Complete genome sequence of a new isolate of *Caper latent virus* in caper

Antonio Tiberini[1,4*], Ignazio Fontana[2], Francesco Mercati[2], Ian Adams[3], Adrian Fox[3], Giuliana Albanese[1] and Laura Tomassoli[4]

[1]Università degli Studi Mediterranea di Reggio Calabria, Dipartimento di AGRARIA, Località Feo di Vito - 89122 Reggio Calabria, Italy. [2]Consiglio Nazionale delle Ricerche, Istituto di Bioscienze e BioRisorse (IBBR) - U.O.S. di Palermo Corso Calatafimi 414, 90129 Palermo, Italy. [3] Fera Science Ltd, Sand Hutton, York, UK, YO41 1LZ; [4]CREA Centro di Ricerca Difesa e Certificazione, Via C.G. Bertero 22 - 00156 Roma, Italy. E-mail: antonio.tiberini@unirc.it, ORCID: 0000-0001-9910-7809

Annotated sequence record

ABSTRACT

A carlavirus, isolated from asymptomatic wild *Capparis spinosa* L. plants in Sicily, was full genome sequenced through a high-throughput sequencing (HTS). Only partial genome sequences related to *Caper latent virus* (genus *Carlavirus*) were available. The genomic sequence was found to be 8,280 nt in length, excluding poly(A) tail, containing five putative Open Reading Frames (ORFs) out the six reported in *Carlavirus* genus, lacking of ORF 6. Molecular characterization was assessed through ORFs phylogenetic comparing, showing a clear close relationship to caper latent virus (CapLV) in ORF 1 and ORF 5 with a nucleotide identity ranging from 87% and 90% in RdRp and CP region respectively. According to molecular criteria for species demarcation, the present isolate (named CapLV-W) could be considered as a variant of CapLV.On the contrary, in ORF 2, ORF 3 and ORF 4 the line of demarcation between carlavirus and foveavirus genera is unclear. Therefore, further studies are needed to clarify the CapLV's taxonomy.

Carlaviruses are plant viruses belonging to the family *Betaflexiviridae*, sharing a distinct lineage of alphavirus-like replication proteins with flexuous filamentous particles 610–700 x 12–15 nm in size and transmitted by aphids in a non-persistent manner. Genome is a monopartite, with a positive-sense, single-stranded RNA of 7.4–9 kb with a 30-poly(A) tail, containing six open reading frames (ORFs) (Adams et al. (2012). A carlavirus member, Caper latent virus (CapLV), was identified for the first time in 1987 in asymptomatic leaves of caper plants (*Capperis spinosa* L.) in Apulia (Italy) on basis of particle morphology, protein and RNA composition and its serological relationship to Helenium virus S (HVS), (Di Franco and Gallitelli, 1987).

In 2006, CapLV was detected by molecular assay in caper plants without any apparent symptoms but with a general decline of plantations, in most of Sicilian small archipelagos: Aeolean islands (Lipari and Salina), Pelagie islands (Pnatelleria and Linosa)), and Ustica island (Tomassoli et al., 2006). The first molecular characterization was achieved in 2007 during a study including newly samples from the Mediterranean basin (Tomassoli et al., 2006). The sequences of two genome regions, i.e. helication/replication related proteins (ORF1 - AEF12642.1) and coat protein (CP) (ORF5 - AEF12643.1) were obtained respectively (Tiberini et al., 2011). To date only preliminary studies and limited genomics information are available despite CapLV infection still represents a serious problem for high-quality and long-productive life of caper plantations.

Recent advent of High-Throughput Sequencing (HTS) technologies allowed generating extensive data in a very cost-effective way, drastically improving research on viral pathogens (Massart et al. 2014). In the present study asymptomatic wild *C. spinosa* L. samples were collected in mainland Sicily in 2016. Collected plants had tested positive for the presence of CapLV using specific RT-PCR (Tiberini et al., 2011) and were subsequently analysed by HTS. To enrich viral RNA, samples were partially purified according to Tiberini et al. (2011) and RNA was extracted from the partial purified solution using the RNeasy Plant Mini Kit (Qiagen, Hiden, Germany). The enriched total RNA was used as template for library preparation using the TruSeq kit (Illumina, SanDiego, CA, USA), following manufacture's procedures.

Raw sequences were subjected to adapter clipping and quality trimming using Trimmomatic version 0.33 (Bolger et al. 2014) removing adapter sequence contamination and low-quality nucleotides (PHRED < 20). *De novo* assembly of cleaned reads was carried out in Trinity (v2.0.6) (Grabherr et al. 2011) with default parameters. All contigs were analysed by BLASTn and BLASTx and inferred to be related to carlaviruses. To confirm the scaffold assembly, and to exclude artefact and chimeras, a rapid amplification of cDNA ends (RACE) (Roche, Mannheim, Germany), with poly(a) tail primer, following producer's instruction, and additional amplification using internal primer sets (Supplementary table S1) were performed. Obtained amplicons were sequenced by Sanger method (Sanger et al., 1977). The genomic sequence of the target virus was found to be 8,280 nt in length, excluding poly(A) tail and deposited on GenBank (Acc. No MT311966).

Using BLAST and ORF Finder tool (http://www.ncbi.nlm.nih.gov/orffinder/) five ORFs were predicted (Table 1). The homology of isolated sequence with available ORFs of CapLV (ORF1 and ORF5) indicated that the inferred virus was an isolate of CapLV, with a nucleotide identity ranging from 87% and 90% in RdRp and CP region respectively (Supplementary material: TableS2). According

to demarcation criteria for members of the family *Betaflexiviridae* (Adams et al., 2012) (Table S1), the assembled sequence was considered as a divergent isolate CapLV (proposed name CapLV-W). A whole-genome-based, alignment-free, and parameter-free method was carried out using CVTree software (Zuo and Hao, 2015), comparing the assembled genome to the available whole genomes of some species belonging to the *Betaflexiviridae* family (Supplementary Figure 1). As expected, CapLV-W clustered among the *Carlavirus* genus members, although lacking the ORF6.

BLASTs and phylogenetic analysis were also performed for each CapLV-W ORF including carlaviruses and members of *Foveavirus* genus, as the genus next-closest related to CapLV-W, and characterized by 5 ORFs genome organization.

In all ORFs, the BLAST hits with greatest homology belong to carlaviruses, except for ORF4, showing the highest identity (41%) with *Grapevine rupestris stem pitting associated virus* - GRSPaV, of the genus *Foveavirus*. Phylogenetic analysis (Figure 1) was performed by Neighbor-Joining (NJ) method (Bootstrap: 1000 reps) implemented in MEGA 7 (Tamura et al. 2011). As expected, CapLV-W clustered with available sequences of CapLV (ORF1 and ORF5), close to the other carlaviruses but distinct from foveaviruses. In ORF2-4 (TGB1, TGB2, TGB3) there is not a clear demarcation between carlaviruses and foveaviruses and the CapLV-W isolate clustered mainly with foveavirus isolates. Since a genome recombination was previously reported in *Carlavirus* genus (Wang et al., 2018), the assembled genome was tested for possible recombination amongst carlavirus and foveavirus species, using Recombination Detection Program version 4 (RDP4) (Martin et al., 2015). No evidence of significant recombination was observed.

Here, we report the first complete genome sequence of an ~~Sicilian~~ isolate of caper latent virus (CapLV- W), molecularly characterized as a member of the genus *Carlavirus*. Unlike other members of this genus, the virus appears to lack ORF6, and in light of this report further molecular studies should be carried out to better understand the taxonomy and evolutionary relationship of CapLV with *Carlavirus* genus.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Research involving human participants or animals

This study did not include experiments with human participants or animals performed by any of the authors.

REFERENCES

Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, Oshikawa N, 2012. Family Betaflexiviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. London: Academic Press/Elsevier; 2012. pp. 920–941.

Bolger AM., Lohse M., Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, Volume 30, Issue 15: 2114-2120

Di Franco A and Gallitelli D. 1985. Rhabdovirus-like particle in caper leaves with vein yellowing. *Phytopathologia Mediterranea* 24: 234-236

Grab Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011 Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 29(7): 644-52 doi: 10.1038/nbt.1883.

Martin DP, Murrell B, Golden M. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol*. https ://doi.org/10.1093/ve/vev00

Massart D., Olmos S., Jijakli H., and Candresse T. 2014. Current impact and future directions of hogh throughput sequencing in plant virus diagnostics. *Virus Res*. 188:90-6

Sanger F., Nicklen S., Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. PNAS 74(12): 5463-5467.

Tiberini A., Mascia T. and Tomassoli L. 2011. First nucleotide sequence of a Carlavirus genus infecting caper. *Phytopathologia Mediterranea* 50, 482-488.

Tomassoli L.; Tiberini A. 2006. Molecular detection of caper viruses. Journal of Plant Pathology 88: supplement 3 p. 60

Wang R., Dong J., Wang Z., Zhou T., Li Y., Ding W. 2018. Complete nucleotide sequence of a new carlavirus in chrysanthemums in China. *Archives of Virology* 163:1973-1976

Zuo G, Hao B (2015) CVTree3 web server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy, Genomics Proteomics & Bioinformatics, 13: 321-331

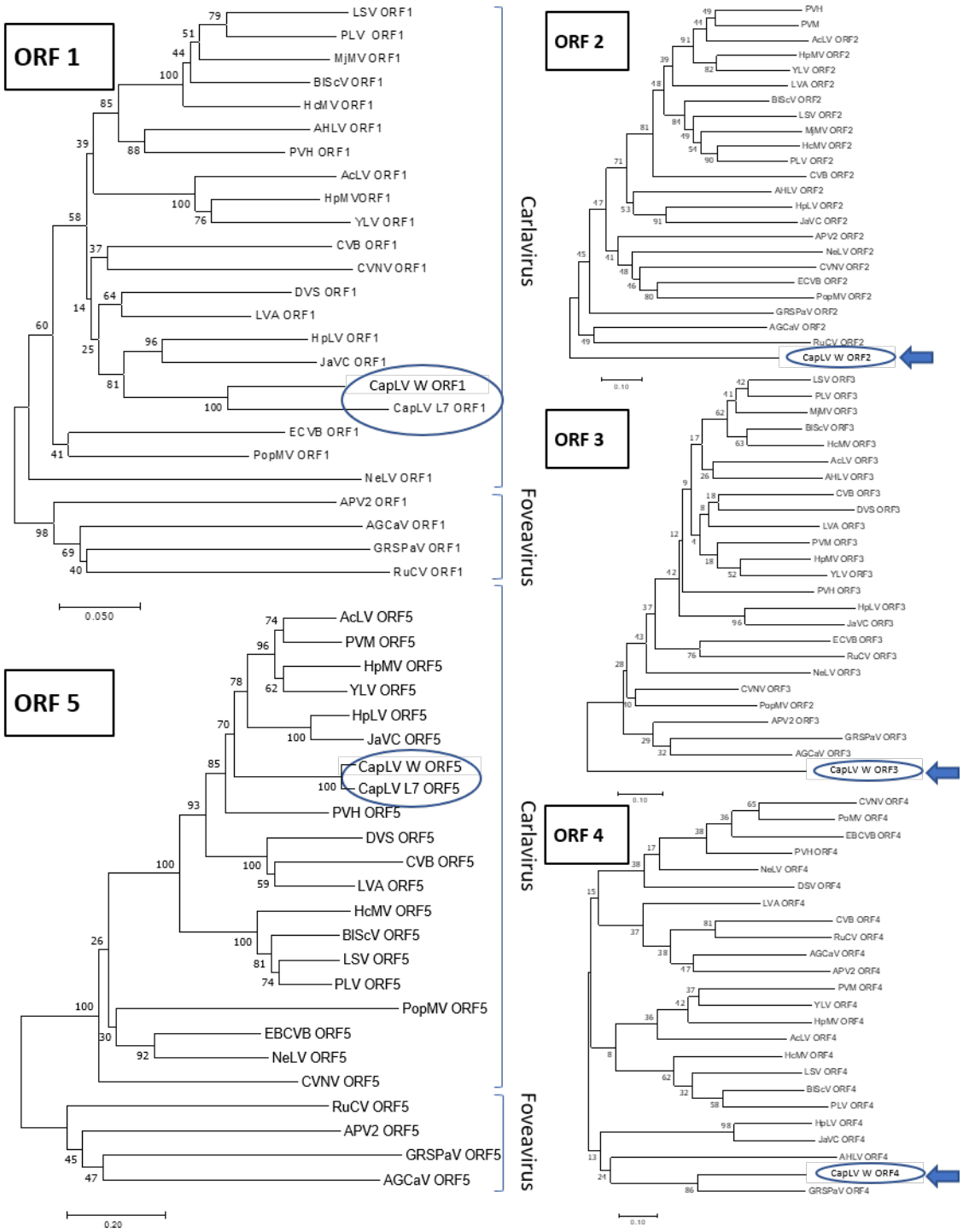


Figure 1 Phylogenetic relationship ORF by ORF of CapLV-W and other member of *Carlavirus* genus based on aminoacidic sequences. In all the analysis member of *Foveavirus* genus were also included as outgroup. Phylogenetic trees were constructed by Neighbor-Joining (NJ) method implemented in MEGA 7 (Tamura et al. 2011). Reliability of the trees was estimated using a bootstrap with 1000 replicates. The viruses included in the analysis were reported in Supplementary Table 3.

Table 1 Open Reading Frame deducted by Open Reading Frame (ORF) Finder tool (http://www.ncbi.nlm.nih.gov/orffinder/). For each ORF is reported the genome position and putative encoded protein and domains identified in the CapLV-W genome using the Conserved Domain Database on the NCBI website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Open reading frame	**Genome position (nt)**	**Protein encoded**
ORF 1	50 - 6034	Replication-associated polyprotein
ORF 2	6060 - 6881	Triple gene block 1
ORF 3	6730 - 7065	Triple gene block 2
ORF 4	7032 - 7238	Triple gene block 3
ORF 5	7255 - 8208	Coat Protein
Deducted domain	**Genome position (nt)**	**Function**
1	177-1115	Viral methyltransferase
2	3099-3362	Carlavirus peptidase
3	3645-4001	RNA Helicase
4	4776-6002	RNA-dependent RNA polymerase II
5	6740-6976	Plant viral movement protein 1
6	7643-8065	Carlavirus coat protein

Supplementary material

Table S1 Primer sets used in RT-PCR assay for Rapid amplification of cDNA ends (RACE) to determine 5' and 3' ends, and to confirm reconstructed scaffold.

Primer Name	Position	Sequence 5'-3'
CapO1R	5'	5'-GGCTCAAGAGAGGTTATGAA-3'
CapO5F	3'	5'-ACCACCATTACATTGGAATATCTAGTTAG-3'
CapORF1F	5715	5'-CGCAAATTGAAACTGAAGGCC-3'
CapTGB1R	6434	5'-GTGTTCTGTCCCAATCTTCGG-3'
CapTGB3F	7010	5'-TAATAAAGAACGCGCCTCTGC-3'
CapCPR	7794	5'-CCTGAGGGTCTAGGTATGCAG-3'

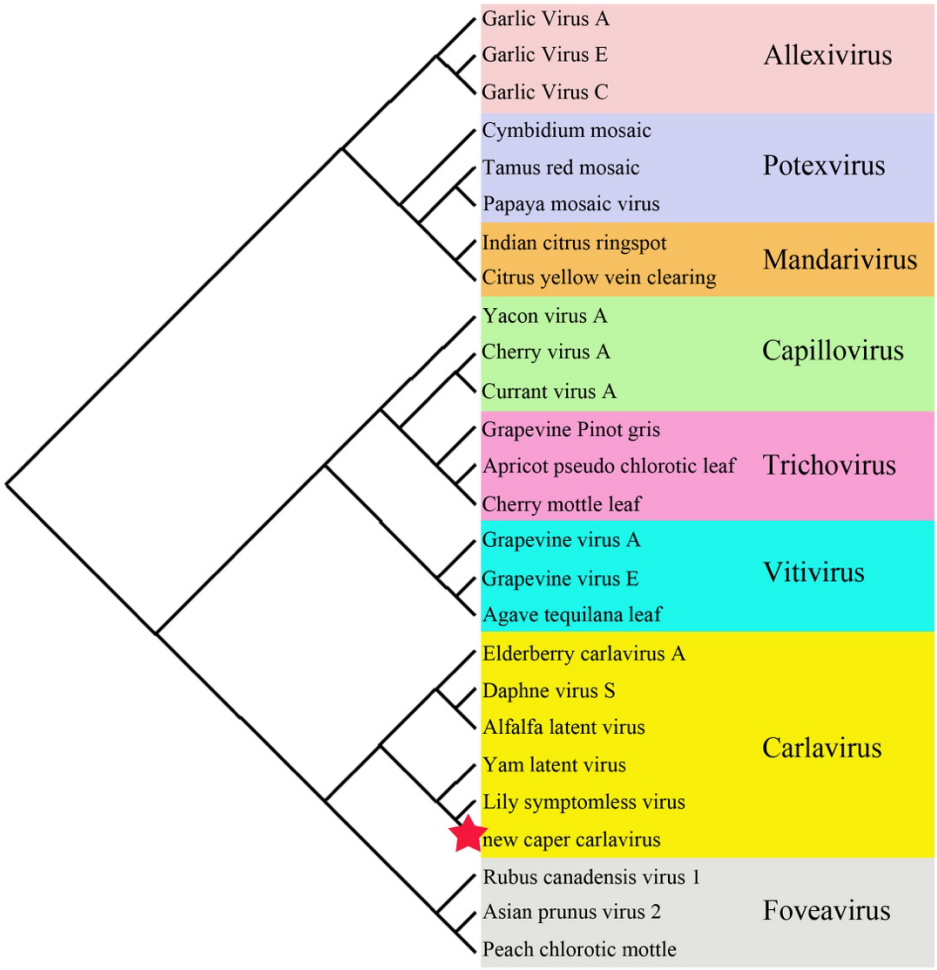
Table S2 BLAST analysis performed in each Open Reading Frame (ORFs). For each are reported virus name showing the highest homology including query coverage, identity percentage, positives percentage and accession number.

Isolates	Query cover	Identity	Positives	Accession
ORF1 - Replication associated polyprotein				
Caper latent virus	69%	87%	95%	AEF12642.1
Ligustrum virus A	81%	54%	69%	YP_009288956.1
Ligustrum virus A	90%	50%	64%	AYC63228.1
Potato virus H	80%	54%	65%	AXL97637.1
Chrysanthemum virus B	79%	54%	67%	CAO78688.1
Potato virus H	76%	54%	66%	YP_006522434.1
Chrysanthemum virus B	79%	54%	65%	CAK54343.1
Potato virus H	77%	53%	66%	AEI55831.1
Chrysanthemum virus B	79%	54%	64%	CAK54345.1
Chrysanthemum virus B	79%	53%	63%	BAF95196.1
ORF2 - Triple gene block 1				
Chrysanthemum virus B	82%	36%	52%	CAF04207.2
Sweet potato chlorotic fleck virus	78%	36%	51%	AKO69641.1
Sweet potato chlorotic fleck virus	78%	36%	50%	ALF99744.1
American hop latent virus	83%	34%	51%	YP_006297587.1
Nerine latent virus	83%	31%	50%	AUW53001.1
Tamus red mosaic virus	80%	34%	50%	YP_004849315.1
Cowpea mild mottle virus	80%	34%	50%	AHG23051.1
Apple stem pitting virus	80%	34%	50%	AGR66479.1
Potato virus M	82%	35%	51%	AGU27039.1
ORF3 - Triple gene block 2				
Potato virus S	97%	38%	60%	AVL25845.1
Yam latent virus	96%	34%	56%	YP_009116870.1
Cowpea mild mottle virus	96%	35%	54%	AGZ15238.1
Carrot carlavirus WM-2008	100%	37%	55%	ACG60022.1
Helleborus net necrosis virus	96%	39%	55%	YP_002574616.1
Hop latent virus	94%	37%	51%	AJR19305.1
Grapevine rupestris stem pitting-associated virus	94%	36%	53%	AXL94671.1
Grapevine rupestris stem pitting-associated virus	94%	35%	52%	AVK43103.1
Cherry virus B	96%	38%	52%	BBD14451.1
Poplar mosaic virus	91%	35%	57%	CAH55775.1
ORF 4 - Triple gene block 3				
Grapevine rupestris stem pitting-associated virus	94%	41%	56%	AXL94829.1
Grapevine rupestris stem pitting-associated virus	94%	39%	56%	AXL94801.1

Grapevine rupestris stem pitting-associated virus	94%	39%	56%	AXL94773.1
Grapevine virus T	89%	41%	61%	AYQ96101.1
Grapevine virus T	85%	40%	60%	AYQ96186.1
Grapevine rupestris stem pitting-associated virus	86%	39%	57%	ARO69984.1
Grapevine rupestris stem pitting-associated virus	86%	39%	57%	ARO70068.1
Grapevine rupestris stem pitting-associated virus	86%	39%	57%	ARO70026.1
Grapevine rupestris stem pitting-associated virus	86%	39%	57%	ABD79045.1
Grapevine rupestris stem pitting-associated virus	86%	39%	57%	AXL94655.1
ORF 5 - Coat Protein				
Caper latent virus	99%	90%	95%	AEF12643.1
Potato virus M	92%	58%	73%	ABM21678.1
Phlox virus M	79%	65%	78%	ACI06093.1
Phlox virus M	79%	65%	72%	ABP68910.1
Potato virus M	92%	58%	78%	ABM21679.1
Hop mosaic virus	78%	62%	78%	BAB72006.1
Potato virus M	79%	64%	70%	ABQ96339.1
Potato virus M	79%	64%	77%	AGU26701.1
Potato virus M	79%	64%	72%	AAP76205.1
Hop mosaic virus	78%	62%	70%	ACS45247.1

Table S3. Virus species included in the phylogenetic analysis. For each virus is reported acronym and Accession number.

Virus	Acronym	Accession number
***Carlavirus* genus**		
Aconitum latent virus	AcLV	AB051848.1
American hop latent virus	AHLV	JQ245696
Blueberry scorch virus	BlScV	L25658.1
Chrysamthemum virus B	CVB	NC009087
Caper latent virus	CapLV	HQ588147.1, HQ588148.1
Coleus vein necrosis virus	CVNV	EF527260.1
Daphne virus S	DVS	NC008020.1
Elderberry carlavirus B	ECVB	KJ572561.2
Hippeastrum latent virus	HpLV	NC011540.1
Hop mosaic virus	HpMV	EU527979.1
Hydrangea chlorotic mottle virus	HcMV	EU754720.2
Jasmine virus C	JaVC	KX364696.1
Lily symptomless virus	LSV	AJ516059
Ligustrum virus A	LVA	KX000914.1
Mirabilis jalapa mottle virus	MjMV	JN039374.1
Nerine latent virus	NeLV	NC_028111.1
Passiflora latent virus	PLV	DQ455582.1
Poplar mosaic virus)	PopMV	AY505475
Potato virus H	PVH	JQ904630.1
Potato virus M	PVM	NC001361
Yam latent virus	YLV	KJ789130.1
***Foveavirus* genus**		
Apple green crinkle associated virus	AGCaV	HE963831.1
Asian prunus virus 2	APV2	KY445748.1
Grapevine rupestris stem pitting associated virus	GRSPaV	KX274274.1
Rubus canadensis virus	RuCV	NC_019025.1



\

Supplementary Figure 1 A phylogenetic tree was done using CVTree softrware based on a method whole-genome-based, alignment-free, and parameter-free. In the analysis were included including the whole genomes of the Carlavirus family isolates as input. In addition to the isolate characterized in this study (new caper virus (MT311966)), the following viruses were included isolated from *Betaflexiviridae* family: *Allexivirus*: *Garlic virus A* – GarV-A (NC_003375.1), *Garlic virus C* – GarV-C (NC003376.1), *Garlic virus E* – GarV-E (NC_004012.1); *Potexvirus*: *Cymbidium mosaic virus* – CymMV (KR185347.1), *Papaya mosaic virus* – PapMV (NC001748.1), *Tamus red mosaic virus* – TRMV (NC016003.1); *Mandarivirus*: *Indian citrus ringspot virus* – ICRSV (NC003093.1), *Citrus yellow vein clearing virus* – CYVCV (NC026592.1); *Capillovirus*: *Cherry virus A* – CVA (KY445749.1), *Currant virus A* – CuVA (NC029301.1), *Yacon virus A* – YVA (NC030657.1); *Trichovirus*: *Apricot pseudo chlorotic leaf virus* – APCLSV (NC006946.1), *Cherry mottle leaf virus* – CMLV (NC002500.1), *Grapevine pinot gris virus* – GPGV (KY747493); *Vitivirus*: *Agave tequilana leaf virus* – ATLV (NC034833.1), *Grapevine virus A* – GVA (KC962564.1), *Grapevine virus E* – GVE (KF588015.1); *Carlavirus*: *Alfalfa latent virus* – ALV (NC026616.2), *Daphne virus S* – DVS (NC008020.1), *Elderberry carlavirus A* – ECVA (NC029085.1), *Lily symptomless virus* – LSV (AJ516059), *Yam latent virus* – YLV (KJ789130.1); *Foveavirus*: *Asian prunus virus* 2 – APV2 (KY445748.1), *Peach chlorotic mottle virus* – PCMV (NC009892.1), *Rubus canadensis virus* 1 – RuCV (NC_019025.1).