Article
Unlocking the Hidden Genetic Diversity of Varicosaviruses, the Neglected Plant Rhabdoviruses

Nicolas Bejerman 1,2,* , Ralf G. Dietzgen 3,* and Humberto Debat 1,2,*

1 Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria (IPAVE—CIAP—INTA), Camino 60 Cuadras Km 5.5, Córdoba X5020ICA, Argentina
2 Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad de Fitopatología y Modelización Agrícola, Camino 60 Cuadras Km 5.5, Córdoba X5020ICA, Argentina
3 Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia, QLD 4072, Australia
* Correspondence: bejerman.nicolas@inta.gob.ar (N.B.); r.dietzgen@uq.edu.au (R.G.D.); debat.humberto@inta.gob.ar (H.D.)

Abstract: The genus Varicosavirus is one of six genera of plant-infecting rhabdoviruses. Varicosaviruses have non-enveloped, flexuous, rod-shaped virions and a negative-sense, single-stranded RNA genome. A distinguishing feature of varicosaviruses, which is shared with dichorhaviruses, is a bi-segmented genome. Before 2017, a sole varicosavirus was known and characterized, and then two more varicosaviruses were identified through high-throughput sequencing in 2017 and 2018. More recently, the number of known varicosaviruses has substantially increased in concert with the extensive use of high-throughput sequencing platforms and data mining approaches. The novel varicosaviruses have revealed not only sequence diversity, but also plasticity in terms of genome architecture, including a virus with a tentatively unsegmented genome. Here, we report the discovery of 45 novel varicosavirus genomes which were identified in publicly available metatranscriptomic data. The identification, assembly, and curation of the raw Sequence Read Archive reads has resulted in 39 viral genome sequences with full-length coding regions and 6 with nearly complete coding regions. The highlights of the obtained sequences include eight varicosaviruses with unsegmented genomes, which are linked to a phylogenetic clade associated with gymnosperms. These findings have resulted in the most complete phylogeny of varicosaviruses to date and shed new light on the phylogenetic relationships and evolutionary landscape of this group of plant rhabdoviruses. Thus, the extensive use of sequence data mining for virus discovery has allowed us to unlock the hidden genetic diversity of varicosaviruses, the largely neglected plant rhabdoviruses.

Keywords: plant rhabdovirus; varicosaviruses; genome architecture; virus taxonomy; metatranscriptomics

1. Introduction
A recently discovered huge number of diverse viruses has revealed the complexities of the evolutionary landscape of replicating entities and the challenges associated with their classification [1], leading to the first comprehensive proposal of the virus world megataxonomy [2]. Nevertheless, a minuscule portion, likely a small fraction of one percent, of the virome has been characterized so far [3]. Therefore, we have a limited knowledge of the vast world virome, with its remarkable diversity, that includes every potential host organism assessed so far [4–6]. Data mining of publicly available transcriptome datasets has become an efficient and inexpensive strategy to unlock the diversity of the plant virome [5]. Data-driven virus discovery relies on the vast number of available datasets on the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). This resource, which is growing at an exceptional rate and includes data of a large and diverse number of organisms, represents a substantial fraction of species that
Pathogens 2022, 11, 1127

2 of 17

population our planet, which makes the SRA database an invaluable source to identify novel viruses [7].

Varicosavirus is one of the six genera that are comprised of plant rhabdoviruses (family Rhabdoviridae, subfamily Betarhabdovirinae), and its members are thought to have a negative-sense, single-stranded, bi-segmented RNA genome [8]. Nevertheless, recently, we described the first apparently unsegmented varicosavirus [9]. In those varicosaviruses with segmented genomes, RNA 1 consists of one to two genes, with one of those encoding the RNA-dependent RNA polymerase L, while RNA 2 consists of three to five genes, with the first open reading frame (ORF) encoding a nucleocapsid protein (N) [8,10]. On the other hand, the only unsegmented varicosavirus described so far has five ORFs, in the order: 3′-N-Protein 2-Protein 3-Protein 4-L-5′ [9]. Varicosaviruses appear to have a diverse host range that includes dicots, monocots, gymnosperms, ferns, and liverworts [6,9]. The vector of a sole member, lettuce big vein-associated virus (LBVaV), has been characterized, which is the chytrid fungus Olpidium spp. [11].

Until 2017, LBVaV was the only identified and extensively characterized varicosavirus [12–14], and then, in 2017 and 2018, two novel varicosaviruses were identified through high-throughput sequencing (HTS) [15,16]. However, in 2021 and 2022, there was a five-fold increase in the number of reported varicosaviruses, with 12 out 15 discovered through data mining of publicly available transcriptome datasets [6,9,17,18], while the other three were identified using HTS [19–21] (Supplementary Figure S1). Nevertheless, only some minor biological aspects, such as mechanical transmissibility, of some of these members were further characterized [15,20]. Therefore, varicosaviruses are, by far, the least-studied plant rhabdoviruses, and many aspects of their epidemiology remain elusive.

In terms of genetic diversity, before this study, while greatly expanded by recent works, the Varicosavirus genus includes only three accepted species and 15 tentative members. In this study, we identified 45 novel varicosaviruses by analyzing publicly available metatranscriptomic data. Thus, the extensive use of data mining for virus discovery has allowed us to unlock some of the hidden diversity of varicosaviruses, the much-neglected plant rhabdoviruses.

2. Material and Methods

2.1. Identification of Plant Rhabdovirus Sequences from Public Plant RNA-seq Datasets

Three strategies were used to detect varicosavirus sequences: (1) Amino acid sequences corresponding to the nucleocapsid and polymerase proteins of known varicosaviruses were used as queries in tBlastn searches with the parameters word size = 6, expected threshold = 10, and scoring matrix = BLOSUM62, against the Viridiplantae (taxid: 33090) Transcriptome Shotgun Assembly (TSA) sequence databases. The obtained hits were manually explored and based on percentage identity, query coverage, and E-value (>1 × 10^{-5}) and shortlisted as likely corresponding to novel virus transcripts, which were then further analyzed. (2) Raw sequence data corresponding to the SRA database associated with the 1K study [22] were explored for varicosa-like virus sequences. (3) The Serratus database was explored, employing the serratus explorer tool [5], and using as queries the sequences of LBVaV, red clover varicosavirus, and black grass varicosavirus. Those SRA libraries that matched the query sequences (alignment identity > 45%; score > 10) were further explored in detail.

2.2. Sequence Assembly and Identification

The nucleotide (nt) raw sequence reads from each SRA experiment, which are associated with different NCBI bioprojects (Table 1), were downloaded and pre-processed by trimming and filtering with the Trimmomatic tool as implemented in http://www.usadellab.org/cms/?page=trimmomatic (accessed on 19 August 2022). The resulting reads were assembled de novo with rnaSPAdes using standard parameters on the Galaxy.org server. The transcripts obtained from the de novo transcriptome assembly were subjected to bulk local BLASTX searches (E-value < 1 × 10^{-5}) against a collection of varicosavirus protein se-
quences available at https://www.ncbi.nlm.nih.gov/protein?term=txid140295[Organism] (accessed on 19 August 2022). The resulting viral sequence hits of each bioproject were visually explored. Tentative virus-like contigs were curated (extended or confirmed) by iteratively mapping each SRA library’s filtered reads. This strategy used BLAST/ohmmer to extract a subset of reads related to the query contig and used the retrieved reads to extend the contig and then repeated the process iteratively using the extended sequence as query. The extended and polished transcripts were reassembled using the Geneious v8.1.9 (Biomatters Ltd., San Diego, CA, USA) alignment tool with high sensitivity parameters. Bowtie2, available at http://bowtie-bio.sourceforge.net/bowtie2/index.shtml (accessed on 26 September 2022), was used with standard parameters for filtered read mapping to calculate the mean coverage of each assembled virus sequence.

Table 1. Summary of the novel varicosaviruses identified from the plant RNA-seq data available in the NCBI database. The acronyms of the best hits are listed in Supplementary Table S1.

| Plant Host | Taxa/Family | Virus Name/Abbreviation | Bioproject ID/ Data Citation | Segment/Coverage | Length (nt) | Accession Number | Protein ID | Length (aa) | Highest Scoring Virus-Protein E-Value/Query Coverage/%Identity% ( Blast P) |
|------------|-------------|-------------------------|-----------------------------|------------------|------------|------------------|------------|------------|---------------------------------------------------------------|
| Trojan fir (Abies nordmanniana) | Gymnosperm/Pinaeaceae | Abies virus 1/ AbV1 | PRJNA83706/ University of Connecticut, USA | RNA1:30.97X | 11,287 | BK016731 | N | 430 | PfAfiV1-N/9e-130/97/50.79 |
| | | | | | | | | 2 | 420 | PfAfiV1-P2/2e-19/97/20.05 |
| | | | | | | | 3 | 317 | PfAfiV1-P2/3e-103/97/47.44 | no hits |
| | | | | | | | 4 | 163 | L | 2060 | PfAfiV1-L/0.0/98/52.68 |
| Dwarf mistletoe (Arceuthobium schiaeasteum) | dicot/ Santalaceae | Arceuthobium virus 8/ AreV8 | PRJNA307380/ [21] | RNA1:9.31X/RNA2:72.33X | 6628/4149 | BK016732/ BK016733 | L | 2013 | AAsV2-L/0.0/98/100 |
| | | | | | | | N | 369 | ZaVV1-N/1e-34/91/28.36 | no hits |
| | | | | | | | 2 | 453 | N | 159 | no hits |
| Bei Wu Tou (Acer truncatum) | dicot/ Ranunculaceae | Aconitum virus 1/ AcoV1 | PRJNA647302/ [21] | RNA1:10.16X/RNA2:105.03X | 6483/5616 | BK016734/ BK016735 | L | 2000 | ZaVV1-L/0.0/97/61.18 |
| | | | | | | | N | 424 | ZaVV1-N/2e-115/99/43.82 | no hits |
| | | | | | | | 2 | 329 | VVFP2-N/4e-36/80/32.13 | no hits |
| | | | | | | | 3 | 311 | VzVV1-P1/3e-105/85/54.51 | no hits |
| | | | | | | | 4 | 204 | VVFP4-N/2e-17/57/33.33 | no hits |
| | | | | | | | 5 | 297 | VVFP5-N/5e-12/97/26.18 | no hits |
| Catkin yew (Amentotaxus argentea) | Gymnosperm/ Cephalotaxaceae | Amentotaxus virus 1/ AmeV1 | PRJNA498605/ [25] | RNA1:109.96X | 10,965 | BK016736 | N | 391 | AsaV2-L/3e-111/94/43.95 |
| | | | | | | | 2 | 431 | PiFleV1-L/2e-66/95/26.98 | no hits |
| | | | | | | | 3 | 314 | AsaV2-P2/4e-83/94/43.42 | no hits |
| | | | | | | | 4 | 187 | AsaV2-P2/3e-45/96/34.57 | no hits |
| | | | | | | | 5 | 160 | PiFleV1-L/0.0/99/46.36 | no hits |
| Common windgrass (Apera spica-venti) | monocot/ Poaceae | Apera virus 1/ ApV1 | PRJNA356380/ [26] | RNA1:11.98X/RNA2:110.50X | 6516/6502 | BK016737/ BK016738 | L | 2027 | MeRoV1-L/0.0/98/52.12 |
| | | | | | | | N | 447 | MeRoV1-N/2e-69/82/54.57 | no hits |
| | | | | | | | 2 | 363 | MeRoV1-P2/4e-17/75/26.37 | no hits |
| | | | | | | | 3 | 298 | MeRoV1-P2/2a-80/97/41.25 | no hits |
| | | | | | | | 4 | 196 | L | 444 | no hits |
| Lace plant (Apophyton nudagascarian- sis) | monocot/ Apophytonaceae | Aponogeton virus 1/ ApoV1 | PRJNA591467/ [27] | RNA1:36.42X/RNA2:81.25X | 6678/5628 | BK016739/ BK016740 | L | 2022 | BrRV1-L/0.0/98/52.7 |
| | | | | | | | N | 454 | BrRV1-N/7e-81/88/57 | no hits |
| | | | | | | | 2 | 300 | BrRV1-P3/2e-45/96/34 | no hits |
| | | | | | | | 3 | 174 | BrRV1-P3/0e-03/73/25 | no hits |
| Warmwood (Artemisia abovinum) | dicot/ Astereaceae | Artemisia virus 1/ ArtV1 | PRJNA731565/ [28] | RNA1:33.06X/RNA2:50.30X | 7373/4497 | BK016741/ BK016742 | L | 2020 | BrRV1-L/0.0/98/49.18 |
| | | | | | | | N | 453 | BrRV1-N/3e-45/76/28.90 | no hits |
| | | | | | | | 2 | 294 | N | 174 | no hits |
| Common milkweed (Asclepias syriaca) | dicot/ Asclepiadaceae | Asclepias syriaca virus 3 AsSyV3 | PRJNA210776/ [29] | RNA1:37.86X/RNA2:138.94X | 6506/6280 | BK016743/ BK016744 | L | 2021 | TTVF-L/0.0/94/42.62 |
| | | | | | | | N | 453 | TTVF-N/3e-39/78/32.13 | no hits |
| | | | | | | | 2 | 370 | TTVF-P2/2e-73/72/29.26 | no hits |
| | | | | | | | 3 | 286 | TTVF-P3/1e-45/87/33.33 | no hits |
| Beautiful tree fern (Asplenium loricatum) | Polypodiophyta/ Aspleniaceae | Asplenium virus 1/ AspV1 | PRJNA281136/ [30] | RNA1:4.51X/RNA2:8.91X | 6287*/4371* | BK016745/ BK016746 | L | 1957* | TTVF-L/0.0/98/43.81 |
| | | | | | | | N | 396 | TTVF-N/2e-79/90/37.82 | no hits |
| | | | | | | | 2 | 490 | N | 294 | no hits |
| | | | | | | | 3 | 294 | TTVF-P3/1e-45/87/33.33 | no hits |
| | | | | | | | 4 | 127* | L | 98 | no hits |
| Plant Host                          | Taxa/Family                  | Virus Name/Abbreviation                  | Bioproject ID/ Data Citation | Segment/Coverage | Length (nt) | Accession Number | Protein ID | Length (aa) | Highest Scoring Virus-Protein E-Value/Query Coverage% | Identity% |
|-----------------------------------|-----------------------------|------------------------------------------|------------------------------|------------------|-------------|------------------|------------|-------------|---------------------------------------------|-----------|
| Short pod mustard                 | Brassica incana              | Brassica virus 2_Ins/BrV2_Ins           | PRJNA287699/ [31]          | RNA2/11.89X     | 6316        | BK016747         | 2          | 591         | TVLV-L/0.0/99.41.86                              | no hits   |
| Indian mustard                    | Brassica juncea var. rugosa | Brassica virus 2_Jun/BrV2_Jun           | PRJNA28942/ [32]           | RNA2/50.91X     | 5537        | BK016750         | 3          | 495         | TVLV-L/0.0/99.41.57                              | no hits   |
| Chinese kale                      | Brassica oleracea var. albenga (Brassica oleracea var. albenga) | Brassica virus 2_Ole/BrV2_Ole         | PRJNA252713/ [33]          | RNA2/11.03X     | 5647        | BK016752         | 3          | 495         | TVLV-L/0.0/99.41.57                              | no hits   |
| Crab lipped spider orchid (Caladenia plicata) | Caladenia virus 1/CalV1 | Caladenia virus 1/CalV1               | PRJNA384675/ [34]          | RNA1/10.51X     | 6454        | BK016755         | 3          | 495         | TVLV-L/0.0/99.41.57                              | no hits   |
| Coreflower (Centaurea cyanus)     | Centaurea cyanus             | Centaurea virus 1/ CenV1               | PRJNA37156/ [26]           | RNA1/63.11X     | 6789        | BK016757         | 2          | 341         | TVLV-L/0.0/99.50.50                              | no hits   |
| Chamomile (Chamaemelum nobilis)  | Chamaemelum nobilis          | Chamaemelum virus 1/ ChaV1             | PRJNA38246/ [35]           | RNA1/21.33X     | 6670        | BK016759         | 3          | 341         | TVLV-L/0.0/99.41.57                              | no hits   |
| Melon (Cucumis melo)              | Cucumis melo                 | Cucumis virus 1/ CuvV1                | PRJNA381300/ [36]          | RNA1/47.9X      | 6199        | BK016761         | 2          | 404         | TVLV-L/0.0/99.41.57                              | no hits   |
| Chen cypress (Cupressus chengensis) | Gymnosperm/ Cupressaceae      | Cupressus virus 1/ CupV1              | PRJNA596937/ [37]          | RNA1/32.13X     | 12143       | BK016763         | 5          | 419         | TVLV-L/0.0/99.41.57                              | no hits   |
| Tree mustard hair fern           | Polytechnium/ Hypostradiaceae | Polytechnium/ Hypostradiaceae         | PRJNA42211/ [38]           | RNA1/8.88X      | 6321        | BK016764         | 2          | 313         | TVLV-L/0.0/99.41.57                              | no hits   |
| Wallflower (Erysimum hirsutum)    | Erysimum hirsutum            | Erysimum virus 1/ EryV1               | PRJNA607615/ [39]          | RNA1/271.24X    | 6678        | BK016766         | 2          | 439         | TVLV-L/0.0/99.41.57                              | no hits   |
| Liverwort (Furcellaria orientalis) | Marchantiophyta Furcellaria | Furcellaria virus 1/ FruV1            | PRJNA60575/ [40]           | RNA1/11.60X     | 6485        | BK016768         | 3          | 289         | TVLV-L/0.0/99.41.57                              | no hits   |
| Noug (Guizotia abyssinica)        | AsaV1/ Cupressaceae          | Cupressus virus 1/ CupV1              | PRJNA37156/ [28]          | RNA1/153.49X    | 6457        | BK016770         | 3          | 340         | TVLV-L/0.0/99.41.57                              | no hits   |
| Common velvet grass (Holcus lanatus) | Holcus virus 1/ HolV1     | Holcus virus 1/ HolV1                | PRJNA62165/ [41]           | RNA1/19.48X     | 6571        | BK016772         | 2          | 476         | TVLV-L/0.0/99.41.57                              | no hits   |
| Oxeye daisy (Leucanthemum vulgarum) | Leucanthemum virus 1/ LeaV1 | Leucanthemum virus 1/ LeaV1            | PRJNA37156/ [28]          | RNA1/141.76X    | 6763        | BK016774         | 2          | 448         | TVLV-L/0.0/99.41.57                              | no hits   |
| Plant Host          | Taxa/Family       | Virus Name/Abbr | Bioproject ID/ Data Citation | Segment/Coverage | Length (nt) | Accession Number | Protein ID | Length (aa) | Highest Scoring Virus/Protein/E-Value/Query Coverage/%Identity% (Plast F) |
|---------------------|-------------------|----------------|-----------------------------|------------------|-------------|------------------|------------|--------------|---------------------------------------------------------------------|
| Downy flax          | dicot/ Linaceae   | Linium virus 1/ LiniV1 | PRJEB21674/1000 Plant (1KP) Transcriptomes Initiative | RNA1:26.47X RNA2:11.90X | 5999 * | BK016176 | 2 | 463 | no hits |
| Yunnan pine         |                   |                |                             |                  |            |                  |            |              | no hits |
| Riverbank linden    | dicot/ Fabaceae   | Lupinus virus 1/ LupV1 | PRJNA318864/41 | RNA1:14.64X RNA2:9.57X | 6688 | BK016182 | 2 | 497 | no hits |
|                     |                   |                |                             |                  |            |                  |            |              | no hits |
| Sweet clover        | dicot/ Fabaceae   | Melilotus virus 1/Alb/MeV1 Alb | PRJNA647665/42 | RNA1:30.69X RNA2:98.21X | 6657 | BK016184 | 3 | 393 | no hits |
| (Melilotus spp)     |                   |                |                             |                  |            |                  |            |              | no hits |
| Sweet clover        | dicot/ Fabaceae   | Melilotus virus 1/Oli/MeV1 Oli | PRJNA751395/43 | RNA1:12.13X RNA2:25.36X | 6433 | BK016186 | 2 | 399 | no hits |
| (Melilotus spp)     |                   |                |                             |                  |            |                  |            |              | no hits |
| Early spider orchid | monocot/ Orchidaceae | Ophrys virus 1/ OphV1 | PRJNA574279/44 | RNA1:7.72X RNA2:20.15X | 6134 * | BK016188 | 3 | 293 | no hits |
| (Ophrys sphegodes)  |                   |                |                             |                  |            |                  |            |              | no hits |
| Purple Grass        | monocot/ Poaceae  | Pennisetum virus 1/ PenV1 | PRJNA282666/23 | RNA1:44.59X RNA2:112.25X | 6284 | BK016190 | 2 | 286 | no hits |
| (Pennisetum vulgare) |                   |                |                             |                  |            |                  |            |              | no hits |
| Qinghai spruce      | Gymnosperm/ Pinaceae | Picea virus 1/ PceV1 | PRJNA037530/23 | RNA1:5.86X | 11,193 | BK016192 | 4 | 382 | no hits |
| (Picea crassifolia) |                   |                |                             |                  |            |                  |            |              | no hits |
| Jack pine           | Gymnosperm/ Pinaceae | Pinus banksiana virus 1/ PinBanV1 | PRJNA254866/45 | RNA1:97.66X | 11276 | BK016193 | 4 | 319 | no hits |
| (Pinus banksiana)   |                   |                |                             |                  |            |                  |            |              | no hits |
| Yunnan pine         | Gymnosperm/ Pinaceae | Pinus yunnanensis virus 1/PyYunnV1 | PRJNA057489/46 | RNA1:36.47X | 12,057 | BK016194 | 3 | 204 | no hits |
| (Pinus yunnanensis) |                   |                |                             |                  |            |                  |            |              | no hits |
| Spenndor primrose   | dicot/ Primulaceae | Primula virus 1/ PuiV1 | PRJNA344868/47 | RNA1:7.72X RNA2:149.23X | 6352 | BK016195 | 2 | 288 | no hits |
| (Primula oreades)   |                   |                |                             |                  |            |                  |            |              | no hits |
| Goldilocks buttercup | dicot/ Ranunculaceae | Ranunculus virus 1/ RanV1 | PRJNA217403/48 | RNA1:29.64X RNA2:163.27X | 6481 | BK016197 | 2 | 145 | no hits |
| (Ranunculus auricomus) |                   |                |                             |                  |            |                  |            |              | no hits |
| Radish              | dicot/ Brassicaceae | Raphanus virus 1/ RapV1 | PRJNA598586/49 | RNA1:365.02X RNA2:521.73X | 6410 | BK016199 | 3 | 175 | no hits |
| (Raphanus sativus)  |                   |                |                             |                  |            |                  |            |              | no hits |
| Siberian currant     | dicot/ Convolvulaceae | Ribs virus 1/ RibV1 | PRJNA047394/50 | RNA1:6.28X RNA2:33.97X | 6323 | BK016181 | 2 | 92 | no hits |
| (Ribes laciniatum)  |                   |                |                             |                  |            |                  |            |              | no hits |
Table 1. Cont.

| Plant Host                     | Taxa/Family            | Virus Name/Abbreviation | Bioproject ID/ Data Citation | Segment/Coverage | Length (nt) | Accession Number | Protein ID | Highest Scoring Virus-Protein/E-Value/Query Coverage %/Identity % |
|--------------------------------|------------------------|-------------------------|-----------------------------|------------------|-------------|------------------|------------|---------------------------------------------------------------|
| Japanese umbrella pine (umbrella pine) | Gymnosperm/ | Sciadopitys virus 1_Chi/Scv1_Chi | PRJNA396655/ Institute of Botany, CAS, China | RNA1/98.99X | 11,224 | BK061803 | N | 389 | ASAV2-N/1e-115/95/41.13 |
| Japanese umbrella pine (umbrella pine) | Gymnosperm/ | Sciadopitys virus 1_Can/Scv1_Can | PRJEB49212/ [51] | RNA1/14.02X | 11,132 | BK061804 | N | 389 | ASAV2-N/1e-115/95/41.67 |
| Woody grassland senecio (Senecio coronarius) | dicot/ | Seneo virus 1_SerV1 | PRJNA312157/ [52] | RNA1/10.59X | 6173 * | BK061805 | 2 | 466 | ASAV2-P2/6e-26/50/30.14 |
| Bladder campion (Silene vulgaris) | dicot/ | Silene virus 1_SilV1 | PRJNA104951/ [53] | RNA1/29.59X | 6391 | BK061807 | N | 389 | ASAV2-N/1e-115/95/41.67 |
| Broadbean daisy (Streptoglossa macrocephala) | dicot/ | Streptoglossa virus 1_StV1 | PRJNA372165/ [28] | RNA1/131.33X | 5130 | BK061814 | L | 2031 * | LBV-L/0.0/98/42.82 |
| Tansy (Tanacetum vulgare) | dicot/ | Tanacetum virus 1_TanV1 | PRJNA464340/ [54] | RNA1/10.19X | 6888 | BK061815 | L | 2019 | SpV1-L/6.0/95/95.91 |
| Hybrid yew (Taxus media) | Gymnosperm/ | Taxus virus 1_TaxV1 | PRJNA497542/ [55] | RNA1/37.28X | 11,174 | BK061817 | L | 2020 | SpV1-L/6.0/95/95.91 |
| Chinese mistletoe yew (Torreya grandis) | Gymnosperm/ | Torreya virus 1_TorV1 | PRJNA498605/ [25] | RNA1/59.04X | 10,253 | BK061818 | L | 379 | TvVV-N/2e-57/93.25 |
| Liverwort (Draba lacino) | Marchantiophyta/ | Trebucha virus 1_TreV1 | PRJNA505755/ [56] | RNA1/364.20X | 6684 | BK061819 | N | 379 | TvVV-L/0.0/97/35.4 |
| Wheat (Triticum aestivum) | monochot/ | Triticum virus 1_TriV1 | PRJNA520830/ [56] | RNA1/10.25X | 6290 | BK061820 | N | 379 | TvVV-L/0.0/97/35.4 |
| Variiegated swallow-wort (Vincenticum versicolor) | dicot/ | Vincenticum virus VV_VinV1 | PRJNA99262/ [57] | RNA1/56.05X | 6598 | BK061823 | N | 379 | TvVV-L/0.0/97/35.4 |
| Corn (Zea mays) | monochot/ | Zea virus 1_ZeaV1 | PRJNA407269/ [58] | RNA1/6.25X | 6407 | BK061826 | L | 2037 | AMV1-L/0.0/99/49.07 |

* partial sequence.

2.3. Bioinformatics Tools and Analyses

2.3.1. Sequence Analyses

ORFs were predicted with ORFFinder (minimal ORF length 150 nt, genetic code 1, https://www.ncbi.nlm.nih.gov/orffinder/, accessed on 22 August 2022) and the functional domains and architectures of translated gene products were determined using InterPro (https://www.ebi.ac.uk/interpro/search/sequence-search, accessed on 22 August 2022)
and the NCBI conserved domain database-CDD v3.19 (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, accessed on 22 August 2022). Further, HHPred and HHBlits, as implemented in https://toolkit.tuebingen.mpg.de/#/tools/ (accessed on 22 August 2022), were used to complement the annotation of divergent predicted proteins by hidden Markov models. Transmembrane domains were predicted using the TMHMM version 2.0 tool (http://www.cbs.dtu.dk/services/TMHMM/, accessed on 22 August 2022).

2.3.2. Pairwise Sequence Identity

Percentage amino acid (aa) sequence identities of the L protein of those varicosaviruses identified in this study, as well as those available in the NCBI database, were calculated using SDTv1.2 [59]. Virus names, abbreviations, and NCBI accession numbers of the varicosaviruses already reported are shown in Supplementary Table S1.

2.3.3. Phylogenetic Analysis

Phylogenetic analysis based on the predicted polymerase protein of all available varicosaviruses was completed using MAFFT 7.505 (https://mafft.cbrc.jp/alignment/software) (accessed on 25 August 2022) with multiple aa sequence alignments and using FFT-NS-i as the best-fit model. The aligned aa sequences were used as inputs to generate phylogenetic trees using the maximum-likelihood method (best-fit model = E-INS-i) with the FastTree 2.1.11 tool (available at http://www.microbesonline.org/fasttree/) (accessed on 25 August 2022). Local support values were calculated with the Shimodaira-Hasegawa test (SH) and 1000 trees were resampled. The L proteins of four selected cytorhabdoviruses were used as outgroups. To explore the potential phylogenetic co-divergence of varicosaviruses with their associated host plants, plant host cladograms were generated in phyloT v.2 (https://phylot.biobyte.de/, accessed on 26 August 2022) based on NCBI Taxonomy. Connections were manually inferred between the viral and plant phylogenetic trees and visually inspected.

3. Results and Discussion

Most varicosaviruses likely do not induce easily discernable disease symptoms. Since their presence is not expected in the sequencing libraries of apparently “healthy” vegetables, they are ideal candidates to be identified through mining publicly available metatranscriptomic data. Accordingly, very recently, 12 novel proposed varicosaviruses were discovered when publicly available transcriptome datasets were mined [6,9,17,18]. Therefore, to unlock the hidden diversity of varicosaviruses, we extensively searched for these viruses in already available plant transcriptome data. This bioinformatics research resulted in the identification of 45 novel varicosaviruses, including the corrected full-length coding genome segments of the previously reported Arceuthobium sinesis-associated virus 2 (ASaV2) [18], which had apparently been reconstructed from the genome segments of two different varicosaviruses. We also identified three novel variants of three recently discovered varicosaviruses, confirming and strengthening the results previously reported by Bejerman et al. [9]. This significant number of newly discovered varicosaviruses represents a 3.5-fold increase in the known varicosaviruses (Supplementary Figure S1), which clearly highlights the importance of data-driven virus discovery to illuminate the landscape of largely overlooked taxonomic groups, such as varicosaviruses.

More details, identification, assembly, and curation of raw SRA reads in this study resulted in 39 viral genome sequences with full-length coding regions and six with nearly complete coding regions. These viruses were associated with 45 plant host species (Table 1). Most of the tentative plant hosts of the novel varicosaviruses are herbaceous dicots (24/45), nine are herbaceous monocots, eight are gymnosperms, and four are liverworts and ferns (Table 1).

The genomes of 37 viruses identified in this study were bisegmented, where the RNA 1 of 36 of them encodes only the L protein, while the RNA 1 of Chamaemelum virus 1 (ChaV1) has an additional ORF 5’ to the L gene, supported by the identification of the conserved
intergenic sequence (see below), encoding a 171 aa putative protein (Table 1, Figure 1), which appears to be the first varicosavirus reported with an ORF in this position. The RNA 2 segments of these 37 viruses have three to five genes in the order 3′-N-PX-5′. Twelve of them have three genes, while 17 have four genes and eight contained five genes (Table 1, Figure 1). Of the previously reported varicosaviruses, six have three genes, four have four genes, and four have five genes; therefore, RNA 2 has a flexible genomic architecture and is apparently the most frequent genomic organization in the RNA 2 of varicosaviruses that includes four genes (21 members) or three genes (18 members).

Figure 1. Left: Maximum-likelihood phylogenetic tree based on the amino acid sequence alignments of the complete L gene of all the varicosaviruses reported thus far and in this study. The scale bar indicates the number of substitutions per site. The node labels indicate fast tree support values. Four cytorhabdoviruses were used as outgroups. Right: Genomic organization of the varicosavirus sequences used in the phylogeny. An asterisk and bold font indicate those viruses identified in this study. The accession numbers of all the viruses are listed in Supplementary Table S1 and Table 1.
The consensus gene junction sequences of the bisegmented varicosaviruses were determined to be 3’ AU(N)_3UUUUUGCU 5’ (Table 2), while the gene junction sequences of all but one of the unsegmented varicosaviruses differed slightly in the 3’ end, being GU(N)_3 instead of AU(N)_3 (Table 2). Strikingly, the consensus gene junction of the unsegmented Torreya virus 1 (TorV1) was similar to that of the bisegmented varicosaviruses. The potential implication of this difference in the gene junctions needs to be explored since it could be linked to the basal evolutionary grouping of TorV1 (see below).

Table 2. Consensus varicosavirus gene junction sequences.

| Virus * | 3’ end mRNA | Intergenic Spacer | 5’ end mRNA |
|---------|-------------|--------------------|-------------|
| AbiV1   | CU(N)_3UUUUU| G                  | CUCU        |
| ArcV8   | AU(N)_3UUUUU| G                  | CUCU        |
| AcoV1   | AU(N)_3UUUUU| G                  | CUCU        |
| AmeV1   | CU(N)_3UUUUU| G                  | CUCU        |
| ApeV3   | AU(N)_3UUUUU| G                  | CUCU        |
| ApoV1   | AU(N)_3UUUUU| G                  | CUCU        |
| ArtV1   | AU(N)_3UUUUU| G                  | CUCU        |
| AscSyV3 | AU(N)_3UUUUU| G                  | CUCU        |
| AspV1   | AU(N)_3UUUUU| G                  | CUCU        |
| BrV2    | AU(N)_3UUUUU| G                  | CUCU        |
| CalV1   | AU(N)_3UUUUU| G                  | CUCU        |
| CenV1   | AU(N)_3UUUUU| G                  | CUCU        |
| ChaV1   | AU(N)_3UUUUU| G                  | CUCU        |
| CucV1   | AU(N)_3UUUUU| G                  | CUCU        |
| CupV1   | CU(N)_3UUUUU| G                  | CUCU        |
| DidV1   | AU(N)_3UUUUU| G                  | CUCU        |
| EryV1   | AU(N)_3UUUUU| G                  | CUCU        |
| FruV1   | AU(N)_3UUUUU| G                  | CUCU        |
| GuiV1   | AU(N)_3UUUUU| G                  | CUCU        |
| HolV1   | AU(N)_3UUUUU| G                  | CUCU        |
| LeuV1   | AU(N)_3UUUUU| G                  | CUCU        |
| LinV1   | AU(N)_3UUUUU| G                  | CUCU        |
| LufV1   | AU(N)_3UUUUU| G                  | CUCU        |
| LupV1   | AU(N)_3UUUUU| G                  | CUCU        |
| MelV1   | AU(N)_3UUUUU| G                  | CUCU        |
| OphV1   | AU(N)_3UUUUU| G                  | CUCU        |
| PenV1   | AU(N)_3UUUUU| G                  | CUCU        |
| PicV1   | CU(N)_3UUUUU| G                  | CUCU        |
| PiBanV1 | CU(N)_3UUUUU| G                  | CUCU        |
| PiYunV1 | CU(N)_3UUUUU| G                  | CUCU        |
| PriV1   | AU(N)_3UUUUU| G                  | CUCU        |
| RanV1   | AU(N)_3UUUUU| G                  | CUCU        |
| RapV1   | AU(N)_3UUUUU| G                  | CUCU        |
| RibV1   | AU(N)_3UUUUU| G                  | CUCU        |
Table 2. Cont.

| Virus *      | 3' end mRNA | Intergenic Spacer | 5' end mRNA |
|--------------|-------------|-------------------|-------------|
| SciV1        | CU(N)UUUUU  | G                 | CUCU        |
| SenV1        | AU(N)UUUUU  | G                 | CUCU        |
| SilV1        | AU(N)UUUUU  | G                 | CUCU        |
| StrV1        | AU(N)UUUUU  | G                 | CUCU        |
| TanV1        | AU(N)UUUUU  | G                 | CUCU        |
| TaxV1        | AU(N)UUUUU  | G                 | CUCU        |
| TorV1        | AU(N)UUUUU  | G                 | CUCU        |
| TreV1        | AU(N)UUUUU  | G                 | CUCU        |
| TriV1        | AU(N)UUUUU  | G                 | CUCU        |
| VinV1        | AU(N)UUUUU  | G                 | CUCU        |
| ZeaV1        | AU(N)UUUUU  | G                 | CUCU        |
| AAnV1        | AU(N)UUUUU  | G                 | CUCU        |
| AMVV1        | AU(N)UUUUU  | G                 | CUCU        |
| BrV1         | AU(N)UUUUU  | G                 | CUCU        |
| LBVaV        | AU(N)UUUUU  | G                 | CUCU        |
| LoV1         | AU(N)UUUUU  | G                 | CUCU        |
| MelRoV1      | AU(N)UUUUU  | G                 | CUCU        |
| MGVV         | AU(N)UUUUU  | G                 | CUCU        |
| MgVV         | AU(N)UUUUU  | G                 | CUCU        |
| PhPiV1       | AU(N)UUUUU  | G                 | CUCU        |
| PiFleV1      | GU(N)UUUUU  | G                 | CUCU        |
| RCaVV        | AU(N)UUUUU  | G                 | CUCU        |
| SpV1         | AU(N)UUUUU  | G                 | CUCU        |
| TiVV         | AU(N)UUUUU  | G                 | CUCU        |
| VVV          | AU(N)UUUUU  | G                 | CUCU        |
| XVV          | AU(N)UUUUU  | G                 | CUCU        |
| ZaVV1        | AU(N)UUUUU  | G                 | CUCU        |

The consensus gene junction sequences of the viruses identified in this study are highlighted in light grey. * Names and abbreviations of newly identified viruses are listed in Table 1; while the names and abbreviations of known viruses are listed in Supplementary Table S1.

There is a great dearth of data on the potential functions of putative proteins, other than N and L, encoded by varicosaviruses, and, intriguingly, there were no conserved domains identified in these proteins. We grasped some shared identities, primarily for the cognate P3 (but also for several P2 proteins) (Table 1), though for most of the encoded proteins, the BlastP results were orphans, with no known signals or domains present and no clues towards their putative (or conserved) function. Thus, further studies should be focused on the functional characterization of these proteins to gain essential knowledge regarding the elusive proteome of varicosaviruses beyond the N and L proteins.

The pairwise aa sequence identities between the L proteins of all the reported varicosaviruses, including those identified in this study, showed great diversity and an overall low identity between the different varicosaviruses (Figure 2, Supplementary Table S2). Relatively low sequence identity is a common feature among rhabdovirus taxa, characterized by a high level of diversity in both the genome sequence and organization [10]. In addition, the overall low sequence identity among the novel viruses detected here and
with the previously described varicosaviruses suggests that despite the many viruses identified in this study, there likely remains a significant amount of virus “dark matter“ for yet-to-be-discovered varicosaviruses.

When we analyzed the diversity between the variants of viruses which are likely members of the same species, we found that proteins encoded by the Brassica virus 2, Spinach virus 1, and Sciadopitys virus 1 variants were very similar. On the other hand, proteins encoded by the Brassica virus 1, Lolium virus 1, and Melilotus virus 1 variants were quite diverse, but, nevertheless, they showed aa identities for the N and L proteins exceeding 80%. Thus, we tentatively propose an aa sequence identity of 80% across the L gene as the threshold for species demarcation in the Varicosavirus genus, a taxonomic criterion which had previously not been fully defined [10]. This threshold is strongly supported by the comparison of the L protein aa sequence of 60 viruses (Figure 2, Supplementary Table S2). Based on this criterion, all 39 novel viruses with their complete coding region assembled in this study should be considered as belonging to novel Varicosavirus species, which would increase the number of members of the genus by more than an order of magnitude.

Bejerman et al. [9] tentatively reported the first unsegmented varicosavirus, Pinus flexilis virus 1 (PiFlxV1), which was associated with the gymnosperm Pinus flexilis. In this study, we complemented that result by the discovery of eight additional unsegmented varicosaviruses which were exclusively associated with gymnosperms (Table 1), some of which
are linked to the same genus Pinus and present a significant co-evolution of viruses and hosts. These results robustly support a clade of gymnosperm-associated varicosaviruses with a distinct genome architecture, requiring the rewriting of a previously proposed key feature and fundamental marker of varicosaviruses: their genomic bisegmented nature. It is tempting to speculate that the unsegmented genomic architecture may be linked to the adaptation to gymnosperm hosts and a shared ancient evolutionary history of these viruses and hosts.

Interestingly, in the BlastP analyses of N, P2, and P3 of the gymnosperm-associated viruses, most of them had, as a best hit to the cognate proteins encoded by the putative bisegmented ASaV2 (Table 1), a virus apparently hosted by a parasitic plant of spruce (Picea, Pinaceae). Furthermore, unexpectedly, the best hit of the putative P5 protein encoded on ASaV2 RNA2 was a fragment of the PiFleV1 L protein, while the deduced L protein on ASaV2 RN1 was not a best hit with PiFleV1, but instead, with the non-gymnosperm-linked MelRoV1 hosted by the Orobanchaceae parasitic plant Melampyrum roseum. Thus, we suspected that ASaV2 was potentially misassembled from fragments belonging to two different viruses. Consequently, we re-analyzed the original SRA data used by Sidhartan et al. [18] and were able to assemble two distinct varicosavirus genomes: one bisegmented genome presumably linked to the parasitic plant and one unsegmented genome most likely linked to spruce, which would support our hypothesis. We believe that there are several reasons that led to the original ASaV2 description: (i) the atypical and unexpected existence at the time of an unsegmented varicosavirus; (ii) the presence of two varicosaviruses in the very same sequencing library, which may be the first tentative evidence in the literature of co-infection of two varicosaviruses; and (iii) the fact that the sequence reads corresponding to the L gene region of the unsegmented varicosavirus were low, which may have affected the assembling pipelines used by the authors. All in all, independently verifying unexpected re-analysed SRA data may lead to a clearer understanding of the genomic structure of the mined RNA virus genomes. Nevertheless, the inability to return to the original biological material to replicate, confirm, and validate the assembled viral genome sequences is a significant limitation of the data mining approach for virus discovery. Thus, researchers must be cautious when analysing SRA public data for virus discovery and understand the preliminary nature of its results.

The phylogenetic analysis based on the deduced L protein aa sequences placed all unsegmented varicosaviruses, except TorV1, into a distinct clade. Interestingly, TorV1 was placed in a clade that was basal to all varicosaviruses (Figure 1). This distinct phylogenetic branching and clustering of the unsegmented viruses suggests that they share a unique evolutionary history among varicosaviruses. Moreover, this may suggest that bisegmented varicosaviruses are evolutionarily younger than unsegmented ones. It may also mean that a genome split in varicosa-like viruses occurred after the radiation of gymnosperms and angiosperms. Bisegmented varicosaviruses did not cluster according to their genomic organization, nor did they cluster with the plant species associated with each virus (Figure 1). For example, brassica virus 1 and brassica virus 2 were placed in distinct clades, while two viruses associated with orchids (Ophius virus 1 and Caladenia virus 1) were placed in different clusters, and monocot-associated viruses were not all grouped together. On the other hand, all varicosaviruses associated with ferns and liverworts belonged to the same cluster, which was also shared with previously reported varicosaviruses from these plant types, while most of the grass-associated varicosaviruses were also clustered together (Figure 1).

We generated a tanglegram to compare the virus phylogram and plant host cladogram to further explore virus–host relationships (Figure 3). This analysis showed that the viruses of some clades clearly co-diverged with their hosts, including the gymnosperm-associated virus clade, the SpV1 and Silene virus 1 clade, the grass-associated virus clade, and the clade of fern and liverworts viruses, suggesting a shared host–virus evolution in those clades (Figure 3). However, the tanglegram topology also indicated that for most of the vari-
cosaviruses, there was no apparent concordant evolutionary history with their plant hosts, similar to what was previously reported for invertebrate and vertebrate rhabdoviruses [60].

Figure 3. Tanglegram showing the phylogenetic relationships of the varicosaviruses (left), which are linked with the associated plant host(s) shown on the right. Links of well-supported clades of viruses to taxonomically related plant species are indicated in blue, orange, and green. A maximum likelihood phylogenetic tree of rhabdoviruses was constructed based on the conserved amino acid sequence of the complete L protein. Plant host cladograms were generated in phylOT v.2 based on NCBI taxonomy. Internal nodes represent the taxonomic structure of the NCBI taxonomy database, including species, genus, family, order, subclass, and sub-kingdom. Viruses identified in the present study are shown in bold font. The scale bar indicates the number of substitutions per site.

Several lines of evidence suggest that varicosaviruses may be vertically transmitted: (i) a close host–virus co-evolution in some clades may reflect species isolation and a lack of horizontal transmission, (ii) some viruses detected in this study were identified from seed transcriptomics databases, and (iii) an emerging characteristic of persistent, chronic infections of several plant viruses which are likely vertically transmitted are latent/asymptomatic infections, a characteristic which appears to be shared with varicosaviruses. Thus, further studies should be carried out to elucidate the transmission mode...
of varicosaviruses beyond the fungal-transmitted LBVaV [11]. It is worth mentioning that even with the availability of thousands of RNAseq libraries of fungi and arthropods, we failed to detect any evidence of varicosaviruses in those organisms, which could suggest that vectors of varicosaviruses are rare or non-existent.

Before the era of data-driven virus discovery, few viruses had been identified in gymnosperms [61–64]. However, when data mining was applied to publicly available transcriptomes, many novel viruses were identified in this large group of higher plants, highlighting the rich and diverse gymnosperm virosphere, which still is largely unexplored. A distinct clade of gymnosperm-associated viruses was recently identified within amalgaviruses [65], while we recently described two distinct caulimovirids and geminivirids linked to the gnetophyte Welwitschia mirabilis [66]. Eight unsegmented varicosaviruses associated with gymnosperms were identified in this study, and another was discovered by Bejerman et al. [9]. Taken together, all of these recently discovered viruses in gymnosperms strongly suggest that they may have evolutionary trajectories that are distinct from those infecting angiosperms. Thus, it is likely that further exploration of additional gymnosperm datasets or new transcriptome studies of other gymnosperms will yield plenty of novel viruses with unique features, highlighting their close evolution with their hosts. The clear association between gymnosperm-associated viruses and their hosts likely indicates a close coevolution, which suggest an early adaptation of this group of viruses to infect gymnosperms. This hypothesis is also supported by the distinct genomic architecture and divergent evolutionary history among varicosaviruses, as shown in the phylogenetic tree, which are characterized by long branches and distinctive clustering. Taken together, the gymnosperm-associated varicosaviruses could be taxonomically classified in a novel genus within the family Rhabdoviridae, subfamily Betarhabdovirinae, for which we suggest the name “Gymnorhavirus”.

In summary, this study highlights the importance of the analysis of SRA public data as a valuable tool not only to accelerate the discovery of novel viruses, but also to gain insight into their evolution and to refine virus taxonomy. Using this approach, we looked for hidden varicosavirus-like virus sequences to unlock the veiled diversity of a largely neglected plant rhabdovirus genus, the varicosaviruses. Our findings, including an approximately 3.5-fold expansion of the current genomic diversity within the genus, resulted in the most complete phylogeny of varicosaviruses to date, and they shed new light on the genomic architecture, phylogenetic relationships, and evolutionary landscape of this unique group of plant rhabdoviruses. Future studies should assess many intriguing aspects of the biology and ecology of these viruses such as potential symptoms, vertical transmission, and putative vectors.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pathogens11101127/s1, Figure S1: Stacked bar chart showing the number of previously reported varicosaviruses and those in this study; Table S1: Virus names, abbreviations, and NCBI accession numbers of the varicosavirus sequences used in this study; Table S2: Amino acid sequence identity of the complete L gene ORF.

**Author Contributions:** Conceptualization, N.B., R.G.D. and H.D.; data analysis, N.B. and H.D.; writing—original draft preparation, N.B.; writing—review and editing, N.B., R.G.D. and H.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. The participation of R.G.D. in this study was jointly supported by the Queensland Government Department of Agriculture and Fisheries and the University of Queensland through the Queensland Alliance for Agriculture and Food Innovation.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The nucleotide sequence data reported are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK061731-BK061826. These sequences are available as Supplementary Materials.
Acknowledgments: The authors would like to express their sincere gratitude to the generators of the underlying data used for this work, which are cited in Table 1. By following open-access practices and supporting accessible raw sequence data in public repositories available to the research community, they have promoted the generation of new knowledge and ideas.

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

References
1. Koonin, E.V.; Krupovic, M.; Agol, V.I. The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? Microbiol. Mol. Biol. Rev. 2021, 85, e000521. [CrossRef]
2. Koonin, E.V.; Dolja, V.V.; Krupovic, M.; Varsani, A.; Wolf, Y.I.; Yutin, N.; Zerbini, F.M.; Kuhn, J.H. Global Organization and Proposed Megataxonomy of the Virus World. Microbiol. Mol. Biol. Rev. 2020, 84, e00061-19. [CrossRef] [PubMed]
3. Geoghegan, J.L.; Holmes, E.C. Predicting virus emergence amid evolutionary noise. Open Biol. 2017, 7, 170–189. [CrossRef] [PubMed]
4. Dolja, V.V.; Krupovic, M.; Koonin, E.V. Deep Roots and Splendid Boughs of the Global Plant Virome. Annu. Rev. Phytopathol. 2020, 58, 23–53. [CrossRef] [PubMed]
5. Edgar, R.C.; Taylor, J.; Lin, V.; Altman, T.; Barbera, P.; Meleshko, D.; Novakovsky, G.; Buchfink, B.; Al-Shayeb, B.; et al. Petabase-scale sequence alignment catalyses viral discovery. Nature 2022, 602, 142–147. [CrossRef]
6. Mifsud, J.C.O.; Gallagher, R.V.; Holmes, E.C.; Geoghegan, J.L. Transcriptome Mining Expands Knowledge of RNA Viruses across the Plant Kingdom. J. Virol. 2022, e00260-22. [CrossRef]
7. Lauber, C.; Seitz, S. Opportunities and Challenges of Data-Driven Virus Discovery. Biomolecules 2022, 12, 1073. [CrossRef]
8. Dietzgen, R.G.; Bejerman, N.E.; Goodin, M.M.; Higgins, C.M.; Huot, O.B.; Kondo, H.; Martin, K.M.; Whitfield, A.E. Diversity and epidemiology of plant rhabdoviruses. Virus Res. 2020, 281, 197942. [CrossRef]
9. Bejerman, N.; Dietzgen, R.; Debat, H. Illuminating the Plant Rhabdovirus Landscape through Metatranscriptomics Data. Viruses 2021, 13, 1304. [CrossRef]
10. Walker, P.J.; Freitas-Astúia, J.; Bejerman, N.; Blasdell, K.R.; Breyta, R.; Dietzgen, R.G.; Fooks, A.R.; Kondo, H.; Kurath, G.; Kuzmin, I.V.; et al. ICTV Virus Taxonomy Profile: Rhabdoviridae 2022. [CrossRef]
11. Campbell, R.N. Fungal Transmission of Plant Viruses. Annu. Rev. Phytopathol. 1996, 34, 87–108. [CrossRef] [PubMed]
12. Sasaya, T.; Ishikawa, K.; Koganezawa, H. The nucleotide sequence of RNA1 of Lettuce big-vein virus, genus Varicosavirus, reveals its relation to nonsegmented negative-strand RNA viruses. Virology 2002, 297, 289–297. [CrossRef] [PubMed]
13. Sasaya, T.; Kusaba, S.; Ishikawa, K.; Koganezawa, H. Nucleotide sequence of RNA2 of Lettuce big-vein virus and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses. J. Gen. Virol. 2004, 85, 2709–2717. [CrossRef]
14. Verbeek, M.; Dullemans, A.M.; van Bekkum, P.J.; van der Vlugt, R.A.A. Evidence for Lettuce big-vein associated virus as the causal agent of a syndrome of necrotic rings and spots in lettuce. Plant Pathol. 2013, 62, 444–451. [CrossRef]
15. Koloniu, I.; Fránová, J.; Sarkisova, T.; Přibylova, J.; Lenz, O.; Petrikz, K.; Špak, J. Identification and molecular characterization of a novel varicosova-like virus from red clover. Arch. Virol. 2018, 163, 2213–2218. [CrossRef] [PubMed]
16. Sabbadin, F.; Glover, R.; Stafford, R.; Rozado-Aguirre, Z.; Boomham, N.; Adams, I.; Mumford, R.; Edwards, R. Transcriptome sequencing identifies novel persistent viruses in herbicide resistant wild-grasses. Sci. Rep. 2017, 7, 18004. [CrossRef]
17. Shin, C.; Choi, D.; Hahn, Y. Identification of the genome sequence of Zostera associated varicosavirus 1, a novel negative-sense RNA virus, in the common eelgrass (Zostera marina) transcriptome. Acta Virol. 2022, 65, 373–380. [CrossRef]
18. Siddharthan, V.K.; Chaturvedi, K.K.; Baranwal, V.K. Diverse RNA viruses in a parasitic owering plant (spruce dwarf mistletoe) revealed through RNA-seq data mining. J. Gen. Plant Pathol. 2022, 88, 138–144. [CrossRef]
19. Chen, Y.-M.; Sadig, S.; Tian, J.-H.; Chen, X.; Lin, X.-D.; Shen, J.-J.; Chen, H.; Hao, Z.-Y.; Wille, M.; Zhou, Z.-C.; et al. RNA viromes from terrestrial sites across China expand environmental viral diversity. Nat. Microbiol. 2022, 7, 1312–1323. [CrossRef]
20. Nabeshima, T.; Abe, J. High-throughput sequencing indicates novel Varicosavirus, Emaravirus and Deltapartitivirus infections in Vitis vinifera. Viruses 2021, 13, 827. [CrossRef]
21. Zhao, F.; Liu, H.; Qiao, Q.; Wang, Y.; Zhang, D.; Wang, S.; Tian, Y.; Zhang, Z. Complete genome sequence of a novel varicosavirus infecting tall morning glory (Ipomoea purpurea). Arch. Virol. 2021, 166, 3225–3228. [CrossRef]
22. Leebens-Mack, J.H.; Barker, M.S.; Carpenter, E.J. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 2019, 574, 679–685. [CrossRef]
23. Wang, Y.; Li, X.; Zhou, W.; Li, T.; Tian, C. De novo assembly and transcriptome characterization of spruce dwarf mistletoe Arceuthobium sitchensanum uncovers gene expression profiling associated with plant development. BMC Genom. 2016, 17, 771. [CrossRef] [PubMed]
24. Tang, M.; Zhao, W.; Xing, M.; Zhao, J.; Jiang, Z.; You, J.; Ni, B.; Ni, Y.; Liu, C.; Li, J. Resource allocation strategies among vegetative growth, sexual reproduction, asexual reproduction and defense during growing season of Aconitum kusnezoffii Reichb. Plant J. 2021, 105, 957–977. [CrossRef]
25. Yu, C.; Zhan, X.; Zhang, C.; Xu, X.; Huang, J.; Feng, S.; Shen, C.; Wang, H. Comparative metabolomic analyses revealed the differential accumulation of taxoids, flavonoids and hormones among six Taxaceae trees. Sci. Hort. 2021, 285, 110196. [CrossRef]
26. Babineau, M.; Mahmood, K.; Mathiassen, S.K.; Kudsk, P.; Kristensen, M. De novo transcriptome assembly analysis of weed Apera spica-venti from seven tissues and growth stages. BMC Genom. 2017, 18, 128. [CrossRef] [PubMed]

27. Rowarth, N.M.; Curtis, B.A.; Einfeldt, A.L.; Archibald, J.M.; Lacroix, C.R.; Gunawardena, A.H. RNA-Seq analysis reveals potential regulators of programmed cell death and leaf remodelling in lace plant (Aponogeton madagascariensis). BMC Plant Biol. 2021, 21, 375. [CrossRef] [PubMed]

28. Jayasena, A.S.; Fisher, M.F.; Panero, J.L.; Secco, D.; Bernath-Levin, K.; Berkowitz, O.; Taylor, N.L.; Schilling, E.E.; Whelan, J.; Mylne, J.S. Stepwise Evolution of a Buried Inhibitor Peptide over 45 My. Mol. Biol. Evol. 2017, 34, 1505–1516. [CrossRef]

29. Weitemier, K.; Straub, S.C.; Fishbein, M.; Bailey, C.D.; Cronn, R.C.; Liston, A. A draft genome and transcriptome of common milkweed (Asclepias syriaca) as resources for evolutionary, ecological, and molecular studies in milkweeds and Apocynaceae. PeerJ 2019, 7, e7649. [CrossRef]

30. Shen, H.; Jin, D.; Shu, J.-P.; Zhou, X.-L.; Lei, M.; Wei, R.; Shang, H.; Wei, H.-J.; Zhang, R.; Liu, L.; et al. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. GigaScience 2017, 7, gix116. [CrossRef]

31. An, H.; Qi, X.; Gaynor, M.L.; Hao, Y.; Gebken, S.C.; Mabry, M.E.; McAlvay, A.C.; Teakle, G.R.; Conant, G.C.; Barker, M.S.; et al. Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid Brassica napus. Nat. Commun. 2019, 10, 2878. [CrossRef]

32. Bisht, D.S.; Chamola, R.; Nath, M.; Bhat, S.R. Molecular mapping of fertility restorer gene of an alloplasmic CMS system in Brassica juncea containing Moricandia arvensis cytoplasm. Mol. Breed. 2015, 35, 14. [CrossRef]

33. Wu, Q.; Wang, J.; Mao, S.; Xu, H.; Wu, Q.; Liang, M.; Yuan, Y.; Liu, M.; Huang, K. Comparative transcriptome analyses of genes involved in sulforaphane metabolism at different treatment in Chinese kale using full-length transcriptome sequencing. BMC Genom. 2019, 20, 377. [CrossRef]

34. Xu, H.; Bohman, B.; Wong, D.C.J.; Rodriguez-Delgado, C.; Scaffidi, A.; Flematti, G.R.; Phillips, R.D.; Pichersky, E.; Peakall, R. Complex Sexual Deception in an Orchid Is Achieved by Co-opting Two Independent Biosynthetic Pathways for Pollinator Attraction. Curr. Biol. 2017, 27, 1867–1877.e5. [CrossRef] [PubMed]

35. Tai, Y.; Hou, X.; Liu, C.; Sun, J.; Guo, C.; Su, L.; Jiang, W.; Ling, C.; Wang, C.; Wang, H.; et al. Phytochemical and comparative transcriptome analyses reveal different regulatory mechanisms in the terpenoid biosynthesis pathways between Matricaria recutita L. and Chamaemelum nobile L. BMC Genom. 2020, 21, 169. [CrossRef]

36. Lü, P.; Yu, S.; Zhu, N.; Chen, Y.-R.; Zhou, B.; Pan, Y.; Tzeng, D.; Fabi, J.P.; Argyris, J.; Garcia-Mas, J.; et al. Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. Nat. Plants 2018, 4, 784–791. [CrossRef]

37. Li, J.; Milne, R.I.; Ru, D.; Miao, J.; Tao, W.; Zhang, L.; Xu, J.; Liu, J.; Mao, K. Allopatric divergence and hybridization within Cupressus chengiana (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China. Mol. Ecol. 2020, 29, 1250–1266. [CrossRef]

38. Huang, C.; Qi, X.; Chen, D.; Qi, J.; Ma, H. Recurrent genome duplication events likely contributed to both the ancient and recent rise of ferns. J. Integr. Plant Biol. 2019, 62, 433–455. [CrossRef]

39. Osuna-Mascaró, C.; de Casas, R.R.; Gómez, J.M.; Loureiro, J.; Castro, S.; Landis, J.B.; Hopkins, R.; Perfetti, F. Hybridization and introgression are prevalent in Southern European Erysimum (Brassicaceae) species. Ann. Bot. 2022. [CrossRef] [PubMed]

40. Young, E.; Carey, M.; Meharg, A.A.; Meharg, C. Microbiome and ecotypic adaption of Melilotus sativa (L.) to extremes of its soil pH range, investigated through transcriptome sequencing. Microbiome 2018, 6, 48. [CrossRef]

41. Nevedo, B.; Atchison, G.W.; Hughes, C.E.; Filatov, D.A. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. Nat. Commun. 2018, 9, 12384. [CrossRef]

42. Wu, F.; Duan, Z.; Xu, P.; Yan, Q.; Meng, M.; Cao, M.; Jones, C.S.; Zong, X.; Zhou, P.; Wang, Y.; et al. Genome and systems biology of Melilotus albus provides insights into coriandrum biosynthesis. Plant Biotechnol. J. 2021, 20, 592–609. [CrossRef] [PubMed]

43. Huang, R.; Snedden, W.; DiCenzo, G. Reference nodule transcriptomes for Melilotus officinalis and Medicago sativa cv. Algonquin. Grassl. Res. 2022, 6, e408. [CrossRef]

44. Piñeiro Fernández, L.; Byers, K.J.R.P.; Cai, J.; Sedeek, K.E.M.; Kellenberger, R.T.; Russo, A.; Qi, W.; Aquino Fournier, C.; Schlüter, P.M. A Phylogenomic Analysis of the Floral Transcriptomes of Sexually Deceptive and Rewarding European Orchids, Ophrys and Gymnadenia. Front. Plant Sci. 2019, 10, 1553. [CrossRef]

45. Peery, R.M.; McAllister, C.H.; Cullingham, C.I.; Mahon, E.L.; Arango-Velez, A.; Cooke, J.E. Comparative genomics of the chitinase gene family in lodgepole and jack pines: Contrasting responses to biotic threats and landscape level investigation of genetic differentiation. Botany 2021, 99, 355–385. [CrossRef]

46. Cai, N.; Xu, Y.; Chen, S.; He, B.; Li, G.; Li, Y.; Duan, A. Variation in seed and seedling traits and their relations to geo-climatic factors among populations in Yunnan Pine (Pinus yunnanensis). J. For. Res. 2016, 27, 1009–1017. [CrossRef]

47. Zhao, Z.; Luo, Z.; Yuan, S.; Mei, L.; Zang, D. Global transcriptome and gene co-expression network analyses on the development of distyly in Primula oboidea. Heredity 2019, 123, 784–794. [CrossRef] [PubMed]

48. Pellino, M.; Højsgaard, D.; Schmutzer, T.; Scholz, U.; Hörandl, E.; Vogel, H.; Sharbel, T.F. Asexual genome evolution in the apomictic Ranunculus auricomus complex: Examining the effects of hybridization and mutation accumulation. Mol. Ecol. 2013, 22, 5908–5921.

49. Yang, Z.; Li, W.; Su, X.; Ge, P.; Zhou, Y.; Hao, Y.; Shu, H.; Gao, C.; Cheng, S.; Zhu, G.; et al. Early Response of Radish to Heat Stress by Strand-Specific Transcriptome and miRNA Analysis. Int. J. Mol. Sci. 2019, 20, 3321. [CrossRef]
50. Zhou, B.; Wang, J.; Lou, H.; Wang, H.; Xu, Q. Comparative transcriptome analysis of dioecious, unisexual floral development in *Ribes diacanthum* pall. *Gene* **2019**, *699*, 43–53. [CrossRef] [PubMed]

51. Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Carpenter, E.; Matasci, N.; Ayyampalayam, S.; Barker, M.S.; Burleigh, J.G.; Gitzendanner, M.A.; et al. Phylogenomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4859–E4868. [CrossRef] [PubMed]

52. Meier, S.K.; Adams, N.; Wolf, M.; Balkwill, K.; Muasya, A.M.; Gehring, C.A.; Bishop, J.M.; Ingle, R.A. Comparative RNA-seq analysis of nickel hyperaccumulating and non-accumulating populations of *Senecio coronatus* (Asteraceae). *Plant J.* **2018**, *95*, 1023–1038. [CrossRef]

53. Baloun, J.; Nevtalova, E.; Kovacova, V.; Hudzieczek, V.; Čegan, R.; Vyskot, B.; Hobza, R. Characterization of the HMA7 gene and transcriptomic analysis of candidate genes for copper tolerance in two *Silene vulgaris* ecotypes. *J. Plant Physiol.* **2014**, *171*, 1188–1196. [CrossRef]

54. Clancy, M.V.; Haberer, G.; Jud, W.; Niederbacher, B.; Niederbacher, S.; Senft, M.; Zytynska, S.E.; Weisser, W.W.; Schnitzler, J.-P. Under fire-simultaneous volatilome and transcriptome analysis unravels fine-scale responses of tansy chemotypes to dual herbivore attack. *BMC Plant Biol.* **2020**, *20*, 551. [CrossRef] [PubMed]

55. Zhou, T.; Luo, X.; Yu, C.; Zhang, C.; Zhang, L.; Song, Y.B.; Dong, M.; Shen, C. Transcriptome analyses provide insights into the expression pattern and sequence similarity of several taxol biosynthesis-related genes in three *Taxus* species. *BMC Plant Biol.* **2019**, *19*, 33. [CrossRef] [PubMed]

56. Hodge, B.A.; Paul, P.A.; Stewart, L.R. Occurrence and High-Throughput Sequencing of Viruses in Ohio Wheat. *Plant Dis.* **2020**, *104*, 1789–1800. [CrossRef]

57. Yu, X.; Wang, W.; Yang, H.; Zhang, X.; Wang, D.; Tian, X. Transcriptome and comparative chloroplast genome analysis of *Vincetoxicum versicolor*: Insights into molecular evolution and phylogenetic implication. *Front. Genet.* **2021**, *12*, 602528. [CrossRef]

58. Lanver, D.; Müller, A.N.; Happel, P.; Schweizer, G.; Haas, F.B.; Franitza, M.; Pellegrin, C.; Reissmann, S.; Altmüller, J.; Rensing, S.A.; et al. The Biotrophic Development of *Ustilago maydis* Studied by RNA-Seq Analysis. *Plant Cell* **2018**, *30*, 300–323. [CrossRef]

59. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [CrossRef]

60. Geoghegan, J.L.; Duchêne, S.; Holmes, E.C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLOS Pathog.* **2017**, *13*, e1006215. [CrossRef]

61. Alvarez-Quinto, R.A.; Lockhart, B.E.L.; Fetzer, J.L.; Olszewski, E.N. Genomic characterization of cycad leaf necrosis virus, the first badnavirus identified in a gymnosperm. *Arch. Virol.* **2020**, *165*, 1671–1673. [CrossRef] [PubMed]

62. Koh, S.H.; Li, H.; Admiraal, R.; Jones, M.G.; Wylie, S. Catharanthus mosaic virus: A potyvirus from a gymnosperm, Welwitschia mirabilis. *Virus Res.* **2015**, *203*, 41–46. [CrossRef] [PubMed]

63. Han, S.S.; Karasev, A.V.; Ieki, H.; Iwanami, T. Nucleotide sequence and taxonomy of *Cycas necrotic stunt virus*. *Arch. Virol.* **2002**, *147*, 2207–2214. [CrossRef] [PubMed]

64. Rastrojo, A.; Núñez, A.; Moreno, D.A.; Alcamí, A. A New Putative Caulimoviridae Genus Discovered through Air Metagenomics. *Microbiol. Resour. Announc.* **2018**, *7*, e00955-18. [CrossRef]

65. Sidharthan, V.K.; Rajeswari, V.; Vanamala, G.; Banarwal, V.K. Revisiting the amalgaviral landscapes in plant transcriptomes expands the host range of plant amalgaviruses. *Available SSRN 4210265* **2022**. [CrossRef]

66. Debat, H.; Bejerman, N. A glimpse into the DNA virome of the unique “living fossil” Welwitschia mirabilis. *Gene* **2022**, *843*, 146806. [CrossRef]