Arch Virol
DOI 10.1007/s00705-008-0043-8

BRIEF REPORT

Genome structure and organization of a member of a novel and distinct species of the genus *Caulimovirus* associated with dahlia mosaic

V. Pahalawatta · K. L. Druffel · S. D. Wyatt · K. C. Eastwell · H. R. Pappu

Received: 21 April 2007 / Accepted: 31 December 2007

Abstract The genome structure and organization of a new and distinct caulimovirus that is widespread in dahlia (*Dahlia variabilis*) was determined. The double-stranded DNA genome was ca. 7.0 kb in size and shared many of the features of the members of the genus *Caulimovirus*, such as the presence of genes potentially coding for the movement protein, the inclusion body protein, and the reverse transcriptase (RT), and an intergenic region consisting of a potential 35S promoter. However, the virus differed from the previously described dahlia mosaic caulimovirus and other known caulimoviruses in that the aphid transmission factor (ATF) was absent and the putative coat protein contained a C-terminal deletion and was fused in-frame with the RT. Sequence identity at the amino acid level with known caulimoviruses including a previously reported caulimovirus from dahlia was low and ranged from 32 to 72%. The absence of an ATF and the highly divergent nature of the genomic sequence are characteristics of this new caulimovirus that is widely associated with dahlia.

Sequences reported here are available in GenBank under the following accession numbers: EU096520, EU096521, EU096522, EU096523, EU096524

V. Pahalawatta · K. L. Druffel · S. D. Wyatt · H. R. Pappu (✉)
Department of Plant Pathology, Washington State University, PO Box 646430, Pullman, WA 99164-6430, USA
e-mail: hrp@wsu.edu

K. C. Eastwell
Department of Plant Pathology, Washington State University, IAREC, 24106 N. Bunn Road, Prosser, WA 99350, USA

Dahlia mosaic caulimovirus (DMV) is an important viral pathogen of dahlia (*Dahlia variabilis*) in the US and several parts of the world. First reported from Germany in 1928, the virus is considered to be one of the most important disease constraints affecting dahlias. DMV is a member of the family Caulimoviridae, genus *Caulimovirus* with a circular double-stranded DNA genome of approximately 8 kb [23]. The symptomatology, propagative hosts and the role of various aphid species in virus transmission have been reported [1, 4, 5]. The most characteristic symptoms of the disease include mosaic and vein-banding accompanied by stunting and leaf distortion. Symptoms vary depending on the cultivar, and cultivars with few or no symptoms are common. The economic impact of the disease is due to its negative affect on the quality and quantity of flowers produced by diseased plants.

The physical map of a DMV isolate collected from Portland, OR, USA (referred to as DMV-Portland in this paper) has been reported [23], and its genomic sequence, containing six open reading frames (ORFs), is available in GenBank (GenBank accession numbers: AY309480, AY309479, AY291588, AY291587, AY291586 and AY291585). However, our surveys of dahlias to determine the incidence of DMV revealed the overwhelming presence of another caulimovirus, designated DMV-D10 that was found to be distinct from DMV-Portland [17–21]. A partial sequence of a caulimovirus isolated from dahlia samples from the Netherlands (DMV-Holland) was recently reported [16]. To better understand the identity and taxonomic position of the caulimoviruses associated with dahlia mosaic, the genome structure and organization of the DMV-D10 genome was deciphered and was compared to those of DMV-Portland and other members of the genus *Caulimovirus*.

D10-infected dahlia plants were collected from the Dahlia Trial Garden, Tacoma, WA, USA. Total genomic DNA was extracted from infected dahlia leaves following a modified Dellaporta method [22] as described previously [21].

Initial cloning of the D10 genome was accomplished by using primers V4739 (5′-TGTCCACAAGGTCACTA-3′) and C5400 (5′-TKTCRTTRCTRTGGTAATT-3′) derived from the conserved regions of the reverse transcriptase (RT) gene of caulimoviruses. A 670-bp fragment was amplified, cloned and sequenced. Based on the sequence obtained, D10-specific primers were designed and the entire genome was cloned by primer walking. The PCR reactions included a final volume of 2 µl 10 × PCR buffer, 0.2 µl *Taq* polymerase (Promega, Madison, WI, USA) and a final concentration of 0.15 mM dNTPs, 2 mM $MgCl_2$ and 0.6 µM of each primer in a total reaction volume of 20 µl. PCR amplification was performed in a DNA thermal cycler (Bio-Rad, Hercules, CA, USA) programmed for 3 min at 94°C for initial denaturation and 50 cycles each consisting of 30 s at 94°C, 20 s at T_m and 1 min extension per 1,000 bp product at 72°C, followed by a final extension for 7 min at 72°C.

PCR products of expected size were cloned using pGEM-T (Promega, Madison, WI, USA) or TOPO TA (Invitrogen, Carlsbad, CA, USA) cloning kits. Some PCR products were sequenced directly after purification using a QIAquick PCR purification kit (QIAGEN, Valencia, CA, USA). The ca. 7-kb DMV-D10 genomic DNA (Fig. 1a) was cloned in six cloning steps (Fig. 1b). All clones were derived from a single plant. In each step, two specific primers were utilized for PCR amplification. Recombinant plasmids were isolated [24] and were sequenced using M13 forward and reverse primers, whereas PCR products were sequenced using the specific primer pairs that were initially used to obtain the PCR product. A minimum of two clones were sequenced from each cloned region to verify sequence data. Nucleotide sequences were determined using the ABI Prism Sequencing System at the Molecular Biology Core Laboratory of the Washington State University, Pullman, WA, USA.

Nucleotide sequences and their putative translation products were compared to other caulimovirus sequences available in GenBank [3] using BLASTN and BLASTX [2]. Pairwise alignment of amino acid sequences to determine sequence identity and similarity were carried out using the Needleman–Wunsch Global Alignment in EMBOSS [14]. DNA fragments were assembled into contigs, and the contigs were assembled to give a complete genome sequence using ContigExpress (Vector NTI Suite 9.0.0). Amino acid sequence alignments and phylogenetic analysis of each DMV open reading frame with ORFs of other caulimoviruses was done using CLUSTALW version 1.83 [28] and MEGA3 [10].

The D10 genomic sequence consisted of 7,046 bp with five complete ORFs. The ORFs were numbered to correspond to those found in members of other *Caulimovirus* species. Interestingly, only part of the ORF corresponding to the coat protein (CP) ORF found in other caulimoviruses (ORF IV) was present in DMV-D10, and it was fused in-frame with the ORF encoding the RT (ORF V). Therefore, this ORF in DMV-D10 has been termed ORF IV/V.

Complete genomes of eight D10-like isolates from various parts of the country were cloned and sequenced, and there was more than 95% sequence identity among them. The size of the DMV-D10 genome was comparable to those of other known caulimoviruses, which range from 7.2 to 8.0 kb. The A + T composition of the genome was 66.4%, slightly higher than that of other known caulimoviruses, for which the A + T composition ranges from 60.0 to 64.6%. Percentage similarity of DMV-D10 proteins to homologous proteins in other caulimoviruses ranged from 25 to 74% (Table 1). Figure 1a is a graphical representation of the DMV-D10 genome, showing the approximate location of each ORF discussed below.

Open reading frame I of DMV-D10 consisted of 990 bp encoding the movement protein (MP). A putative transport domain (GNLAYGKFMFTVY; amino acids 207-219), identified in other caulimoviruses to be important for virus cell-to-cell movement, was present within this region [7, 8]. Phylogenetic analysis showed that the MP of DMV-D10 was closest to that of figwort mosaic virus (FMV), whereas the MP of DMV-Portland clustered with that of mirabilis mosaic virus (MMV) (Fig. 2a).

The ORF II of caulimoviruses encodes an aphid transmission factor (ATF). A homologous ORF was not identified in DMV-D10. A smaller ORF found in the DMV-D10 genome did not show any significant homology with those available in GenBank. Previous studies on caulimovirus ATF have identified the amino acid sequence motif IXG, X being any amino acid, to be necessary for the interaction between the ATF and virus particles for aphid transmission [25]. The IXG motif is present in the ATF of DMV-Portland, cauliflower mosaic virus (CaMV) and carnation etched ring virus but absent in DMV-D10 ORF II.

Open reading frame III of DMV-D10 potentially encodes a DNA binding (DNAb) protein. The DNAb has a C-terminal proline-rich domain that is considered to be necessary for its non-sequence-specific nucleic-acid-binding activity [9, 13, 27]. This ORF contains the coiled coil structure that is conserved among caulimoviruses at the N-terminus and was shown to assemble as a tetramer to form a functional protein *in planta* [27]. The DNAb of DMV-

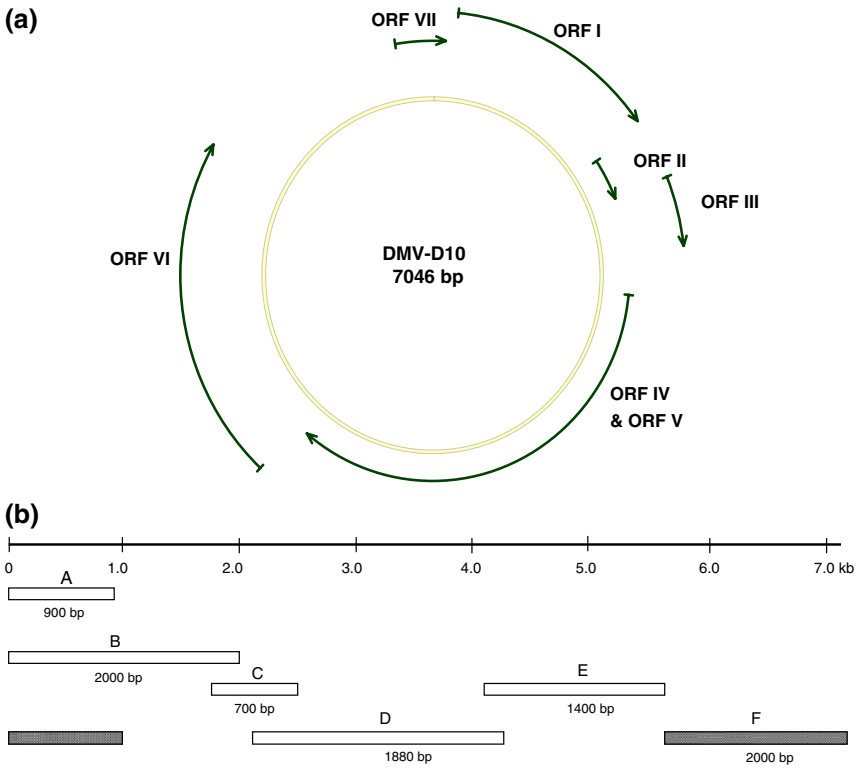

Fig. 1 **a** Dahlia mosaic virus genome organization. The *double-lined closed circle* represents the viral genome (not to scale). *Lines with arrows* represent the various open reading frames. ORFs I, III, IV, V and VI code for putative movement protein, DNA binding protein, coat protein, reverse transcriptase and inclusion body protein, respectively. The aphid transmission factor coded by ORF III in caulimoviruses is absent in the D10 genome. **b** A linear representation of the locations of the clones used in the sequencing of the entire genome of dahlia mosaic virus-D10. *Letters A–F* represent clones. The size of each clone is indicated in base pairs (*bp*). The *hatched box* represents a single clone

D10 DNAb formed a cluster with those of DMV-Portland, MMV and FMV (Fig. 2b).

Open reading frames IV and V of DMV-D10 occur as a single ORF due to the absence of a stop codon for ORF IV. The ORF IV of DMV-D10 consists of 100 aa representing only the amino terminal part of the viral CP, in contrast to ca. 480 aa ORF IV for known caulimoviruses. DMV-D10 does not have a conserved motif identified in other caulimovirus CPs, which include a RNA-binding domain that is consistent with a cysteine motif or "zinc finger" ($CX_2CX_4HX_4C$) [8]. Only part of this conserved motif (HX_4C) could be seen in the DMV-D10 sequence as HYANECP (aa 708-714). Furthermore, DMV-D10 CP also lacked a lysine-rich core upstream of the cysteine motif that was found in other caulimoviruses. The percentage amino acid sequence identity of DMV-D10 CP with other caulimovirus CPs was low (5.8–7.7%), and the DMV-D10 formed a clade with that of DMV-Portland (Fig. 2c).

Open reading frame V encodes a polyprotein containing all of the motifs conserved in caulimovirus replicases. These motifs include aspartic protease, RT and ribonuclease H (RNase H). The amino acid sequence YVDTGASLC (aa 820-828) in DMV D-10 was similar to the putative protease domain reported in other caulimoviruses [7, 8, 29]. A conserved amino acid sequence representing the RT domain of caulimoviruses was present in DMV-D10 as YVDDIIVF (aa 1161-1168) [7, 8]. Amino acid sequences IIETDASNDFWG (aa 1307-1318), EEVCRYTSGSFK (aa 1329-1340) and EHLEGSKNVLADCL (aa 1413-1426) in DMV-D10 are homologous to the RNase H motifs reported in other caulimoviruses [26]. The amino acid sequence identity of DMV-D10 with other caulimovirus polyproteins

Table 1 Percentage identity and percentage similarity of dahlia mosaic virus—D10 open reading frames (ORFs) to other caulimoviruses

Genus *Caulimovirus*	ORF I	ORF III	ORF IV	ORF V	ORF VI
Dahlia mosaic virus (DMV-Portland)	37.0 (53.0)	35.2 (50.8)	6.9 (11.6)	54.4 (64.8)	29.3 (40.1)
Cauliflower mosaic virus	41.7 (57.1)	25.9 (47.4)	7.5 (11.4)	58.3 (71.2)	18.6 (28.4)
Carnation etched ring virus	36.0 (49.5)	27.7 (33.0)	8.5 (11.3)	60.3 (71.5)	19.1 (31.8)
Figwort mosaic virus	39.4 (53.4)	32.8 (50.4)	7.2 (11.6)	70.1 (79.2)	26.6 (36.3)
Horseradish latent virus	39.3 (54.6)	35.9 (45.0)	7.3 (10.5)	56.2 (68.0)	19.1 (29.0)
Mirabilis mosaic virus	35.7 (51.7)	31.3 (45.5)	6.5 (12.7)	62.7 (73.0)	26.0 (37.3)
Strawberry vein banding virus	24.8 (41.6)	21.6 (32.1)	6.2 (11.9)	46.3 (58.8)	18.2 (29.6)

Percentage similarities are listed within brackets

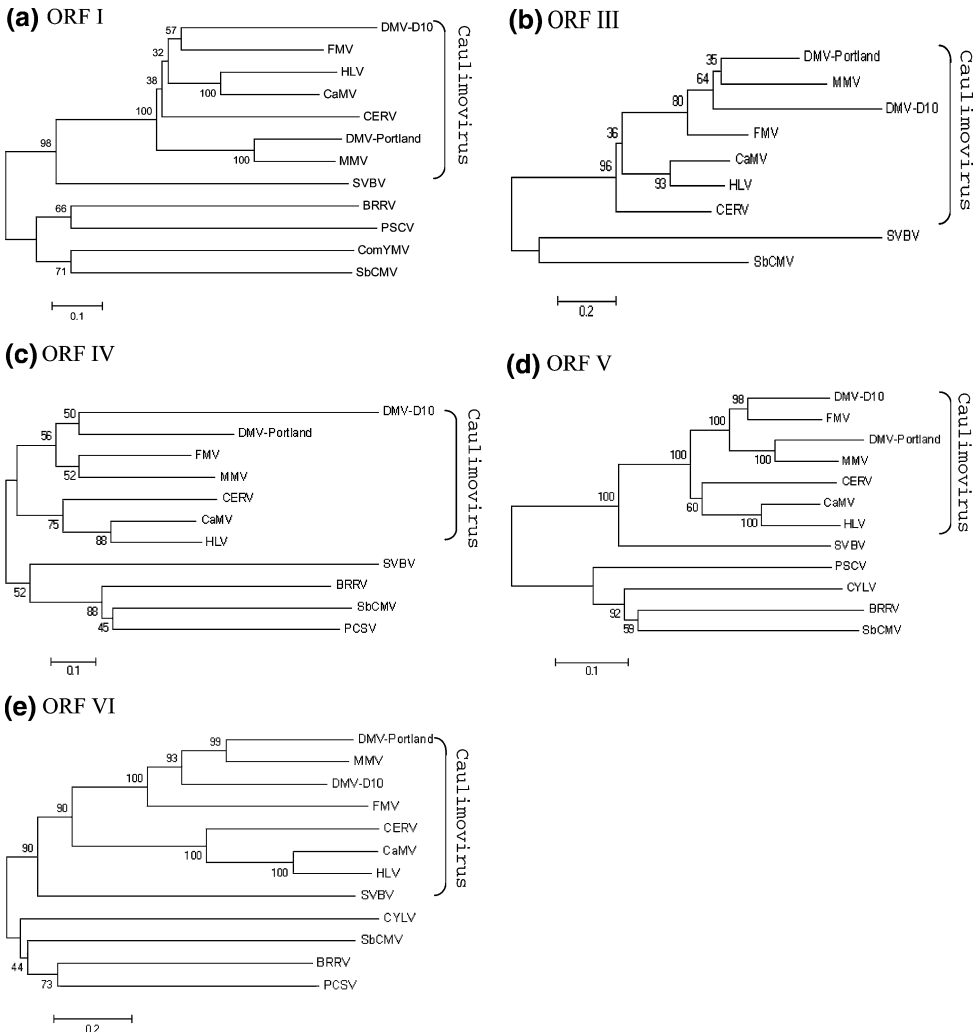

Fig. 2 Phylograms drawn from Clustal W alignments of the different ORFs of selected members of the family Caulimoviridae compared with those of dahlia mosaic virus—D10 and DMV-Portland. *BRRV* blueberry red ringspot virus, *CaMV* cauliflower mosaic virus, *CERV* carnation etched ring virus, *CYLV* cestrum yellow leaf virus, *ComYMV* commelina yellow mottle virus, *DMV* dahlia mosaic virus, *FMV* figwort mosaic virus, *HLV* horseradish latent virus, *MMV* mirabilis mosaic virus, *PCSV* peanut chlorotic streak virus, *SbCMV* soybean chlorotic streak virus, *SVBV* strawberry vein banding virus. Bootstrap values are indicated at branching points in the phylogram as a percentage of 1,000 iterations

ranged from 40 to 70%. The phylogram based on the protein sequences showed clustering of DMV-D10 with FMV, while DMV-Portland formed a cluster with MMV (Fig. 2d).

The ORF-VI of DMV-D10 encodes the viral inclusion body (IB) protein. Phylogenetic relationships among the IBs of caulimoviruses showed clustering of DMV-Portland and MMV, while DMV-D10 did not form a single cluster with any of the representative caulimoviruses (Fig. 2e). An amino acid sequence motif, GLTKYIY, that is well conserved in other caulimoviruses was also seen in DMV-D10 (aa 1717-1723) [7, 8].

The ORF-VII of DMV-D10 encodes a 73-aa protein following a 990-nucleotide intergenic region between ORF VI and ORF VII. A similar ORF is present in all known caulimoviruses, but functional proteins have not been found in infected plants [30].

The large intergenic region between ORFs VI and VII of DMV-D10 contains a putative 35S promoter homolog. The TATA box as well as the conserved residues TC around the TATA box, and the TGACG motif upstream of the histone genes and the CaMV 35S promoter [6] were also found in DMV-D10. However, a CAAT box that was found upstream of the TATA box in the 35S promoter of CaMV was not found in the DMV-D10 sequence.

The genome organization, sequence relationships and phylogenetic analysis of DMV-D10 suggested that it is a member of a distinct species within the genus *Caulimovirus*. DMV-D10 was found to be the most prevalent DMV sequence in dahlias in the US [20, 21]. While the overall genome organization of DMV-D10 shared the features common among known members of in the genus *Caulimovirus*, major differences are the truncated CP that is fused with the RT and the absence of the gene coding for the ATF.

Despite extensive sequencing of numerous D10-like isolates from several parts of the country, the gene coding for the ATF could not be found in the genome. Absence of the gene coding for the ATF could be explained by the fact that DMV-D10 is seed transmitted at a very high frequency

and could be detected in pollen [19]. Moreover, nearly 100% of the samples collected from several parts of the US and Europe were positive for D10 [20]. This widespread occurrence of D10 suggests the possibility of the integration of the D10 sequence in the host plant genome. Our studies showed that D10 could exist as an endogenous pararetrovirus [18]. There is also a possibility that the ATF of DMV-Portland may act in trans and facilitate virus transmission in instances where there are mixed infections. Dahlia is primarily vegetatively propagated, and the loss of ATF may be an indication of the influence of cropping procedures on the evolution and selection of DMV. A similar loss of aphid transmission after consecutive mechanical inoculations was reported for cucumber mosaic virus [15]. Leclerc et al. [12] showed that the C-terminal half of the CaMV CP is important for virus transmission through its interaction with the C-terminus of the ATF. It remains to be seen if the observed deletion of the C-terminal part of the CP in the D10 genome along with the absence of the ATF could affect the transmissibility of D10 by aphids. In contrast, the N-terminus of the CP was reported to be critical for viral infection in CaMV [11]. Severe symptoms were present in plants infected with only D10, supporting the critical role of the amino terminal part of the CP in the infection process by DMV-D10. An infectious DNA clone of D10 would facilitate studies on structure-function relationships. Despite our repeated attempts, no virions could be seen in either partially purified virus preparations or leaf dip preparations. This could be explained by the lack of a complete ORF IV, suggesting that functional CP molecules may not be synthesized in D10-infected cells. Whether this unique organization is a result of viral evolution brought about by continuous vegetative propagation of dahlias is of particular interest. The possibility of a satellite, extra genomic DNA or a helper virus that could code for the complete CP and ATF, although not known for caulimoviruses, however, could not be ruled out.

Studies on CaMV suggest that nucleic acid binding activity is the major function of the ORF III protein product and that it is related to aphid transmission via its interaction with ATF [13]. However, other reports have shown that the ORF III protein product is essential for infectivity within the plant host [9]. Therefore, it is possible that in DMV-D10, where ATF appears to be absent, the ORF III product potentially plays a role in virus infectivity in dahlia. DMV-D10 appears to be closely related to DMV-Portland and FMV when considering the phylogenetic relationships based on sequence relationships of ORFs I, IV and V. Overall, the complete molecular characterization of the DMV-D10 genome provides evidence of the existence of at least two distinct caulimoviruses that are associated with dahlia mosaic and highlights the need for developing management options for these viruses.

Acknowledgments Research was supported in part by the Samuel and Patricia Smith Endowment for Dahlia Virus Research, created by the American Dahlia Society, and funding from the USDA Northwest Nursery Crop Research Center. PPNS No. 0480, Department of Plant Pathology, College of Agricultural, Human and Natural Resource Sciences, Agricultural Research Center, Project # WNPO 0545, Washington State University, Pullman, WA 99164-6430, USA.

References

1. Albouy J, Lemattre M, Wang WC, Mevel A (1992) Production of pathogen tested dahlia in France. Acta Hortic 325:781–784
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
3. Benson DA, Karsch-Mizrach I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. Nucleic Acids Res 33:D34–D38
4. Brierley P, Smith FF (1950) Some vectors, hosts and properties of dahlia mosaic virus. Plant Dis Rep 34:363–370
5. Brunt AA (1971) Some hosts and properties of dahlia mosaic virus. Ann Appl Biol 67:357–368
6. Fang RX, Nagy F, Sivasubramaniam S, Chua NH (1989) Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell 1:141–150
7. Glasheen BM, Polashock JJ, Lawrence DM, Gillet JM, Ramsdell DC, Vorsa N, Hillman BI (2002) Cloning, sequencing, and promoter identification of Blueberry red ringspot virus, a member of the family Caulimoviridae with similarities to the "Soybean chlorotic mottle-like" genus. Arch Virol 147:2169–2186
8. Hasegawa A, Verver J, Shimada A, Saito M, Goldbach R, Van Kammen A, Miki K, Kameya-Iwaki, Hibi T (1989) The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter. Nucleic Acids Res 17:9993–10013
9. Jacquot E, Geldreich A, Keller M, Yot P (1998) Mapping regions of the cauliflower mosaic virus ORF III product required for infectivity. Virology 242:395–402
10. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163
11. Leclerc D, Chapdelaine Y, Hohn T (1999) Nuclear targeting of the cauliflower mosaic virus coat protein. J Virol 73:553–560
12. Leclerc D, Stavolone L, Meier E, Guerra-Peraza O, Herzog E, Hohn T (2001) The product of ORF III in cauliflower mosaic virus interacts with the viral coat protein through its C-terminal proline rich domain. Virus Genes 22:159–165
13. Leh V, Jacquot E, Geldrich A, Hermann T, Leclerc D, Ceruttio M, Yot P, Keller M, Blanc S (1999) Aphid transmission of cauliflower mosaic virus requires the viral PIII protein. EMBO J 18:7077–7085
14. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequences of two proteins. J Mol Biol 48:443–453
15. Ng J, Perry KL (1999) Stability of the aphid transmission phenotype of cucumber mosaic virus. Plant Pathol 48:388–394
16. Nicolaisen M (2003) Partial molecular characterization of dahlia mosaic virus and its detection by PCR. Plant Dis Rep 87:945–948
17. Pahalawatta V (2007) Biological and molecular properties of dahlia mosaic virus. Ph.D. thesis, Washington State University, Pullman

18. Pahalawatta V, Druffel K, Pappu HR (2008) A new and distinct species in the genus *Caulimovirus* exists as an endogenous plant pararetroviral sequence in its host, *Dahlia variabilis*. Proceedings of the international symposium on virus diseases of ornamental plants. Haarlem (in press)
19. Pahalawatta V, Druffel K, Pappu HR (2007) Seed transmission of dahlia mosaic virus in *Dahlia pinnata*. Plant Dis 91:88–91
20. Pahalawatta V, Miglino R, Druffel KL, Jodlowska A, van Schadewijk AR, Pappu HR (2007) Incidence and relative distribution of distinct caulimoviruses (genus *Caulimovirus*, family Caulimoviridae) associated with dahlia mosaic in *Dahlia pinnata*. Plant Dis 91:1194–1197
21. Pappu HR, Wyatt SD, Druffel KL (2005) Dahlia mosaic virus: molecular detection and distribution in dahlia in the US. HortScience 40:697–699
22. Presting GG, Smith OP, Brown CR (1995) Resistance to potato leaf roll virus in potato plants transformed with the coat protein gene or with vector control constructs. Phytopathol 85:436–442
23. Richins RD, Shepherd RJ (1983) Physical maps of the genomes of dahlia mosaic virus and mirabilis mosaic virus—two members of the caulimovirus group. Virology 124:208–214
24. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York
25. Schmidt I, Blanc S, Esperandieu P, Kuhl G, Devauchelle G, Louis C, Cerutti M (1994) Interaction between the aphid transmission factor and virus particles is a part of the molecular mechanisms of cauliflower mosaic virus aphid transmission. Proc Natl Acad Sci USA 91:8885–8889
26. Stavolone L, Ragozzino A, Hohn T (2003) Characterization of cestrum yellow leaf curling virus: a new member of the family Caulimoviridae. J Gen Virol 84:3459–3464
27. Stavolone L, Villani ME, Leclerc D, Hohn T (2005) A coiled-coil interaction mediates cauliflower mosaic virus cell-to-cell movement. Proc Natl Acad Sci USA 102:6219–6224
28. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
29. Torruella M, Gordon K, Hohn T (1989) Cauliflower mosaic virus produces an aspartic proteinase to cleave its polyproteins. EMBO J 8:2819–2825
30. Wurch T, Kirchherr D, Mesnard J, Lebeurier G (1990) The cauliflower mosaic virus open reading frame VII product can be expressed in *Saccharomyces cerevisiae* but is not detected in infected plants. J Virol 64:2594–2598