Archives of Virology: Brief Reports

Cymbidium chlorotic mosaic virus, a new sobemovirus isolated from the spring orchid (*Cymbidium goeringii*) in Japan

Hideki Kondo*, Shogo Takemoto, Kazuyuki Maruyama, Sotaro Chiba**, Ida Bagus Andika and Nobuhiro Suzuki

Institute of Plant Science and Resources (IPSR), Okayama University, Kurashiki 710-0046, Japan

*Corresponding author: Hideki Kondo

Tel. +81(86) 434-1232; Fax. +81(86) 434-1232

e-mail: hkondo@rib.okayama-u.ac.jp

** Present address, Asia Satellite Campus Institute, The Graduate School of Agricultural Science, Nagoya University, Nagoya 464-8601 Japan

Running Title: Characterization of Cymbidium chlorotic mosaic virus

Abstract

Cymbidium chlorotic mosaic virus (CyCMV), isolated from a spring orchid (*Cymbidium goeringii*), was characterized molecularly. CyCMV isometric virions comprise a single, positive-strand RNA genome of 4,083 nucleotides and 30-kDa coat protein. The virus genome potentially encodes for five overlapping open reading frames with a similar genomic organization to sobemoviruses. BLAST searches and phylogenetic analyses revealed that CyCMV is most closely related to papaya lethal yellowing virus, a proposed dicot-infecting sobemovirus (58.8% nucleotide sequence identity), but has a relatively distant relationship to monocot-infecting sobemoviruses, with only modest sequence identities. These suggest that CyCMV is a new monocot-infecting member of the floating genus *Sobemovirus*.

Main text

Cymbidium is one of the most commercially important cultivated orchids (*Orchidaceae*), comprising nearly 52 species that are widely distributed in the tropical and subtropical regions of Asia [6]. Cymbidium mosaic virus (CymMV, genus *Potexvirus*), odontoglossum ringspot virus (ORSV, genus *Tobamovirus*) and orchid fleck virus (proposed genus *Dichorhavirus*) are the most prevalent and economically important *Cymbidium* infecting viruses [1, 3]. Some minor virus diseases of *Cymbidium* have also been reported, *i.e.,* cymbidium ringspot virus (genus *Tombusvirus*) [11], tomato ringspot virus (genus *Nepovirus*) [8], cymbidium mild mosaic virus (CyMMV, a putative member of genus *Carmovirus*) [4] and an unassigned potyvirus (a possible strain of calanthe mild mosaic virus) [21]. In a previous study, we reported a hitherto undescribed virus, cymbidium chlorotic mosaic virus (CyCMV) causing plant stunting and chlorotic streaks on newly developed leaves of *Cymbidium goeringii* Reichenbach ill. (Spring Orchid) (Fig. 1a) collected from Yamaguchi Prefecture in the west part of the island of Honshu in Japan [12]. CyCMV is sap-transmissible to a *Cymbidium* cultivar (Fig. 1b), but not to other orchids or several experimental plant species, suggesting that this virus has a limited host range ([12] and H Kondo, unpublished results). Viral particles are isometric with a diameter of ca. 28 nm and no serological relationship to CyMMV was observed ([12] and H Kondo, unpublished results). In this study, we report the molecular characterization of CyCMV isolate Cym92-20 (CyCMV-JP) collected from Yamaguchi Prefecture [12] (Supplementary Table S1). Our data suggest that this virus could be considered a new species of the floating genus *Sobemovirus*.

Viral particles were isolated from virus-infected leaves (*Cymbidium* cv. Kenny “Wine Color”) using the method as described by Thomas [23] with the following modifications. 1) The initial supernatant after low speed centrifugation was treated with Driselase (0.5%, w/v) instead of chloroform for clarification, 2) before the first cycle of differential centrifugation, the virus was precipitated using polyethylene glycol 6000 in the presence of 2% Triton X-100 (w/v), 3) the second differential centrifugation was omitted, and 4) further purification of the virions was performed by centrifugation on a sucrose-cesium chloride gradient. Using electron microscopy with negative staining techniques,

spherical virus particles approximately 28 nm in diameter were observed in the purified virus preparations of the Cym92-20 isolate (Fig. 1c). The viral RNA that was obtained by the phenol/chloroform extraction of the purified particles was single stranded RNA (ssRNA) in nature of about 4 kb in length on the basis of indistinguishable migrations of the position of the genomic RNA (4.1 kb) of the southern bean mosaic virus (SBMV) (Fig. 1d and data not shown). A single polypeptide (CP) of approximately 30 kDa was detected in the SDS polyacrylamide gel electrophoresis (PAGE) of the purified particles (Fig. 1e). The presence of double-stranded RNA (dsRNA) in infected plants is often used for the diagnosis of unknown viruses [5]. Two viral dsRNA species were detected from the virus-infected *Cymbidum* plants by PAGE (Fig. 1f) and their sizes were approximately 4.1 and 0.9 kbp, respectively (Supplementary Fig. S1a). A minor, faster-migrating band of unknown origin was detectable in both healthy and CyCMV-infected plants. The 4.1-kbp dsRNA is most likely the replicative form (RF) of the CyCMV genomic RNA, while the 0.9-kbp dsRNA band is probably the RF of a putative coat protein subgenomic RNA (see below). The virion morphology, size of the coat protein, ssRNA genome type, the presence of two dsRNA species, and the limited host range of this virus are typical characteristics of members in the genus *Sobemovirus* [24]. Moreover, genus-specific PCR analysis using the degenerate primers [19] also suggested that CyCMV is a sobemovirus (Supplementary Fig. S1b). However, indirect ELISA and immuno-electron microscopy observations indicated that CyCMV is serologically unrelated to two sobemoviruses occurring in Japan, SBMV (supplied by S.T. Ohki, Osaka Prefecture University, Japan) and cocksfoot mottle virus (CfMV, supplied by S. Toriyama, National Institute for Agro-Environmental Sciences, Japan) (data not shown).

To sequence the CyCMV genomic RNA, viral cDNA fragments were amplified by a random-PCR method using a universal primer (5′-GCCGGAGCTCTGCAGAATTCNNNNNN-3′) [7]. Two regions (nucleotide positions 76–1081 and 994–2884) consisting of some large gaps and ambiguous sequences were amplified by RT-PCR using virus-specific primers. The 5′ and 3′ ends of the RNA genome were determined by rapid amplification of cDNA ends (RACE) as described previously [14]. All of cDNA and RACE fragments were cloned into the pZErO-2 vector (Invitrogen, San Diego, CA, USA) and sequenced using the ABI3100 sequencer (Applied Biosystems, Foster City, CA, USA).

The full-length CyCMV genomic RNA (GenBank accession no. LC019764) was 4,038 nucleotides and the 5′- and 3′-untranslated regions (UTRs) were 91 and 105 nucleotides, respectively (Fig. 2a). The genome organization of CyCMV is similar to that of sobemoviruses, containing four open reading frames (ORFs) (ORF1, ORF2a, ORF2b and ORF3) and the fifth non-AUG ORF (ORFx), as described recently [15, 24] (Fig. 2a). The first ORF, ORF1, encodes a polypeptide of 131 amino acids (aa) with a putative systemic movement and/or RNA silencing suppressor functions (15.1-kDa, P1) [16]. The start codon of ORF1 has a poor Kozak context, with pyrimidine (U) at positions -3 and A at position +4

(^{89}UGU**AUG**A^{95}, a start codon is shown in bold), which is similar to other sobemoviruses [16, 24]. The second ORF, ORF2a, encodes a 59.1-kDa polyprotein (545 aa, P2a) containing a serine protease motif, ^{174}H (X_{35})D(X_{62})TXXGWSG^{278}, and two predicted transmembrane domains in the N-terminal region (Fig. 2b). According to a multiple alignment of P2a encoded by sobemoviruses, the CyCMV genome-linked protein (VPg, 9.0 kDa) may be processed by viral protease at the consensus cleavage sites E^{317}/T^{318} and E^{397}/T^{398} (Fig. 2b and Supplementary Fig. S2). This domain contains ^{384}WAE^{386} and a short E/D sequences (^{389}EDE^{391}) instead of the relatively conserved W(A/G)D motif and the E/D-rich region, respectively [17, 22]. In addition, a proteolytic cleavage site located between the N-terminal region and protease domain of sobemovirus P2a proteins [17, 22] is also conserved in the sequence of CyCMV P2a (Fig. 2b and Supplementary Fig. S2). The third ORF, ORF2b, encodes a 56.6-kDa protein (506 aa, P2b) containing a putative RNA-dependent RNA polymerase (RdRp) motif, ^{327}T(X_3)N(X_{19})GDD^{353} and is probably translated as a P2a–P2b fusion protein (940 aa, 103.5 kDa) via a -1 frameshift mechanism (Fig. 2b). As in the case of all characterized sobemoviruses, the heptanucleotide slippery sequence (^{1743}UUUAAAC^{1749}) with a following stem-loop, which is considered an essential structure for -1 frameshifting, was predicted at the overlapping region of ORF2a and b [18] (data not shown). The fourth ORF, ORF3, encodes the 28.3 kDa (262 aa) viral coat protein (CP). A conserved ^{3181}ACAAA^{3185} sequence, which is identical to 5'-end of the genome, is located just upstream of the CP-initiation codon, as reported for other sobemoviruses [16] (Fig. 2a). This sequence was mapped to the stem region of a predicted stem–loop structure (data not shown) that is assumed to be important for the CP-subgenomic RNA transcription [9]. The fifth ORF of unknown function (ORFx), which partially overlaps with both ORF1 and ORF2a, was predicted to encode a 15.4 kDa protein (144 aa) when the translation is initiated at a non-AUG initiation site ^{294}GAG**AUC**G^{300} [15].

BLASTp searches with the deduced aa sequences of CyCMV-P2a, -P2b and -CP revealed that the virus shares the highest sequence identities (40% [P2a], 66% [P2b] and 46% [CP]) with papaya lethal yellowing virus (PLYV, a proposed sobemovirus) isolated in Brazil [18], while it shares sequence identity of around 26–37% (P2a), 50–63% (P2b) and 23–41% (CP) with other sobemoviruses (Supplementary Table S2). CyCMV P1 showed 29% sequence identity with PLYV P1, but not with P1 of other sobemoviruses. Furthermore, CyCMV ORFx-protein had no homology with its counterparts encoded in any other sobemoviruses or any known proteins in the NCBI database (data not shown).

Maximum-likelihood (ML) phylogenetic analysis based on the complete nucleotide sequences showed that CyCMV is closely related to the dicot-infecting PLYV, but is distantly related to classical monocot (*Gramineae*)-infecting sobemoviruses (CfMV and rice yellow mottle virus) and ryegrass mottle virus (RGMoV), a more recent sequenced monocot sobemovirus, although RGMoV is also distantly related to those classical monocot sobemoviruses (Fig. 3a). A similar topology was seen in

ML trees constructed with the P2a–P2b fusion protein or CP of CyCMV and other sobemoviruses (Fig. 3b and 3c). No recombination event was observed between CyCMV and other sobemoviruses using the Recombination Detection Program, RDP4 (http://web.cbio.uct.ac.za/~darren/rdp.html) (data not shown). Based on the analysis using pairwise global alignment in the PASC (Pairwise Sequence Comparison) site at NCBI (http://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi) [2], the complete genome sequence of CyCMV has moderate levels of nucleotide sequence identities with PLYV (58.8%) and some other sobemoviruses (56.2–58.6%). These values are much lower than the current species demarcation criteria within the genus *Sobemovirus*, with less than 75% overall genome sequence identity among species [24]. Taken together, CyCMV could be considered a novel species in the genus *Sobemovirus*.

We have previously reported that in the period of 1991-1994, *C. goeringii* and some other unknown oriental *Cymbidium* species grown in some locations in Yamaguchi Prefecture showed similar symptoms to those typically induced by CyCMV [12]. In 2009, a similar disease in *C. goeringii* was discovered in Osaka Prefecture, an area about 400 km away from Yamaguchi Prefecture (T. Kawano, Research Institute of Japan Plant Protection, personal communication). To further confirm the occurrence of CyCMV in these areas, RT-PCR detection using primers specific for the RdRp core region (nucleotide position 2531−2873, 343 nt) was performed on four symptomatic leaf samples (frozen leaves) of *C. goeringii* collected from Yamaguchi Prefecture (samples, Cym92-7, Cym92-15 and Cym92-16) and Osaka Prefecture (Cym09-1, supplied by T. Kawano) (Supplementary Table S1). CyCMV was detected in all these samples (Supplementary Fig S1c). The nucleotide sequences of the amplified fragments (GenBank accessions LC019765–LC019768) were highly similar to the sequence of the CyCMV Cym92-20 isolate (91–99%). A phylogenetic analysis using these RdRp sequences revealed two distinct CyCMV groups, subgroup I (Cym92-20 and Cym92-16) and subgroup II (Cym92-7, Cym92-15 and Cym09-1) (Supplementary Fig. S1d). The nucleotide sequence homology between these two subgroups ranged from 91% to 92%, while the amino acid sequence homology ranged from 98% to 100%. Only a few nucleotide substitutions were observed within each subgroup. These data show that there are at least two CyCMV strains in Japan but their biological differences are unknown.

C. goeringii has been cultivated as an ornamental plant for more than ten centuries in China [10], and is one of the most popular traditional ornamental plants in China, Japan, Korea, and Southeast Asia. *C. goeringii* is usually propagated by the division of pseudobulbs [20]. Therefore, it is likely that CyCMV primarily spread through vegetative propagation, as in the case of major orchid viruses (CymMV and ORSV) [1], although insect vector(s) might also exist, as in the case of some sobemoviruses [24]. The data suggest that CyCMV is widely dispersed into the western part of Honshu Island, Japan. It is

therefore interesting to further investigate the geographical distribution and sequence variation of CyCMV strains infecting oriental *Cymbidium* plants in Japan and other East Asian countries.

Acknowledgments

We thank K. Mitsuhata and Dr. T. Tanaka for their technical assistance, Drs. M. Tahara and N. Inouye for collecting plants, Drs. T. Kawano, S.T. Ohki and S. Toriyama for providing viruses and antisera, and Drs. T, Maeda, N. Inouye and T. Tamada for their helpful comments. This work was supported by the Ohara Foundation for Agricultural Research and Yomogi Inc.

References

1. Ajjikuttira PA, Lim-Ho CL, Woon MH, Ryu KH, Chang CA, Loh CS, Wong SM (2002) Genetic variability in the coat protein genes of two orchid viruses: *Cymbidium mosaic viru*s and *Odontoglossum ringspot virus*. Arch Virol 147:1943-1954

2. Bao YM, Chetvernin V, Tatusova T (2014) Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. Arch Virol 159:3293-3304

3. Blanchfield AL, Mackenzie AM, Gibbs A, Kondo H, Tamada T, Wilson CR (2001) Identification of *Orchid fleck virus* by reverse transcriptase-polymerase chain reaction and analysis of isolate relationships. J Phytopathol 149:713-718

4. Chang MU, Yora K (1978) Isolation of Cymbidium mild mosaic virus. Korean J Plant Protection 17:131-138

5. Dodds JA, Morris TJ, Jordan RL (1984) Plant viral double-stranded-RNA. Ann Rev Phytopathol 22:151-168

6. Du Puy D, Cribb P (1988) The genus *Cymbidium*. Christopher Helm, London

7. Froussard P (1992) A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. Nucleic Acids Res 20:2900

8. Goff LM, Corbett MK (1977) Association of tomato ringspot virus with a chlorotic leaf streak of *Cymbidium* orchids. Phytopathology 67:1096-1100.

9. Hacker DL, Sivakumaran K (1997) Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs. Virology 234:317-327

10. Hew CS (2001) Ancient Chinese orchid cultivation - A fresh look at an age-old practice. Sci Hortic 87:1-10

11. Hollings M, Stone OM, Barton RJ (1977) Pathology, soil transmission and characterization of cymbidium ringspot, a virus from *Cymbidium* orchids and white clover (*Trifolium repens*). Ann Appl Biol 85:233-248

12. Kondo H, Maeda T, Mitsuhata K, Inouye N (1996) Detection of the viruses occurring in oriental *Cymbidium* in Japan. Bull Res Inst Bioresour Okayama Univ 4:149-162 (in Japanese with English abstract)

13. Kondo H, Maeda T, Gara IW, Chiba S, Maruyama K, Tamada T, Suzuki N (2014) Complete genome sequence of Habenaria mosaic virus, a new potyvirus infecting a terrestrial orchid (*Habenaria radiata*) in Japan. Arch Virol 159:163-166

14. Kondo H, Maruyama K, Chiba S, Andika IB, Suzuki N (2014) Transcriptional mapping of the messenger and leader RNAs of orchid fleck virus, a bisegmented negative-strand RNA virus. Virology 452:166-174

15. Ling R, Pate AE, Carr JP, Firth AE (2013) An essential fifth coding ORF in the sobemoviruses. Virology 446:397-408

16. Meier M, Olspert A, Sarmiento C, Truve E (2008) *Sobemovirus*. In: Mahy B, van Regenmortel M (eds) Encyclopedia of virology. Elsevier, Oxford, UK, pp 644-652

17. Nair S, Savithri HS (2010) Processing of SeMV polyproteins revisited. Virology 396:106-117

18. Pereira AJ, Alfenas-Zerbini P, Cascardo RS, Andrade EC, Zerbini FM (2012) Analysis of the full-length genome sequence of papaya lethal yellowing virus (PLYV), determined by deep sequencing, confirms its classification in the genus Sobemovirus. Arch Viroly 157:2009-2011

19. Sereme D, Lacombe S, Konate M, Pinel-Galzi A, Traore VSE, Hebrard E, Traore O, Brugidou C, Fargette D, Konate G (2008) Biological and molecular characterization of a putative new sobemovirus infecting *Imperata cylindrica* and maize in Africa. Arch Virol 153:1813-1820

20. Shimasaki K, Uemoto S (1991) Rhizome induction and plantlet regeneration of *Cymbidium goeringii* from flower bud cultures *in vitro*. Plant Cell Tissue Organ Cult 25:49-52

21. Singh MK, Sherpa AR, Hallan V, Zaidi AA (2007) A potyvirus in *Cymbidium* spp. in northern India. Australasian Plant Dis Notes 2:11-13

22. Somera M, Truve E (2013) The genome organization of lucerne transient streak and turnip rosette sobemoviruses revisited. Arch Virol 158:673-678

23. Thomas JE (1986) Purification and properties of ginger chlorotic fleck virus. Ann Appl Biol 108:43-50

24. Truve E, Fargette D (2011) Genus *Sobemovirus*. In: King AM, Adams MJ, Lefkowitz EJ, Carstens EB (eds) Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, CA, USA, pp 1185-1189

25. Zhang R, Liu SX, Chiba S, Kondo H, Kanematsu S, Suzuki N (2014) A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, *Rosellinia necatrix*, with similarity to hypo-like viruses. Frontiers in Microbiology 5:360. doi: 10.3389/fmicb.2014.00360

Figures

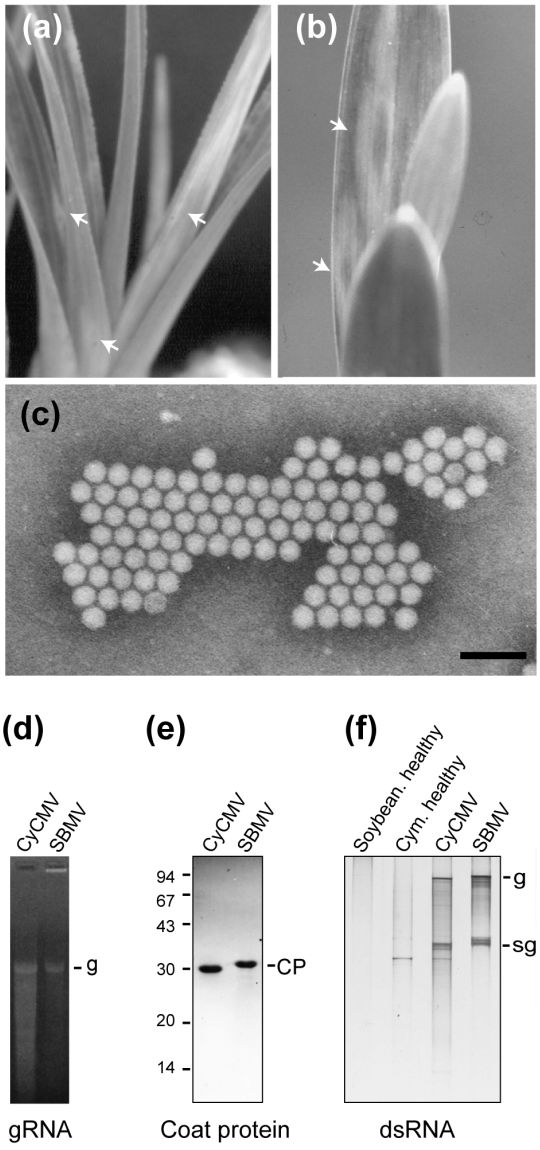


Fig. 1. (a) Symptoms in *Cymbidium goeringii* naturally infected with cymbidium chlorotic mosaic virus (CyCMV). Chlorotic streaks on newly developed shoots are indicated by white arrows. (b) *Cymbidium* sp. cv. Melody Fair "Marilyn Monroe" inoculated with the crude sap from leaves. Chlorotic streaks on the newly developed shoots are indicated by white arrows. (c) Electron micrograph of a purified preparation of CyCMV particles stained with 1% phosphotungstic acid. Bar represents 100 nm. (d) Denaturing agarose gel (1.0%) electrophoresis of the CyCMV virion-associated RNA. Gel was stained with ethidium bromide. Southern bean mosaic virus (SBMV) genomic RNA (approximately 4.1 kb) was used as a size control. (e) SDS-PAGE (15%) of CyCMV coat protein (CP) staining with Coomassie brilliant blue-R250. The positions of the marker proteins (Pharmacia) are shown on the left. (f) PAGE (6%) of double-stranded RNAs (dsRNAs) extracted from healthy and viral-infected *Cymbidium* leaves. Gel was stained with silver nitrate. Two dsRNA species extracted from SBMV infecting soybean plants (approximately 4.1 and 1.0 kbp) were used as size controls (d-f). g: replicative form (RF) of the genomic RNA; sg:. RF of a putative CP subgenomic RNA.

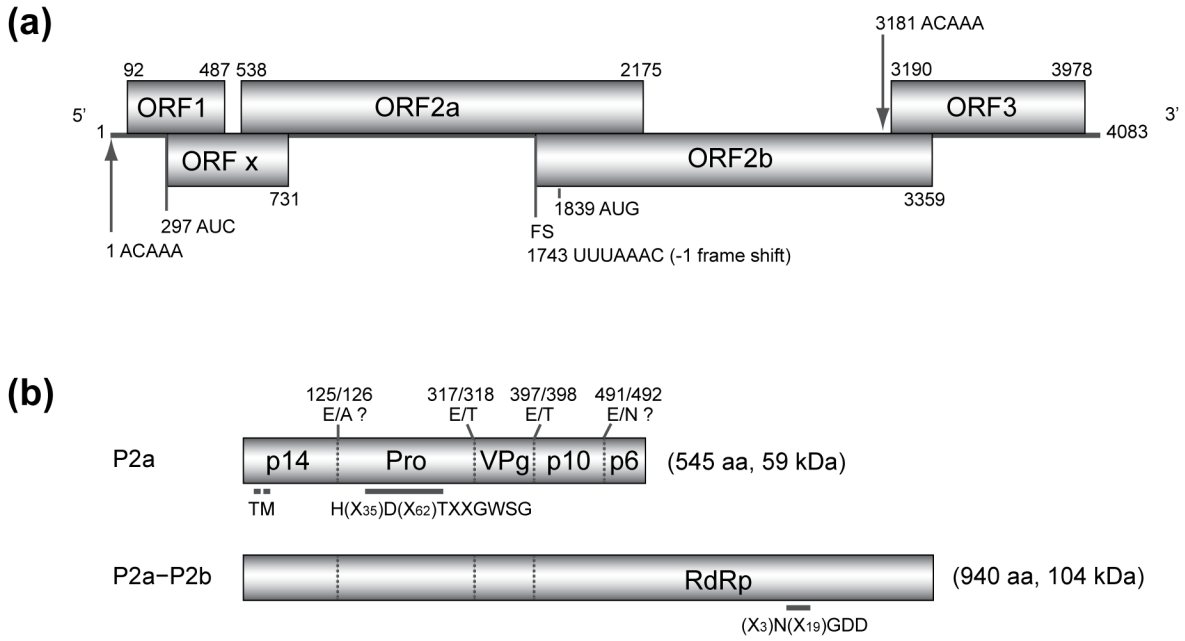


Fig. 2. Schematic representation of the CyCMV genome. (a) Genomic organization of CyCMV. The 5′ and 3′ untranslated regions (UTRs) are shown as bold lines, while open boxes represent open reading frames (ORFs). The numbers represent the nucleotide sequence positions of the start and stop codons for four ORFs (ORF1, ORF2a, ORF2b and ORF3) and a recently described fifth non-AUG ORF (ORFx). The conserved nucleotide motifs (ACAAA) and their nucleotide positions present at the 5′-terminal and upstream of the CP-ORF (ORF3) are indicated by arrows. The heptanucleotide motif (UUUAAAC), where a −1 ribosomal frameshift (FS) might occur, is denoted as FS. (b) The cleavage sites and domains in polyproteins P2a and P2a–P2b. The 2a polyprotein is predicted to encode five mature proteins [17]: p14 (membrane anchor), Pro (serine protease), VPg (viral protein genome linked), p10 (unknown function), and p6 (unknown function). The locations of putative cleavage sites by the viral proteinases are indicated as dashed vertical lines in the coding region with the peptide sequence at each site (E/A, E/T and E/N). The forth putative cleavage site appears to be unconserved among sobemoviruses (see Supplementary Fig. S2). The putative domains for TM (trans-membrane domain), Pro and RdRp (RNA-dependent RNA polymerase) are indicated by thick bars below P2a and P2a–P2b along with consensus amino acid sequences.

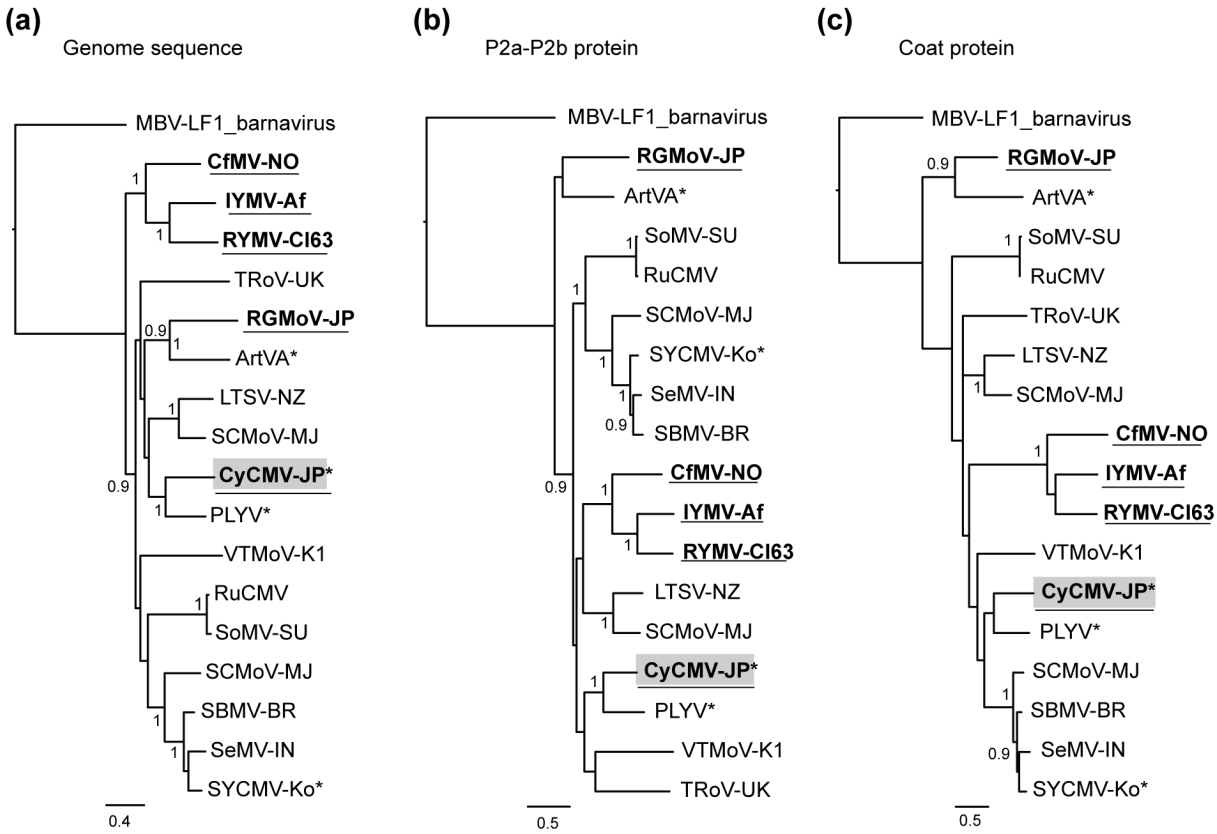


Fig. 3. Phylogenetic analyses of CyMV and other sobemoviruses. Phylogenetic trees were calculated using the nucleotide sequences of entire genome (a), the P2a–P2b fusion protein (b) and the coat protein (c) of CyCMV-JP (Cym92-20 isolate) and other sobemoviruses. Each tree was constructed using the maximum-likelihood (ML) method as described previously [13, 25] with the appropriate substitution mode. The sequence data of sobemoviruses (presented as acronyms) used for the analyses were obtained from the EMBL/DDBJ/Genbank and are listed in Supplementary Table S1. Viruses marked with an asterisk are unassigned species, while CyCMV is highlighted. Monocot-infecting sobemoviruses are bolded and underlined. Mushroom bacilliform virus (MBV-LF1, *Barnavirus*, *Barnaviridae*) was used as an outgroup. The branch support values were estimated using the approximate likelihood ratio test and shown at the nodes (aLRT, only values greater than 0.9 are indicated).

Supplementary Table S1.

List of virus genome sequences compared in Figure 3 and Figure S1.

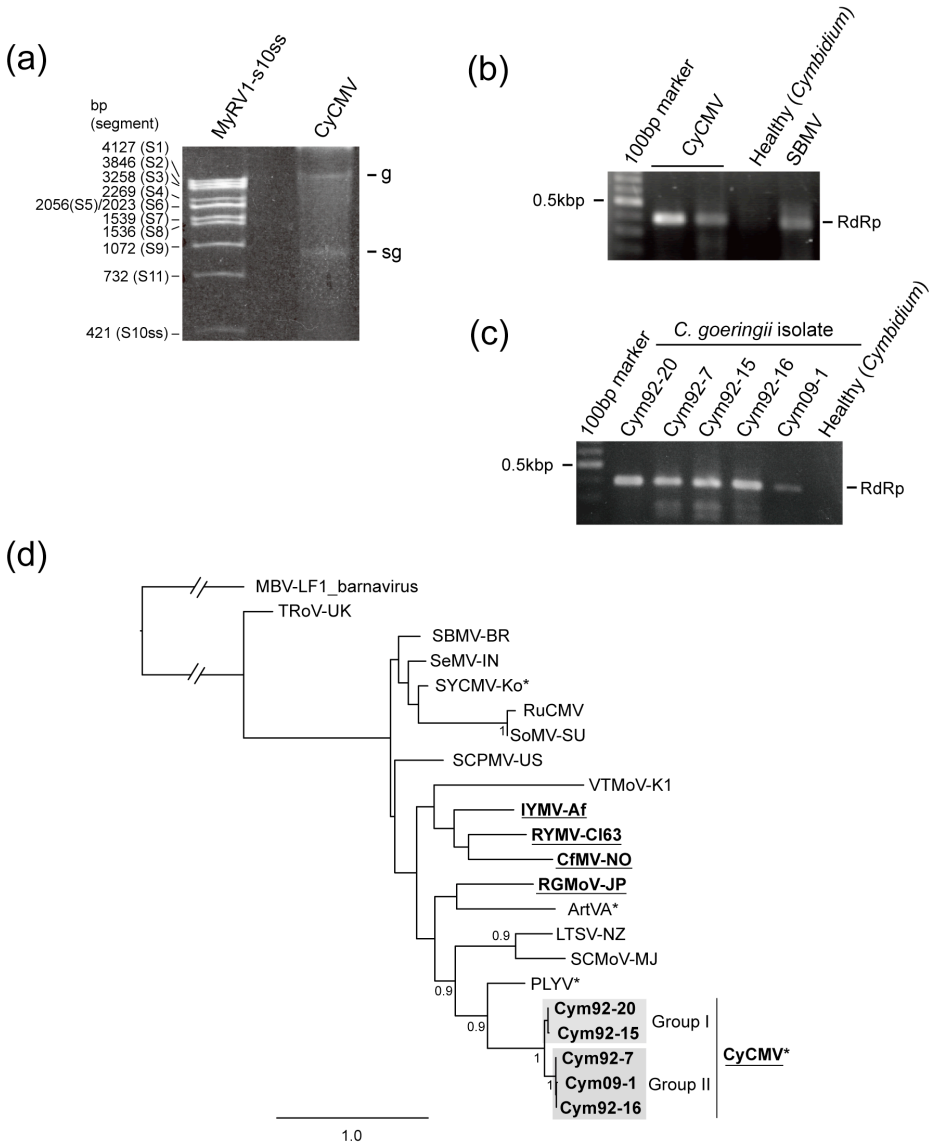
Species names -Names of isolates (Sample name)	Abbreviations	Size (nt)	NCBI Acc. /RefSeq
Cymbidium chlorotic mosaic virus - Japan (Cym92-20)	**CyCMV-JP**	**4,083**	LC019764
- Cym92-15		(343)	LC019765
- Cym92-7		(343)	LC019766
- Cym92-16		(343)	LC019767
- Cym09-1		(343)	LC019768
Genus *Sobemovirus*			
Cocksfoot mottle virus - Norway	CfMV-NO	4,082	NC_002618
Imperata yellow mottle virus-Africa	IYMV-Af	4,575	NC_011536
Lucerne transient streak virus - New Zealand	LTSV-NZ	4,279	NC_001696
Rice yellow mottle virus - Côted'Ivoire: CI63	RYMV-CI63	4,449	NC_001575
Ryegrass mottle virus – Japan	RGMoV-JP	4,212	NC_003747
Sesbania mosaic virus - India	SeMV-IN	4,449	NC_002568
Southern bean mosaic virus – Brazil	SBMV-BR	4,132	NC_004060
Southern cowpea mosaic virus - USA	SCPMV-US	4,193	NC_001625
Sowbane mosaic virus – USA Synonym: Rubus chlorotic mottle virus	SoMV-US	4,003	NC_014608
Subterranean clover mottle virus - Australia	SCMoV-MJ	4,258	NC_004346
Turnip rosette virus - UK	TRoV-UK	4,035	NC_004553
Velvet tobacco mottle virus - Australia-K1	VTMoV-K1	4,247	NC_014509
Unassigned sobemoviruses			
Artemisia virus A	ArtVA	4,138	NC_017914
Papaya lethal yellowing virus	PLYV	4,145	NC_018449
Soybean yellow common mosaic virus	SYCMV	4,152	NC_016033
Genus *Barnavirus* (outgroup)			
Mushroom bacilliform virus-LF1	MBV-LF1	4,009	NC_001633

Supplementary Table S2.

Comparisons of amino acid sequence identity of the viral proteins between CyCMV and sobemovirus sequences available in the GenBank database.

Virus	P1 (ORF1)	P2a (ORF2a)	P 2b (ORF2b)	CP (ORF3)
ArtV*,[a]	nh [b]	30% (78, 9e-38) [b]	57% (86, 3e-161)	28% (59, 8e-07)
CfMV	nh	29% (84, 1e-35)	54% (86, 1e-161)	30% (87, 1e-10)
IYMV	nh	26% (87, 2e-30)	54% (92, 2e-163)	31% (98, 3e-12)
LTSV-NZ	nh	35% (92, 2e-67)	59% (86, 1e-169)	34% (86, 9e-30)
PLYV*	**29% (83%, 3.3)**	**40% (92, 1e-109)**	**66% (88, 0.0)**	**46% (90, 2e-65)**
RYMV	nh	30% (78, 9e-28)	55% (89, 2e-163)	27% (68, 4e-08)
RGMoV	nh	31% (92, 8e-48)	53% (98, 2e-163)	23% (97, 1-06)
SeMV	nh	33% (84, 7e-57)	59% (91, 0.0)	36% (100, 2e-46)
SBMV	nh	36% (76, 5e-61)	62% (88, 0.0)	36% (100, 4e-45)
SCPMV	nh	32% (76, 5e-61)	63% (64, 3e-144)	33% (99, 2e-32)
SoMV	nh	32% (89, 5e-55)	56% (87, 1e-161)	32% (91, 3e-26)
SCMoV	nh	31% (83, 2e-53)	56% (97, 2e-180)	41% (81, 2e-47)
SYCMV*	nh	35% (96, 1e-66)	62% (88, 0.0)	34% (100, 4e-48)
TRoV	nh	37% (79, 6e-66)	53% (89, 5e-147)	34% (91, 1e-33)
VTMoV	nh	31% (81, 4e-40)	50% (99, 1e-153)	32% (83, 4e-22)

[a] Unassigned viruses are indicated by their abbreviation with asterisks (see Supplementary Table S1).
[b] The percent of amino acid sequence identity (%) in the sequence comparison analysis is shown with their query coverage (%) and expect value in parenthesis. nh: no hit against the corresponding target.

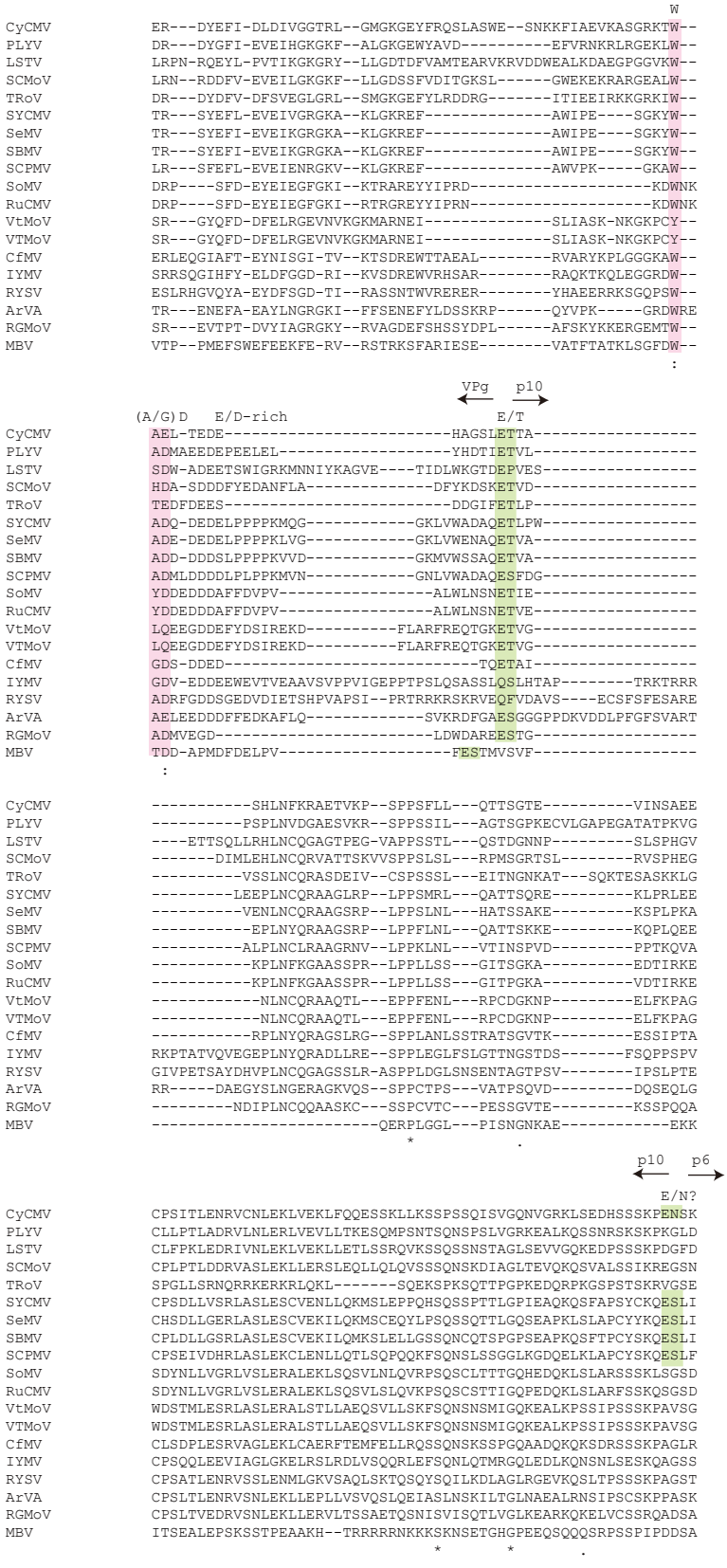


Supplementary Fig. S1. dsRNA analysis and RT-PCR detection of CyCMV. (a) Agarose gel (1.2%) electrophoresis of dsRNAs isolated from CyCMV-infected leaves. The dsRNA genome of mycoreovirus 1 variant carrying an internal deletion of the S10 segment (MyRV1-S10ss strain) was used as a size standard (Tanaka et al., 2011, J. Gen. Virol. 92: 1949-1959). Gel was stained with ethidium bromide. g: replicative form (RF) of the genomic RNA; sg: RF of putative subgenomic RNAs. (b and c) RT-PCR detection of CyCMV using *Cymbidium* leaves materials with a set of primers specific for the sobemovirus-RdRp domain (sobAS: 5′-RTCNCCCATNGCDATRCACCA-3′ and sobA: 5′-CCNTCNAARCCNGGNATGGG-3′). PCR products were agarose gel electrophoresed (1.2%) and stained with ethidium bromide. Southern bean mosaic virus (SBMV) was used as a positive control (b). The 100 bp DNA size marker is on the left. (d) A phylogenetic tree calculated from the nucleotide sequences of RdRp core

regions of CyCMV (nucleotide positions 2531−2873, 343 nt) and sobemoviruses. The sequence data of sobemoviruses (presented as acronyms) used for the analysis are listed in Table S1. Viruses marked with an asterisk are unassigned species. Monocot-infecting sobemoviruses are underlined, while CyCMV isolates are highlighted in gray. Mushroom bacilliform virus (MBV-LF1) was used as an outgroup. The branch support values were estimated using the approximate likelihood ratio test (aLRT, only values greater than 0.9 are indicated) and shown at the nodes.

Supplementary Fig. S2. Amino acid sequence alignment of the P2a of CyCMV and sobemoviruses using the MAFFT program version 7 (http://mafft.cbrc.jp/alignment/server/). The alignments of the N- and C-terminus regions were not shown. Identical and similar residues are indicated by asterisks and dots below the alignment, respectively. The putative catalytic residues (red), cleavage sites (green) and other conserved motifs (pink) in sobemovirus polyproteins are highlighted. The sequence data of sobemoviruses (presented as acronyms) used for the analysis are listed in Supplementary Table S1.


p14
Pro
E/A?
H
CyCMV 118 ISRVDGNEASILGNMMNKVNPGSEPASLVSISNGT-DVVGMGARVNY-NG-TTWLLTASH
PLYV ISNKYGTECPVMGVSLSHVAPGKEPSSLVAISNGS-EIIGFGSRVKF-NG-VTYLLTANH
LSTV ISKNGEDEAAVLDSVYSSVLTGTEPASLVLLKSGS-QTVGFGARVSY-EGCSDYLLTAHH
SCMoV ISKDGEKECAIFNAGYSSVLPGTEPTSLVMLKAKD-LAVGFGARVRF-NGCSDYLLTAYH
TRoV -SQDGNKETVCIDRMSTVTPAGKEPPSLVTLKVGD-RVVGMGSRVSW-GG-NTYLLTAAH
SYCMV NGDEHSNESAILGSFYSAVKPGDEPASLVAIKSGP-HTIGFGCRTKI-DG-EDALLTANH
SeMV NVGDASNESAVLGSFYSPVKAGDEPASLVAIKSGP-TTIGFGCRTKI-DG-EDCLLTAHH
SBMV NGDDASNESAVLGSFYSSVKPGDEPASLVAIKSGP-QTIGFGCRTKI-DG-DDCLFTANH
SCPMV NKVEVGQESAILGSTYSVVETGGEPKSLVAVKSGD-STLGFGARVYH-EG-MDVLMVPHH
SoMV VNQDEDNECAILGNSYPKVEPHREPGSFVLLKVNE-EVVGAGCRVIY-DG-GDYLLTAHH
RuCMV INQDEDNECAILGNSYSQVVPRSEPGSFVLLKVNE-EVVGAGCRVIY-DG-GDYLLTAHH
VtMoV ISQGVVKEGAMIGVPFTSLPSGAEPKSLVVLYNDG-VRIGMGSRVNW-QG-ADYLLTASH
VTMoV ISQGVVKEGAMIGVPFTSLPSGAEPKSLVVLYNDG-VRIGMGSRVNW-QG-ADYLLTASH
CfMV STRKETVEMAVENSRLQPLESSQTPKSLVALYSQD-LLSGWGSRIKGPDG-QEYLLTALH
IYMV LYRKTVTEAYAGGSMPTPVAPDKLPGCVVTLYHGE-TFLGFGSRVRTPNG-QDLLMTNHH
RYSV VNPGVSMEASRGGLPPTSVKLEDVPPSVVVLYHDS-VRLGLGTRVRTPTG-RDLLMTNHH
ArVA --LRSEKEGACIGSPVSLIEPGKEPNYLVVIQDENGFTRGMASRVQI-GR-NDVMLTAFH
RGMoV --NIDRDEAAVANSIVSSVAPGKEPGSLVCIQAKDGKVIGMGARVHC-GP-ATVLVTAGH
MBV GRKGTVLEAAQPGSDF--VECEQWPNGLVAIRRHDGRIVGMGFLVVL-NG-KWRLVTAAH
D
CyCMV VWNGTS-PVLYLAK------GGLQTEVS-TEWPIGLSCTHRTADFVMVKVPDRVWSRLGV
PLYV VWNGRY-TALKIVK------GSQEVSVDASMYTAKYFCEDLRVDFAMVPIPEPIWTKLGV
LSTV VIEPH--EKLDLCK------RGKVIPDL--DLTTTYDCEDKFAEFAMIKVPSNYWSRLGV
SCMoV VIKPH--EKLNLCK------GGYMVEDV--DLAVTCGSDHDAVDFALIKVPPAVWSKLKV
TRoV VCALH--KDIYIYK------NAIGTPLG-AGWTRRYGATHKTADFTLIEVPPTVWAKLGV
SYCMV VWNGAA-KPDALAK------NGKQVAVE--DWAVPVSSDHQMLDFVAVRVPSCVWSKLGV
SeMV VWCNSM-RPTGLAK------AGKQVSVE--DWEISMSSSDKMLDFAIVRVPTHVWSKLGV
SBMV VWNNSM-RPTALAK------RGKQVAIE--DWETPLSCDHKMLDFVVVRVPKHVWSKLGV
SCPMV VWYNDK-PHTALAK------NGRSVDTE--DWEVEAACADPRIDFVLVKVPTAVWAKLAV
SoMV VWSQA---PNHIAK------GGKTVEIS-TEMKPHLSSKNKVLDFCLVPIPAAVWSNLGV
RuCMV VWSQA---PNHIAK------GGKTVEIS-MEMKPYLSSKNKVLDFCLVPVPAAVWSNLGV
VtMoV VWSLVT-GDFQMAK------ASKMVSVK--DAKCYVEAAHAKLDFALIKVPNKYWSSIGV
VTMoV VWSLVT-GDFQMAK------ASKMVSVK--DAKCYVEAAHAKLDFALIKVPNKYWSSIGV
CfMV VWETN---ISHLCK------DGKKVPIS--GCPIVASSADSDLDFVLVSVPRNAWSVLGV
IYMV IALLE---PNAIAK------GGDMVECD--FGPMKYACHDQRMDCCFYNVSPKIWSLLKV
RYSV IAALE---PNGIAY------KGHLKKVA--LDAPVIACDHPHIDCAFYEVPPKIWSLLGV
ArVA VIDTA--EKLYMAKFSNNERTGRRVEIDFQNWKLDFASRDARIDVASVFVPQKVWSSLGV
RGMoV VLKKGMIADLYLAKYSVSSKEGKRVLMD-PTWKIEYGSLNKEADVISVQVPAAVWSRLGG
MBV VARECK-RGIMLSA-GI---DSKTVTFQ--DLDVVL---QTQVDACIMNVPAGTAASLGV
GwSG
CyCMV KSAPLS-AMAKTSIVTIFSDSNGTML-SSSGRAVKG-EYSHDIQHTCSTTNGWSGSPLYY
PLYV KASNLS-TMSRTSLINVYGGADPTKLKCSSAKAVKA-EYSHQIVHYATTGGGWSGTPLYY
LSTV GVAKLS-ALSKKSTVSLYGGTSSTGLTCSSGFAYKG-KSGYAIIHEASTTKGWSGTPLYV
SCMoV GVGKLE-PMTKKTHITVYGGSDSTRLLSSSGPAYKG-KAGYAIIHEASTTKGWSGTPLYS
TRoV KAASLQ-PLNKLSVVTVYSANSSTVITSSSSRAVTQ-EFRHVIIHSCNTTAGTSGSPLYS
SYCMV KSTPLV-CPSAKDAVTCYGGPSSDELLSSVGNCSPT-EFAWKVTHNCPTAAGWSGTPLYS
SeMV KSTPLV-CPSSKDVITCYGGSSSDCLMSGVGSSSTS-EFTWKLTHTCPTAAGWSGTPLYS
SBMV KATQLV-CPSDKDAVTCYGGSSSDNLLSGTGVCSKV-DFSWKLTHSCPTAAGWSGTPIYS
SCPMV RSTKVL-APVHGTAVQTFGGQDSKQLFSGLGKAKAL-DNAWEFTHTAPTAKGWSGTPLYT
SoMV KSSKIA-SLHQRSNVTVYGGTASKMLLSSFGVAEAD-DNPLRIIHKASTARAWSGSPLYN
RuCMV KSSKIA-SLHQRSNVTVYGGTASTMLLSSFGIAEAD-DNPLRIIHKASTARAWSGSPLYN
VtMoV GSAKLL-WHKPGQVVKVYGGR-SDELVSSVGRAEKDPDLSLRLTHNASTAPGWSGSPLYN
VTMoV GSAKLL-WHKPGQVVKVYGGR-SDELVSSVGRAEKDPDLSLRLTHNASTAPGWSGSPLYN
CfMV GVARLE-LLKRRTVVTVYGGLDSKTTYCATGVAEL--ENPFRIVTKVTTTGGWSGSPLYH
IYMV SSAPLK-PMLRPMACNLFGGVSSTSISTSLAFVRPE-SNPWFLQHNATTGRGWSGTPLIH
RYSV KSASLK-PLVKQTAVSLFGGSSSTDFSSCVGIAQIG-DNPFLIRHQSTTCSGWSGSPLYH
ArVA KSAKVKIPTAERKPVQVFGADSSSAFKSSVGLGTFV--SEFTGEHSATTTKGWSGSPVIS
RGMoV PPPACG-SLLLKYQSSPMVGRLPGSYNPRKVLRLQM-AICLWPIH-AQPVQGGAGTPLYA
MBV RKVVINRTPSESKVVRTYGYN-SGKFCMSEGLVGTT-SANMGFRHGCSTLRGWSGTPIYR
Pro
VPg
E/T
CyCMV KGA-VVGIHCGLKDFGVSNRGVNVGVLLTA-SAGLETVY-------SEISNTLISPEEAD
PLYV NGA-VVGMHTGSIKLGESNRGVNVAMLLNL---ACETVF-------SEITYTEIDSDLAL
LSTV GNN-IVGVHTGCGKAGETNRGTNVRVLLDL-SSGYESDF-------SEISYGEIDLDNFR
SCMoV GNT-IVGVHTGSGQVGYSNRAVNVKLLLTA-VSKFETIF-------SEISYGELDEDNYL
TRoV GDN-VVGVHLGTEVTMHSNRACNVGLVLGA---FHESII-------SQGTLSEISADEAA
SYCMV SRG-VVGMHTGFENIGKMNRGVNMFYVANYLLRSQETLP-------PELSVIEIPFEDVE
SeMV SRG-VVGMHVGFEEIGKLNRGVNMFYVANYLLRSNETLP-------PELSIIEIPFDDVE
SBMV SRG-VVGMHVGFEDIGKLNRGVNAFYVSNYLLRSQETLP-------PELSVIEIPFEDVE
SCPMV RDG-IVGMHTGYVDIGTSNRAINMHFIMSCLVSKMETLP-------PELGYREISLEDVG
SoMV SNGLVLGVHLGYDQLGSTNRAVNIGYVLRF-TTSNETAP-------PDLNFIEITEDEAV
RuCMV SNGLVLGVHLGYDQLGSTNRAVNIGYVLRT-TSSNETAP-------PDLNFVEITEDEAV
VtMoV SENFVVGLHTGFSAAEQRNEAVDVAKLLHLALRTKETTF-------SEIGVSLIDEDEIE
VTMoV SENFVVGLHTGFSAAEQRNEAVDVAKLLHLALRTKETTF-------SEIGVSLIDEDEIE
CfMV KDA-IVGLHLGARPSAGVNRACNVAMAFRVVRKFVTVENSELYPDQSSGPARELDAETYT
IYMV KGA-VYGLHIGA--LGSVNVASNPALYFLTYRREVEVES----PELWKSVPNEITNEEYE
RYSV KGC-VVGLHIGA--ADGYNVASNVAWYFHTFKKEVIVES----PFEIYGKFREANSEEYD
ArVA NGC-VIGVHRGVDLDKVNSNKFTIIHQSFF-PPGLETMY-------DYGHIRELDEEQFE
RGMoV GSD-IVAIHRRWEDIGVKNLATNLSIFHAN-CESSEN---------GEQGAREIDAEEWI
MBV DNK-VVGIHSRCNGI-YENFGLSLDLLVGR-LESEETDR-------YARTMEEFNTEDRP


VPg
p10
(A/G)D E/D-rich
E/T
p10
p6
E/N?
CyCMV
PLYV
LSTV
SCMoV
TRoV
SYCMV
SeMV
SBMV
SCPMV
SoMV
RuCMV
VtMoV
VTMoV
CfMV
IYMV
RYSV
ArVA
RGMoV
MBV