CrossRef] [PubMed]
96. Firth, A.E.; Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **2012**, *93*, 1385–1409. [CrossRef] [PubMed]
97. Gordon, K.; Fütterer, J.; Hohn, T. Efficient initiation of translation at non-aug triplets in plant cells. *Plant J. Cell Mol. Biol.* **1992**, *2*, 809–813.
98. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8301–8305. [CrossRef] [PubMed]
99. Fütterer, J.; Potrykus, I.; Bao, Y.; Li, L.; Burns, T.M.; Hull, R.; Hohn, T. Position-dependent att initiation during plant pararetrovirus rice tungro bacilliform virus translation. *J. Virol.* **1996**, *70*, 2999–3010. [CrossRef] [PubMed]
100. Turina, M.; Maruoka, M.; Monis, J.; Jackson, A.O.; Scholthof, K.B. Nucleotide sequence and infectivity of a full-length cdna clone of panicum mosaic virus. *Virology* **1998**, *241*, 141–155. [CrossRef]
101. Castaño, A.; Ruiz, L.; Hernández, C. Insights into the translational regulation of biologically active open reading frames of pelargonium line pattern virus. *Virology* **2009**, *386*, 417–426. [CrossRef]
102. Hammond, R.W.; Hammond, J. Maize rayado fino virus capsid proteins assemble into virus-like particles in *Escherichia coli*. *Virus Res.* **2010**, *147*, 208–215. [CrossRef]
103. Natilla, A.; Murphy, C.; Hammond, R.W. Mutations in the alpha-helical region of the amino terminus of the maize rayado fino virus capsid protein and cp:Rna ratios affect virus-like particle encapsidation of RNAs. *Virus Res.* **2015**, *196*, 70–78. [CrossRef]
104. Natilla, A.; Hammond, R.W. Analysis of the Solvent Accessibility of Cysteine Residues on Maize rayado fino virus Virus-like Particles Produced in *Nicotiana benthamiana* Plants and Cross-linking of Peptides to VLPs. *J. Vis. Exp.* **2013**. [CrossRef]
105. Espinoza, A.M.; Ramírez, P.; León, P. Cell-free translation of maize rayado fino virus genomic RNA. *J. Gen. Virol.* **1988**, *69*, 757–762. [CrossRef]
106. Patel, A.; McBride, J.A.M.; Mark, B.L. The endopeptidase of the maize-affecting marafivirus type member maize rayado fino virus doubles as a deubiquitinase. *J. Biol. Chem.* **2021**, *297*, 100957. [CrossRef] [PubMed]
107. Lombardi, C.; Ayach, M.; Beaurepaire, L.; Chenon, M.; Andreani, J.; Guerois, R.; Jupin, I.; Bressanelli, S. A compact viral processing proteinase/ubiquitin hydrolase from the otu family. *PLoS Pathog.* **2013**, *9*, e1003560. [CrossRef] [PubMed]
108. Bailey-Elkin, B.A.; van Kasteren, P.B.; Snijder, E.J.; Kikkert, M.; Mark, B.L. Viral OTU deubiquitinases: A structural and functional comparison. *PLoS Pathog.* **2014**, *10*, e1003894. [CrossRef] [PubMed]
109. Komander, D.; Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **2012**, *81*, 203–229. [CrossRef] [PubMed]
110. Mann, K.S.; Sanfaçon, H. Expanding repertoire of plant positive-strand RNA virus proteases. *Viruses* **2019**, *11*, 66. [CrossRef] [PubMed]
111. Fieulaine, S.; Witte, M.D.; Theile, C.S.; Ayach, M.; Ploegh, H.L.; Jupin, I.; Bressanelli, S. Turnip yellow mosaic virus protease binds ubiquitin suboptimally to fine-tune its deubiquitinase activity. *J. Biol. Chem.* **2020**, *295*, 13769–13783. [CrossRef]
112. Decroly, E.; Ferron, F.; Lescar, J.; Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.* **2012**, *10*, 51–65. [CrossRef]
113. Gorbalenya, A.E.; Koonin, E.V. Viral proteins containing the purine ntp-binding sequence pattern. *Nucleic Acids Res.* **1989**, *17*, 8413–8440. [CrossRef]

114. Gorbalenya, A.E.; Koonin, E.V.; Donchenko, A.P.; Blinov, V.M. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* **1988**, *235*, 16–24. [CrossRef]
115. Wu, B.; Grigull, J.; Ore, M.O.; Morin, S.; White, K.A. Global organization of a positive-strand RNA virus genome. *PLoS Pathog.* **2013**, *9*, e1003363. [CrossRef] [PubMed]
116. Agranovsky, A.A.; Dolja, V.V.; Gorbulev, V.G.; Kozlov, Y.V.; Atabekov, J.G. Aminoacylation of barley stripe mosaic virus RNA: Polyadenylate-containing RNA has a 3′-terminal tyrosine-accepting structure. *Virology* **1981**, *113*, 174–187. [CrossRef]
117. Giegé, R.; Florentz, C.; Dreher, T.W. The TYMV tRNA-like structure. *Biochimie* **1993**, *75*, 569–582. [CrossRef]
118. Al Rwahnih, M.; Daubert, S.; Golino, D.; Rowhani, A. Deep sequencing analysis of RNAs from a grapevine showing syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* **2009**, *387*, 395–401. [CrossRef]
119. Loughran, G.; Firth, A.E.; Atkins, J.F. Ribosomal frameshifting into an overlapping gene in the 2b-encoding region of the cardiovirus genome. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, E1111–E1119. [CrossRef]
120. Staple, D.W.; Butcher, S.E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005**, *3*, e213. [CrossRef]