ECRONICON OPEN ACCESS

EC MICROBIOLOGY
Research Article

Phylogenetic Comparison Among *Caulimoviruses*

Ali Mahmoudpour*

Department of Plant Pathology, University of California, Davis, Davis, CA, 95616, USA

***Corresponding Author:** Ali Mahmoudpour, Ph.D., Norgen Biotek Corp., 3430 Schmon Parkway, Thorold, ON L2V 4Y6, Canada.

Received: November 25, 2016; **Published:** December 19, 2016

Abstract

The focus of this article is to touch up on the molecular biology and genetics of Strawberry vein banding virus (SVBV) and the distinctions among members of Caulimoviruses. Mode of virus transmission has divided these viruses into two subgroups. Viruses carrying aphid transmissibility factor are transmitted by various species of aphids. Other members lacking the gene responsible for aphid acquisition are generally seed transmissible. Due to similarity of virus replication to that of retroviruses, in absence of integration into host genome, Caulimoviruses were considered as pararetroviruses. Reports indicating evidence of genomic integration of Dahlia Mosaic Virus into the host genome corresponds with its seed transmissibility. Further studies are needed to trace the virus integration among other members especially those of seed transmissible members. Differences among the sizes of coat protein genes and observations on the multiple subunits of truncated coat protein that lead into differences of particle size introduce further distinctions among the members. The corresponding data in alignment analysis of amino acid sequences of similar genes among these viruses along with variable genomic organization provide evidences of further distinctions. These data demonstrate clear phylogenetic distinction among members of this group of plant viruses and ruling out that they have evolved from ancestral entity.

Keywords: *Caulimoviridae; Caulimoviruses; Replication and Expression; Phylogenetic Comparison*

Abbreviations

(SVBV): Strawberry Vein Banding Virus; (Camv): Cauliflower Mosaic Virus; (ORF): Open Reading Frame; (ATF): Aphid Transmission Factor; (TAP): Transactivator Protein; (Psvbv-E3): Name of the SVBV DNA Clone Inserted in Puc8 Plasmid

Introduction

Caulimoviridae and *Caulimoviruses*

Caulimoviridae is the small family of plant viruses consisting of the single genus of *Caulimovirus* comprised of 9 - 10 viruses which were grouped together due to particle morphology and genomic structure composed of dsDNA. Further studies provided dissimilar characteristics including host range, mode of viral transmission and distinct genomic properties hence distanced these members from each other. The initial virus infection was defined through mosaic symptoms in cauliflower hence the name of virus family and genus were derived from and represented by CaMV. Unique characteristics of this virus distanced it significantly from other common plant viruses. Molecular biology of *Caulimoviruses* was studied extensively by numerous scientists in recent decades. Significant number of reports was made on CaMV and a brief review recently reported by this author [1].

Viral Transmission among *Caulimoviruses*

Members of the *Caulimoviridae* are divided by mode of transmission. Those *Caulimoviruses* including CaMV, FMV, CERV, and SVBV car-

rying a gene so called aphid transmission factor (ATF) are naturally transmitted by different species of aphids in semi-persistent manner from feeding on infected plants and spreading the virus to other plants. The ATF is not found in other members including SbCMV, PCSV, PVCV, DMV, and MMV or CVMV. These members are not aphid transmissible and their seed transmissibility has been described [2-4]. PVCV with an unusual genomic organization (encoding only two ORFs) is differentiated from the other *Caulimoviruses* significantly. CVMV, which has spherical virions and intermediate characteristics, has been suggested to be considered as an intermediate member between families of *Caulimoviridae* and *Badnaviridae* [5,6].

Dahlia mosaic virus (DMV) infecting Dahila variabilis has been reported to be a *Caulimovirus* which integrates into host genome and uses seed and pollen transmissibility. It was considered to be endogenous para-retrovirus (EPRV) while having similar ORFs to CaMV except truncated coat protein (ORF IV) and lacking ATF [4]. This description (integration of viral genome into host chromosome seen among retroviruses) remains to be studied among other *Caulimoviruses* especially those members which are seed transmissible and lack gene for ATF.

A Brief History of Genetic Studies in SVBV

Stenger and coworkers purified and cloned an approximately 8.0-kb ds-DNA genome of the virus into a pUC8 cloning vector to generate pSVBV-E3. SVBV was detected in the Czech Republic [8] and the complete nucleotide sequence of the pSVBV-E3 was determined [9]. Unlike CaMV genome, which was shown to consist of 8031 bp [10], the SVBV sequence consisted of 7,876 bp derived from sequencing SVBV-E3 clone [9]. The sequence data was confirmed by this author while demonstrating the infectivity of the Stenger's pSVBV-E3 clone [11]. Comparison of SVBV sequence with that of CaMV [12] provides that the SVBV genome is 155 bp shorter. Unlike CaMV, where an extra ORF of VII with an undetectable protein has been introduced, such unknown ORFs have not been seen or reported in SVBV sequence illustrated in Figure 1 [1,13].

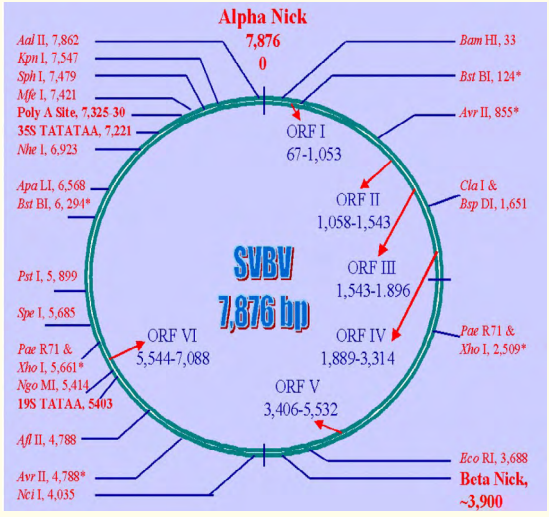

Figure 1: *Physical map of the strawberry vein banding virus genome, showing locations of restriction enzymes that cleaves once or twice (as marked with *). Putative promoter sequences of gene VI (19 S RNA) and the full-length genomic (35 S RNA), transcriptional termination site (Poly A), and single-stranded nicks in the genome are shown in bold.*

Using the sequence of the clone along with other previous data the genomic organization was delineated and the assignment of SVBV as a distinct *Caulimovirus* species substantiated [14]. Comparison of a 431-bp PCR-amplified DNA fragment corresponding with coat protein from six SVBV isolates collected in the Czech Republic and the United States did not yield significant sequence differences. So, it was concluded that they were all isolates of the same virus. It was speculated that since the virus was first described in California and Oregon, it must have been introduced into Europe via importation of infected plant materials [15].

Materials and Methods

Phylogenetic Analysis of SVBV among its close relatives

"Pile up" program (SeqWeb version 1.1 of Genetics Computer Group (GCG), version 10, Wisconsin package, Madison, WI) was used to construct multiple amino acid sequence alignments of SVBV genes with those of all other available *Caulimoviruses* until 2000 [including Cauliflower mosaic virus (CaMV), accession J02048, Franck and coworkers Carnation etched ring virus (CERV), accession Z71511, Palkovics & balazs, [17]; Cassava vein clearing virus (CVMV), accession U59751, De Kochko and coworkers; Peanut chlorotic stunt virus (PCSV), accession U13988, Richins, [18]; Figwort mosaic virus (FMV), accession X06166, Richins and coworkers; Petunia vein clearing (PVCV), accession U95208, Richert-Poeggeler & Shepherd, [3]; Soybean chlorotic mosaic virus (SbCMV), accession X15828, Hasegawa and coworkers; Strawberry vein banding virus (SVBV), accession X97304, [14] and alignment data are presented at the supplement. Cassava vein clearing virus (CVMV), the closest relative of the *Caulimoviridae*, was included as an out-group member in this analysis. The entire amino acid sequences of movement, capsid, and replicase proteins of these viruses and their conserved core region (determined as the longest stretch of aligning amino acids without any gap or interruption) were aligned separately. To compute pairwise distances and generate phylogenetic dendrograms, multiple sequence alignments were analyzed using PAUP (version 4.0b2 for Macintosh™, Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts). "GAP" (global alignment program, GCG version 10) was used for pairwise comparison of entire amino acid and DNA sequences of all SVBV genes with those of other *Caulimovirus* sequences. In GAP analysis, percent identity, percent similarity and number of gaps were demonstrated. Default penalty values of programs were applied in all analyses. The amino acid sequence inferred from translation of DNA sequence of a petunia vein clearing virus (PVCV) clone was delineated in this study and its genomic organization compared to other viruses.

Results and Discussion

Genetic Organization of SVBV

The double-stranded circular-DNA SVBV genome consists of 7,876 bp and six open reading frames (ORF) is illustrated in Figure 1. According to close similarity to genome of CaMV, ORF I to ORF VI encodes movement protein (MP), aphid transmission factor (ATF), non-sequence specific DNA-binding protein (also defined as second ATF), capsid protein (CP), multi-functional enzymatic protein acting as RNA/DNA dependent DNA polymerase (replicase)/RNAase H/proteinase [21-23] and lastly transactivator protein (TAP) through ORF VI. The 19S promotor or TATAA box at position 5,403 is responsible for synthesis of 19S mRNA which includes only ORF VI, apparently for over-expression of TAP and initiation of viral infection, which is the site of virus replication and encapsidation in the cytoplasm of infected cells. The well-known 35S promotor starts with the TATATAA box positioned at 7,221-7 of the genome along with a stretch of poly A at 7,325-30. The promotor has been shown to overlap with ORF VII in CaMV as well as including several short open reading frames, apparently associated with gene expression in terms of ribosomal shunt, but no protein or activity has been documented so far for these open reading frames [24].

Sequence analysis of SVBV also revealed transcriptional elements known to be present in other Caulimoviruses (Figure 1). A 35S promoter sequence (TATATAA) was identified at position 7,221, and transcriptional termination sequences or poly A signals (AATAAA) were located at positions 7,325 and 7,590. In contrast, a single such sequence is known for CaMV. The termination sequence is presumed (see below) to be same for both the 19S and 35S RNAs. If this is the case, the messenger RNA (mRNA) for gene VI would have a 3' non-coding

sequence of 337 nucleotides, and the genomic RNA (35S RNA) would have 115 nt at 3'-terminal redundancy [1,13]. If the second termination sequence is active, then these 3' noncoding sequences would be longer (502 and 370 nt, respectively). In CaMV, the 3' non-coding sequences are 268 and 207 nt, respectively. In FMV, these sequences are 162 and 174 nt, respectively [25]. Promoter sequences for the SVBV 19S RNA (TATAA) were located at positions 5,335 and 5,404. The second site is the more likely promoter in SVBV, because the coding sequence of gene VI starts at position 5,544 and this would result in a small (141 nt) 5' non-coding sequence. In CaMV, the 19S RNA leader sequence is 45 nt, whereas that of FMV is 77 nt.

Based upon examination of aligned amino acid sequences of SVBV and other *Caulimoviruses* it was determined that the replicase gene was the most highly conserved sequence. Therefore, this sequence was selected for phylogenetic analyses. The phylogenetic relationships among SVBV and other *Caulimoviruses* determined by parsimony analysis of the entire and the conserved core regions of the replicase protein amino acid sequence (Figure 2). Phylogenetic analysis of the full-length nucleotide or amino acid sequences of the movement protein (ORF I) revealed more distant relationships among these viruses compared with the replicase gene (data not shown). Clusters identified from the ORF I analyses, included CaMV/FMV/CERV, SVBV, and SbCMV/PCSV. PVCV and CVMV ORF I sequences were too divergent, and could not be aligned for inclusion in these analyses. Analysis of the conserved core region of the movement protein (not including PVCV) provided results that were similar to that observed for the ORF V analysis (data not shown).

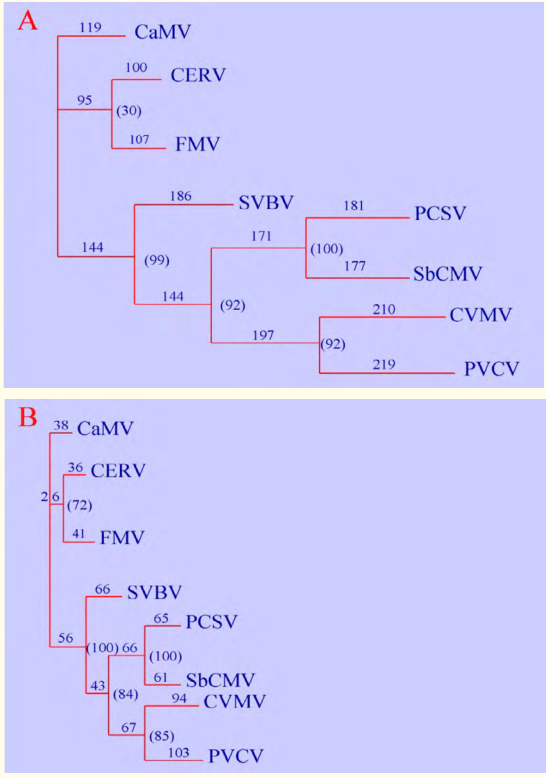

Figure 2: *Phylogenetic dendrograms generated by parsimony analysis of amino acid sequences of the entire replicase gene/ORF V (Figure 2A) or conserved core regions (Figure 2B) from various Caulimoviruses with CVMV as an out group. The coordinates of conserved regions are shown in the figure. For this analysis, the most parsimonious trees were generated using a heuristic search of all possible trees. Numerical values for each branch length are provided and reflect the number of changes between branch nodes. Bootstrap analysis was performed to provide statistical (P = 0.05%) support for the topology of the most parsimonious trees. The percent bootstrap values for each phylogenic clad is shown in parenthesis.*

The phylogenetic analysis conducted with ORF IV (CP) produced similar results to those from the ORF I analysis, except that the relative positions of FMV and CERV were switched in the first cluster (data not shown). The CP sequences of PVCV and CVMC did not align sufficiently with those of any of the other viruses and were excluded from the analysis. The aphid transmission gene (ORF II) is present only in CaMV, FMV, CERV, and SVBV (i.e., aphid-transmissible *Caulimoviruses*). SbCMV, PCSV, PVCV and CVMV lack this gene or had no homologous domain in their polyproteins (i.e., non-aphid transmissible viruses).

Phylogenetic analysis of SVBV based on replicase (ORF V) amino acid sequence placed SVBV in an intermediate position between two *Caulimoviruses* clusters, one with CaMV/CERV/FMV, and another with PCSV/SbCMV. The newly described PVCV is a tentative member of the *Caulimoviridae* that encodes only two polyproteins (analyzed in this study). Together with CVMV, PVCV had a more distant relationship to SVBV and other *Caulimoviruses*. Sequence comparison showed that movement and coat protein genes were more divergent than the replicase gene. Moreover, the divergence in these genes separated PVCV and CVMV further from the other *Caulimoviruses* in the phylogenetic analyses. However, similar results were obtained with other genes, suggesting that the relationships shown here are real.

Data in Table 1 describe the percent similarity and number of gaps created in pairwise comparisons between nucleotide and amino acid sequences of all SVBV ORFs and those of other *Caulimoviruses*. The overall DNA and protein identity/similarity between SVBV ORFs and those of other viruses was low (e.g., ranging from 18 to 63%). The large number of gaps within the aligned sequences further reveals the high degree of diversity among these viruses. Number of amino acid differences of SVBV movement, capsid and replicase proteins (intact gene and conserved core region) compared to other members of *Caulimoviridae* is described in Table 2.

ORF I	**SVBV**	**CaMV**	**CERV**	**FMV**	**SbCMV**	**PCSV**	**PVCV**	**CVMV**
DNA Similarity, %	100	48.7	48.2	49.2	51.4	47.3	39.2	45.1
Gaps in DNA: DNA	0	4	5	3	7	6	7	7
Amino Acid Residues	328	327	319	323	303	320	1105	1199
AA Identity, %	100	35.3	31.2	33.7	29.5	28	20	24.1
AA Similarity, %	100	45.1	45.3	45.7	42.4	41.6	20	35.4
Gaps in AA: AA	0	4	6	3	2	5	0	6
ORF II	SVBV	CaMV	CERV	FMV	SbCMV	PCSV	PVCV	CVMV
DNA Similarity, %	100	42.6	45	42.3	N/A	N/A	N/A	N/A
Gaps in DNA: DNA	0	3	3	4	N/A	N/A	N/A	N/A
Amino Acid Residues	161	159	168	164	N/A	N/A	N/A	N/A
AA Identity, %	100	20.1	25.2	15.7	N/A	N/A	N/A	N/A
AA Similarity, %	100	36.4	35.5	27.7	N/A	N/A	N/A	N/A
Gaps in AA: AA	0	4	4	2	N/A	N/A	N/A	N/A
ORF III	SVBV	CaMV	CERV	FMV	SbCMV	PCSV	PVCV	CVMV
DNA Similarity, %	100	42.8	37.5	40.2	N/A	N/A	44.3	N/A
Gaps in DNA: DNA	0	3	1	2	N/A	N/A	3	N/A
Amino Acid Residues	117	129	128	115	N/A	N/A	1105	N/A
AA Identity, %	100	20.5	25	19.8	N/A	N/A	18.1	N/A
AA Similarity, %	100	34	33	32.4	N/A	N/A	28.4	N/A
Gaps in AA: AA	0	4	0	4	N/A	N/A	3	N/A
ORF IV	SVBV	CaMV	CERV	FMV	SbCMV	PCSV	PVCV	CVMV
DNA Similarity, %	100	46.4	50.6	49.4	44.5	45.4	42.7	44

Gaps in DNA: DNA	0	10	14	11	11	14	15	13
Amino Acid Residues	474	489	494	489	440	462	1075	1199
AA Identity, %	100	33	30.6	32.1	27.5	26.7	26.8	42.8
AA Similarity, %	100	46.5	41.9	43.6	38.8	37.3	36.3	42.8
Gaps in AA: AA	0	11	13	12	13	15	2	0
ORF V	SVBV	CaMV	CERV	FMV	SbCMV	PCSV	PVCV	CVMV
DNA Similarity, %	100	58.7	57.9	58.2	49.9	51.6	45	51.6
Gaps in DNA: DNA	0	7	4	3	16	17	16	13
Amino Acid Residues	708	679	659	666	741	694	1075	652
AA Identity, %	100	53.8	51.8	51.9	36	35	27.3	35.5
AA Similarity, %	100	61.7	62.6	60.8	47	45.1	37	46.7
Gaps in AA: AA	0	7	5	2	14	14	11	14
ORF VI	SVBV	CaMV	CERV	FMV	SbCMV	PCSV	PVCV	CVMV
DNA Similarity, %	100	44.6	41	44.3	42.9	42.6	N/A	40.8
Gaps in DNA: DNA	0	17	9	14	16	12	N/A	8
Amino Acid Residues	514	520	496	512	463	420	N/A	392
AA Identity, %	100	23.8	21.5	20.6	18.7	23.1	N/A	36.2
AA Similarity, %	100	32.7	32.2	30.8	37.5	32.2	N/A	22
Gaps in AA: AA	0	11	11	7	0	12	N/A	19

Table 1: *Gap Analysis. Comparison of DNA and amino acid (AA) sequences of SVBV open reading frames with those of other viruses generated with GCG default penalty values.*

ORF I (movement protein)					
	Residue	Difference	Core Region	Residue	Difference
SVBV	328	0	112 thru 266	155	0
CaMV	327	198	121 thru 275	155	92
FMV	323	194	118 thru 272	155	88
CERV	319	208	113 thru 267	155	93
SbCMV	303	209	96 thru 249	154	92
PCSV	320	233	107 thru 260	154	106
CVMV	898-1199	225	N/A	N/A	N/A
ORF IV (capsid protein)					
	Residue	Difference	Core Region	Residue	Difference
SVBV	474	0	152 thru 361	210	0
CaMV	489	279	187 thru 386	200	127
FMV	489	303	179 thru 386	208	130
CERV	494	286	192 thru 392	201	125
SbCMV	440	317	149 thru 358	210	151
PCSV	462	339	156 thru 368	213	158
CVMV	298 thru 822	332	N/A	N/A	N/A

ORF V (replicase protein)					
	Residue	Difference	Core Region	Residue	Difference
SVBV	708	0	282 thru 588	307	0
CaMV	679	322	250 thru 555	306	123
FMV	666	321	243 thru 548	306	127
CERV	659	310	232 thru 539	308	128
SbCMV	741	383	207 thru 516	310	160
PCSV	694	370	195 thru 501	307	163
CVMV	652	390	207 thru 521	315	170
PVCV	43 thru 759	462	285 thru 588	304	184

Table 1: *Phylogeny of SVBV. Number of amino acid differences of SVBV movement, capsid and replicase proteins (intact gene and conserved core region) compared to other members of Caulimoviridae.*

The replicase gene, as has been noted in a previous report [14], was the most conserved sequence among the *Caulimoviruses* investigated in this study. Furthermore, by removing the more divergent 5' and 3' ends of the replicase gene from the analysis, higher percent identities were obtained, suggesting a greater degree of conservation for the core sequence. This is consistent with the core of the replicase being conserved among *Caulimoviruses*. Sporadic studies concerning members of *Caulimoviruses* have been reported since 2000. New genomic sequences of DMV and MMV have been reported which were not available during phylogenetic studies conducted in comparing SVBV with available data. DMV was reported to be integrated into host genome [4]. Number of seed transmissible virus was increased after sequence data confirmed the lack of ATF in half of *Caulimoviruses*.

CaMV has been widely studied by numerous distinguished scientists while they had access to herbaceous hosts and all kinds of natural and laboratory means to delineate the representative member of *Caulimoviridae*. Reviewing these publications demonstrates that CaMV could only represent itself with limited similarities to other members. Half of the members have been shown to be seed transmissible and lacked aphid transmission factor. Reports have been made that DMV or possibly other viruses of the family are integrated into host genome. This investigation has not generally included other members yet. The *Caulimoviridae* span a range of different genomic organizations, encoding from two to seven open reading frames. The details of the expression of those genes are not yet well characterized even for the CaMV type member.

Lack of a specific antibody to SVBV hindered the identification of the coat protein by Western-blot analysis. The same virion products used in DNA isolation and Southern blot hybridization analysis were used in SDS-PAGE analysis of SVBV coat protein [13,26]. Three 45-60 kDa-protein bands were observed in preparations obtained from the infected plants. These were presumably subunits of SVBV capsid protein produced by proteolytic cleavages of the full-length polypeptide. Multiple coat protein bands (37, 39, 44, and 57 kDa), all cross reacting with antibody raised against 37-kDa protein [27] have been reported for CaMV [28]. The predicted molecular weight of the full-length SVBV coat protein is 56.0 kDa (474 residues), compared with 56.7 kDa (489 residues) for CaMV. In addition to the presence of numerous basic amino acids in the coat protein of SVBV (55 lysine and 30 arginine residues) and CaMV (57 lysine and 22 arginine residues), it has been shown in CaMV that coat protein subunits and virion surface are heavily glycosylated [29] and phosphorylated [30]. Therefore, these subunits move slower in the gel, and estimation of exact sizes is difficult. Proteolytic activity of the replicase protein (protease domain), found also in virus shells, cleaves the native coat protein into required subunits [31] used in assembly of virions consisting of 420 subunits [32]. The exact sizes of these subunits and locations of cleavage sites remain unclear.

Conclusion

The sizes of coat protein genes are not consistent across the family. The observation of several sizes of CP in CaMV or other members indicate that cleavage events influence the sizes of virion particles, which range from 45 to 53 nm in diameter. The *Caulimoviridae* could not be considered as para-retroviruses, yet reports have been made implying viral integration into host genome. Improving techniques and the ease of sequencing through NGS and targeted NGS will be able to elucidate some of questions regarding the molecular biology of this family of viruses, members of which are possibly far apart from each other. Regardless of basic similarities, grouping them into a single family with initial findings including their ds-DNA genome and icosahedral structure, they are not necessarily evolved from a single ancestral virus. Like other biological entities, they are composed of similar building blocks. As a matter of fact, gene homology and the concept of evolution and the molecular clock may not necessarily organize the creatures of the world into the particular orders we may wish. There are too many un-known mysteries in the creation of the universe to allow us to understand the life using our simplified vision. The more we find out, we discover our lack of knowledge in describing the uncertainty of nature as controlled by the Mighty Creator.

Acknowledgements

This study was accomplished at the Department of Plant Pathology and Foundation Plant Service (FPS), University of California, Davis. Hereby, I express my deepest gratitude to Dr. Adib Rowhani for kindly sponsoring the project and Dr. Stephan D. Daubert for being my mentor throughout conducting my project.

Conflict of Interest

I declare that there is no financial interest or any conflict of interest exists.

Supplementary Data

These raw data illustrate multiple sequence alignment of amino acid of corresponding genes (intact gene or the core part) of viruses used in phylogeny analysis. The amino acid sequences were deduced from the published data of DNA sequences available at the time of analysis.

Multiple Sequence Alignment

ORF I Full-Length

Gap Weight: 8

GapLengthWeight: 2

```
          1                                                  50
CaMV MDLYPEENTQ SEQSQNSE.. ..NNMQIFKS ENSDGFSSDL MISNDQLKNI
 FMV ...MCSTRKT SVMDEKVI.. ..ENEEIHFQ ENSNGFSADL TIHQDKLKQI
CERV .......MNS SVEKQNSEIP EKENEEFTFQ DNSQGFELEF STNKKTLSKI
SVBV .........M SEEEIRMDQP QGGHDEYIFE EEGT.YAHDV AIDSTLLKEI
SbCMV .......... .......... ...MEIVELK DDNQEYFLDA LLG....KEI
PCSV .......... ...MSREESS MSEIQEIVSH EEEGSYLVDI LLDNKLVTKI
CVMV .......... .......... .....IKEYN KTEPEKINKI IFTSEKFKQI
```

```
  51                                                  100
CaMV SKTQ.LTLEK EKIFKMPNVL SQVMKKAFSR KNEILYCVST KELSVDIHDA
 FMV SKTG.LNLEK EHIFNMPSSL TKAFKTAFKR KNEIFYCVST KEMSVDIKDV
CERV QKAN.LSLKT NDAFNI.... ..SFLKAFSR KNHIYYHVNY KEFSVDICDT
SVBV EKKD.LELTS EEVFKTPS.L WKKFLKA..R KNICISCVSS REYPIEIAQA
SbCMV EKTD.FSITE KENFK..QNK FKELRNVFSR DNILKFGLMT GEVQIPIEQT
PCSV KNNGHLDLES NEGIN..... .QKILEKINR KNIIYNGIMC GEQSVPIEQA
CVMV MEND.LNMTK DKIFH..... NNKLKKLFGK KEIEYYIVTD IEHPIDVKYV

 101                                                  150
CaMV TGK...VYLP LITKEEINKR LSSLKPE.VR KTMSMVHLGA VKILLKAQFR
 FMV SGQ...VYLP LITKQEIQQK LMKIDPS.VR SKISMIHLGA VKILLTAQFR
CERV HGK...NYLP LVTKSEIKKN LDKIKDEKVR STISDIHFGA IKVLIKARFR
SVBV NGL...TEIP FFNREEIESK KRVLKPED.R KKIDFIHIGS VKIMIKSTFR
SbCMV DGS...VFLA TINKEQITKR ISKIE.EKQR RLIRYVHIST LQVLIKSTFL
PCSV TAK...FQAP VVKISQVQER LKKIK.ETDR KKIGFIHVNV VQIVIRSTFR
CVMV QNQDKIINLP LYNQEIFENE IQKI.PDKDQ NKIRNIHLAA VEIVVKAYFR

 151                                                  200
CaMV NGIDTPIKIA LIDDRINSRR D.CLLGAAKG NLAYGKFMFT VYPKFGISLN
 FMV QGIDTSVKMA LIDDRIVNRK D.SLLGAARG NLAYGKFMFT VYPKFALSLQ
CERV EGINSPIKMA LIDDRITDRQ D.SILGAAHG NLVYGKFMFT VYPKYTTSIL
SVBV TGIDSPISVA LLDRRMKNAK D.AVFGGVKG NLSYGKLIFT YNPKISVSLR
SbCMV KGLDTPLELT LRDNRLLNLE E.SKIAVGHG NLKYGKMKFD VNLQLGLSLK
PCSV EGITTPVIIR VEDNRIQDKR Y.SLLGQIEG DLGYGVIKFN VTLQYPIPLI
CVMV EGIDTPFEII LCDDRITYPQ EGSLVEVLIG NLIYQKVKFT KIINYSISIE

 201                                                  200
CaMV TQRLNQTLSL IHDFENKNLM NKGDKVMTIT YVVGYALTNS HHSIDYQS.N
 FMV SKNLDKTLSF IHQFERKDLM KTGDKVFTVT YLIGYALTNS HHSIEYRK.N
CERV DQRLDRTLAF IHHFERNDLM RKGDKVFSIT YLVAYALANS HHSIDYKE.K
SVBV DPTINKTLTL AHFFEKEELM HEGNHPYTIS YKIGYTLSNS HHSLEFRP.K
SbCMV DLDLDRSIIL NYKFLRRNFM KEGNHAFSIS YRINYALSNS HHSVEFKQ.K
PCSV TRSFNNCIGV ICEFRKQDLM KQGDIPLVVC YRIAYALTNS AISLQYKH.L
CVMV DKNLDKSLVM YWNLEGIKMI KD.SKIFSIR LRNLYVLSNK HIVKNKKQYN

 251                                                  300
CaMV ATIELEDVFQ EIGN.VQQSE FCTIQNDECN WAIDIAQ.NK ALLGAKTKTQ
 FMV SNIEIEEVFK DIGQ.IEESP FCDISPIDEN WTMDIAR.NK KSIATGSSSR
CERV DAIEIDDVFS EIGS.VKSPT FTELDPEPNS WAIDIAQ.GK QPIGFKPKPT
SVBV EPICIDDLFS TVGK.ISQAP LQEITPIENY WRMSLGNTSR RLLGERPRMI
SbCMV EKIYIDELFS EVLE.LKHPV FSKLTKSQSL RIEPSPVFEK PLISFKENQK
PCSV DRLYTNKIFS ETSTIIRTDQ VQKNFLREHS QRFPSLQIGE S..SSQNRLI
CVMV GNIIIEPIFQ DVIQNNNRNY IEYGKP.... .....GKFDR TKLKSYSRRF
```

```
        301                                                  350
CaMV IGNN...... L.QIG.NSAS SSNTE..... ..NELARVSQ NIDLLKNKLK
 FMV ..RN...... F.RID.ESLL ENKDE..... ..NLLRSMST KIDTLGKKLS
CERV VSNN...... FLRFDKETSP SSSHQ..... ..KSLEEISD KIDTLVVKLN
SVBV VIEDDEEETP LDRHTERALA RSQSRMLGLR PLEDLRTTSR KINTLAHQL.
SbCMV TEEKTVFKPP ..KRDFELTE TSKLKSMISD LTQKVVNLDK KI........
PCSV PTESGSQLAP PTR..KDNLY KEDREDQIEK IRKQVNELST VVTSLDKRI.
CVMV NEPLRLDDRT NIQREKDQIE KADHNLELQK ELNNLNYYSQ QGQS......

        351
CaMV EICGE.
 FMV LIYDNE
CERV NIS...
SVBV ......
SbCMV ......
PCSV ......
CVMV ......
```

ORF I core region

Gap Weight: 8

GapLengthWeight: 2

```
   1                                                    50
CaMV RKTMSMVHLG AVKILLKAQF RNGIDTPIKI ALIDDRINSR RDCLLGAAKG
 FMV RSKISMIHLG AVKILLTAQF RQGIDTSVKM ALIDDRIVNR KDSLLGAARG
CERV RSTISDIHFG AIKVLIKARF REGINSPIKM ALIDDRITDR QDSILGAAHG
SVBV RKKIDFIHIG SVKIMIKSTF RTGIDSPISV ALLDRRMKNA KDAVFGGVKG
SbCMV RRLIRYVHIS TLQVLIKSTF LKGLDTPLEL TLRDNRLLNL EESKIAVGHG
PCSV RKKIGFIHVN VVQIVIRSTF REGITTPVII RVEDNRIQDK RYSLLGQIEG

  51                                                   100
CaMV NLAYGKFMFT VYPKFGISLN TQRLNQTLSL IHDFENKNLM NKGDKVMTIT
 FMV NLAYGKFMFT VYPKFALSLQ SKNLDKTLSF IHQFERKDLM KTGDKVFTVT
CERV NLVYGKFMFT VYPKYTTSIL DQRLDRTLAF IHHFERNDLM RKGDKVFSIT
SVBV NLSYGKLIFT YNPKISVSLR DPTINKTLTL AHFFEKEELM HEGNHPYTIS
SbCMV NLKYGKMKFD VNLQLGLSLK DLDLDRSIIL NYKFLRRNFM KEGNHAFSIS
PCSV DLGYGVIKFN VTLQYPIPLI TRSFNNCIGV ICEFRKQDLM KQGDIPLVVC

  101                                                  150
CaMV YVVGYALTNS HHSIDYQSNA TIELEDVFQE IGNVQQSEFC TIQNDECNWA
 FMV YLIGYALTNS HHSIEYRKNS NIEIEEVFKD IGQIEESPFC DISPIDENWT
CERV YLVAYALANS HHSIDYKEKD AIEIDDVFSE IGSVKSPTFT ELDPEPNSWA
SVBV YKIGYTLSNS HHSLEFRPKE PICIDDLFST VGKISQAPLQ EITPIENYWR
SbCMV YRINYALSNS HHSVEFKQKE KIYIDELFSE VLELKHPVFS KLTKSQ.SLR
PCSV YRIAYALTNS AISLQYKHLD RLYTNKIFSE TSTIIRTDQV Q.KNFLRFPS
```

```
  151
CaMV IDIAQ
 FMV MDIAR
CERV IDIAQ
SVBV MSLGN
SbCMV IEPSP
PCSV LQIGE
```

ORF IV Full Length

Gap Weight: 8

GapLengthWeight: 2

```
        1                                                  50
CaMV MAESILDRTI NR...FWYNL GEDCLSESQF DLMIRLMEES LD.GDQIIDL
CERV .........M NREAILWKNI ..NSIPEEP. DLIKSLEVLS MEQNDREREL
 FMV .......... .......... .......... .....MATKK MRISEKLWDS
SVBV .......... .......... .......... .......... ..........
PCSV .......... .......... .......... .......... ..........
SbCMV .......... .......... .......... .......... ..........
CVMV EKQKKEDMML EMIKNLQNEL EQLKIQRHKE HEKQAELTKI QMLEEELEEE

        51                                                 100
CaMV TSLPSDNLQV .EQVMT.TTE ..DSISEEES EFLLAIGETS E.........
CERV EHNLILNKQI SEQIPEWIIP ..DSLSELSS GIDLNFVLEE Q.........
 FMV LEKDECNIDE VVQLMSLDEE ..ELITKFAQ EVSLRIRYLD EKPNEISFIA
SVBV ...MVSRRER LEQLFDEDHP EMDIIIQYLS LLDHELDDCQ EEKLQVLVAK
PCSV .......... .MDNIDRLTQ LLEKM.NLGN MAKLDEEDAE GIIRGIDSDD
SbCMV .......... MEETQQELTQ QLKELETLMA AINLDDSKQN QPIYQNSSES
CVMV LDPDNLEKEV LNNIQN..IQ ISSDISE.SS EINEISDNET EQISGSDSDY

        101                                                150
CaMV EESDS...GE EPEFE..Q.. .....VRMDR TGGTEIPKEE DGEGPSRYNE
CERV EVNDN...NS QPSLE..EEV VSESDVESMR SFNVAMNRGE VGESS...NK
 FMV EATED...YS EPETESSDEE TYFQQIRMER GESSETKREQ QDLGATRKRK
SVBV EMIDY...SS SEDDEPTRNV ..KVNIKEED EETYRPNRKR RGNSSSQPNY
PCSV EG...PTFNM EKSFRNCSPS GNLTPSKRPK VEEGETSQQE DFQAKLDKSF
SbCMV EASETPTKNF IYDFSSAEDF EE..PVKVPI AVEAETSNKR KFDK..NPQF
CVMV NNEQINVKIE GEEYEYKDNY RYYKPQPPYY KKDIRRERQY KGQSSQRADY
```

```
  151                                                  200
 CaMV RK..RKTPED RYFPTQPKTI PGQ..KQTSM GMLNIDC... QTNRRTLIDD
 CERV RP..KREPDL FTSFGKIRED IGD..KNPSL NILNLDCVNS PSDRKNKIDK
  FMV IE..ERNP.. FYTPPVHKGI PSSTGRGTEI STLNLDCISS FEERKVMIDK
 SVBV TR..YDIPQE Y........I PNRKTGLTNT KSLDLDCA.. .SNRRQLIEE
 PCSV RDYSYKKIHG KTDKKVPFKY LQKERGIED. GILNNHCA.. .ENQEELILA
SbCMV TRFTYQKI.. ......PKEI LPAHQTTSTI GVLDIDCV.. .ANTEKIIKE
 CVMV IKNRREEFES TYQANMNTTI ......NDSG EILNLDC.TT PEEAEDRIQK

  201                                                  250
 CaMV WAAEIGLIVK TNREDYLD.P ETI...LLLM EHKTSGIAKE LIRNTRWN..
 CERV WAAELGLVFL TNPEAYTTAP NAARARLAYM EHKSLGIVNR FIKSTQWT..
  FMV WFNEISLIIQ TNKESF.... DTSLKVLTLM EHRTEGIAKS FIKQATWDIL
 SVBV WDNEMRLIIK TEK....ALT NDFDLILTLA KSKTVGNAKQ LLESLHTEAF
 PCSV WINRLKLKVQ SGDETIRNL. .ALSDFVTYL QYQTTGVLAD WAQNTQIPFP
SbCMV WFNHHSILIT INEE.LKNL. .SSLDTFYYL VYKTRGIAHA YLSN....LP
 CVMV WTQSMSIALV KQQLSNEQAK Q.......FI RRTFIGNVKE WYKNLTNEAK

  251                                                  300
 CaMV R......TTG DIIEQVIDAM YTMFLGLNYS DNKVAEKIDE QEKAK..I..
 CERV Q......MNG DILLNVVSGL YTMFLGEDYT GNQ..EKTLE QERAKASL..
  FMV I......TPE KIIEEVLTGF YTMFIGLDYA ..LSAEKEEE KRIKKAEE..
 SVBV KQ..ASTTGE EFLTTLTNSF YTIFVGTNYL TQGTREK..E KAVQEARN..
 PCSV QGLTGSETTA EAKLQYLNR. YAQEIFSEFI G..YGPDEAR KEDKPNII..
SbCMV SEVL.SRIPA DRK..QVDD. WVYNLSREFV GRLERPESEE AFSQNNYY..
 CVMV QKLEGNAPLL SLTHMELG.. ....LRAEF. GKLGIESDVE KHEKKTSIAR

  301                                                  350
 CaMV RMT.KLQLCD ICY..LEEFT CDYEKNMYKT ELADFPGYIN QYL..SKIPI
 CERV RLI.NLQLCD ICS..LQSFF CDYESNLYKL PQNEYPSLVK QYL..AKIPI
  FMV LLI.KSQLCN ICE..LDNFT CFYEKQINQL KFEDFPKWIE LYL..GKIPI
 SVBV RLV.KLQICN LCS..LESFF CDYETNLLKL PIEEWPKYIE EYI..RKIPF
 PCSV RILSNMKLCD PCY..IENFF CQVEANYYKV T..DRTGL.L DIVL.SKMPP
SbCMV KLI.NLEICN MCY..LENFL CEFQSRYYGI NPIDRENLKV DLLLYAKLPE
 CVMV HKILQLQICS MDHQNLNAYL CEFQEYYYSA NYTEAESENI LNMFYSKLPE

  351                                                  400
 CaMV I.GEKALTRF RHEANGTSI. .YSLGFAAKI .VKEELSKIC DLSKKQKKLK
 CERV V.GEKASKRF EEEASAATS. .YSLGFAHKL .VNEELAKIC ELSKKQKKLK
  FMV I.GKQSKERW DNEKSFTTK. .YSLAFAKRI .IQEEIAKYC DFQRTSKKLK
 SVBV V.GMEVLEEY SKQDS.ITK. .GSLGYAHNL .IKAYMEKKC KSLKIKKEIR
 PCSV P.MVTYIQTQ INDPSRTRRI .LTLGLIRRF AIE.YRQNLC LRKIEKNYIR
SbCMV Y.VRTQVEAY FNASITSNKL DNTLG.GRIT ALKLWQTEQC NQKLAKRQ.A
 CVMV PWGQQVLNGY LSEIKGKNLL D.SIG.ARMT YLQEFISDKC KENWTQKQAR
```

```
  401                                                  450
 CaMV KFNKKCCS.. IGEASTEYGC KKT.ST..KK YHK.KRYKKK YKAYKP..YK
 CERV RFNKNCCS.. TFEKPYEYGC KPSYSK..KK KYS.KKYKPK YTKYKV..IR
  FMV NFSKKCCSKN SLDPLVSFGC RDTKKKDFKK SSKYKAYKKK KTLKKL..WK
 SVBV R..NMCCPKF S.SPETQYGC KPISHKKAKK QKYKQYYKKK YRLRKPKRWT
 PCSV NVDPKCCKQL D.DVPQEYGC ..SQPYRYKK K.......RKP FRRIKARKYP
SbCMV SVG.LCCSKI E.DKIGKYGC RKSNP.RAKK P.......KKK FRKIK..KYP
 CVMV KIQLSKNLDC SYYEVGKYGC KQIRPHKRKR .......YYKK YIPIKRKYFN

  451                                                  500
 CaMV KKK.KFRSGK YFKPKEKKG. .........SK QKYCPKGKKD CRCWICNIEG
 CERV KKK.KFSPGK YFKPKDKKS. .........EK AKYCPKGKKT CRCWVCNIEG
  FMV KKKRKFTPGK YF....SKK. .........KP EKFCPQGRKK CRCWICTEEG
 SVBV NSRRKYSGRK LFRRKRDQKE ETSQQSPEEK KKFCPQGKTT CRCWICNEIG
 PCSV ...YR....K WKPKYRYKVR RKGYSKQQNK QKTCPRGKKT CRCWICQEEG
SbCMV KKNFW....K WNNQRKKKTF RK..KRPFRK QQTCPTGKKK CQCWLCHEEG
 CVMV KKRYK....K YYRPKKFLKR KNPH...... .........KA CKCYNCGEEG

  501                                                  550
 CaMV HYANEC..PN RQS.SEKAHI LQ....QAEK LGLQPIEEPY EGVQEVFILE
 CERV HYANEC..PN RQT.SEKFKL IQ....IAEN YGLEPIENPY EDQQEICLL.
  FMV HYANEC..PN RKSHQEKVKI LI....HGMN EGYYPLEDAY TGNLEVFSME
 SVBV HFAKDC..RN KSANHNK..I IE....ELQS LQLEPVFDLN ELKIEEKFWE
 PCSV HYANEC..PN RKINQKKDKY VR....MLYS VGYEPIEEDY E.TDESLDFD
SbCMV HYANEC..P. .KKDNKKAQT LK....LIFD LGFEPVESDI E.TDE....E
 CVMV HISPNCKKPK KKTRINNLEA LEFKNTEMEN LEFETNKNDI IWVEEIEVIQ

  551                            579
 CaMV YKEEEEETST EESDGSSTSE DSDSD....
 CERV ...EQIQLSS SDSELDDTCE ESSSEESE.
  FMV IIEE..TTSE EESTTDSDSS SSDDEQLSF
 SVBV L.KEVSESSE SESEISSD.E SSDSEDLE.
 PCSV IYSLTSETDS ETESENEFEE .........
SbCMV LFELTSEDSS EDEY...... .........
 CVMV PLHYEEEEKY KGNYSDRILQ NPYY.....
```

ORF IV core region

Gap Weight: 8

GapLengthWeight: 2

```
  1                                                  50
CaMV LLLMEHKTSG IAKELIRNTR WN..R..... .TTGDIIEQV IDAMYTMFLG
CERV LAYMEHKSLG IVNRFIKSTQ WT..Q..... .MNGDILLNV VSGLYTMFLG
 FMV LTLMEHRTEG IAKSFIKQAT WDILI..... .TPEKIIEEV LTGFYTMFIG
SVBV LTLAKSKTVG NAKQLLESLH TEAFKQ..AS TTGEEFLTTL TNSFYTIFVG
PCSV VTYLQYQTTG VLADWAQNTQ IPFPQGLTGS ETTAEAKLQY LNR.YAQEIF
SbCMV FYYLVYKTRG IAHAYLSN.. ..LPSEVL.S RIPADRK..Q VDD.WVYNLS

  51                                                 100
CaMV LNYSDNKVAE KIDEQEKAK. .IRMT.KLQL CDICYLEEFT CDYEKNMYKT
CERV EDYTGNQ..E KTLEQERAKA SLRLI.NLQL CDICSLQSFF CDYESNLYKL
 FMV LDYA..LSAE KEEEKRIKKA EELLI.KSQL CNICELDNFT CFYEKQINQL
SVBV TNYLTQGTRE K..EKAVQEA RNRLV.KLQI CNLCSLESFF CDYETNLLKL
PCSV SEFIG..YGP DEARKEDKPN IIRILSNMKL CDPCYIENFF CQVEANYYKV
SbCMV REFVGRLERP ESEEAFSQNN YYKLI.NLEI CNMCYLENFL CEFQSRYYGI

  101                                                150
CaMV ELADFPGYIN QYL..SKIPI IGEKALTRFR HEANGTSI.. YSLGFAAKI.
CERV PQNEYPSLVK QYL..AKIPI VGEKASKRFE EEASAATS.. YSLGFAHKL.
 FMV KFEDFPKWIE LYL..GKIPI IGKQSKERWD NEKSFTTK.. YSLAFAKRI.
SVBV PIEEWPKYIE EYI..RKIPF VGMEVLEEYS KQDS.ITK.. GSLGYAHNL.
PCSV T..DRTGL.L DIVL.SKMPP PMVTYIQTQI NDPSRTRRI. LTLGLIRRFA
SbCMV NPIDRENLKV DLLLYAKLPE YVRTQVEAYF NASITSNKLD NTLG.GRITA

  151                                                200
CaMV VKEELSKICD LSKKQKKLKK FNKKCCS..I GEASTEYGCK KT.ST..KKY
CERV VNEELAKICE LSKKQKKLKR FNKNCCS..T FEKPYEYGCK PSYSK..KKK
 FMV IQEEIAKYCD FQRTSKKLKN FSKKCCSKNS LDPLVSFGCR DTKKKDFKKS
SVBV IKAYMEKKCK SLKIKKEIRR ..NMCCPKFS .SPETQYGCK PISHKKAKKQ
PCSV IE.YRQNLCL RKIEKNYIRN VDPKCCKQLD .DVPQEYGC. .SQPYRYKKK
SbCMV LKLWQTEQCN QKLAKRQ.AS VG.LCCSKIE .DKIGKYGCR KSNP.RAKKP

  201                          226
CaMV HK..KRYK.K KYKAYKP..Y KKKKKF
CERV YS..KKYK.P KYTKYKV..I RKKKKF
 FMV SK.YKAYK.K KKTLKKL..W KKKKRK
SVBV KY.KQYYK.K KYRLRKPKRW TNSRRK
PCSV RKPFRRIKAR KYPYRK.WKP KYRYKV
SbCMV KKKFRKIK.. KYPKKNFWKW NNQRKK
```

ORF V Full Length

Gap Weight: 8

GapLengthWeight: 2

```
  1                                                  50
CaMV KPMKYSPMDR EEFDKQIKEL LDLKVIKPSK S....PHMAP AFLVNNEAEK
CERV KPMSYSPSDR EEFDRQIKEL LELKVIKPSK S....THMSP AFLVENEAER
 FMV KPMSYSPQDR EGFAKQIKEL LDLGLIIPSK S....QHMSP AFLVENEAER
SVBV KPMQYSPQDR EEFKTQIEEL LKLGIIRPSK S....PHSSP FMVRNHAEI
PCSV NRIPYTQKDI DEFREETSKQ IELGILRQSK S....PHSAP AFYVENHNEI
SbCMV NRIPYTMRDV QEFEEECEDL LKKGLIRESQ S....PHSAP AFYVENHNEI
CVMV PPMLYQETDL PEFKMHIEEM IKEGFIEEKT NFEDKKYSSP AFIVNKHSEQ
PVCV SHSGMNPEHL QLALKECDEL QQFDLIEPSD SQWACE.... AFYVNKRSEQ

  51                                                 100
CaMV RRGKKRMVVN YKAMNKATVG DAYNLPNKDE LLTLIRGKKI FSSFDCKSGF
CERV RRGKKRMVVN YKAMNKATKG DAHNLPNKDE LLTLVRGKKI YSSFDCKSGL
 FMV RRGKKRMVVN YKAINQATIG DSHNLPNMQE LLTLLRGKSI FSSFDCKSGF
SVBV KRGKARMVIN YKKLNDHTKG DGYLLPNKEQ LLQRIGGKTF YSSFDCKSGF
PCSV KRGKRRLVIN Y.KMNKATKG DAYNLDRL.Y LTD..RESNW FSTLDAKSGF
SbCMV KRGKRRMVIN YKKMNEATIG DSYSYQEK.I LSEKIKGSLW FSSLDAKSGY
CVMV KRGKTRMVID YKDLNKKAKV VKYPIPNKDT LIHRSIQARY YSKFDCKSGF
PCVV VRGKLRLVIN YQPLNHFLQD DKFPIPNKLT LFSHLSKAKL FSKFDLKSGF

 101                                                 150
CaMV WQVLLDQESR PLTAFTC.PQ GHYEWNVVPF GLKQAPSIFQ R.HMDEAFRV
CERV WQVLLDKESQ LLTAFTC.PQ GHYQWNVVPF GLKQAPSIFP KTYANSHSNQ
 FMV WQVVLDEESQ KLTAFTC.PQ GHFQWKVVPF GLKQAPSIFQ R.HMQTALNG
SVBV WQVRLAPETI QLTAFSC.PQ GHYEWLVMPF GLKQAPAIFQ RHMDESLSNM
PCSV LQLRLDEETK PLTAFSCPPQ MHLEYNVMPM GLKQAPSQFQ R.FMDNNLRG
SbCMV YQLRLHENTK PLTAFSCPPQ KHYEWNVLSF GLKQAPCIYQ R.FMDQSLKG
CVMV YHIKLEEDSK KYTAFT.VPQ GYYQWKVLPF GYHNSPSIFQ Q.FMDRIFRP
PCVV WQLGIHPNER PKTGF.CIPD RHFQWKVMPF GLKTAPSLFQ KAMI.KIFQP

 151                                                 200
CaMV FRKFCCVYVD DILVFSN.NE EDHLLHVAMI LQKCNQHGII LSKKKAQLFK
CERV YSKYCCVYVD DILVFSNTGR KEHYIHVLNI LRRCEKLGII LSKKKAQLFK
 FMV ADKFCMVYVD DIIVFSN.SE LDHYNHVYAV LKIVEKYGII LSKKKANLFK
SVBV YPQFCAVYVD DIIVFSKT.E EEHLGHVKIV LNRCKALGIV LSKKKAQLCK
PCSV LEDISLAYID DIIVFTKGTK DYHLKQVARV LIQLGNHGVI LSKEKAKIAF
SbCMV LDHIYLAYID DILIFTKGSK EQHVNDVRIV LQRIKEQGII ISKKKSKLIQ
CVMV YYDFIIVYID DILVFSKTIE E.HKIHIAKF RDITLANGLI ISKKKTELCK
PCVV ILFSALVYID DILLFSETLE .DHIKLLNQF ISLVKKFGVM LSAKKMILAQ
```

```
  201                                                  250
CaMV KKINFLGLEI .DEGTHKPQG HILEHINKFP DTLEDKKQLQ RFLGILTYA.
CERV EKINFLGLEI .DQGTHCPQN HILEHIHKFP DRIEDKKQLQ RFLGILTYA.
 FMV EKINFLGLEI .DKGTHCPQN HILENIHKFP DRLEDKKHLQ RFLGVLTYA.
SVBV TTINFLGLVI .ERGNLKVQS HIGLHLVAFP DQLSDRNALQ RFLGLLNYI.
PCSV EEIEFLGLKI LKNGFIEPQK HLLEKIAEFP DQLQDRKQIQ KFLGCLNYIG
SbCMV QEIEYLGLKI QGNGEIDLSP HTQEKILQFP DELADRKQIQ RFLGCINYIA
CVMV EKIDFLGVQI .EQGGIELQP HIINKILEKH TKIKNKTELQ SILGLLNQIR
PVCV NKIQFLGMDF AD.GTFSPAG HISLELQKFP DTNLSVKQIQ QFLGIVNYIR

  251                                                  300
 CaMV .SDYIPKLAQ IRKPLQAKL. .KENVPWRWT KEDTLYMQKV KKNLQGFP.P
 CERV .SDYIPKLAS IRKPLQSKL. .KEDSTWTWN DTDSQYMAKI KKNLKSFP.K
  FMV .ETYIPKLAE IRKPLQVKL. .KKDVTWNWT QSDSDYVKKI KKNLGSFP.K
 SVBV .SAYFPKIAN LRSPLQVKL. .KKEITWSWT EKDTETVRKI KSLVKTLP.D
 PCSV EKGFFKELAK ERKVLQKML. .SEKLPWKWN DLATLAVKRL KQVCKNLP.R
SbCMV PEGFFRTLAL ERKHLQKKI. .SVKNPWKWD TIDTKMVQSI KGKIQSLP.K
 CVMV H..FIPHLAQ ILLPIQKKLK IKDEEIWTWT KEDEEKIKLI QDYSKNLVIK
 PVCV ..DFIPEVTE HISPLSDMLK KK...PPAWG KCQDNAVKQL KQLAQQV.KS

  301                    321
 CaMV LHHPLPEEKL ..IIETDASD D
 CERV LYHPEPNDKL ..VIETDASE E
  FMV LYLPKPEDHL ..IIETDASD S
 SVBV LYNPSPEDKP ..IIECDASD D
 PCSV LYVAKPSDLL ..ILTTDASD T
SbCMV LYNASIQDFL ..IVETDASQ H
 CVMV MKYPINKEDM NWIIEVDASN N
 PVCV LHIPSEGKK. ..ILQTDASD Q
```

Bibliography

1. Mahmoudpour A. 'Host range determination and relative quantification of susceptibility to strawberry vein banding virus infection using Agroinoculation and PCR". *Journal of EC Microbiology* 4.1 (2016): 608-616.

2. Iwaki M., *et al*. "Soybean chlorotic mottle, a new caulimovirus on soybean". *Plant Disease* 68 (1984): 1009-1011.

3. Richert-Poeggeler K R and R J Shepherd. "Petunia vein-clearing virus: A plant pararetrovirus with the core sequences for an integrase function". *Virology* 236.1 (1997): 137-146.

4. Pahalawatta V., *et al*. "A new and distinct species in the genus Caulimovirus exists as an endogenous plant pararetroviral sequence in its host, Dahlia variabilis". *Virology* 376.2 (2008): 253-257.

5. Calvert L A., *et al*. "Characterization of cassava vein mosaic virus: a distinct plant pararetrovirus". *Journal of General Virology* 76.5 (1995):1271-1276.

6. De Kochko A., *et al*. "Nucleotide sequence and genomic organization of the cassava vein mosaic virus (CVMV), an intermediate between caulimo and badna viruses" (1996).

7. Stenger D C., *et al*. "Isolation, molecular cloning, and detection of strawberry vein banding virus DNA". *Phytopathology* 78.2 (1988): 154-159.

8. Honetšlegrová J., *et al*. "Detection and isolation of strawberry vein banding virus in the Czech Republic". *Acta Horticulturae* 385 (1995): 29-32.

9. Petrzik K., *et al*. "Strawberry vein banding virus complete genome" (1996).

10. Pique M., *et al*. "Sequence of a cauliflower mosaic virus strain infecting Solanaceous plants". *Journal of Gene* 155.2 (1995): 305-306.

11. Mahmoudpour A. "Infectivity of recombinant strawberry vein banding virus DNA". *Journal of General Virology* 84.6 (2003): 1377-1381.

12. Donson J and R Hull. "Physical mapping and molecular cloning of caulimovirus DNA". *Journal of General Virology* 64.10 (1983): 2281-2288.

13. Mahmoudpour A. "Molecular biology of strawberry vein banding caulimovirus: Molecular Characterization, Diagnosis and Host Range of SVBV". *Lambert Academic Publishing. Köln, Germany* (2009): 108.

14. Petrzik K., *et al*. "Strawberry vein banding virus-definitive member of the genus caulimovirus". *Virus Genes* 16.3 (1998): 303-305.

15. Mráz I., *et al*. "Variability in coat protein sequence homology among American and European sources of strawberry vein banding virus". *Plant Disease* 82.5 (1998): 544-546.

16. Franck A., *et al*. "Nucleotide sequence of cauliflower mosaic virus DNA". *Cell* 21.1 (1980): 285-294.

17. Palkovics L and E Balazs. "Simple and rapid detection of carnation etched ring virus by polymerase chain reaction". *Acta Phytopathologica et Entomologica Hungarica* 31 (1996): 161-168.

18. Richins RD. "Organization and transcription of the peanut chlorotic streak caulimovirus genome" (1994).

19. Richins RD., *et al*. "Sequence of figwort mosaic virus DNA (Caulimovirus group)". *Nucleic Acids Research* 15.20 (1987): 8451-8466.

20. Hasegawa A., *et al*. "The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter". *Nucleic Acids Research* 17.23 (1989): 9993-10013.

21. Takahashi H., *et al*. "DNA sequence of gene VI of cauliflower mosaic virus Japanese strain S (CaMV S-Japan)". *Nucleic Acids Research* 17.19 (1989): 7981-7989.

22. Leisner S M., *et al*. "Long distance movement of cauliflower mosaic virus in infected turnip plants". *Molecular Plant-Microbe Interactions* 5.1 (1992): 41-47.

23. Haas M., *et al*. "Cauliflower mosaic virus: still in the news". *Molecular Plant Pathology* 3.6 (2002): 419-429.

24. Hemmings-Mieszczak M and T Hohn. "A stable hairpin preceded by a short open reading frame promotes nonlinear ribosome migration on a synthetic mRNA leader". *RNA* 5.9 (1999): 1149-1157.

25. Hull R and J Donson. "Physical mapping of the DNAs of carnation etched ring and figwort mosaic viruses". *Journal of General Virology* 60 (1982): 125-134.

26. Mahmoudpour M M A. "Strawberry Vein Banding Caulimovirus: Biology and Characterization". Ph.D. Dissertation. University of California, Davis (2000): 124.

27. Fütterer J and T Hohn. "Involvement of nucleocapsids in reverse transcription: a general phenomenon? *Trends in Biochemical Sciences* 12 (1987): 92-95.

28. Al Ani R., *et al*. "The structure of cauliflower mosaic virus". *Virology* 93 (1979): 175-187.

29. Du Plessis D H and P Smith. "Glycosylation of cauliflower mosaic virus capsid polypeptide". *Virology* 109.2 (1981): 403-408.

30. Hahn P and R J Shepherd. "Phosphorylated proteins in cauliflower mosaic virus". *Virology* 107.1 (1980): 295-297.

31. Hohn T and J Fütterer. "The proteins and functions of plant pararetroviruses: knowns and unknowns". *Critical Reviews in Plant Sciences* 16.1 (1997): 133-161.

32. Cheng RH., *et al*. "Cauliflower mosaic virus: A 420 subunit (T= 7), multilayer structure". *Virology* 186.2 (1992): 655-668.

Volume 4 Issue 6 December 2016