Acta virologica **66**: 206 – 215, 2022 doi: 10.4149/av_2022_302

A novel RNA virus, Thesium chinense closterovirus 1, identified by high-throughput RNA-sequencing of the parasitic plant *Thesium chinense*

Chaerim Shin[1], Dongjin Choi[1], Ken Shirasu[2,3], Yasunori Ichihashi[4], Yoonsoo Hahn[1*]

[1]Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; [2]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; [3]Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan; [4]RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Received March 15, 2022; accepted May 20, 2022

Summary. – The genome sequence of a closterovirus (genus *Closterovirus*, family *Closteroviridae*), tentatively named Thesium chinense closterovirus 1 (TcCV1), was identified by performing high-throughput RNA-sequencing of the haustoria and root tissues of *Thesium chinense*, a parasitic plant. The TcCV1 genome was predicted to encode nine proteins, eight of which have orthologs in previously identified closteroviruses. The TcCV1 RNA-dependent RNA polymerase (RdRp) and heat shock protein 70 homolog (Hsp70h) showed 27.8–68.2% and 23.8–55.1% amino acid identity, respectively, to orthologous proteins of known closteroviruses. The putative +1 ribosomal frameshifting site required for producing RdRp was identified as GUU**UAG**C with UAG stop codon and the skipped nucleotide U. Phylogenetic trees based on RdRp and Hsp70h show that TcCV1 is a novel member of the genus *Closterovirus*, forming a subclade with a group of known closteroviruses, including mint virus 1 and carnation necrotic fleck virus. The genome sequence of TcCV1 may be useful for studying the genome evolution of closteroviruses.

Keywords: Thesium chinense closterovirus 1; Closterovirus; Closteroviridae; Thesium chinense

Introduction

Closteroviridae is a family of RNA viruses that infect plants, leading to considerable yield losses in commercially important plants (beet, carrot, citrus, grape, raspberry, strawberry, tobacco, and tomato) (Agranovsky *et al*., 1994; Karasev *et al*., 1995; Tzanetakis *et al*., 2007; Tzanetakis and Martin, 2007; Menzel *et al*., 2009; Martelli *et al*., 2012; Wang *et al*., 2016; Fiallo-Olive and Navas-Castillo, 2019; Martelli, 2019; Liu *et al*., 2021). The family *Closteroviridae* comprises of four known genera (*Ampelovirus*, *Closterovirus*, *Crinivirus*, and *Velarivirus*) and a group of unassigned species (Agranovsky, 2016; Martelli, 2019; Fuchs *et al*., 2020).

Members of the family *Closteroviridae* have long filamentous virions and positive-sense, single-stranded RNA genomes that are 13–19 kilobases (kb) long (Martelli, 2019; Fuchs *et al*., 2020). They have either monopartite (*Ampelovirus*, *Closterovirus*, *Velarivirus*, and the unassigned group) or di/tripartite (*Crinivirus*) genomes, which both contain open reading frames (ORFs) for seven conserved proteins and additional proteins. For example, beet yellows virus (BYV; species *Beet yellows virus*; genus *Closterovirus*) encodes nine proteins: ORF1a polyprotein, ORF1ab fusion protein, 6 kilodalton (kDa) protein (p6), heat shock protein 70 homolog (Hsp70h), 64 kDa protein (p64; also known as heat shock protein 90 homolog or Hsp90h), minor coat protein (CPm), major coat protein (CP), 20 kDa protein (p20), and 21 kDa protein (p21) (Agranovsky *et al*., 1994; Dolja, 2003). The first seven proteins are shared

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

Abbreviations: CP = coat protein; CPm = minor coat protein; CNFV = carnation necrotic fleck virus; CTV = citrus tristeza virus; ER = endoplasmic reticulum; Hsp70h = heat shock protein 70 homolog; MV1 = mint virus 1; ORF = open reading frame; RdRp = RNA-dependent RNA polymerase; RF = ribosomal frameshifting; RSS = RNA silencing suppressor; TcCV1 = Thesium chinense closterovirus 1

among most *Closteroviridae* family members, with different molecular masses according to species (Martelli, 2019). Additional ORFs are present in many viruses that are unique to a species or group of species (Agranovsky, 2016; Fuchs *et al*., 2020; Park *et al*., 2021; Song *et al*., 2021).

ORF1a polyproteins and ORF1ab fusion proteins are translated directly from viral genomic RNA, whereas other proteins are produced using subgenomic RNAs (Dolja, 2003; Agranovsky, 2016; Fuchs *et al*., 2020; Song *et al*., 2021). The ORF1a polyprotein, which contains a papain-like leader protease, RNA methyltransferase, Zemlya, and RNA helicase domain, is proteolytically processed into at least four mature peptides. ORF1b contains an RNA-dependent RNA polymerase (RdRp) domain that is produced as an ORF1ab fusion protein via +1 ribosomal frameshifting (RF), which occurs at a low rate near the end of ORF1a. The 7-nucleotide (nt) long consensus sequences of GUU**UAG**C, GUU**UAA**C, and GUU**CGG**C, where the stop codons (UAG and UAA) and a rare arginine codon (CGG) are in bold font and nucleotides skipped during the +1 RF are underlined, have been proposed to cause the +1 RF during the translation of ORF1a (Karasev *et al*., 1995; Agranovsky, 2016; Atkins *et al*., 2016; Zheng *et al*., 2018; Park *et al*., 2021). ORF1a polyproteins and ORF1ab fusion proteins are involved in viral genome replication and the transcription of subgenomic RNAs (Dolja *et al*., 2006; Song *et al*., 2021).

Five proteins (p6, Hsp70h, p64, CPm, and CP) or their analogous counterparts are involved in the intercellular movement of viruses and assembly of virions (Dolja *et al*., 2006; Song *et al*., 2021). The p6 protein is a small integral membrane protein that mediates cell-to-cell movement of the virus and is associated with the endoplasmic reticulum (ER) (Alzhanova *et al*., 2000). Hsp70h, which may be derived from the cellular molecular chaperone HSP70 via horizontal gene transfer, is involved in viral genome replication, virion assembly, and cell-to-cell movement (Peremyslov *et al*., 2004). CP is the major capsid protein that forms a long helical body of filamentous virions and encapsidates approximately 95% of the viral genomic RNA (Dolja *et al*., 2006). CPm forms a terminal structure at one end of the virion together with Hsp70h and p64 (Dolja, 2003).

Extra ORFs encoding small proteins are located in the 3′ region or between the conserved ORFs. One of these extra ORFs is the RNA silencing suppressor (RSS) (Dolja *et al*., 2006; Song *et al*., 2021). Double-stranded RNA molecules generated during viral genome replication trigger plant defense responses that utilize RNA-silencing mechanisms (Waterhouse *et al*., 2001). Viral RSS proteins interfere with and suppress the host defense system (Wang *et al*., 2012; Flores *et al*., 2013). Proteins with RSS function have been identified in some *Closteroviridae* members, including BYV p21, citrus tristeza virus (CTV) p23, and grapevine leafroll-associated virus 2 (GLRaV2) p24 (Song *et al*., 2021).

Transcriptome data from plant tissues often contain sequences originating from RNA viruses that symptomatically or asymptomatically infect the host plant (He *et al*., 2015; Susi *et al*., 2019; Liu *et al*., 2021; Quintanilha-Peixoto *et al*., 2021). Novel viral genome sequences can be identified by sequence comparisons of assembled plant transcriptome contigs and known viral protein sequences (Orilio *et al*., 2018; Bejerman *et al*., 2020, 2021; Choi *et al*., 2021; Goh *et al*., 2021; Shin *et al*., 2021). In this study, the genome sequence of a novel positive-sense, single-stranded RNA virus belonging to the genus *Closterovirus* (family *Closteroviridae*) was identified using the transcriptome data obtained from the haustoria and roots of the parasitic plant *Thesium chinense* (Ichihashi *et al*., 2018).

Materials and Methods

Transcriptome data. Transcriptome data were previously obtained from the haustoria and roots of *T. chinense* (Ichihashi *et al*., 2018). The sequence data of 14 sequencing runs from seven *T. chinense* individuals are available in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the SRA Project Acc. No. SRP114897. High-quality reads were collected by filtering the raw reads using sickle (version 1.33; https://github.com/najoshi/sickle), with the parameter "-q 30 -l 55," and assembled into contigs using the SPAdes Genome Assembler (version 3.15.3; http://cab.spbu.ru/software/spades), with the parameter "--rnaviral" (Bushmanova *et al*., 2019).

Identification of viral genome contigs. The assembled *T. chinense* transcriptome contigs were compared with known viral RdRp domain sequences using the DIAMOND program to identify the putative viral genome contigs. Known viral RdRp sequences were extracted from 20 Pfam families (release 35.0; https://pfam.xfam.org; Pfam Acc. Nos. PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05273, PF05788, PF05919, PF06317, PF06478, PF07925, PF08467, and PF12426).

Putative ORFs of the viral genome contigs were predicted using the getorf program of the EMBOSS package (version 6.6.0.0; http://emboss.sourceforge.net), with the parameter "-reverse N -find 1 -minsize 150." The molecular weights of the predicted viral proteins were calculated using the pepstats program in the EMBOSS package. Functional domains in the viral proteins were predicted using the Pfam database sequence search tool with an E-value cutoff of 0.001. The transmembrane domains were predicted using TMHMM (version 2.0; https://services.healthtech.dtu.dk/service.php?TMHMM-2.0).

Phylogenetic analysis. NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to collect known viral genome

and protein sequences. Pairwise identities of the viral protein sequences were calculated using the needle program in the EMBOSS package. Multiple sequence alignments were visualized using BOXSHADE (version 3.31; https://launchpad.net/ubuntu/focal/+package/boxshade). Sequence logo representation was created using WebLogo (version 3; http://weblogo.threeplusone.com).

Multiple alignments of viral protein sequences were performed using MAFFT (version 7.475; https://mafft.cbrc.jp/alignment/software), with the parameter "--auto" (Nakamura *et al.*, 2018). Poorly aligned regions from the multiple sequence alignments were removed using trimAl (version 1.4.rev22; http://trimal.cgenomics.org), with the parameter "-automated1" (Capella-Gutierrez *et al.*, 2009). Maximum likelihood phylogenetic trees were inferred using IQ-TREE (version 2.1.3; http://www.iqtree.org), with the parameter "-B 1000" (Minh *et al.*, 2020), and visualized using MEGA software (version 10.2.4; https://www.megasoftware.net) (Kumar *et al.*, 2018).

Results and Discussion

Genome sequence of Thesium chinense closterovirus 1 (TcCV1)

T. chinense RNA-sequencing reads previously obtained from the haustoria and roots of seven individual plants were assembled into transcriptome contigs (Ichihashi *et al.*, 2018). Sequence similarity searches of the transcriptome contigs against the known viral RdRp sequences revealed two putative viral genome contigs with the lengths of 15293 nt and 15289 nt. These contigs, named C1 and C2, respectively, shared 94.0% nucleotide identity with each other, indicating that they are two variants of the same virus. The C1 and C2 contigs showed the highest amino acid (aa) sequence identity with the RdRp domain sequence of mint virus 1 (MV1), which was previously identified in mint (*Mentha* spp.) (Tzanetakis *et al.*, 2005). MV1 is a member of the genus *Closterovirus*, supporting the notion that the C1 and C2 contigs are genome sequences of viruses closely related to the genus *Closterovirus*.

Sequence similarity searches of the NCBI protein database using C1 and C2 contigs as queries indicated that they had the highest sequence similarities (approximately 60–70% aa identity) with viruses of the genus *Closterovirus*, including MV1, Rehmannia virus 1 (ReV-1), tobacco virus 1 (TV1), carnation necrotic fleck virus (CNFV), carrot closterovirus 1 (CtCV1), and carrot yellow leaf virus (CYLV) (Karasev *et al.*, 1994; Tzanetakis *et al.*, 2005; Menzel *et al.*, 2009; Adams *et al.*, 2014; Wang *et al.*, 2016; Kwon *et al.*, 2018). Therefore, contigs C1 and C2 were considered to be genome sequences of a putative novel closterovirus (the genus *Closterovirus*; the family *Closteroviridae*) and named Thesium chinense closterovirus 1 (TcCV1). The TcCV1 genome sequences were deposited in NCBI under Acc. Nos. OM801605 (C1) and OM801606 (C2). The C1 and C2 contig sequences share 94.0% nucleotide identity with each other and encode an identical set of proteins exhibiting 94.1–99.5% aa identities with their respective orthologs. The longer

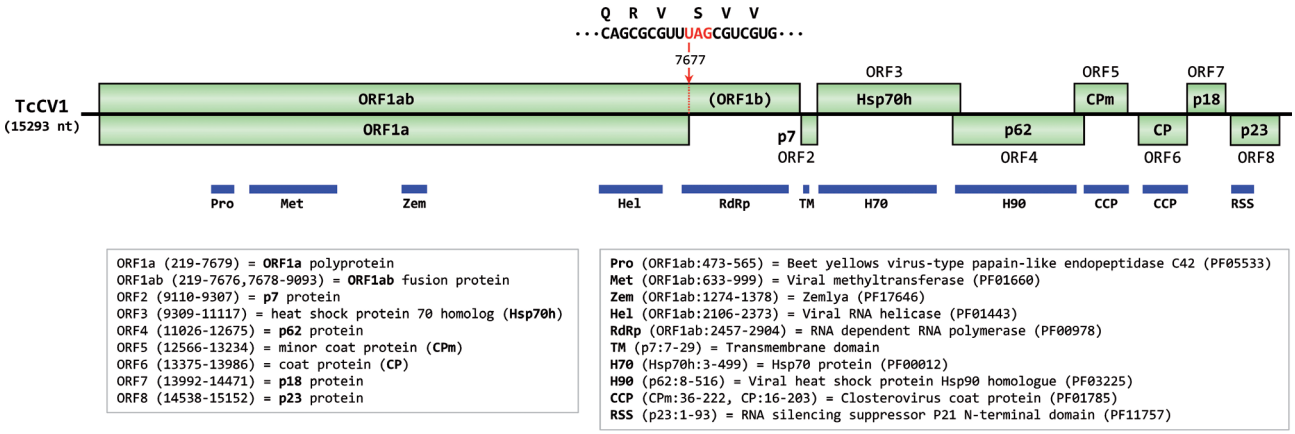

Fig. 1

Genomic structure of Thesium chinense closterovirus 1 (TcCV1)

Schematic representation of the TcCV1 genome (C1 contig). The predicted open reading frames (ORFs) are shown as boxes with ORF numbers and encoded protein names. The coordinates of the ORFs are listed in the box at the bottom left. The predicted functional domains are indicated by the blue lines below the ORFs. Coordinates and Pfam names (with Acc. Nos.) are presented in the box at the bottom right. Nucleotide sequences surrounding the stop codon (red) of ORF1a are shown above the genome, with the corresponding amino acid sequence of the ORF1ab fusion protein at the top. The +1 ribosomal frameshifting by skipping the uridine (U) residue at genome position 7677 produces the ORF1ab fusion protein.

contig, C1, was chosen to represent the TcCV1 genome sequence for further analysis.

The TcCV1 genome sequence (the C1 contig with 15293 nt length) was predicted to encode nine ORFs: ORF1a, ORF1b, and ORF2 to ORF8 (Fig. 1). The ORF1a polyprotein is 2486 aa long and was predicted to have a papain-like leader protease, RNA methylase, Zemlya, and RNA helicase domain. The Zemlya domain is a recently described functional domain identified only in the genus *Closterovirus* (Gushchin *et al.*, 2017). No homologs of the Zemlya domain have been found in proteins of other genera of the family *Closteroviridae*, other viruses, or cellular organisms. The Zemlya domain was proposed to mediate remodeling of ER membranes to form viral replication factories during closterovirus infection (Gushchin *et al.*, 2017). The presence of the Zemlya domain further supports the inclusion of TcCV1 in the genus *Closterovirus*.

TcCV1 ORF1b encodes an RdRp domain that is produced at a low rate, presumably via the +1 RF during translation of ORF1a (Atkins *et al.*, 2016; Martelli, 2019). A tentative +1 RF sequence predicted in the TcCV1 genome is GUU**UAG**C with the UAG stop codon of ORF1a and the skipped U nucleotide. The +1 RF may occur by skipping the U residue, the first nucleotide of the stop codon, at the nt position 7677 (C1 contig), which causes the translational machinery to continue to translate ORF1b. The predicted ORF1ab fusion protein is 2957 aa long.

TcCV1 ORF2 encodes a small protein with a length of 65 aa and a calculated molecular weight of 7 kDa, which is hereafter referred to as p7. TcCV1 p7 showed sequence similarities with conserved p6-like proteins of other closteroviruses. Closterovirus p6-like proteins are associated with the ER membrane and are involved in cell-to-cell movement of the virus (Alzhanova *et al.*, 2000). TcCV1 p7 was predicted to have a transmembrane domain in the N-terminal half (aa positions 7–29), indicating that it may have the same function as other p6-like proteins.

TcCV1 ORF3 encodes a 602 aa protein that contains a viral Hsp70 domain and shows sequence similarity with Hsp70h proteins of members of the family *Closteroviridae*. The viral Hsp70h protein may be transferred to an ancestral *Closteroviridae* virus from the cellular molecular chaperon HSP70 (Satyanarayana *et al.*, 2000; Dolja *et al.*, 2006). Hsp70h plays multiple roles during the viral life cycle including genome replication, virion formation, and cell-to-cell movement (Peremyslov *et al.*, 2004).

The ORF4 sequence of TcCV1 encodes a 549 aa protein with a calculated molecular mass of 62 kDa. This protein, named p62, showed sequence similarities with p64-like proteins of approximately 60 kDa that are seen in members of the family *Closteroviridae*. TcCV1 p62 and orthologous p64-like proteins share a viral Hsp90h domain and are also known as Hsp90h proteins (Song *et al.*, 2021).

ORF5 and ORF6 of TcCV1 encode 222 aa and 203 aa proteins that are orthologous to closterovirus CPm and CP, respectively. CPm and CP share a closterovirus-specific coat protein domain and may have originated from the duplication of an ancestral gene in the common ancestor of the family *Closteroviridae* (Dolja *et al.*, 2006). CP encapsidates most of the viral genomic RNA, whereas CPm together with Hsp70h and p64-like proteins form the terminal structure at one end of the virion (Dolja, 2003; Dolja *et al.*, 2006).

ORF7 predicted in the TcCV1 genome sequence may encode a 159 aa protein, named p18, with a calculated molecular weight of 18 kDa. However, no homologous protein sequences were detected in the current NCBI sequence databases or in the genomes of known closteroviruses. Therefore, the hypothetical protein p18 is specific to TcCV1. ORFs encoding proteins with unknown functions are commonly found in many closteroviruses and are specific to a virus or shared only among closely related viruses (Agranovsky, 2016; Park *et al.*, 2021; Song *et al.*, 2021).

ORF8, the last ORF of TcCV1, encodes a 23 kDa (204 aa) protein named p23. TcCV1 p23 exhibits sequence similarities with p21-like proteins found in many closteroviruses, including BYV p21, CTV p23, and GLRaV2 p24. TcCV1 p23 and its orthologous proteins have an RSS domain that is essential for interference with and suppression of the host RNA-silencing mechanism (Dolja, 2003; Wang *et al.*, 2012; Flores *et al.*, 2013).

Analysis of the +1 RF sequences of closterovirus genomes

The closterovirus RdRp proteins encoded by ORF1b are produced via the +1 RF during the translation of ORF1a (Atkins *et al.*, 2016; Martelli, 2019; Park *et al.*, 2021). The +1 RF was first proposed to occur at the ORF1a stop codon or at a rare codon preceded by two uridines (UU) (Agranovsky *et al.*, 1994; Karasev *et al.*, 1995). The most frequently observed motif is GUU_stop_C (underscores mark codon boundaries of ORF1a and 'stop' indicates a stop codon), whereas CTV has a rare arginine codon (CGG) instead of a stop codon (Karasev *et al.*, 1995; Park *et al.*, 2021).

To determine whether there were additional conserved sequences at the +1 RF sites, the amino acid and nucleotide sequences of 22 closteroviruses were comparatively analyzed (Fig. 2). First, multiple alignments of the C-termini of ORF1a polyprotein sequences and N-termini of ORF1b protein sequences were conducted (Fig. 2a). For ORF1b, the N-terminus was extended to the first sense codon. The +1 RF must occur within the overlapping region of ORF1a and the extended ORF1b, which is a +1 reading frame relative to ORF1a.

The positions of the ORF1a stop codons were conserved in 19 of 22 viruses. Three viruses, CTV, Elephantopus scaber closterovirus (ESCV), and rose leaf rosette-associated virus (RLRaV), have an extended C-terminus of approximately 20 aa compared to the other nineteen viruses. A highly conserved arginine-valine (RV) dipeptide was observed at the end of ORF1a in all examined viruses except CNFV, which has an arginine-alanine dipeptide, suggesting that the dipeptide sequence is important and must be included in the ORF1ab fusion protein. Therefore, it is highly likely that the +1 RF may occur after the RV dipeptide position in all members of the genus *Closterovirus*, regardless of the presence of a stop codon after the RV dipeptide.

The extended N-termini of ORF1b ORFs of 22 closteroviruses varied in length and did not contain any notably conserved sequences (Fig. 2a). However, the ORF1b sequence after the proposed +1 RF site displayed very high sequence conservation, which further supported the notion that all known closteroviruses may use the same +1 RF site for the production of the ORF1ab fusion protein. When the TcCV1 ORF1ab fusion protein sequence deduced using the proposed +1 RF site was analyzed, the RdRp domain was predicted to be at aa positions 2457–2904, which spans across the ORF1ab boundary at position 2486–2487. The RdRp domains of all ORF1ab fusion proteins predicted from known closteroviruses span their predicted +1 RF sites.

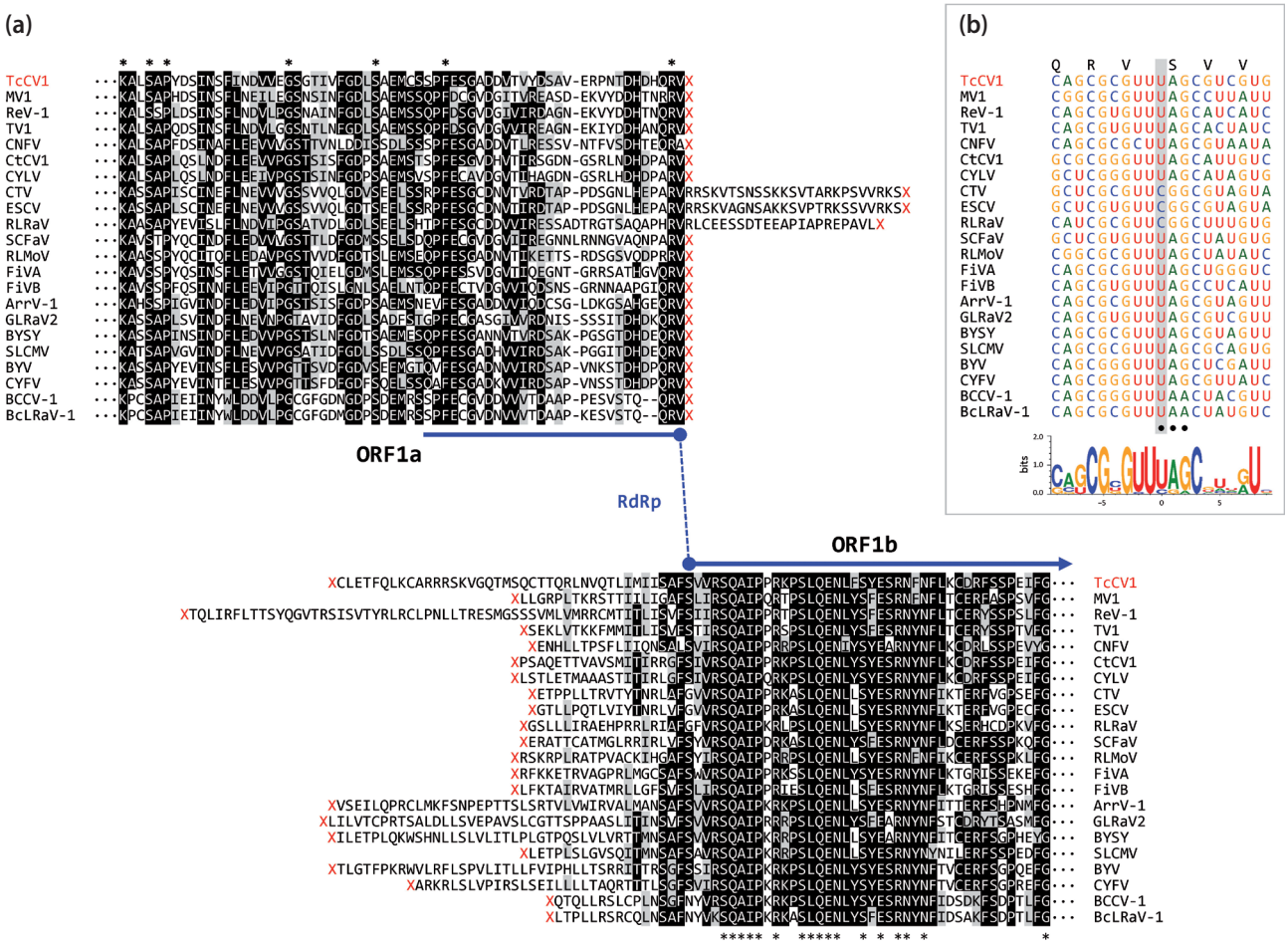

Fig. 2

Sequence comparisons of closterovirus +1 ribosomal frameshifting (RF) sites

(a) Multiple alignments of ORF1a C-terminal (top) and ORF1b N-terminal (bottom) sequences of 22 closteroviruses are shown. ORF1b N-terminal sequences were extended up to the first in-frame stop codon. Stop codons are shown as red Xs. Amino acid residues that are identical or similar in ≥ 50% of the viruses are highlighted in black or gray, respectively. Asterisks above or below the alignments indicate residues that are identical in all viruses. Predicted RdRp domain is marked by blue lines. The dotted blue line denotes the +1 RF that produces ORF1ab fusion proteins. **(b)** Nucleotide sequences at the proposed +1 RF sites of 22 closteroviruses are presented. Stop codons of ORF1a are marked by the three dots (note that CTV, ESCV, and RLRaV have a CGG codon at this position). Nucleotide residues that are presumably skipped by +1 RF are shaded in gray and marked as position 0. The corresponding TcCV1 ORF1ab fusion protein sequence is shown at the top. Sequence logo representation of the alignment is presented at the bottom.

Table 1. Sequence identities of the RdRp and Hsp70h proteins of TcCV1 and related viruses

No.	Genus	Virus	Acronym	Genome[a]	RdRp[b]	Hsp70h[b]
1	*Closterovirus*	Mint virus 1	MV1	NC_006944	68.2% (322/472)	50.1% (305/609)
2		Rehmannia virus 1	ReV-1	NC_040572	66.2% (312/471)	50.0% (307/614)
3		Tobacco virus 1	TV1	NC_027712	65.9% (313/475)	49.5% (304/614)
4		Carnation necrotic fleck virus	CNFV	NC_038419	64.8% (307/474)	45.1% (274/607)
5		Carrot closterovirus 1	CtCV1	KF533697	65.1% (308/473)	52.8% (328/621)
6		Carrot yellow leaf virus	CYLV	NC_013007	62.5% (295/472)	55.1% (340/617)
7		Citrus tristeza virus	CTV	NC_001661	54.5% (274/503)	41.0% (254/619)
8		Elephantopus scaber closterovirus	ESCV	MN814306	55.7% (268/481)	42.3% (262/620)
9		Rose leaf rosette-associated virus	RLRaV	NC_024906	52.1% (256/491)	41.0% (250/610)
10		Strawberry chlorotic fleck-associated virus	SCFaV	NC_008366	57.7% (277/480)	39.2% (240/612)
11		Raspberry leaf mottle virus	RLMoV	NC_008585	55.7% (264/474)	42.4% (263/620)
12		Fig virus a	FiVA	MN817232	44.1% (241/546)	33.7% (211/626)
13		Fig virus b	FiVB	MN817233	45.5% (245/538)	32.0% (196/613)
14		Arracacha virus 1	ArrV-1	NC_040570	56.0% (265/473)	39.8% (243/611)
15		Grapevine leafroll-associated virus 2	GLRaV2	NC_007448	56.1% (267/476)	42.4% (257/606)
16		Beet yellow stunt virus	BYSY	NC_043106	54.8% (259/473)	45.9% (283/616)
17		Soybean leaf crinkle mottle virus	SLCMV	LC601607	56.4% (266/472)	47.0% (287/610)
18		Beet yellows virus	BYV	NC_001598	56.4% (268/475)	46.0% (281/611)
19		Carnation yellow fleck virus	CYFV	NC_022978	55.6% (262/471)	46.1% (281/609)
20		Blackcurrant closterovirus 1	BCCV-1	NC_040834	51.5% (250/485)	38.8% (242/623)
21		Blackcurrant leafroll-associated virus 1	BcLRaV-1	NC_040722	52.0% (251/483)	38.8% (242/624)
22	Unassigned	Persimmon virus b	PeVB	NC_025967	33.9% (177/522)	30.3% (197/651)
23		Actinidia virus 1	AcV1	NC_035453	34.5% (183/530)	29.8% (197/662)
24		Olive leaf yellowing-associated virus	OLYaV	MT809205	40.0% (202/505)	29.7% (195/657)
25		Blueberry virus a	BVA	NC_018519	39.5% (218/552)	26.4% (179/678)
26	*Ampelovirus*	Little cherry virus 2	LChV2	NC_005065	32.5% (156/480)	26.5% (169/638)
27		Grapevine leafroll-associated virus 1	GLRaV1	NC_016509	34.2% (162/474)	28.3% (182/643)
28		Grapevine leafroll-associated virus 4	GLRaV4	NC_016416	32.5% (158/486)	25.8% (164/635)
29		Plum bark necrosis stem pitting-associated virus	PBNSPaV	NC_009992	31.2% (148/474)	27.0% (172/638)
30	*Velarivirus*	Areca palm velarivirus 1	ArPV1	NC_027121	28.0% (144/515)	23.8% (161/677)
31		Little cherry virus 1	LCV1	NC_001836	30.0% (145/484)	25.7% (177/690)
32		Cordyline virus 1	CV1	NC_038421	27.9% (135/484)	25.5% (164/642)
33		Cordyline virus 3	CV3	NC_043107	29.0% (144/497)	27.0% (170/630)
34	*Crinivirus*	Blackberry yellow vein-associated virus	BYVaV	NC_006962, NC_006963	29.5% (146/495)	27.9% (180/645)
35		Cucurbit yellow stunting disorder virus	CYSDV	NC_004809, NC_004810	27.8% (133/479)	27.1% (170/628)
36		Sweet potato chlorotic stunt virus	SPCSV	NC_004123, NC_004124	30.0% (146/486)	27.0% (174/644)
37		Lettuce infectious yellows virus	LIYV	NC_003617, NC_003618	29.0% (141/487)	27.5% (176/641)

[a]NCBI Acc. Nos. for genome sequences (note that criniviruses have two genomic segments); [b]amino acid sequence identities to the TcCV1 RdRp and Hsp70h proteins in the format of "percent identity (identical residues/aligned length)".

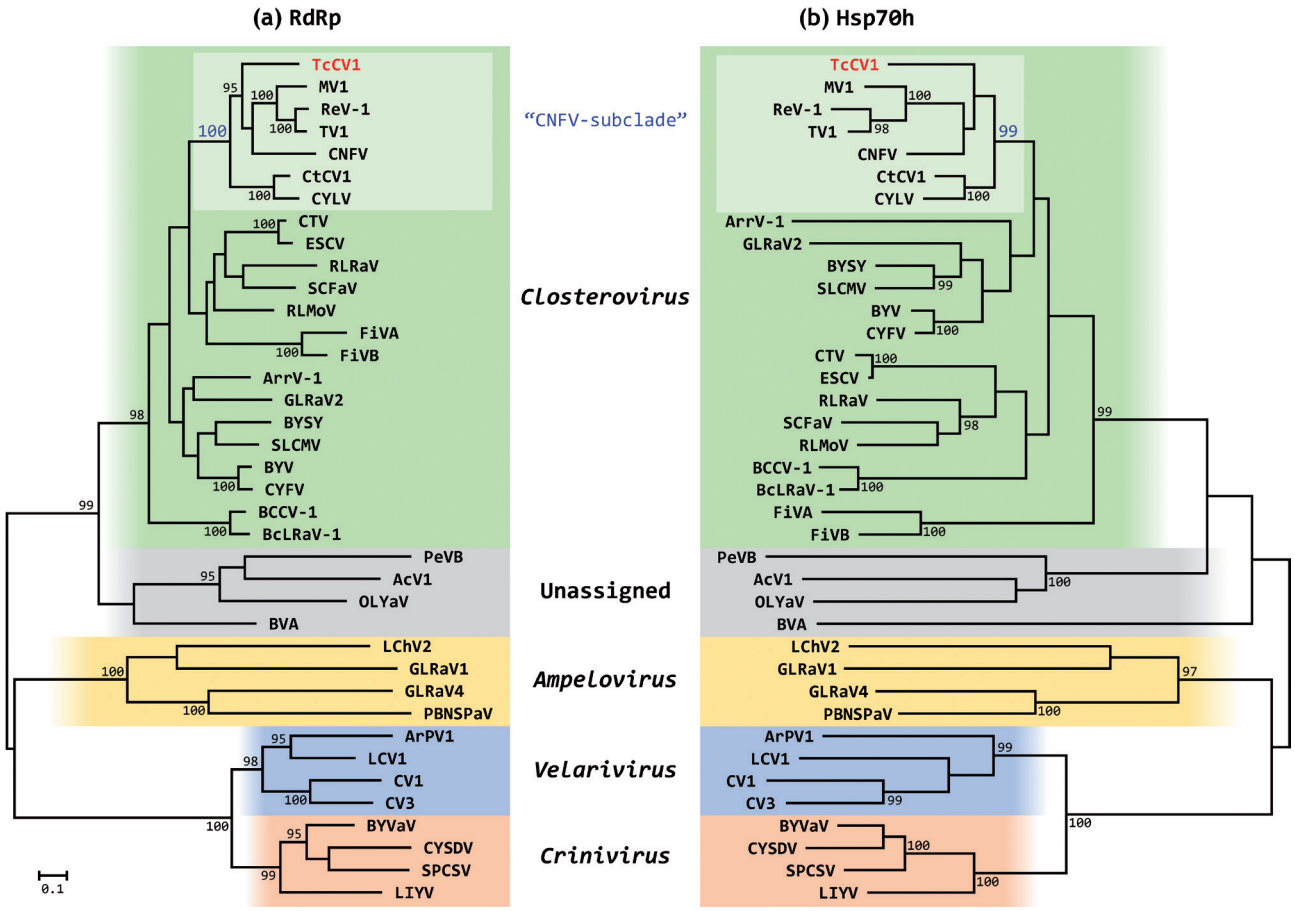

Fig. 3

Phylogenetic relationships of TcCV1 and related viruses

Maximum likelihood phylogenetic trees were inferred based on the RdRp **(a)** and Hsp70h **(b)** protein sequences. TcCV1 forms a subclade named the "CNFV-subclade" (light green box) with six previously known viruses within the genus *Closterovirus* (green background). Bootstrap branch support values of 95% or higher, calculated from 1000 replicates, are shown at the nodes. See Table 1 for detailed information of these viruses.

Nucleotide sequences surrounding the putative +1 RF site collected from 22 closteroviruses were analyzed (Fig. 2b). Four different sequences were observed at the 7 nt long +1 RF site: GUU**UAG**C (16 viruses), GUU**CGG**C (3 viruses), GUU**UAA**C (2 viruses), and GCU**UAG**C (1 virus), where the skipped residues are underlined. In addition to the 7 nt positions, three positions were found to be conserved. When the skipped residue position was set to 0, the C, G, and U nucleotides at positions −6, −5, and +8, respectively, were identical in all 22 closteroviruses. The CG dinucleotide at positions −6 and −5 are the first two nucleotides of the CGN (where N represents A, C, G, or U) codon for arginine, which is conserved in all ORF1ab fusion proteins. The U nucleotide at position +8 is the second nucleotide of the codons for valine (GUN) or isoleucine (AUH; H represents A, C, or U), which are similar amino acids found at this position in all ORF1ab fusion proteins. Therefore, it is not clear whether these sequences are conserved in the +1 RF site or encoded amino acid.

Phylogenic position of TcCV1 among closteroviruses

The phylogenetic position of TcCV1 within the genus *Closterovirus* was inferred using the RdRp (ORF1b) and Hsp70h (ORF3) sequences. A total of 37 known viruses of the family *Closteroviridae* (21 from the genus *Closterovirus* and 4 representatives from each of the three other genera and the unassigned group) were collected (Table 1). The RdRp and Hsp70h protein sequences of the known viruses were compared with those of TcCV1. The TcCV1 RdRp and Hsp70h showed 27.8–68.2% and 23.8–55.1% aa identity, respectively, to the orthologous proteins of known viruses.

Multiple alignments were separately generated from the RdRp and Hsp70h protein sequences and filtered to

remove phylogenetically uninformative sites. Maximum likelihood phylogenetic trees constructed with 1000 bootstrap replicates for both RdRp and Hsp70h showed that TcCV1 is a novel member of the genus *Closterovirus* (family *Closteroviridae*) (Fig. 3). In both trees, TcCV1 formed a strong subclade (bootstrap support values of 100% and 99%, respectively) tentatively referred here as the "CNFV-subclade" after CNFV, which was identified first in the group. The "CNFV-subclade" includes TcCV1, MV1, ReV-1, TV1, CNFV, CtCV1, and CYLV. Branching patterns of the "CNFV-subclade" members showed strong concordance between the RdRp and Hsp70h phylogenetic trees, indicating that the evolutionary history of this group can be unambiguously resolved. However, other closteroviruses showed largely discordant branching patterns between RdRp and Hsp70h phylogenetic trees, suggesting the possibility of complex genomic evolutionary processes involving genomic recombination or lineage-specific evolutionary rates.

Conclusion

The genome sequence of a novel closterovirus (genus *Closterovirus*), TcCV1, was identified in transcriptome data obtained from the haustoria and root tissues of *T. chinense*. The TcCV1 genome encodes nine proteins, eight of which have orthologs in closely related closteroviruses. Comparisons of amino acid and nucleotide sequences of the putative +1 RF sites required for ORF1ab fusion protein production implied that the +1 RF occurs at a conserved location in all closteroviruses. Phylogenetic analyses showed that TcCV1 is a novel closterovirus that forms a strong subclade with a group of previously known closteroviruses. The genome sequence of TcCV1 may be useful for studying the evolution of closterovirus genome organization.

Acknowledgment. This work was supported by National Research Foundation of Korea (NRF) grants funded by the Government of Korea (grant numbers 2018R1A5A1025077 and 2020R1A2C1013403).

References

Adams IP, Skelton A, Macarthur R, Hodges T, Hinds H, Flint L, Nath PD, Boonham N, Fox A (2014): Carrot yellow leaf virus is associated with carrot internal necrosis. PLoS One 9, e109125. https://doi.org/10.1371/journal.pone.0109125

Agranovsky AA (2016): Closteroviruses: molecular biology, evolution and interactions with cells. In Gaur RK, Petrov NM, Patil BL, and Stoyanova MI (Eds): Plant Viruses: Evolution and Management. Springer Singapore, Singapore, pp. 231–252. https://doi.org/10.1007/978-981-10-1406-2_14

Agranovsky AA, Koonin EV, Boyko VP, Maiss E, Frotschl R, Lunina NA, Atabekov JG (1994): Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. Virology 198, 311–24. https://doi.org/10.1006/viro.1994.1034

Alzhanova DV, Hagiwara Y, Peremyslov VV, Dolja VV (2000): Genetic analysis of the cell-to-cell movement of beet yellows closterovirus. Virology 268, 192–200. https://doi.org/10.1006/viro.1999.0155

Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV (2016): Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. Nucleic Acids Res. 44, 7007–7078. https://doi.org/10.1093/nar/gkw530

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

Bejerman N, Dietzgen RG, Debat H (2021): Illuminating the plant rhabdovirus landscape through metatranscriptomics data. Viruses 13, 1304. https://doi.org/10.3390/v13071304

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019): rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience 8. 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

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

Dolja VV (2003): Beet yellows virus: the importance of being different. Mol. Plant Pathol. 4, 91–98. https://doi.org/10.1046/j.1364-3703.2003.00154.x

Dolja VV, Kreuze JF, Valkonen JP (2006): Comparative and functional genomics of closteroviruses. Virus Res. 117, 38–51. https://doi.org/10.1016/j.virusres.2006.02.002

Fiallo-Olive E, Navas-Castillo J (2019): Tomato chlorosis virus, an emergent plant virus still expanding its geographical and host ranges. Mol. Plant Pathol. 20, 1307–1320. https://doi.org/10.1111/mpp.12847

Flores R, Ruiz-Ruiz S, Soler N, Sanchez-Navarro J, Fagoaga C, Lopez C, Navarro L, Moreno P, Pena L (2013): Citrus tristeza virus p23: a unique protein mediating key virus-host interactions. Front. Microbiol. 4, 98. https://doi.org/10.3389/fmicb.2013.00098

Fuchs M, Bar-Joseph M, Candresse T, Maree HJ, Martelli GP, Melzer MJ, Menzel W, Minafra A, Sabanadzovic S, ICTV Report Consortium (2020): ICTV virus taxonomy profile: Closteroviridae. J. Gen. Virol. 101, 364–365. https://doi.org/10.1099/jgv.0.001397

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

Gushchin VA, Karlin DG, Makhotenko AV, Khromov AV, Erokhina TN, Solovyev AG, Morozov SY, Agranovsky AA (2017): A conserved region in the closterovirus 1a polyprotein drives extensive remodeling of endoplasmic reticulum membranes and induces motile globules in Nicotiana benthamiana cells. Virology 502, 106–113. https://doi.org/10.1016/j.virol.2016.12.006

He Y, Yang Z, Hong N, Wang G, Ning G, Xu W (2015): Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. Mol. Plant Pathol. 16, 449–458. https://doi.org/10.1111/mpp.12202

Ichihashi Y, Kusano M, Kobayashi M, Suetsugu K, Yoshida S, Wakatake T, Kumaishi K, Shibata A, Saito K, Shirasu K (2018): Transcriptomic and metabolomic reprogramming from roots to haustoria in the parasitic plant, Thesium chinense. Plant Cell Physiol. 59, 729–738. https://doi.org/10.1093/pcp/pcx200

Karasev AV, Boyko VP, Gowda S, Nikolaeva OV, Hilf ME, Koonin EV, Niblett CL, Cline K, Gumpf DJ, Lee RF et al. (1995): Complete sequence of the citrus tristeza virus RNA genome. Virology 208, 511–20. https://doi.org/10.1006/viro.1995.1182

Karasev AV, Nikolaeva OV, Koonin EV, Gumpf DJ, Garnsey SM (1994): Screening of the closterovirus genome by degenerate primer-mediated polymerase chain reaction. J. Gen. Virol. 75, 1415–1422. https://doi.org/10.1099/0022-1317-75-6-1415

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018): MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kwon SJ, Jin M, Cho IS, Yoon JY, Choi GS (2018): Identification of rehmannia virus 1, a novel putative member of the genus Closterovirus, from Rehmannia glutinosa. Arch. Virol. 163, 3383–3388. https://doi.org/10.1007/s00705-018-4014-4

Liu Q, Zhang S, Mei S, Zhou Y, Wang J, Han GZ, Chen L, Zhou C, Cao M (2021): Viromics unveils extraordinary genetic diversity of the family Closteroviridae in wild citrus. PLoS Pathog. 17, e1009751. https://doi.org/10.1371/journal.ppat.1009751

Martelli GP (2019): A brief historical account of the family Closteroviridae. Methods Mol. Biol. 2015, 7–13. https://doi.org/10.1007/978-1-4939-9558-5_2

Martelli GP, Aboughanem N, Agranovsky A, Al Rwahnih M, Dolja VV, Dovas C, Fuchs M, Gugerli P, Hu JS, Jelkmann W, Katis N, Maliogka V, Melzer MJ, Menzel W, Minafra A, Rott M, Rowhani A, Sabanadzovic S, Saldarelli P (2012): Taxonomic revision of the family Closteroviridae with special reference to the grapevine leafroll-associated members of the genus Ampleovirus and the putative species unassigned to the family. J. Plant Pathol. 94, 7–19.

Menzel W, Goetz R, Lesemann DE, Vetten HJ (2009): Molecular characterization of a closterovirus from carrot and its identification as a German isolate of carrot yellow leaf virus. Arch. Virol. 154, 1343–1347. https://doi.org/10.1007/s00705-009-0428-3

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

Orilio AF, Blawid R, Costa GA, Gomes SSVSF, Nagata T, Madeira NR, Inoue-Nagata AK, Resende RO (2018): High-throughput sequencing reveals a novel closterovirus in arracacha (Arracacia xanthorrhiza). Arch. Virol. 163, 2547–2550.

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/10.1007/s00705-018-3863-1

Peremyslov VV, Andreev IA, Prokhnevsky AI, Duncan GH, Taliansky ME, Dolja VV (2004): Complex molecular architecture of beet yellows virus particles. Proc. Natl. Acad. Sci. USA 101, 5030–5035. https://doi.org/10.1073/pnas.0400303101

Quintanilha-Peixoto G, Fonseca PLC, Raya FT, Marone MP, Bortolini DE, Mieczkowski P, Olmo RP, Carazzolle MF, Voigt CA, Soares ACF, Pereira GAG, Goes-Neto A, Aguiar E (2021): The sisal virome: uncovering the viral diversity of Agave varieties reveals new and organ-specific viruses. Microorganisms 9, 1704. https://doi.org/10.3390/microorganisms9081704

Satyanarayana T, Gowda S, Mawassi M, Albiach-Marti MR, Ayllon MA, Robertson C, Garnsey SM, Dawson WO (2000): Closterovirus encoded HSP70 homolog and p61 in addition to both coat proteins function in efficient virion assembly. Virology 278, 253–265. https://doi.org/10.1006/viro.2000.0638

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

Song Y, Hanner RH, Meng B (2021): Probing into the effects of grapevine leafroll-associated viruses on the physiology, fruit quality and gene expression of grapes. Viruses 13, 593. https://doi.org/10.3390/v13040593

Susi H, Filloux D, Frilander MJ, Roumagnac P, Laine AL (2019): Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. PeerJ 7, e6140. https://doi.org/10.7717/peerj.6140

Tzanetakis IE, Halgren A, Mosier N, Martin RR (2007): Identification and characterization of raspberry mottle virus, a novel member of the Closteroviridae. Virus Res. 127, 26–33. https://doi.org/10.1016/j.virusres.2007.03.010

Tzanetakis IE, Martin RR (2007): Strawberry chlorotic fleck: identification and characterization of a novel Closterovirus associated with the disease. Virus Res. 124, 88–94. https://doi.org/10.1016/j.virusres.2006.10.005

Tzanetakis IE, Postman JD, Martin RR (2005): Characterization of a novel member of the family Closteroviridae from Mentha spp. Phytopathology 95, 1043–8. https://doi.org/10.1094/PHYTO-95-1043

Wang F, Qi S, Gao Z, Akinyemi IA, Xu D, Zhou B (2016): Complete genome sequence of tobacco virus 1, a closterovirus from Nicotiana tabacum. Arch. Virol. 161, 1087–1090. https://doi.org/10.1007/s00705-015-2739-x

Wang MB, Masuta C, Smith NA, Shimura H (2012): RNA silencing and plant viral diseases. Mol. Plant-Microbe Interact. 25, 1275–1285. https://doi.org/10.1094/MPMI-04-12-0093-CR

Waterhouse PM, Wang MB, Lough T (2001): Gene silencing as an adaptive defence against viruses. Nature 411, 834–842. https://doi.org/10.1038/35081168

Zheng L, Wu L, Postman J, Liu H, Li R (2018): Molecular characterization and detection of a new closterovirus identified from blackcurrant by high-throughput sequencing. Virus Genes 54, 828–832. https://doi.org/10.1007/s11262-018-1598-4