

Identification of dicistro-like viruses in the transcriptome data of *Striga asiatica* and other plants

Chaerim Shin¹, Dongjin Choi¹, Ken Shirasu^{2,3}, Yoonsoo Hahn^{1*}

¹Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; ²RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; ³Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

Received January 7, 2022; accepted April 11, 2022

Summary. – Dicistroviruses (the family *Dicistroviridae*) are positive-sense single-stranded RNA viruses of the order *Picornavirales*, which is a rapidly growing viral group. They have been detected in a wide range of animals, predominantly in insects and crustaceans. In this study, we identified the genome sequences of 14 dicistro-like viruses in the transcriptome data from 12 plant species, including *Striga asiatica* dicistro-like virus 1 and 2 identified in the transcriptome data of *Striga asiatica*. Sequence comparison and phylogenetic analysis indicated that these 14 plant-associated dicistro-like viruses were novel members of the family *Dicistroviridae*, five of which are placed within the genera *Aparavirus* and *Cripavirus*, which mainly consist of viruses infecting animals, including insects. The other nine plant dicistro-like viruses formed clades with unclassified dicistroviruses. Our study implies that a wide range of plant species may serve as hosts for dicistroviruses or reservoirs for their transmission.

Keywords: dicistrovirus; *Dicistroviridae*; plant; transcriptome; *Striga asiatica*

Introduction

The family *Dicistroviridae* belongs to the order *Picornavirales*, a rapidly growing group of RNA viruses (Carrillo-

Tripp *et al.*, 2015; Yinda *et al.*, 2017). There are three genera in the family *Dicistroviridae* that are approved by the International Committee on Taxonomy of Viruses, namely *Aparavirus*, *Cripavirus*, and *Triatovirus* (Valles *et al.*, 2017). Dicistroviruses (the family *Dicistroviridae*) mainly infect arthropods, including insects and crustaceans, and cause viral diseases in infected hosts (Reinganum *et al.*, 1970; Mari *et al.*, 2002; de Miranda *et al.*, 2010; Warsaba *et al.*, 2020). Notable dicistroviruses are acute bee paralysis virus, cricket paralysis virus (CrPV), and triatoma virus, which are the type species of the genera *Aparavirus*, *Cripavirus*, and *Triatovirus*, respectively (Czibener *et al.*, 2000; Govan *et al.*, 2000; Wilson *et al.*, 2000).

Dicistroviruses have a monopartite, single-stranded, positive-sense RNA genome of 8–10 kilobases (kb) in length with two open reading frames (ORF1 and ORF2) (Bonning and Miller, 2010; Warsaba *et al.*, 2020). ORF1 encodes a non-structural polyprotein that is processed by proteases into individual mature viral proteins, including a helicase, protease, RNA-dependent RNA polymerase (RdRp), and several viral genome-linked proteins that are involved in viral RNA replication. ORF2 encodes a struc-

*Corresponding author. E-mail: hahnny@cau.ac.kr; phone: +82-2-820-5812.

Abbreviations: AsDLV1 = *Alloteropsis semialata* dicistro-like virus 1; AhDLV1 = *Arabidopsis halleri* dicistro-like virus 1; CaDLV1 = *Camptotheca acuminata* dicistro-like virus 1; CbDLV1 = *Cosmos bipinnatus* dicistro-like virus 1; CqDLV1 = *Colobanthus quitensis* dicistro-like virus 1; CqDLV2 = *Colobanthus quitensis* dicistro-like virus 2; CrPV = cricket paralysis virus; DcDLV1 = *Dalbergia cochinchinensis* dicistro-like virus 1; HvDLV1 = *Hordeum vulgare* dicistro-like virus 1; IGR IRES = intergenic region internal ribosomal entry site; KBV = Kashmir bee virus; NCBI = National Center for Biotechnology Information; NoDLV1 = *Nasturtium officinale* dicistro-like virus 1; ORF = open reading frame; SaDLV1 = *Striga asiatica* dicistro-like virus 1; SaDLV2 = *Striga asiatica* dicistro-like virus 2; SASV3 = *Sanxia atyid* shrimp virus 3; ShDLV1 = *Saccharum hybrid* dicistro-like virus 1; SIDLV1 = *Silene latifolia* dicistro-like virus 1; SRA = Sequence Read Archive; TbDLV1 = *Trapa bispinosa* dicistro-like virus 1

tural protein precursor that is cleaved into four viral capsid proteins that are required for the encapsidation of viral RNA molecules.

Two distinct internal ribosomal entry site (IRES) elements, one in the 5'-untranslated region and the other in the intergenic region (IGR), direct cap-independent translation of ORF1 and ORF2, respectively (Bonning and Miller, 2010; Warsaba *et al.*, 2020). The domain organization, structure, and function of IGR IRESs have been extensively studied in CrPV, Kashmir bee virus (KBV), and related viruses (Hertz and Thompson, 2011a; Hertz and Thompson, 2011b; Khong *et al.*, 2016). IGR IRES has a domain that structurally resembles a tRNA molecule bound to the peptidyl site of the ribosome, which mediates the translation initiation of ORF2 at a non-AUG codon (Bonning and Miller, 2010). For example, the first codon of both CrPV and KBV ORF2 is GCU, which encodes an alanine residue (Hertz and Thompson, 2011a).

Dicistroviruses were first identified in arthropods, including insects and crustaceans (Reinganum *et al.*, 1970; Czibener *et al.*, 2000; Govan *et al.*, 2000; Guo *et al.*, 2013; Valles *et al.*, 2017). However, recent studies have demonstrated that dicistroviruses or related viruses are associated with non-arthropod animals, including diverse invertebrates, birds, mammals, and humans (Greninger and Jerome, 2016; Shi *et al.*, 2016; Phan *et al.*, 2018; Dastjerdi *et al.*, 2021). Moreover, dicistrovirus genome sequences have been identified in samples derived from plants and single-celled eukaryotes, suggesting that they might be associated with a wide range of organisms (Chiapello *et al.*, 2020; Nery *et al.*, 2020).

Plant RNA-Seq data often contain sequences originating from RNA viruses that are present in plant tissue samples (Bejerman *et al.*, 2020; Park and Hahn, 2021). Previously, we identified many novel plant RNA virus genome sequences via comprehensive analysis of publicly available plant transcriptome data (Choi *et al.*, 2021; Goh *et al.*, 2021; Park *et al.*, 2021; Park and Hahn, 2021; Shin *et al.*, 2021). In this study, we identified 14 genome sequences of novel dicistro-like viruses in transcriptome data obtained from diverse plants, including *Striga asiatica* (Yoshida *et al.*, 2019).

Materials and Methods

Plant RNA-Seq data analyzed in this study were downloaded from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) (as of July 2021). The plants and SRA project Acc. Nos. are as follows: *Alloteropsis semialata*, SRP072730; *Arabidopsis halleri* subsp. *halleri*, SRP111320; *Camptotheca acuminata*, SRP006330; *Colobanthus quitensis*, SRP051241; *Cosmos bipinnatus*, SRP074877; *Dalbergia*

cochinchinensis, SRP234829; *Hordeum vulgare* subsp. *vulgare*, SRP101929; *Nasturtium officinale*, SRP058520; *S. asiatica*, DRP005455; *Saccharum* hybrid cultivar SP80-3280, SRP041206; *Silene latifolia*, SRP007582; and *Trapa bispinosa*, SRP082135. The detailed SRA Acc. Nos. are presented in Supplementary Data S1. High-quality reads from each plant were obtained using the sickle program (version 1.33; <https://github.com/najoshi/sickle>) with the parameter “-q 30 -l 55” and then assembled into contigs using the rnaviralSPAdes pipeline of the SPAdes assembler (version 3.15.2; <http://cab.spbu.ru/software/spades>) (Bushmanova *et al.*, 2019).

S. asiatica transcriptome contigs were compared with viral RdRp domain sequences using the DIAMOND program (version 2.0.4; <http://www.diamondsearch.org/index.php>) to identify putative viral genome contigs. The known viral RdRp domain sequences (2620) were collected from the Pfam database (release 34.0; <https://pfam.xfam.org>). Pfam database Acc. Nos. are PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF06478, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501.

Putative protein-coding sequences in the viral genomes were predicted using the getorf program of the EMBOSS package (version 6.6.0.0; <http://emboss.open-bio.org>). Functional domains of viral protein sequences were predicted using the InterPro web server (version 86.0; <https://www.ebi.ac.uk/interpro>). Pairwise sequence comparisons were performed using the standalone BLAST software (version 2.9.0+).

Multiple sequence alignments were conducted using the MAFFT program (version 7.475; <https://mafft.cbrc.jp/alignment/software>) with the parameter “-auto” (Nakamura *et al.*, 2018). Phylogenetically informative segments in multiple sequence alignments were collected using the trimAl program (version 1.4. rev22; <http://trimal.cgenomics.org>) with the parameter “-automated1” (Capella-Gutierrez *et al.*, 2009). A maximum-likelihood phylogenetic tree was constructed using the IQ-TREE program (version 2.1.3; <http://www.iqtree.org>) with the substitution model “LG+R8” (Minh *et al.*, 2020). Bootstrap analysis was performed from 1000 replicates using the UFBoot2 method implemented in the IQ-TREE program (the parameter “-B 1000”).

Results and Discussion

S. asiatica, commonly known as red witchweed, is the most widespread species of the genus *Striga*, which consists of obligate hemiparasitic root parasites of various grass crop plants, including rice (Spallek *et al.*, 2013). In this study, the *S. asiatica* RNA-Seq transcriptome data originally generated for the investigation of gene expression during haustorium formation were re-analyzed to identify potentially latently infected RNA viruses (Yoshida *et al.*, 2019; Choi *et al.*, 2021). When the assembled *S. asiatica*

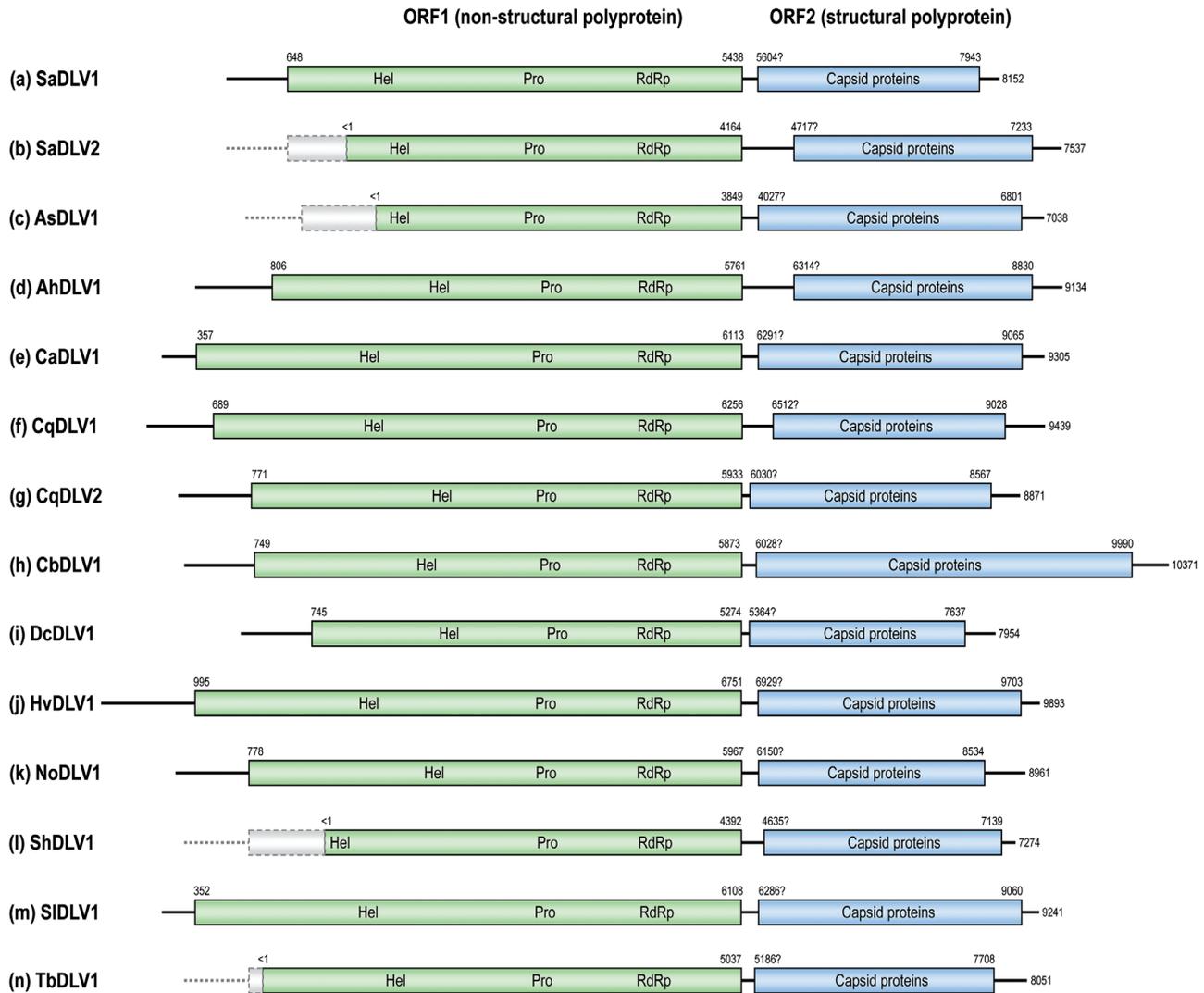


Fig. 1

Genome organizations of plant dicistro-like viruses

Genome organizations of 14 dicistro-like viruses identified in the plant transcriptome data are depicted here. Two open reading frames (ORF1 and ORF2) are marked by boxes with the start and end positions at the top. The start positions of ORF2 were not accurately determined. The helicase, protease, and RdRp domains in ORF1, predicted by InterPro, are indicated as Hel, Pro, and RdRp, respectively. Incomplete 5' regions are represented by dotted lines.

transcript contigs were compared with known viral RdRp domain sequences, several distinct contigs were identified to be derived from the putative RNA viral genomes. Two of them, which were 8152 nucleotides (nt) and 7537 nt, respectively, contained an RdRp domain that showed a strong sequence similarity to those of dicistroviruses. Sequence similarity searches of the NCBI protein database revealed that each of the two contigs encoded two ORFs, similar to ORF1 and ORF2 of known dicistroviruses. Therefore, the two putative RNA viruses, 8152 and SaDLV2

were tentatively named *Striga asiatica* dicistro-like virus 1 (SaDLV1) and 2 (SaDLV2), respectively. Their genome sequences were deposited in the NCBI nucleotide database under the Acc. Nos. MZ598483 and MZ598484.

SaDLV1 ORF1 encodes a full-length non-structural polyprotein with 1596 amino acids (aa), which was predicted to contain a helicase, protease, and RdRp domain (Fig. 1a and Table 1). The SaDLV2 genome was found to be incomplete at its 5' end and contained a partial ORF1 with at least 1387 aa (Fig. 1b). The SaDLV1 and SaDLV2

Table 1. Novel dicistroviruses identified in the plant transcriptome data

Virus	Acronym	NCBI Acc. No.	Genome (nt)	ORF1 (aa)	ORF2 (aa) ^b	ORF1 best hit ^c
Striga asiatica dicistro-like virus 1	SaDLV1	MZ598483	8152	1596	<779	Perth bee virus 4 (MG995733, AWK77898), 964/1450 (66%)
Striga asiatica dicistro-like virus 2	SaDLV2	MZ598484	>7537 ^a	>1387 ^a	<838	Picornavirales sp. (MN035381, QDH90241), 788/1389 (57%)
Alloteropsis semialata dicistro-like virus 1	AsDLV1	BK059245	>7038 ^a	>1282 ^a	<924	Weivirus-like virus sp. (MT138416, QJI53766), 832/1288 (65%)
Arabidopsis halleri dicistro-like virus 1	AhDLV1	BK059246	9134	1651	<838	Picornavirales sp. (MN035381, QDH90241), 784/1390 (56%)
Camptotheca acuminata dicistro-like virus 1	CaDLV1	BK059247	9305	1918	<924	Weivirus-like virus sp. (MT138416, QJI53766), 1203/1921 (63%)
Colobanthus quitensis dicistro-like virus 1	CqDLV1	BK059248	9439	1855	<838	Dicistroviridae sp. (MN905971, QJI52026), 291/807 (36%)
Colobanthus quitensis dicistro-like virus 2	CqDLV2	BK059249	8871	1720	<845	Picornavirales sp. (MN035381, QDH90241), 594/1412 (42%)
Cosmos bipinnatus dicistro-like virus 1	CbDLV1	BK059250	10371	1707	<1320	Picornavirales sp. (MN035381, QDH90241), 702/1452 (48%)
Dalbergia cochinchinensis dicistro-like virus 1	DcDLV1	BK059251	7954	1509	<757	Bivalve RNA virus G2 (NC_032113, YP_009329819), 327/1179 (28%)
Hordeum vulgare dicistro-like virus 1	HvDLV1	BK059252	9893	1918	<924	Weivirus-like virus sp. (MT138416, QJI53766), 1202/1921 (63%)
Nasturtium officinale dicistro-like virus 1	NoDLV1	BK059253	8961	1729	<794	Hubei picorna-like virus 20 (NC_033000, YP_009337055), 275/1279 (22%)
Saccharum hybrid dicistro-like virus 1	ShDLV1	BK059254	>7274 ^a	>1463 ^a	<834	Dicistroviridae sp. (MN905970, QJI52024), 514/810 (63%)
Silene latifolia dicistro-like virus 1	SlDLV1	BK059255	9241	1918	<924	Weivirus-like virus sp. (MT138416, QJI53766), 1203/1921 (63%)
Trapa bispinosa dicistro-like virus 1	TbDLV1	BK059256	>8051 ^a	>1678 ^a	<840	Sanxia atyid shrimp virus 3 (NC_033228.1:YP_009337727.1), 1663/1678 (99%)

^a5' partial. ^bThe first in-frame sense codon of ORF2 was used as the starting position. ^cThe best BLAST hit of the ORF1 aa sequence is presented in the format of "virus name (genome and protein Acc. Nos.), identical aa/aligned aa (% identity)."

ORF1 polyprotein sequences shared 26% identity with 595 aa overlaps, indicating that these two viruses were evolutionarily distant from each other.

Prediction of ORF2 in the SaDLV1 or SaDLV2 genome sequences is highly challenging because translational initiation of the dicistrovirus ORF2 occurs at a non-AUG codon guided by the IGR IRES (Hertz and Thompson, 2011 b). The IGR IRES sequence must be accurately determined to correctly identify the first codon of ORF2. However, prediction of IGR IRES of a novel dicistrovirus is difficult because IGR IRES sequences generally do not exhibit notable sequence similarities among distantly related dicistroviruses. Since there was no sequence similarity between the putative intergenic region of the SaDLV1 or SaDLV2 genome and known IGR IRES sequences of CrPV and KBV, it was not possible to predict the true start codon of ORF2. Therefore, we decided to use the first sense co-

don as the provisional first codon of ORF2. The predicted SaDLV1 and SaDLV2 ORF2 polyproteins were 779 aa and 838 aa, respectively.

Dicistroviruses were first isolated in insects and crustaceans, although recent studies have suggested that more diverse organisms might be their natural hosts (Bonning and Miller, 2010; Chiapello *et al.*, 2020; Dastjerdi *et al.*, 2021). The presence of dicistro-like viruses in the *S. asiatica* sample and other plants raised the possibility that these viruses might be more widely associated with various plants (Nery *et al.*, 2020). To identify their potential viral genome sequences in other plant transcriptome data, the plant transcript contig sequences deposited in the NCBI Transcriptome Sequence Assembly database were searched using the SaDLV1 and SaDLV2 ORF1 polyprotein sequences as queries. Initially, low stringent conditions, such as an E-value cut-off of 1e-5 and minimum length

	SaDLV1	SaDLV2	AhDLV1	CqDLV2	CbDLV1	TbDLV1	CqDLV1	DcDLV1	NoDLV1	ShDLV1	AsDLV1	CaDLV1	HvDLV1	SiDLV1
SaDLV1		25%	26%	25%	26%	26%	27%	24%	24%	21%	25%	25%	25%	25%
SaDLV2	26%		77%	38%	39%	37%	23%	25%	24%	23%	22%	22%	22%	23%
AhDLV1	25%	91%		39%	40%	37%	24%	25%	25%	22%	23%	22%	23%	23%
CqDLV2	26%	41%	42%		41%	39%	25%	27%	26%	22%	22%	22%	22%	22%
CbDLV1	28%	49%	48%	47%		41%	24%	24%	27%	23%	25%	25%	25%	25%
TbDLV1	26%	32%	33%	34%	32%		23%	25%	26%	21%	23%	23%	23%	23%
CqDLV1	27%	31%	30%	30%	31%	28%		22%	22%	24%	25%	25%	25%	25%
DcDLV1	28%	26%	26%	26%	27%	25%	28%		23%	25%	22%	23%	23%	23%
NoDLV1	21%	23%	24%	27%	25%	25%	25%	25%		24%	23%	23%	23%	23%
ShDLV1	26%	27%	28%	28%	26%	24%	27%	27%	23%		27%	22%	23%	22%
AsDLV1	26%	26%	26%	29%	26%	28%	28%	27%	22%	27%		99%	99%	99%
CaDLV1	26%	25%	25%	29%	25%	28%	28%	27%	23%	27%	99%		99%	99%
HvDLV1	26%	25%	26%	29%	26%	28%	28%	26%	23%	27%	99%	99%		99%
SiDLV1	26%	26%	26%	29%	26%	28%	28%	27%	22%	27%	99%	99%	99%	

Fig. 2

Sequence comparison of plant dicistro-like viruses

Pairwise identities of ORF1 (lower left triangle) and ORF2 (upper right triangle) polyprotein sequences of 14 plant dicistro-like viruses are presented.

of 3 kb, were applied to collect 202 putative viral contigs. Then, these contigs were compared with all viral proteins in the NCBI protein database to identify 15 contigs from 14 plant species showing a strong similarity (E-value cut-off of $1e-10$) to known dicistrovirus ORF1 polyproteins. To generate improved contigs, we downloaded raw RNA-Seq data of the 14 plant species from the NCBI SRA database and re-assembled them into contigs using the RNA viral genome assembly mode of the SPAdes Genome Assembler (Bushmanova *et al.*, 2019). Finally, contigs longer than 5 kb were selected and further analyzed to obtain 12 dicistro-like virus contigs from 11 plant species.

The newly identified dicistro-like virus genomes were named after the associated plant species: *Alloteropsis semialata* dicistro-like virus 1 (AsDLV1) from *A. semialata* (cockatoo grass); *Arabidopsis halleri* dicistro-like virus 1 (AhDLV1) from *A. halleri* subsp. *halleri*; *Camptotheca acuminata* dicistro-like virus 1 (CaDLV1) from *C. acuminata* (happy tree); *Colobanthus quitensis* dicistro-like virus 1 (CqDLV1) and 2 (CqDLV2) from *C. quitensis* (Antarctic pearlwort); *Cosmos bipinnatus* dicistro-like virus 1 (CbDLV1) from *C. bipinnatus* (garden cosmos); *Dalbergia cochinchinensis* dicistro-like virus 1 (DcDLV1) from *D. cochinchinensis* (Thailand rosewood); *Hordeum vulgare* dicistro-like virus 1 (HvDLV1) from *H. vulgare* subsp. *vulgare* (common barley); *Nasturtium officinale* dicistro-like virus 1 (NoDLV1) from *N. officinale* (watercress); *Saccharum hybrid* dicistro-like virus 1 (ShDLV1) from *Saccharum hybrid* cultivar SP80-3280 (sugarcane);

Silene latifolia dicistro-like virus 1 (SiDLV1) from *S. latifolia* (white campion); and *Trapa bispinosa* dicistro-like virus 1 (TbDLV1) from *T. bispinosa* (water chestnut). The lists of plant species and SRA Acc. Nos. are presented in Supplementary Data S1. The genome sequences were deposited in the NCBI nucleotide database under the Acc. Nos. BK059245–BK059256.

The ORF1 and ORF2 polyproteins were predicted in each of the 12 dicistro-like virus genome sequences (Fig. 1c–n and Table 1), and nine (AhDLV1, CaDLV1, CqDLV1, CqDLV2, CbDLV1, DcDLV1, HvDLV1, NoDLV1, and SiDLV1) contained complete ORF1, while three (AsDLV1, ShDLV1, and TbDLV1) were incomplete at their 5' ends and lacked the ORF1 start codon. The sizes of the nine complete ORF1 polyproteins ranged from 1509–1918 aa. ORF2 polyproteins were predicted in all 12 viruses and their sizes ranged from 757–1320 aa, when the first in-frame sense codon was regarded as the first codon.

We performed pairwise sequence comparisons of ORF1 and ORF2 polyproteins of all 14 dicistro-like viruses identified in this study (Fig. 2 and Supplementary Data S2). Both ORF1 and ORF2 polyproteins of AsDLV1, CaDLV1, HvDLV1, and SiDLV1 shared 99% identity, indicating that these four viruses belong to the same species. These four viruses have been identified in different plant samples from unrelated studies (Chibalina and Filatov, 2011; Gongora-Castillo *et al.*, 2012; Dunning *et al.*, 2017; Bettgenhaeuser *et al.*, 2021). In the case of SaDLV2 and AhDLV1, the ORF1 and ORF2 polyproteins had 91% and 77% aa identities,



Fig. 3

Phylogenetic relationships among various plant dicistro-like viruses of the family Dicistroviridae

A maximum-likelihood phylogenetic tree was constructed from the ORF1 polyprotein sequences of 14 plant dicistro-like viruses identified in this study (red letters) and related dicistroviruses. Bootstrap supporting values of 90% or higher are shown at the nodes. The lowest taxonomic ranks, as denoted in the NCBI records, are shown in parentheses. The categories of sources from which the viruses were isolated are depicted as symbols. The known and probable genera to which the viruses belong are marked by solid and dotted lines, respectively, on the right.

respectively, suggesting that these two viruses belong to different species but are closely related to each other. These two viruses have also been identified in different plant samples from independent studies (Schvartzman *et al.*, 2018; Yoshida *et al.*, 2019). Detection of almost identical or very similar viruses in different plant species implied that they could spread very rapidly over long distances through highly mobile vectors, such as insects. Other viral pairs exhibited lower identities (21%–49%). Therefore, 14 dicistro-like viruses can be grouped into 11 different species.

We compared the ORF1 polyproteins of the 14 plant dicistro-like viruses identified in this study with those of known dicistroviruses as of July 2021 (Table 1). The TbDLV1 ORF1 polyprotein showed 99% aa identity with that of Sanxia atyid shrimp virus 3 (SASV3), indicating that these two belong to the same species. In contrast, ORF1 polyproteins from other 13 plant dicistro-like viruses exhibited 28%–66% identities with those of known dicistroviruses, suggesting that they are novel viruses. Interestingly, TbDLV1 and SASV3, which seemed to be the same species, were identified in different sources. TbDLV1 was identified in the *T. bispinosa* (a freshwater plant commonly known as water chestnut) transcriptome data obtained from a sample collected in the Wuhan Botanical Garden, Hubei, China (Li *et al.*, 2017), while SASV3 was isolated from a freshwater atyid shrimp collected in Hubei, China (Shi *et al.*, 2016). Detection of the same virus in different organisms in a shared habitat suggested that TbDLV1/SASV3-related viruses might be widespread in freshwater environments in Hubei, China.

For phylogenetic analysis, ORF1 polyprotein sequences in the NCBI protein database showing significant identity to 14 plant dicistro-like virus ORF1 proteins were collected. After filtering and clustering, 91 known viruses were selected, totaling 105 sequences in the final dataset. Detailed information on viruses, including the genome Acc. Nos., virus names, lowest taxonomic ranks, isolation sources, and source categories are presented in Supplementary Data S3.

A maximum-likelihood phylogenetic tree was inferred from multiple alignments of the ORF1 protein sequences of 105 viruses (Fig. 3). The plant dicistro-like viruses identified in this study were broadly distributed within the family *Dicistroviridae*. Four viruses (AsDLV1, CaDLV1, HvDLV1, and SIDLV1), which seemed to belong to the same species, formed a single subclade within the genus *Aparavirus*. ShDLV1 might be a novel member of the genus *Cripavirus* or a closely related taxon. Interestingly, these five plant-associated viruses were placed within the genera *Aparavirus* and *Cripavirus*, which mainly consisted of viruses isolated from arthropods and other animals.

The other nine plant dicistro-like viruses were placed within unclassified clades. SaDLV2 and AhDLV1 formed a strong subclade because they were closely related to each other. TbDLV1 and the previously reported SASV3 also formed a single subclade because they seemed to belong to the same species. The remaining six viruses (SaDLV1, CqDLV1, DcDLV1, NoDLV1, CbDLV1, and CbDLV2) were scattered with unclassified dicistroviruses. The unclassified clades included many previously known plant dicistroviruses, implying that there might be a large taxonomic group consisting of plant-associated dicistroviruses.

In conclusion, 14 plant-associated dicistro-like viruses were identified in the transcriptome data of 12 plant species. Sequence comparison and phylogenetic analysis showed that they belonged to 11 virus species in the family *Dicistroviridae*, 10 of which might be novel species. The presence of dicistro-like viruses in diverse plant samples suggests that these viruses may infect a wide range of plant species or may use plants as reservoirs for transmission.

Acknowledgments. This work was supported by grants from the National Research Foundation of Korea (2018R1A5A1025077 and 2020R1A2C1013403) funded by the Government of Korea.

Supplementary information is available in the online version of the paper.

References

- Bejerman N, Debat H, Dietzgen RG (2020): The plant negative-sense RNA virosphere: virus discovery through new eyes. *Front. Microbiol.* 11, 588427. <https://doi.org/10.3389/fmicb.2020.588427>
- Bettgenhaeuser J, Hernandez-Pinzon I, Dawson AM, Gardiner M, Green P, Taylor J, Smoker M, Ferguson JN, Emmrich P, Hubbard A, Bayles R, Waugh R, Steffenson BJ, Wulff BBH, Dreiseitl A, Ward ER, Moscou MJ (2021): The barley immune receptor Mla recognizes multiple pathogens and contributes to host range dynamics. *Nat. Commun.* 12, 6915. <https://doi.org/10.1038/s41467-021-27288-3>
- Bonning BC, Miller WA (2010): Dicistroviruses. *Annu. Rev. Entomol.* 55, 129–150. <https://doi.org/10.1146/annurev-ento-112408-085457>
- Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019): rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8, giz100. <https://doi.org/10.1093/gigascience/giz100>
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009): trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>

- Carrillo-Tripp J, Bonning BC, Miller WA (2015): Challenges associated with research on RNA viruses of insects. *Curr. Opin. Insect Sci.* 8, 62–68. <https://doi.org/10.1016/j.cois.2014.11.002>
- Chiapello M, Rodriguez-Romero J, Ayllon MA, Turina M (2020): Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus Evol.* 6, veaa058. <https://doi.org/10.1093/ve/veaa058>
- Chibalina MV, Filatov DA (2011): Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* 21, 1475–1479. <https://doi.org/10.1016/j.cub.2011.07.045>
- Choi D, Shin C, Shirasu K, Hahn Y (2021): Two novel poty-like viruses identified from the transcriptome data of purple witchweed (*Striga hermonthica*). *Acta Virol.* 65, 365–372. https://doi.org/10.4149/av_2021_402
- Czibener C, La Torre JL, Muscio OA, Ugalde RA, Scodeller EA (2000): Nucleotide sequence analysis of triatoma virus shows that it is a member of a novel group of insect RNA viruses. *J. Gen. Virol.* 81, 1149–1154. <https://doi.org/10.1099/0022-1317-81-4-1149>
- Dastjerdi A, Everest DJ, Davies H, Denk D, Zell R (2021): A novel dicistrovirus in a captive red squirrel (*Sciurus vulgaris*). *J. Gen. Virol.* 102. <https://doi.org/10.1099/jgv.0.001555>
- de Miranda JR, Cordoni G, Budge G (2010): The acute bee paralysis virus-Kashmir bee virus-Israeli acute paralysis virus complex. *J. Invertebr. Pathol.* 103, S30–S47. <https://doi.org/10.1016/j.jip.2009.06.014>
- Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P, Osborne CP, Christin PA (2017): Introgression and repeated co-option facilitated the recurrent emergence of C4 photosynthesis among close relatives. *Evolution* 71, 1541–1555. <https://doi.org/10.1111/evo.13250>
- Goh CJ, Park D, Hahn Y (2021): A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (*Agave tequilana*). *Acta Virol.* 65, 68–71. https://doi.org/10.4149/av_2021_107
- Gongora-Castillo E, Childs KL, Fedewa G, Hamilton JP, Liscombe DK, Magallanes-Lundback M, Mandadi KK, Nims E, Runguphan W, Vaillancourt B, Varbanova-Herde M, Dellapenna D, McKnight TD, O'Connor S, Buell CR (2012): Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS One* 7, e52506. <https://doi.org/10.1371/journal.pone.0052506>
- Govan VA, Leat N, Allsopp M, Davison S (2000): Analysis of the complete genome sequence of acute bee paralysis virus shows that it belongs to the novel group of insect-infecting RNA viruses. *Virology* 277, 457–463. <https://doi.org/10.1006/viro.2000.0616>
- Greninger AL, Jerome KR (2016): Draft genome sequence of goose dicistrovirus. *Genome Announc.* 4, e00068–16. <https://doi.org/10.1128/genomeA.00068-16>
- Guo ZX, He JG, Xu HD, Weng SP (2013): Pathogenicity and complete genome sequence analysis of the mud crab dicistrovirus-1. *Virus Res.* 171, 8–14. <https://doi.org/10.1016/j.virusres.2012.10.002>
- Hertz MI, Thompson SR (2011a): In vivo functional analysis of the Dicistroviridae intergenic region internal ribosome entry sites. *Nucleic Acids Res.* 39, 7276–7288. <https://doi.org/10.1093/nar/gkr427>
- Hertz MI, Thompson SR (2011b): Mechanism of translation initiation by Dicistroviridae IGR IRESs. *Virology* 411, 355–361. <https://doi.org/10.1016/j.virol.2011.01.005>
- Khong A, Bonderoff JM, Spriggs RV, Tammperre E, Kerr CH, Jackson TJ, Willis AE, Jan E (2016): Temporal regulation of distinct internal ribosome entry sites of the Dicistroviridae cricket paralysis virus. *Viruses* 8, 25. <https://doi.org/10.3390/v8010025>
- Li J, Yang Y, Yang S, Zhang Z, Chen S, Zhong C, Zhou R, Shi S (2017): Comparative transcriptome analyses of a mangrove tree *Sonneratia caseolaris* and its non-mangrove relatives, *Trapa bispinosa* and *Duabanga grandiflora*. *Mar. Genomics* 31, 13–15. <https://doi.org/10.1016/j.margen.2016.10.007>
- Mari J, Poulos BT, Lightner DV, Bonami JR (2002): Shrimp taura syndrome virus: genomic characterization and similarity with members of the genus cricket paralysis-like viruses. *J. Gen. Virol.* 83, 915–926. <https://doi.org/10.1099/0022-1317-83-4-915>
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020): IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Nakamura T, Yamada KD, Tomii K, Katoh K (2018): Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34, 2490–2492. <https://doi.org/10.1093/bioinformatics/bty121>
- Nery FMB, Melo FL, Boiteux LS, Ribeiro SG, Resende RO, Orilio AF, Batista JG, Lima ME, Pereira-Carvalho RC (2020): Molecular characterization of *Hovenia dulcis*-associated virus 1 (HDAV1) and 2 (HDAV2): New tentative species within the order Picornavirales. *Viruses* 12, 950. <https://doi.org/10.3390/v12090950>
- Park D, Goh CJ, Hahn Y (2021): Two novel closteroviruses, fig virus A and fig virus B, identified by the analysis of the high-throughput RNA-sequencing data of fig (*Ficus carica*) latex. *Acta Virol.* 65, 42–48. https://doi.org/10.4149/av_2021_104
- Park D, Hahn Y (2021): Identification of genome sequences of novel partitiviruses in the quinoa (*Chenopodium quinoa*) transcriptome datasets. *J. Gen. Plant Path.* 87, 236–241. <https://doi.org/10.1007/s10327-021-01002-z>
- Phan TG, Del Valle Mendoza J, Sadeghi M, Altan E, Deng X, Delwart E (2018): Sera of Peruvians with fever of unknown origins include viral nucleic acids from non-vertebrate hosts. *Virus Genes* 54, 33–40. <https://doi.org/10.1007/s11262-017-1514-3>
- Reinganum C, O'Loughlin GT, Hogan TW (1970): A nonoccluded virus of the field crickets *Teleogryllus oceanicus* and *T. commodus* (Orthoptera: Gryllidae). *J. Invertebr. Pathol.* 1, 1–10.

- tebr. *Pathol.* 16, 214–220. [https://doi.org/10.1016/0022-2011\(70\)90062-5](https://doi.org/10.1016/0022-2011(70)90062-5)
- Schwartzman MS, Corso M, Fataftah N, Scheepers M, Nouet C, Bosman B, Carnol M, Motte P, Verbruggen N, Hanikenne M (2018): Adaptation to high zinc depends on distinct mechanisms in metal-tolerant populations of *Arabidopsis halleri*. *New Phytol.* 218, 269–282. <https://doi.org/10.1111/nph.14949>
- Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ (2016): Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543. <https://doi.org/10.1038/nature20167>
- Shin C, Choi D, Hahn Y (2021): 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.* 65, 373–380. https://doi.org/10.4149/av_2021_404
- Spallek T, Mutuku M, Shirasu K (2013): The genus *Striga*: a witch profile. *Mol. Plant Pathol.* 14, 861–869. <https://doi.org/10.1111/mpp.12058>
- Valles SM, Chen Y, Firth AE, Guerin DMA, Hashimoto Y, Herrero S, de Miranda JR, Ryabov E, ICTV Report Consortium (2017): ICTV virus taxonomy profile: Dicistroviridae. *J. Gen. Virol.* 98, 355–356. <https://doi.org/10.1099/jgv.0.000756>
- Warsaba R, Sadasivan J, Jan E (2020): Dicistrovirus-host molecular interactions. *Curr. Issues Mol. Biol.* 34, 83–112. <https://doi.org/10.21775/cimb.034.083>
- Wilson JE, Powell MJ, Hoover SE, Sarnow P (2000): Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol. Cell. Biol.* 20, 4990–4999. <https://doi.org/10.1128/MCB.20.14.4990-4999.2000>
- Yinda CK, Zell R, Deboutte W, Zeller M, Conceicao-Neto N, Heylen E, Maes P, Knowles NJ, Ghogomu SM, Van Ranst M, Matthijssens J (2017): Highly diverse population of Picornaviridae and other members of the Picornavirales, in Cameroonian fruit bats. *BMC Genomics* 18, 249. <https://doi.org/10.1186/s12864-017-3632-7>
- Yoshida S, Kim S, Wafula EK, Tanskanen J, Kim YM, Honaas L, Yang Z, Spallek T, Conn CE, Ichihashi Y, Cheong K, Cui S, Der JP, Gundlach H, Jiao Y, Hori C, Ishida JK, Kasahara H, Kiba T, Kim MS, Koo N, Laohavisit A, Lee YH, Lumba S, McCourt P, Mortimer JC, Mutuku JM, Nomura T, Sasaki-Sekimoto Y, Seto Y, Wang Y, Wakatake T, Sakakibara H, Demura T, Yamaguchi S, Yoneyama K, Manabe RI, Nelson DC, Schulman AH, Timko MP, dePamphilis CW, Choi D, Shirasu K (2019): Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Curr. Biol.* 29, 3041–3052.e4. <https://doi.org/10.1016/j.cub.2019.07.086>

SUPPLEMENTARY INFORMATION

Identification of dicistro-like viruses in the transcriptome data of *Striga asiatica* and other plants

Chaerim Shin¹, Dongjin Choi¹, Ken Shirasu^{2,3}, Yoonsoo Hahn^{1*}

¹Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; ²RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; ³Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

Received January 7, 2022; accepted April 11, 2022

*Corresponding author. E-mail: hahnyc@cau.ac.kr; phone: +82-2-820-5812.

Data S1. List of plant transcriptome data

Plant	Virus	Project	Sequencing run
<i>Striga asiatica</i>	Striga asiatica dicistro-like virus 1 (SaDLV1), Striga asiatica dicistro-like virus 2 (SaDLV2)	DRP005455	DRR175773, DRR175774, DRR175775, DRR175776, DRR175777, DRR175778, DRR175779, DRR175780
<i>Alloteropsis semialata</i>	Alloteropsis semialata dicistro-like virus 1 (AsDLV1)	SRP072730	SRR3323127, SRR3323128, SRR3323129
<i>Arabidopsis halleri</i> subsp. <i>halleri</i>	Arabidopsis halleri dicistro-like virus 1 (AhDLV1)	SRP111320	SRR5810196
<i>Camptotheca acuminata</i>	Camptotheca acuminata dicistro-like virus 1 (CaDLV1)	SRP006330	SRR173032, SRR173033, SRR173034, SRR173035, SRR173036, SRR173044, SRR173045, SRR173046, SRR173047, SRR173048, SRR173049, SRR173050, SRR173052, SRR173053, SRR173054
<i>Colobanthus quitensis</i>	Colobanthus quitensis dicistro-like virus 1 (CqDLV1), Colobanthus quitensis dicistro-like virus 2 (CqDLV2)	SRP051241	SRR1720758, SRR1720760, SRR1720762, SRR1720763, SRR1720765, SRR1720767
<i>Cosmos bipinnatus</i>	Cosmos bipinnatus dicistro-like virus 1 (CbDLV1)	SRP074877	SRR3546768, SRR3546769
<i>Dalbergia cochinchinensis</i>	Dalbergia cochinchinensis dicistro-like virus 1 (DcDLV1)	SRP234829	SRR10592617, SRR10592618
<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	Hordeum vulgare dicistro-like virus 1 (HvDLV1)	SRP101929	SRR5345034
<i>Nasturtium officinale</i>	Nasturtium officinale dicistro-like virus 1 (NoDLV1)	SRP058520	SRR2035328, SRR2035361
<i>Saccharum</i> hybrid cultivar SP80-3280	Saccharum hybrid dicistro-like virus 1 (ShDLV1)	SRP041206	SRR1974519, SRR1979656, SRR1979657, SRR1979658, SRR1979659, SRR1979660, SRR1979661, SRR1979662, SRR1979663, SRR1979664, SRR1979665, SRR1979666, SRR1979667, SRR1979668, SRR1979669
<i>Silene latifolia</i>	Silene latifolia dicistro-like virus 1 (SiDLV1)	SRP007582	SRR316281, SRR316283, SRR316286, SRR316287, SRR316288, SRR316289
<i>Trapa bispinosa</i>	Trapa bispinosa dicistro-like virus 1 (TbDLV1)	SRP082135	SRR4031066

Data S2. Pairwise sequence comparison of proteins of plant dicistro-like viruses

Protein sequence identities are presented in the format of “identical residues/aligned length (percent identity).” Lower left triangle, ORF1; upper right triangle, ORF2.

	SaDLV1	SaDLV2	AhDLV1	CqDLV2	CbDLV1	TbDLV1	CqDLV1	DcDLV1	NoDLV1	ShDLV1	AsDLV1	CaDLV1	HvDLV1	SIDLV1
SaDLV1		149/587 (25%)	126/484 (26%)	140/555 (25%)	142/553 (26%)	143/543 (26%)	170/622 (27%)	122/513 (24%)	138/585 (24%)	110/512 (21%)	185/731 (25%)	185/731 (25%)	185/731 (25%)	185/735 (25%)
SaDLV2	153/595 (26%)		647/838 (77%)	321/849 (38%)	227/579 (39%)	308/842 (37%)	206/890 (23%)	151/612 (25%)	172/709 (24%)	177/769 (23%)	129/578 (22%)	129/578 (22%)	130/578 (22%)	131/578 (23%)
AhDLV1	151/594 (25%)	1268/1386 (91%)		332/850 (39%)	233/578 (40%)	312/846 (37%)	214/896 (24%)	180/728 (25%)	155/627 (25%)	168/762 (22%)	168/741 (23%)	165/734 (22%)	168/741 (23%)	169/740 (23%)
CqDLV2	162/622 (26%)	583/1416 (41%)	613/1463 (42%)		233/572 (41%)	325/838 (39%)	168/673 (25%)	151/568 (27%)	160/626 (26%)	149/680 (22%)	109/492 (22%)	108/493 (22%)	108/492 (22%)	108/492 (22%)
CbDLV1	170/617 (28%)	709/1447 (49%)	718/1495 (48%)	369/779 (47%)		237/579 (41%)	129/539 (24%)	133/552 (24%)	159/589 (27%)	131/562 (23%)	120/487 (25%)	120/487 (25%)	120/487 (25%)	121/487 (25%)
TbDLV1	169/655 (26%)	280/876 (32%)	286/874 (33%)	273/798 (34%)	290/903 (32%)		172/758 (23%)	168/663 (25%)	146/566 (26%)	134/652 (21%)	183/796 (23%)	176/767 (23%)	182/796 (23%)	182/796 (23%)
CqDLV1	162/607 (27%)	232/754 (31%)	234/785 (30%)	240/808 (30%)	228/745 (31%)	217/778 (28%)		169/767 (22%)	113/510 (22%)	208/875 (24%)	157/638 (25%)	157/638 (25%)	157/638 (25%)	159/636 (25%)
DcDLV1	161/584 (28%)	323/1254 (26%)	331/1255 (26%)	315/1233 (26%)	199/741 (27%)	181/713 (25%)	217/767 (28%)		170/735 (23%)	61/241 (25%)	112/499 (22%)	113/499 (23%)	113/499 (23%)	113/499 (23%)
NoDLV1	142/669 (21%)	317/1355 (23%)	205/842 (24%)	164/614 (27%)	210/835 (25%)	149/603 (25%)	195/782 (25%)	149/586 (25%)		144/611 (24%)	100/442 (23%)	100/442 (23%)	100/442 (23%)	100/442 (23%)
ShDLV1	171/658 (26%)	163/594 (27%)	167/593 (28%)	164/588 (28%)	198/750 (26%)	169/719 (24%)	200/743 (27%)	200/751 (27%)	171/731 (23%)		171/749 (23%)	190/848 (22%)	170/752 (23%)	169/752 (22%)
AsDLV1	203/787 (26%)	197/771 (26%)	202/784 (26%)	171/589 (29%)	206/800 (26%)	169/613 (28%)	211/742 (28%)	198/743 (27%)	142/635 (22%)	216/802 (27%)		915/924 (99%)	915/924 (99%)	917/924 (99%)
CaDLV1	203/787 (26%)	196/771 (25%)	204/804 (25%)	170/589 (29%)	205/805 (25%)	169/609 (28%)	210/741 (28%)	198/743 (27%)	145/635 (23%)	217/802 (27%)	1276/1282 (99%)		919/924 (99%)	916/924 (99%)
HvDLV1	204/787 (26%)	196/771 (25%)	202/784 (26%)	169/589 (29%)	205/800 (26%)	169/609 (28%)	210/741 (28%)	194/742 (26%)	146/635 (23%)	218/802 (27%)	1274/1282 (99%)	1909/1918 (99%)		916/924 (99%)
SIDLV1	203/787 (26%)	197/771 (26%)	202/784 (26%)	171/589 (29%)	206/800 (26%)	169/613 (28%)	209/741 (28%)	198/743 (27%)	142/635 (22%)	217/802 (27%)	1277/1282 (99%)	1904/1918 (99%)	1902/1918 (99%)	

Data S3. List of viruses used for the phylogenetic analysis

Acc. No.	Virus	Taxonomy	Source	Category
NC_003005	Taura syndrome virus	<i>Aparavirus</i>	<i>Penaeus vannamei</i>	Crustacean
NC_032464	Beihai mantis shrimp virus 4	<i>Riboviria</i>	Mantis shrimp	Crustacean
NC_032791	Wenling picorna-like virus 3	<i>Riboviria</i>	Crustacean	Crustacean
NC_032638	Beihai picorna-like virus 91	<i>Riboviria</i>	Hermit crab	Crustacean
NC_032579	Beihai shrimp virus 1	<i>Riboviria</i>	Penaeid shrimp	Crustacean
NC_032989	Wenzhou shrimp virus 4	<i>Riboviria</i>	<i>Charybdis</i> crab	Crustacean
NC_032830	Wenling crustacean virus 2	<i>Riboviria</i>	Crustacean	Crustacean
MF189971	Caledonia beadlet anemone dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Actinia equina</i>	Invertebrate
NC_032858	Wenling crustacean virus 3	<i>Riboviria</i>	Crustacean	Crustacean
NC_032112	Bivalve RNA virus G1	<i>Dicistroviridae</i>	Bivalve gills	Invertebrate
NC_032441	Beihai mantis shrimp virus 3	<i>Riboviria</i>	Mantis shrimp	Crustacean
MG995700	Bundaberg bee virus 2	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect
MT138416	Weivirus-like virus sp.	<i>Riboviria</i>	<i>Abornis proregulus</i> , anal swab	Bird
KY973643	Human blood-associated dicistrovirus	<i>Dicistroviridae</i>	<i>Homo sapiens</i> , blood	Mammal
MT138417	Weivirus-like virus sp.	<i>Riboviria</i>	<i>Phylloscopus</i> , anal swab	Bird
BK059252	Hordeum vulgare dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	Plant
BK059247	Camptotheca acuminata dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Camptotheca acuminata</i>	Plant
BK059245	Alloteropsis semialata dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Alloteropsis semialata</i>	Plant
BK059255	Silene latifolia dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Silene latifolia</i>	Plant
NC_033173	Sanxia water strider virus 9	<i>Riboviria</i>	Water striders	Insect
NC_032209	Hubei picorna-like virus 46	<i>Riboviria</i>	Spiders	Arthropod
MN722411	Sedum sarmentosum dicistroviridae	<i>Dicistroviridae</i>	<i>Sedum sarmentosum</i>	Plant
NC_032877	Wenzhou channeled applesnail virus 2	<i>Riboviria</i>	Channeled applesnail	Invertebrate
NC_032975	Wenzhou picorna-like virus 28	<i>Riboviria</i>	<i>Charybdis</i> crab	Crustacean
MN551111	Plasmopara viticola lesion associated dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Plasmopara viticola</i>	Single-celled eukaryote
MW023863	Soybean thrips picorna-like virus 10	<i>Picornavirales</i>	<i>Neohydatothrips variabilis</i>	Insect
MH727530	Solenopsis invicta virus 13	<i>Aparavirus</i>	<i>Solenopsis invicta</i> (fire ant)	Insect
MH727529	Solenopsis invicta virus 12	<i>Aparavirus</i>	<i>Solenopsis invicta</i> (fire ant)	Insect
NC_006559	Solenopsis invicta virus 1	<i>Aparavirus</i>	<i>Solenopsis invicta</i>	Insect
NC_033460	Wuhan insect virus 11	<i>Riboviria</i>	Crayfish	Crustacean
NC_009025	Israeli acute paralysis virus	<i>Aparavirus</i>	Honey bee	Insect
NC_002548	Acute bee paralysis virus	<i>Aparavirus</i>	<i>Apis mellifera</i>	Insect
BK059254	Saccharum hybrid dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Saccharum</i> hybrid cultivar SP80-3280	Plant
MN905970	Dicistroviridae sp.	<i>Dicistroviridae</i>	Anal swab	Bird
MT757492	Shenzhen dicistro-like virus	<i>Dicistroviridae</i>	<i>Dermatophagoides pteronyssinus</i>	Arthropod
MT757506	Scotland dicistro-like virus	<i>Dicistroviridae</i>	<i>Psoroptes ovis</i>	Arthropod
MG995714	Robinvale bee virus 1	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect
MT747996	Ohio dicistro-like virus	<i>Dicistroviridae</i>	<i>Dermatophagoides pteronyssinus</i>	Arthropod
MF190024	Barns Ness breadcrumb sponge dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Halichondria panicea</i>	Invertebrate
NC_031687	Centovirus AC	<i>Dicistroviridae</i>	Mosquito	Insect
KY607723	Macrobrachium rosenbergii dicistrovirus 2	<i>Dicistroviridae</i>	<i>Macrobrachium rosenbergii</i>	Crustacean
MW208809	Bactrocera tryoni dicistrovirus 2	<i>Dicistroviridae</i>	<i>Bactrocera tryoni</i> strain bent wings	Insect
NC_029052	Goose dicistrovirus	<i>Dicistroviridae</i>	Goose, feces	Bird

Data S3. Continued 1

Acc. No.	Virus	Taxonomy	Source	Category
MF189973	Caledonia beadlet anemone dicistro-like virus 2	<i>Dicistroviridae</i>	<i>Actinia equina</i>	Invertebrate
NC_035115	Apis dicistrovirus	<i>Dicistroviridae</i>	<i>Apis mellifera</i>	Insect
NC_030115	Anopheles C virus	<i>Cripavirus</i>	<i>Anopheles coluzzii</i>	Insect
NC_001834	Drosophila C virus	<i>Cripavirus</i>	<i>Drosophila</i>	Insect
NC_003924	Cricket paralysis virus	<i>Cripavirus</i>	Cricket	Insect
NC_008029	Homalodisca coagulata virus 1	<i>Triatovirus</i>	<i>Homalodisca coagulata</i>	Insect
NC_003779	Plautia stali intestine virus	<i>Triatovirus</i>	<i>Plautia stali</i>	Insect
NC_003783	Triatoma virus	<i>Triatovirus</i>	<i>Triatoma infestans</i>	Insect
NC_003782	Himetobi P virus	<i>Triatovirus</i>	<i>Laodelphax striatellus</i>	Insect
MG995710	Renmark bee virus 1	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect
NC_003784	Black queen cell virus	<i>Triatovirus</i>	<i>Apis mellifera</i>	Insect
MH213235	Linepithema humile picorna-like virus 1	<i>Picornavirales</i>	<i>Linepithema humile</i>	Insect
MT108240	Anoplolepis gracilipes virus 2	<i>Triatovirus</i>	<i>Anoplolepis gracilipes</i>	Insect
NC_033437	Wuhan arthropod virus 2	<i>Riboviria</i>	pillworm	Arthropod
MT108239	Anoplolepis gracilipes virus 1	<i>Triatovirus</i>	<i>Anoplolepis gracilipes</i>	Insect
MH727526	Solenopsis invicta virus 9	<i>Triatovirus</i>	<i>Solenopsis invicta</i> (fire ant)	Insect
NC_032879	Changjiang picorna-like virus 13	<i>Riboviria</i>	Crayfish	Crustacean
KX884335	Wenzhou picorna-like virus 34	<i>Riboviria</i>	Channeled applesnail	Invertebrate
NC_032582	Beihai picorna-like virus 80	<i>Riboviria</i>	Razor shell	Invertebrate
MG995711	Renmark bee virus 5	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect
MN832470	Kummerowia striata picorna-like virus	<i>Picornavirales</i>	<i>Kummerowia striata</i>	Plant
MZ598483	Striga asiatica dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Striga asiatica</i>	Plant
MG995733	Perth bee virus 4	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect
MN735445	Forsythia suspensa dicistrovirus	<i>Dicistroviridae</i>	<i>Forsythia suspensa</i>	Plant
MN832463	Trichosanthes kirilowii picorna-like virus	<i>Picornavirales</i>	<i>Trichosanthes kirilowii</i>	Plant
MN905971	Dicistroviridae sp.	<i>Dicistroviridae</i>	Anal swab	Bird
MT745991	Dicistroviridae sp. squirrel/UK/2011	<i>Dicistroviridae</i>	<i>Sciurus vulgaris</i> (Red squirrel); adult male, intestinal content	Mammal
BK059248	Colobanthus quitensis dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Colobanthus quitensis</i>	Plant
MN906001	Dicistroviridae sp.	<i>Dicistroviridae</i>	Anal swab	Bird
MT134328	Kummerowia striata dicistrovirus	<i>Dicistroviridae</i>	<i>Ciliophora</i> sp., <i>Kummerowia striata</i> sample	Single-celled eukaryote
BK059251	Dalbergia cochinchinensis dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Dalbergia cochinchinensis</i>	Plant
BK059253	Nasturtium officinale dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Nasturtium officinale</i>	Plant
NC_032113	Bivalve RNA virus G2	<i>Picornavirales</i>	Bivalve gills	Invertebrate
MN734251	Trichosanthes kirilowii dicistrovirus	<i>Dicistroviridae</i>	<i>Trichosanthes kirilowii</i>	Plant
MZ598484	Striga asiatica dicistro-like virus 2	<i>Dicistroviridae</i>	<i>Striga asiatica</i>	Plant
BK059246	Arabidopsis halleri dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Arabidopsis halleri</i> subsp. <i>halleri</i>	Plant
MN035381	Picornavirales sp.	<i>Picornavirales</i>	Grassland soil	Environment
BK059250	Cosmos bipinnatus dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Cosmos bipinnatus</i>	Plant
NC_033000	Hubei picorna-like virus 20	<i>Riboviria</i>	Myriapoda	Arthropod
MT138206	Weivirus-like virus sp.	<i>Riboviria</i>	<i>Grus japonensis</i> , anal swab	Bird
MT138207	Weivirus-like virus sp.	<i>Riboviria</i>	<i>Grus japonensis</i> , anal swab	Bird
NC_032823	Changjiang picorna-like virus 10	<i>Riboviria</i>	Crayfish	Crustacean

Data S3. Continued 2

Acc. No.	Virus	Taxonomy	Source	Category
BK059249	Colobanthus quitensis dicistro-like virus 2	<i>Dicistroviridae</i>	<i>Colobanthus quitensis</i>	Plant
NC_033069	Hubei picorna-like virus 21	<i>Riboviria</i>	Freshwater shellfish	Invertebrate
MF372755	Antarctic picorna-like virus	<i>Picornavirales</i>	<i>Anser anser</i> , cloacal swabs	Bird
NC_033266	Sanxia picorna-like virus 13	<i>Riboviria</i>	Freshwater atyid shrimp	Crustacean
BK059256	Trapa bispinosa dicistro-like virus 1	<i>Dicistroviridae</i>	<i>Trapa bispinosa</i>	Plant
NC_033228	Sanxia atyid shrimp virus 3	<i>Riboviria</i>	Freshwater atyid shrimp	Crustacean
MN841302	Lactuca sativa dicistroviridae	<i>Dicistroviridae</i>	<i>Lactuca sativa</i> L. var. <i>ramosa</i> Hort.	Plant
MN841301	Ginkgo biloba dicistrovirus	<i>Dicistroviridae</i>	<i>Ginkgo biloba</i>	Plant
MT317172	Rice Picorna-like virus 2	<i>Picornavirales</i>	<i>Oryza sativa</i> , green house plants Rby1 or Rby2	Plant
MN722412	Lactuca sativa dicistroviridae	<i>Dicistroviridae</i>	<i>Lactuca sativa</i> L. var. <i>ramosa</i> Hort.	Plant
NC_030232	Antarctic picorna-like virus 1	<i>Picornavirales</i>	Unknown, freshwater	Environment
MN729613	Ginkgo biloba dicistrovirus	<i>Dicistroviridae</i>	<i>Ginkgo biloba</i>	Plant
MN729612	Ginkgo biloba dicistrovirus	<i>Dicistroviridae</i>	<i>Ginkgo biloba</i>	Plant
MN722415	Lactuca sativa dicistroviridae	<i>Dicistroviridae</i>	<i>Lactuca sativa</i> L. var. <i>ramosa</i> Hort.	Plant
MN720005	Ginkgo biloba dicistrovirus	<i>Dicistroviridae</i>	<i>Ginkgo biloba</i>	Plant
NC_032850	Changjiang crawfish virus 2	<i>Riboviria</i>	Crayfish	Crustacean
MN722413	Lactuca sativa dicistroviridae	<i>Dicistroviridae</i>	<i>Lactuca sativa</i> L. var. <i>ramosa</i> Hort.	Plant
NC_033221	Sanxia picorna-like virus 11	<i>Riboviria</i>	Freshwater atyid shrimp	Crustacean
MN606295	Ginkgoaceae-associated picorna-like virus 1	<i>Picornavirales</i>	<i>Ginkgo biloba</i>	Plant
MG995726	Perth bee virus 2	<i>Picornavirales</i>	<i>Apis mellifera</i>	Insect