Bot. Bull. Acad. Sin. (1997) **38:** 77-84

Nucleotide sequence and phylogenetic analysis of a bamboo mosaic potexvirus isolate from common bamboo (*Bambusa vulgaris* McClure)

Chi-Chen Yang[1,2], Jih-Shiou Liu[1], Chan-Pin Lin[2] and Na-Sheng Lin[1,3]

[1]*Institute of Botany, Academia Sinica, Taipei, Taiwan 115, Republic of China*

[2]*Department of Plant Pathology and Entomology, National Taiwan University, Taipei, Taiwan, Republic of China*

(Received September 20, 1996; Accepted January 30, 1997)

Abstract. The complete cDNA sequence corresponding to the genomic RNA of an isolate bamboo mosaic potexvirus (BaMV-V) from common bamboo (*Bambusa vulgaris* McClure) was determined. This isolate is the first potexvirus with which a satellite RNA has been associated. The genome organization of BaMV-V, similar to those of other potexviruses, contained five open reading frames (ORFs 1–5) coding for polypeptides with molecular weight of 156 kDa, 28 kDa, 13 kDa, 6 kDa, and 25 kDa, respectively. Nucleotide sequence analysis showed a 10.0% difference from the BaMV-O isolate previously described whereas the amino acid comparison showed a difference of 3.2%. When three conservative domains of RNA dependent RNA polymerase (RdRp) were used for phylogenetic analysis, the greatest variation between two strains of each virus was only 12.8% of that between the two closest members of the potexvirus group. The grouping of potexviruses distinct from other groups of plant viruses was also confirmed by a comparision of three conservative motifs of RdRp.

Keywords: Bamboo mosaic virus; Nucleotide sequence; Potexvirus.

Introduction

Bamboo mosaic virus (BaMV), a member of the potexvirus group, has a flexuous rod-shaped morphology of 490 nm in length and contains a single-stranded positive-sense RNA genome (Lin et al., 1977; 1992; 1994). The complete nucleotide sequence of BaMV from green bamboo (*Bambusa oldhamii* Munro) (isolate BaMV-O) has been reported (Lin et al., 1994). The RNA genome of BaMV-O is 6366 nucleotides long (excluding poly (A) tail) and contains six putative open reading frames (ORFs 1–6) coding for polypeptides with molecular weights of 155 kDa, 28 kDa, 13 kDa, 6 kDa, 25 kDa, and 14 kDa, respectively. The ORFs 1–5 are common in potexviruses while ORF 6 is divergent.

The ORF 1 encodes RNA dependent RNA polymerase (RdRp), which contains three conservative domains—methyltransferase, nucleotide-binding, and polymerase—common in positive strand RNA viruses (Koonin and Dolja, 1993). ORFs 2–4 constitute the triple gene block and are involved in movement of the virus (Beck et al., 1991; Angell et al., 1996). Triple gene block occurs not only in potexviruses but also in carlaviruses (e.g. potato virus M, PVM) (Zavriev et al., 1991), hordeivirus (e.g. barley stripe mosaic virus, BSMV) (Gustafson and Armour, 1986), furovirus (e.g. beet necrotic yellow vein virus, BNYVV) (Bouzoubaa et al., 1986) and *Nicotiana velutina* mosaic virus (NVMV) (Randles and Rohde, 1990). ORF 5 encodes the capsid protein of the virus. An ORF 6 has been proposed within the coding region of ORF 1 in papaya mosaic virus (PMV) (Sit et al., 1989), clover yellow mosaic virus (CYMV) (Sit et al., 1990), foxtail mosaic virus (FMV) (Bancroft et al., 1991), potato aucuba mosaic virus (PAMV) (Xu et al., 1994) and BaMV-O (Lin et al., 1994) whereas in narcissus mosaic virus (NMV) (Zuidema et al., 1989), white clover mosaic virus (WClMV) (Beck et al., 1990) and strawberry mild yellow edge-associated virus (SMYEAV) (Jelkmann et al., 1992) an ORF 6 is proposed within or downstream of the coding region of the capsid protein gene. However, the function of the putative ORF 6 remains unknown.

Within the potexvirus group, the overall nucleotide sequence difference among strains of potato virus X (PVX) ranges from 3.7% between X3 (Huisman et al., 1988) and Russian (Skryabin et al., 1988a) strains to 22.6% between Russian and Andean (Orman et al., 1990) strains. On the other hand, the nucleotide sequences of the two strains of WClMV, strain M (Forster et al., 1988) and strain O (Beck et al., 1990), have a 12% divergency. The overall differences in their encoded proteins range from 1.8% to 11.2% among strains of PVX and 4.4% in the two strains of WClMV.

[3]Corresponding author. Fax: 886-2-782-7954; E-mail: bo124lin@ccvax.sinica.edu.tw

[4]Sequence data reported from this article have been deposited to the EMBL/GenBank Data Libraries under Accession No. L77962.

Recently a satellite RNA (satBaMV) was identified in a BaMV isolate of common bamboo (*B. vulgare* McClure) (designated BaMV-V) (Lin and Hsu, 1994). SatBaMV was encapsidated with the BaMV-V capsid protein, which serologically reacted with anti-BaMV-O capsid protein serum (Lin and Hsu, 1994). Moreover, both genomic RNAs of BaMV-O and BaMV-V support the replication of satBaMV (Lin and Hsu, 1994). It was also found that the capsid protein of BaMV-V encapsidates its subgenomic RNAs whereas the BaMV-O isolate does not (Lin et al., unpubl. data). In this report, the complete nucleotide sequence of genomic RNA of BaMV-V was determined. The comparison of potexvirus strains, the grouping of members within the potexvirus group, and their phylogenetic relationships to three conservative domains of RdRp are also presented.

Materials and Methods

Virus Purification and RNA Extraction

BaMV-V is an isolate with satellite RNA from common bamboo (*B. vulgaris*) (Lin and Hsu, 1994). BaMV-L (previously designated BaMV-V/S$^-$) (Lin and Hsu, 1994) is the satellite-free isolate derived from BaMV-V. The purification of BaMV and the viral RNA were as described previously (Lin and Chen, 1991; Lin et al., 1992).

RNA Sequencing

The method used to determine RNA direct sequencing was as described in Fichot and Girard (1990). Primer B14 (CCGTTCAAGCGTGTTTGC) complementary to BaMV-O nucleotide 236–253 (Lin et al., 1994) was used to determine the 5' terminal sequence of BaMV-L RNA. Sequence data were then used for the synthesis of primer B33 (see next section) for the construction of the 5' end cDNA clone pBaL1-8.

cDNA Synthesis, Cloning and Sequencing

Reverse-transcription polymerase chain reaction (RT-PCR) was performed for the generation of cDNA of BaMV-L RNA (Sambrook et al., 1989). Primer B21 (ACGG**GAGCTC**$(T)_{15}$, with a SacI site at the 5' end) and B36 (GCTATGTGCTTGCCAGCGGA, nts 3113–3132 based on BaMV-O RNA) (Lin et al., 1994) were used to amplify the 3' half fragment; whereas primer B33 (TGC**GGATCC***TAATACGACTCACTATA*GAAAAGCA TTCCAAAC, 5' terminal nt 1–16 with a BamHI site at the 5' end and a T7 promoter in italic) and B35 (TTCCAATGCGTGTCCACCCA, complementary to nts 3473–3492 based on BaMV-O RNA) were used to amplify the 5' half fragment (Figure 1). The amplified fragments were gel-eluted and then cloned into a TA cloning vector (TA Cloning Kit, Invitrogen Co., San Diego, CA). A common BstXI site in the overlapping region of these two clones was used to cut the fragments and the subsequent ligation reaction for the full-length cDNA construction. The full-length cDNA was then cloned into the SacI site of pUC119. Sequencing of the cDNA insert was performed as described by Wang et al. (1988).

Sequence Data and Phylogenetic Analysis

The nucleotide sequence of BaMV-V was aligned with BaMV-O manually. Gaps were introduced to optimize the alignment. Putative ORFs were analyzed by the computer program DNA Strider 1.2 (Marck, 1988). Three conservative motifs in ORF 1, i.e. methlytransferase domain (Rozanov et al., 1992), NTP-binding domain (Hodgman, 1988; Gorbalenya et al., 1989), and polymerase domain (Argos, 1988; Koonin, 1991) were used for phylogenetic analysis among strains and members of the potexvirus group. Other plant virus groups with the triple gene block and the closely related tymovirus group (Koonin and Dolja, 1993) were used to compare the solidity of the grouping

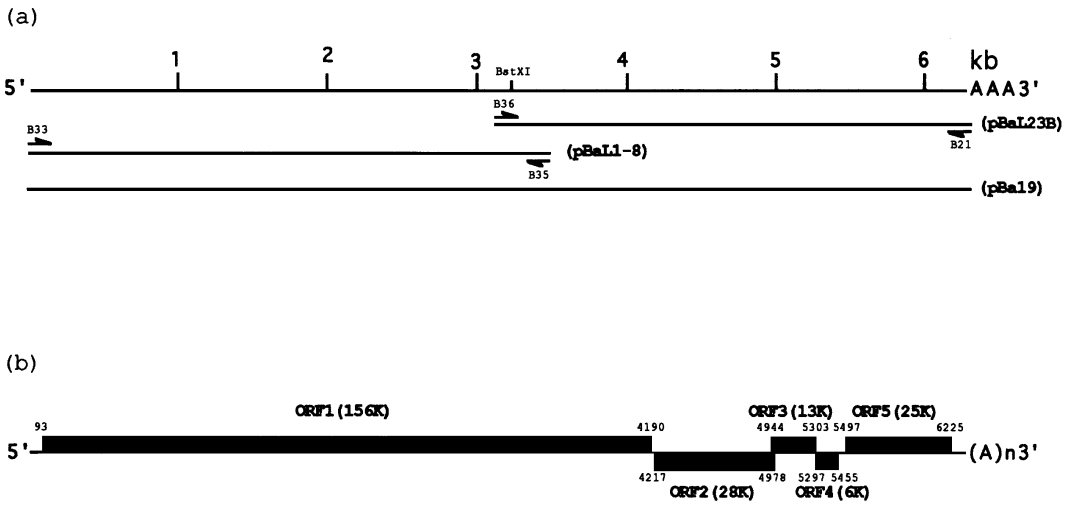


Figure 1. Cloning strategy and genome organization of BaMV-V genomic RNA. (a) Synthetic primers used for the construction of cDNA clones of the 5' end (pBaL1–8), the 3' end (pBaL23B) and the full-length fragment (pBa19) of BaMV-V genomic RNA. (b) Genome organization and the putative five ORFs. The start and end codons for each ORF are indicated.

of potexviruses. Alignment of amino acid residues and phylogenetic tree calculations were done by the program Clustal W (Thompson et al., 1994). Tree-drawings were done by the program NJPlot (Gouy, 1995).

Results and Discussion

Direct RNA Sequencing and Full-Length cDNA Clone

RNA direct sequencing of the 5' extreme using primer B14 revealed the 5' terminal sequence, 5'-GAAAAGCAUUCCAAAC-3'. This sequence was used to synthesize the primer B33 for the construction of 5'-half cDNA clones. Using paired primers (primers B33 and 35, B36 and B21), the cDNA fragments corresponding to the 5' and 3' half were synthesized by RT-PCR and estimated to be 3.4 and 3.2 kb, respectively. Recombinant clones containing the 3.4 kb or 3.2 kb fragments were designated as pBaL1-8 or pBaL23B, respectively (Figure 1a). Clone pBa19 containing the full-length cDNA of BaMV-V genomic RNA in pUC119 was generated by a combination of pBaL1-8 and pBaL23B.

Sequence Data, Genome Organization of BaMV-V and Comparisons Among Strains in Potexviruses

Sequence data from pBa19 revealed that the full-length BaMV-V genome is 6365 nucleotides in length, excluding the poly (A) tail. Computer analysis of the putative ORFs of BaMV-V isolate also revealed five ORFs with calculated molecular weight of 156 kDa (ORF 1), 28 kDa (ORF 2), 14 kDa (ORF 3), 6 kDa (ORF 4), and 25 kDa (ORF 5) (Figure 1b). The putative 14 kDa ORF 6 in BaMV-O was not found in BaMV-V.

The alignment of isolate BaMV-V with BaMV-O showed a 10% nucleotide variation with 630 base substitutions (524 transitions and 106 transversions) and 5 gaps (Figure 2). Base substitutions occur throughout the whole genome whereas gaps occur mainly in the 5' untranslated region. The sequence variation between BaMV-V and BaMV-O is within the range among PVX strains, with the two most divergent strains (Russian and Andean) differing by 22.6% and the two most similar strains (X3 and Russian) by only 3.7% at the nucleotide level. The variation between the two WClMV strains (12%) is similar to that of the two BaMV isolates.

The comparison of the ORFs 1 to 5 between BaMV-O and BaMV-V isolates shows differences of 2.9%, 3.6%, 4.2%, 5.8%, and 4.1%, respectively. The overall amino acid differences among strains of potexviruses are comparable with those of their nucleotide sequences, i.e., 3.2% between two isolates of BaMV, 4.4% between two strains of WClMV, and 1.8% (X3 and Russian) to 11.2% (X3 and Andean) among strains of PVX.

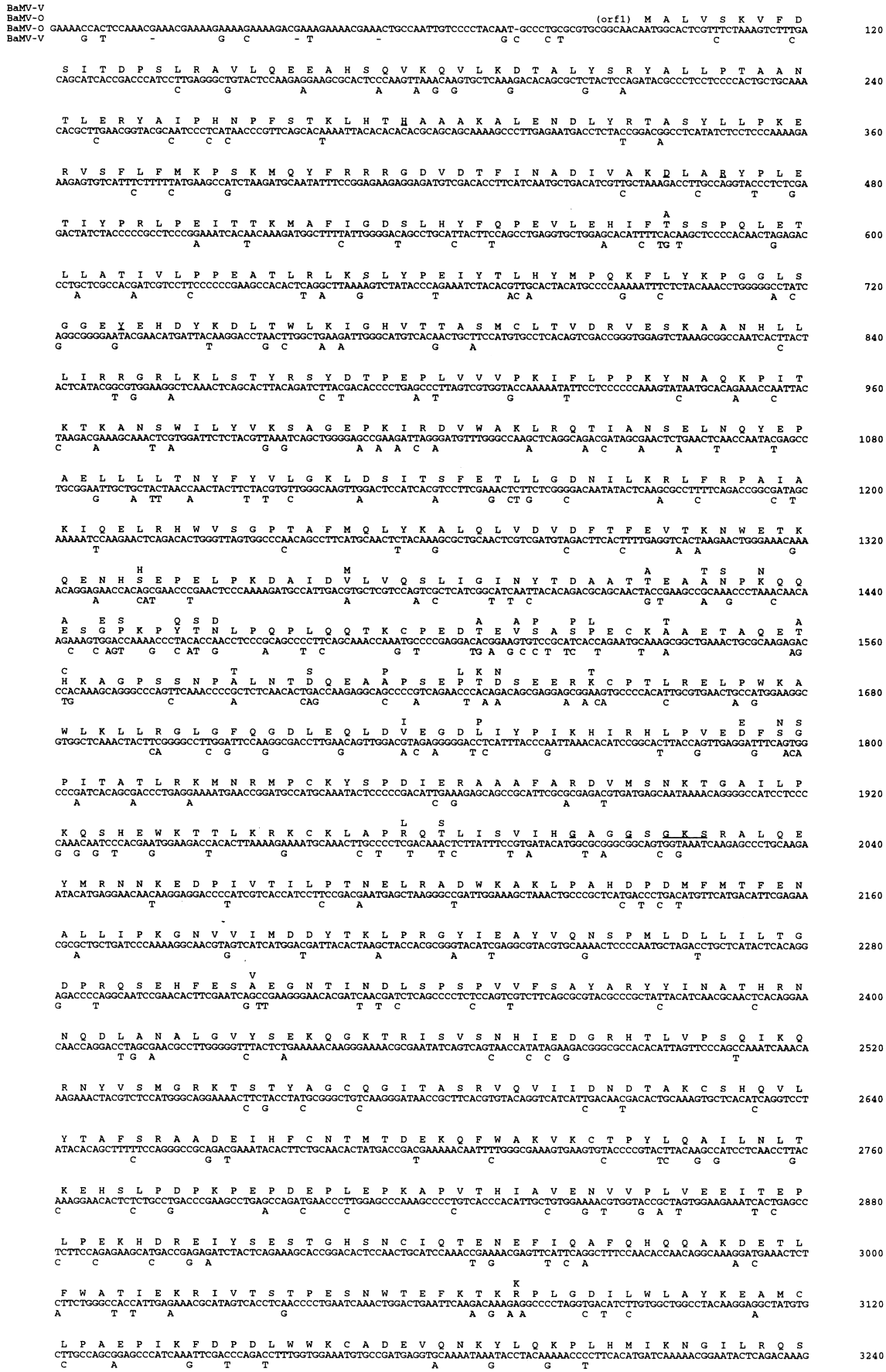
Conserved Motifs in the Noncoding Regions

The 5' noncoding region of BaMV-O RNA contains 94 nucleotides (Lin et al., 1994), but there are only 92 nucleotides in the BaMV-V 5' noncoding region. The alignment showed four gaps in this region. The imperfect repeat of the pentanucleotide GAAAA, which is proposed as an encapsidation signal found in PMV (Lok and AbouHaidar, 1986), is not present in the two BaMV isolates. However, the adenosine-adenosine dinucleotides at the third and fourth positions of the 5' end pentanucleotide repeat, an important feature for assembly in PMV (Sit et al., 1994), are present. Interestingly, satBaMV, which shares 63% sequence homology with BaMV-O in their 5' end untranslated regions, does retain the three consecutive pentanucleotides with AA dinucleotides in its third and fourth positions and was encapsidated with capsid protein of BaMV into rod-shaped particles (Lin and Hsu, 1994). No obvious RNA secondary structures such as those proposed for WClMV (Forster et al., 1988) were observed in the 5' untranslated region of the two isolates of BaMV or satBaMV. Moreover, it was found that BaMV-V isolate contains encapsidated subgenomic RNAs (unpubl. data), as is also the case for NMV (Short and Davies, 1983) and WClMV (M strain) (Forster et al., 1987), while the O isolate does not (Lin et al., 1992). The difference in subgenomic RNA encapsidation in the two isolates of BaMV needs further investigation.

The hexanucleotide ACC/UUAA, which is conserved in the 3' non-coding region of potexviruses and carlaviruses (White et al., 1992), also occurs in both BaMV isolates. When possible secondary structures for the 3' end noncoding region were analyzed, the conserved hexanucleotides (ACCUAA) and polyadenylation signal (AAUAAA) all occurred in loop regions. This was also the case for the 3' end noncoding region of satBaMV (data not shown). However, the conserved ACUUAA was also found in the negative strand upstream of ORF 2 and ORF 5. It occurred in a pseudoknot structure before a long stem-loop structure (data not shown).

Conserved Motifs in the Coding Regions

All the conserved sequence motifs found in the translated regions of BaMV-O and other potexviruses are present in the BaMV-V. In ORF 1, the methyltransferase motif is located at amino acid residues 59 to 223, while the nucleotide-binding and polymerase motifs span the regions from 635 to 863 and 1139 to 1274, respectively. The consensus amino acid sequences for the methyltransferase motif (His residue, DXXR signature and tyrosine residue) (Rozanov et al., 1992), the nucleotide binding (G/AX_2GXGKS/T) (Hodgman, 1988; Gorbalenya et al., 1989) and viral polymerase motif (S/TGX_3TX_3NS/TX_{22}GDD) (Argos, 1988; Koonin, 1991) are underlined in Figure 2. The nucleotide binding motif of ORF 2 (Bancroft et al., 1991; Skryabin et al., 1988b) is located between amino acid residues 31 to 38. Also present is the conserved sequence in the central region of ORF 3, PX_2GDNXHXLPXGGXYXDGXKX_3YX_2P, which occurs not just in all potexviruses (including the two isolates of BaMV) but also in carlavirus (PVM), hordeivirus (BSMV), furovirus (BNYVV), and NVMV (Skryabin et al., 1988b; Randles and Rohde, 1990). Apart from the conserved central stretch, conserved proline and glycine residues


BaMV-V
BaMV-O
BaMV-O GAAAACCACTCCAAACGAAACGAAAAGAAAAGAAAAGACGAAAGAAAACGAAACTGCCAATTGTCCCCTACAAT-GCCCTGCGCGTGCGGCAACAATGGCACTCGTTTCTAAAGTCTTTGA 120
BaMV-V
(orf1) M A L V S K V F D
S I T D P S L R A V L Q E E A H S Q V K Q V L K D T A L Y S R Y A L L P T A A N
CAGCATCACCGACCCATCCTTGAGGGCTGTACTCCAAGAGGAAGCGCACTCCCAAGTTAAACAAGTGCTCAAAGACACAGCGCTCTACTCCAGATACGCCCTCCTCCCCACTGCTGCAAA 240
T L E R Y A I P H N P F S T K L H T H A A A K A L E N D L Y R T A S Y L L P K E
CACGCTTGAACGGTACGCAATCCCTCATAACCCGTTCAGCACAAAATTACACACACACGCAGCAGCAAAAGCCCTTGAGAATGACCTCTACCGGACGGCCTCATATCTCCTCCCAAAAGA 360
R V S F L F M K P S K M Q Y F R R R G D V D T F I N A D I V A K D L A R Y P L E
AAGAGTGTCATTTCTTTTTATGAAGCCATCTAAGATGCAATATTTCCGGAGAAGAGGAGATGTCGACACCTTCATCAATGCTGACATCGTTGCTAAAGACCTTGCCAGGTACCCTCTCGA 480
T I Y P R L P E I T T K M A F I G D S L H Y F Q P E V L E H I F T S S P Q L E T
GACTATCTACCCCCGCCTCCCGGAAATCACAACAAAGATGGCTTTTATTGGGGACAGCCTGCATTACTTCCAGCCTGAGGTGCTGGAGCACATTTTCACAAGCTCCCCACAACTAGAGAC 600
L L A T I V L P P E A T L R L K S L Y P E I Y T L H Y M P Q K F L Y K P G G L S
CCTGCTCGCCACGATCGTCCTTCCCCCCGAAGCCACACTCAGGCTTAAAAGTCTATACCCAGAAATCTACACGTTGCACTACATGCCCCAAAAATTTCTCTACAAACCTGGGGGCCTATC 720
G G E Y E H D Y K D L T W L K I G H V T T A S M C L T V D R V E S K A A N H L L
AGGCGGGGAATACGAACATGATTACAAGGACCTAACTTGGCTGAAGATTGGGCATGTCACAACTGCTTCCATGTGCCTCACAGTCGACCGGGTGGAGTCTAAAGCGGCCAATCACTTACT 840
L I R R G R L K L S T Y R S Y D T P E P L V V V P K I F L P P K Y N A Q K P I T
ACTCATACGGCGTGGAAGGCTCAAACTCAGCACTTACAGATCTTACGACACCCCTGAGCCCTTAGTCGTGGTACCAAAAATATTCCTCCCCCCAAAGTATAATGCACAGAAACCAATTAC 960
K T K A N S W I L Y V K S A G E P K I R D V W A K L R Q T I A N S E L N Q Y E P
TAAGACGAAAGCAAACTCGTGGATTCTCTACGTTAAATCAGCTGGGGAGCCGAAGATTAGGGATGTTTGGGCCAAGCTCAGGCAGACGATAGCGAACTCTGAACTCAACCAATACGAGCC 1080
A E L L L L T N Y F Y V L G K L D S I T S F E T L L G D N I L K R L F R P A I A
TGCGGAATTGCTGCTACTAACCAACTACTTCTACGTGTTGGGCAAGTTGGACTCCATCACGTCCTTCGAAACTCTTCTCGGGGACAATATACTCAAGCGCCTTTTCAGACCGGCGATAGC 1200
K I Q E L R H W V S G P T A F M Q L Y K A L Q L V D V D F T F E V T K N W E T K
AAAAATCCAAGAACTCAGACACTGGGTAGTGGCCCAACAGCCTTCATGCAACTCTACAAAGCGCTGCAACTCGTCGATGTAGACTTCACTTTTGAGGTCACTAAGAACTGGGAAACAAA 1320
Q E N H S E P E L P K D A I D V L V Q S L I G I N Y T D A A T T E A A N P K Q Q
ACAGGAGAACCACAGCGAACCCGAACTCCCAAAAGATGCCATTGACGTGCTCGTCCAGTCGCTCATCGGCATCAATTACACAGACGCAGCAACTACCGAAGCCGCAAACCCTAAACAACA 1440
A E S Q S D A A P P L T A
E S G P K P Y T N L P Q P L Q Q T K C P E D T E V S A S P E C K A A E T A Q E T
AGAAAGTGGACCAAAACCCTACACCAACCTCCCGCAGCCCCTTCAGCAAACCAAATGCCCCGAGGACACGGAAGTGTCCGCATCACCAGAATGCAAAGCGGCTGAAACTGCGCAAGAGAC 1560
C C AGT G C AT G A T C G T TG A G C C T TC T T A AG
C T S P L K N T
H K A G P S S N P A L N T D Q E A A P S E P T D S E E R K C P T L R E L P W K A
CCACAAAGCAGGGCCCAGTTCAAACCCCGCTCTCAACACTGACCAAGAGGCAGCCCCGTCAGAACCCACAGACAGCGAGGAGCGGAAGTGCCCCACATTGCGTGAACTGCCATGGAAGGC 1680
I P E N S
W L K L L R G L G F Q G D L E Q L D V E G D L I Y P I K H I R H L P V E D F S G
GTGGCTCAAACTACTTCGGGGCCTTGGATTCCAAGGCGACCTTGAACAGTTGGACGTAGAGGGGGACCTCATTTACCCAATTAAACACATCCGGCACTTACCAGTTGAGGATTTCAGTGG 1800
P I T A T L R K M N R M P C K Y S P D I E R A A A F A R D V M S N K T G A I L P
CCCGATCACAGCGACCCTGAGGAAAATGAACCGGATGCCATGCAAATACTCCCCCGACATTGAAAGAGCAGCCGCATTCGCGCGAGACGTGATGAGCAATAAAACAGGGGCCATCCTCCC 1920
L S
K Q S H E W K T T L K R K C K L A P R Q T L I S V I H G A G G S G K S R A L Q E
CAAACAATCCCACGAATGGAAGACCACACTTAAAAGAAAATGCAAACTTGCCCCTCGACAAACTCTTATTTCCGTGATACATGGCGCGGGCGGCAGTGGTAAATCAAGAGCCCTGCAAGA 2040
Y M R N N K E D P I V T I L P T N E L R A D W K A K L P A H D P D M F M T F E N
ATACATGAGGAACAACAAGGAGGACCCCATCGTCACCATCCTTCCGACGAATGAGCTAAGGGCCGATTGGAAAGCTAAACTGCCCGCTCATGACCCTGACATGTTCATGACATTCGAGAA 2160
A L L I P K G N V V I M D D Y T K L P R G Y I E A Y V Q N S P M L D L L I L T G
CGCGCTGCTGATCCCAAAAGGCAACGTAGTCATCATGGACGATTACACTAAGCTACCACGCGGGTACATCGAGGCGTACGTGCAAAACTCCCCAATGCTAGACCTGCTCATACTCACAGG 2280
V
D P R Q S E H F E S A E G N T I N D L S P S P V V F S A Y A R Y Y I N A T H R N
AGACCCCAGGCAATCCGAACACTTCGAATCAGCCGAAGGGAACACGATCAACGATCTCAGCCCCTCTCCAGTCGTCTTCAGCGCGTACGCCCGCTATTACATCAACGCAACTCACAGGAA 2400
N Q D L A N A L G V Y S E K Q G K T R I S V S N H I E D G R H T L V P S Q I K Q
CAACCAGGACCTAGCGAACGCCTTGGGGGTTTACTCTGAAAAACAAGGGAAAACGCGAATATCAGTCAGTAACCATATAGAAGACGGGCGCCACACATTAGTTCCCAGCCAAATCAAACA 2520
R N Y V S M G R K T S T Y A G C Q G I T A S R V Q V I I D N D T A K C S H Q V L
AAGAAACTACGTCTCCATGGGCAGGAAAACTTCTACCTATGCGGGCTGTCAAGGGATAACCGCTTCACGTGTACAGGTCATCATTGACAACGACACTGCAAAGTGCTCACATCAGGTCCT 2640
Y T A F S R A A D E I H F C N T M T D E K Q F W A K V K C T P Y L Q A I L N L T
ATACACAGCTTTTTCCAGGGCCGCAGACGAAATACACTTCTGCAACACTATGACCGACGAAAAACAATTTTGGGCGAAAGTGAAGTGTACCCCGTACTTACAAGCCATCCTCAACCTTAC 2760
K E H S L P D P K P E P D E P L E P K A P V T H I A V E N V V P L V E E I T E P
AAAGGAACACTCTCTGCCTGACCCGAAGCCTGAGCCAGATGAACCCTTGGAGCCCAAAGCCCCTGTCACCCACATTGCTGTGGAAAACGTGGTACCGCTAGTGGAAGAAATCACTGAGCC 2880
L P E K H D R E I Y S E S T G H S N C I Q T E N E F I Q A F Q H Q Q A K D E T L
TCTTCCAGAGAAGCATGACCGAGAGATCTACTCAGAAAGCACCGGACACTCCAACTGCATCCAAACCGAAAACGAGTTCATTCAGGCTTTCCAACACCAACAGGCAAAGGATGAAACTCT 3000
K
F W A T I E K R I V T S T P E S N W T E F K T K R P L G D I L W L A Y K E A M C
CTTCTGGGCCACCATTGAGAAACGCATAGTCACCTCAACCCCTGAATCAAACTGGACTGAATTCAAGACAAAGAGGCCCCTAGGTGACATCTTGTGGCTGGCCTACAAGGAGGCTATGTG 3120
L P A E P I K F D P D L W W K C A D E V Q N K Y L Q K P L H M I K N G I L R Q S
CTTGCCAGCGGAGCCCATCAAATTCGACCCAGACCTTTGGTGGAAATGTGCCGATGAGGTGCAAAATAAATACCTACAAAAACCCCTTCACATGATCAAAAACGGAATACTCAGACAAAG 3240

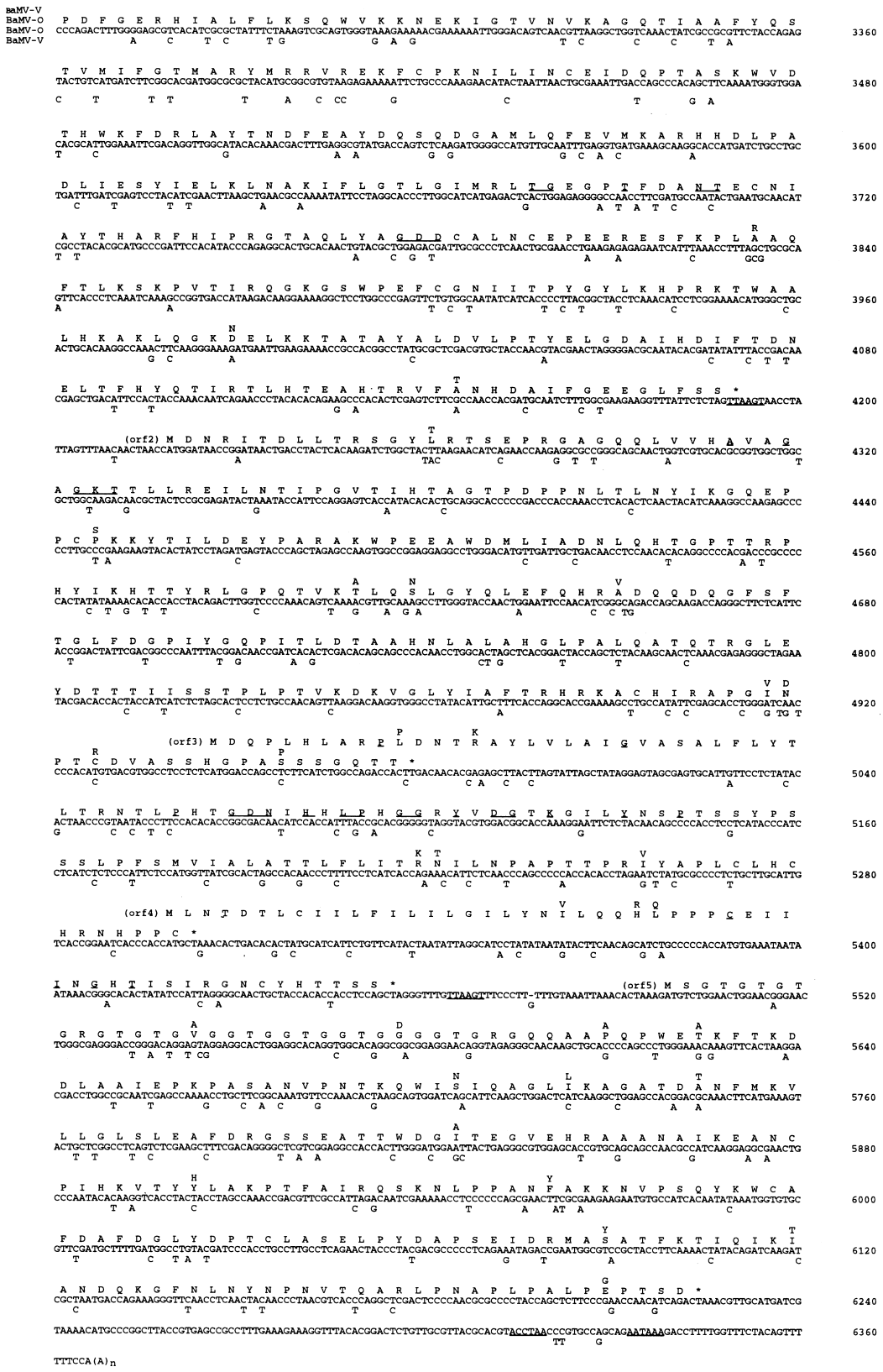


Figure 2. The complete nucleotide sequence of BaMV-V genomic RNA expressed as cDNA sequence with the translated amino acid sequences of ORFs 1–5. Nucleotide and amino acid residues different from those of the previously reported isolate (BaMV-O) are indicated in lower and upper lines, respectively. Gaps are introduced to optimize the alignment. Start codons of each ORF are indicated in parenthesis and stop codons are marked by asterisks. Conserved amino acid residue(s) with other potexviruses in different ORFs are underlined (see text). Conserved hexanucleotides (ACC/TTAA) on both strands and polyadenylation signal are also underlined.

($PX_{10-12}G$) were also found in the N-terminal region of ORF 3 (Figure 2). In ORF 4 of BaMV the conserved motif CX_3IXGXS/T (Neo et al., 1992) was identified following a highly hydrophobic stretch. The functional importance for the conservation of these amino acid residues in ORF 3 and ORF 4 is not yet clear.

Secondary structure predictions based on Chou and Fasman (1978) and ROG (Garnier et al., 1978) methods revealed a β-sheet for the highly hydrophobic region at the N- and C-terminal regions of ORF 3 protein (data not shown). Although the highly hydrophobic regions are a possible transmembrane domain (Claros and von Heijne, 1994), their predicted secondary structure folding does not follow the α-helix or β-sheet transmembrane structures in membrane proteins (Jähnig, 1990). The underlying mechanism for the possible involvement of these structural domains remains to be determined.

Phylogenetic Analysis

The three conservative domains of RdRp common in positive strand RNA viruses were used to infer the phylogenetic relationship among potexviruses. The result showed that, not only are the two isolates of BaMV grouped together, different strains of PVX and WClMV are grouped together as well (Figure 3). Although two strains of PVX (X3 and Andean strains) differ in their overall amino acids by up to 11.2%, they differ in only 24 amino acid residues in the above three conserved regions (533 amino acid residues). On the other hand, the two most closely related potexviruses (BaMV-O and FMV) differ in 188 amino acid residues in these domains. Therefore there is only 12.8% as much variation in strains of potexvirus as in members of the potexvirus group. The results also indicated that members of the potexvirus group have three main lineages; i.e., the BaMV-FMV lineage, the PVX-WClMV lineage, and the CYMV-PMV lineage.

When the amino acid residues from the same domains of RdRp from a carlavirus (PVM) (Zavriev et al., 1991), a furovirus (BNYVV) (Bouzoubea et al., 1987), a hordeivirus (BSMV) (Gustafson et al., 1987; 1989), and a turnip yellow mosaic virus (TYMV) (Morch et al., 1988) were compared with potexviruses, including BaMV-V, their phylogenetic relationships were summarized as shown in Figure 4. The results indicated that all the potexviruses group together when members of other plant virus groups are used in comparison. The difference of the two most divergent potexviruses (BaMV and NMV, 284 amino acid differences) is 67.6% of that between PVX (X3) (the potexvirus member closest to the other virus groups used in comparison) and TYMV (the virus in other groups closest to potexvirus, 420 amino acid differences). Therefore, this comparison supports the notion that potexviruses are most appropriately grouped based on the three conserved domains of RdRp.

Acknowledgements. We thank Mr. Shih-Ting Chang-Chien for some sequencing work. This research was supported in part by grants from Academia Sinica and a National Science Council Project NSC 84-2311-B-001-008-B11.

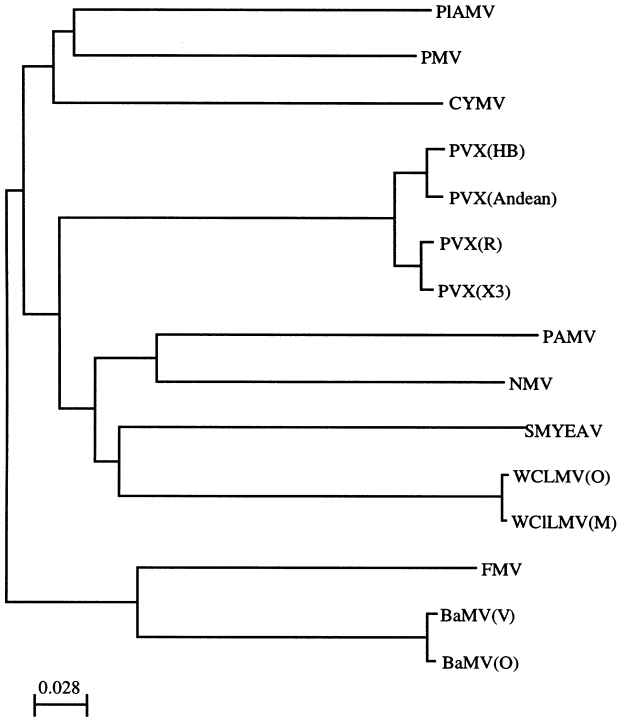


Figure 3. Phylogenetic relationships of strains and members from the potexvirus group. Amino acid residues of three conservative domains (methyltransferase, nucleotide-binding, and polymerase domains) of RdRp (Koonin et al., 1993) are used to infer their phylogenetic relationship.

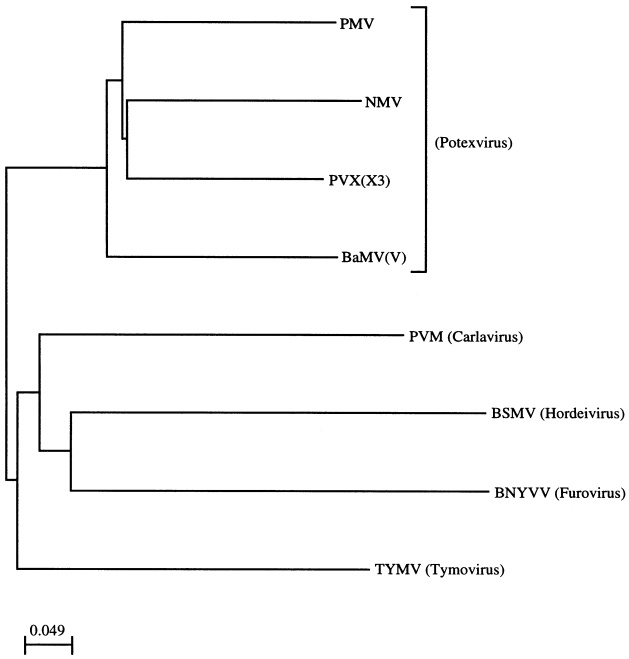


Figure 4. Phylogenetic relationship of members of potexvirus, carlavirus, hordeivirus, furovirus, and tymovirus groups using amino acid residues of three conservative domains of RdRp.

Literature Cited

Angell, S.M., C. Davies, and D.C. Baulcombe. 1996. Cell-to-cell movement of potato virus X is associated with a change in the size-exclusion limit of plasmodesmata in trichome cells of *Nicotiana clevelandii*. Virology **216:** 197–201.

Argos, P. 1988. A sequence motif in many polymerase. Nucleic Acids Res. **16:** 9909–9915.

Atabekov, J.G. 1993. Genebank accession X05198.

Bancroft, J.B., M. Rouleau, R. Johnston, L. Prins, and G.A. Mackie. 1991. The entire nucleotide sequence of foxtail mosaic virus RNA. J. Gen. Virol. **72:** 2713–2181.

Beck, D.L., R.L.S. Forster, M.W. Bevan, K.A. Boxen, and S. Lowe. 1990. Infectious transcript and nucleotide sequence of cloned cDNA of the potexvirus white clover mosaic virus. Virology **177:** 152–158.

Beck, D.L., P.J. Guilford, D.M. Voot, M.T. Andersen, and R.L.S. Forster. 1991. Triple gene block protein of white clover mosaic potexvirus are required for transport. Virology **183:** 695–702.

Bouzoubaa, S., V. Ziegler, D. Beck, H. Guilley, K. Richard, and G. Jonard. 1986. Nucleotide sequence of beet necrotic yellow vein virus RNA-2. J. Gen. Virol. **67:** 1689–1700.

Bouzoubaa, S., L. Quillet, H. Guilley, G. Jonard, and K. Richard. 1987. Nucleotide sequence of beet necrotic yellow vein virus RNA-1. J. Gen. Virol. **68:** 615–626.

Chou, P.Y. and G.D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47:** 45–148.

Claros, M.G. and G. von Heijne. 1994. TopPred II: an improved software for membrane protein structure predictions. CABIOS **10:** 685–686.

Fichot, O. and M. Girard. 1990. An improved method for sequencing of RNA template. Nucleic Acids Res. **18:** 6162.

Forster, R.L.S., P.J. Guilford, and D.V. Faulds. 1987. Characterization of the coat protein subgenomic RNA of white clover mosaic virus. J. Gen. Virol. **68:** 181–190.

Forster, R.L.S., M.W. Bevan, S.A. Harbison, and R.C. Gardner. 1988. The complete nucleotide sequence of the potexvirus white clover mosaic virus. Nucleic Acids Res. **16:** 219–303.

Garnier, J., D.J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:** 97–120.

Gorbalenya, A.E., V.M. Blinov, A.P. Donchenko, and E.V. Koonin. 1989. An NTP–binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. J. Mol. Evol. **28:** 256–268.

Gouy, M. 1995. NJPlot software. *In* Clustal W 1.5 software package.

Gustafson, G. and S.L. Armour. 1986. The complete nucleotide sequence of RNA beta from the type strain of barley stripe mosaic virus. Nucleic Acids Res. **14**: 3895–3909.

Gustafson, G., S.L. Armour, G.C. Gamboa, S.G. Burgett, and J.W. Shephard. 1989. Nucleotide sequence of barley stripe mosaic virus RNA-alpha: RNA-alpha encodes a single polypeptide with homology to proteins from other viruses. Virology **170:** 370–377.

Gustafson, G., B, Hunter, R. Hanau, S.L. Armour, and A.O. Jackson. 1987. Nucleotide sequence and genetic organization of barley stripe mosaic virus RNA-gamma. Virology **158:** 394–406.

Hodgman, T.C. 1988. A new superfamily of replicase protein. Nature (London) **333:** 22–23.

Huisman, M.J., H.J.M. Linthorst, J.F. Bol, and B.J.C. Cornelissen. 1988. The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus–stranded RNA virus. J. Gen. Virol. **69:** 1789–1798.

Jähnig, F. 1990. Structure predictions of membrane proteins are not that bad. Trends Biochem. Sci. **15:** 93–95.

Jelkmann, W., E. Maiss, and R.R. Martin. 1992. The nucleotide sequence and genome organization of strawberry mild yellow edge-associated potexvirus. J. Gen. Virol. **73:** 475–479.

Koonin, E.V. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-stranded RNA virus. J. Gen. Virol. **72:** 2197–2206.

Koonin, E.V. and V.V. Dolja. 1993. Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. **28:** 375–430.

Lin, M.T., E.W. Kitajima, F.P. Cupertino, and C.L. Costa. 1977. Partial purification and some properties of bamboo mosaic virus. Phytopathology **67:** 1439–1443.

Lin, N.S. and C.C. Chen. 1991. Association of bamboo mosaic virus (BaMV) and BaMV-specific electron–dense crystalline bodies with chloroplasts. Phytopathology **81:** 1551–1555.

Lin, N.S., F.Z. Lin, T.Y. Huang, and Y.H. Hsu. 1992. Genome properties of bamboo mosaic virus. Phytopathology **82:** 731–734.

Lin, N.S., B.Y. Lin, N.W. Lo, C.C. Hu, T.Y. Chow, and Y.H. Hsu. 1994. Nucleotide sequence of the genomic RNA of bamboo mosaic potexvirus. J. Gen. Virol. **75:** 2513–2518.

Lin, N.S. and Y.H. Hsu. 1994. A satellite RNA associated with bamboo mosaic potexvirus. Virology **202:** 707–714.

Lok, S. and M.G. AbouHaidar. 1986. The nucleotide sequence of the 5' end of papaya mosaic virus RNA: site of in vitro assembly initiation. Virology **153:** 289–296.

Marck, C. 1988. 'DNA strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:** 1829–1836.

Mentaberry, A.N. 1992. Genebank accession X55802.

Morch, M.D., J.C. Boyer, and A.L. Laenni. 1988. Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. Nucleic Acids Res. **16:** 6157–6173.

Neo, K.K., S.M. Wong, and M. Wu. 1992. Nucleotide sequences of the two ORFs upstream to the coat protein gene of cymbidium mosaic virus. Plant Mol. Biol. **18:** 1027–1029.

Orman, B.E., R.M. Celnik, A.M. Mandel, H.N. Torres, and A.N. Mentaberry. 1990. Complete cDNA sequence of a South American isolate of potato virus X. Virus Res. **16:** 293–306.

Randles, J.W. and W. Rohde. 1990. *Nicotiana velutina* mosaic virus: evidence for a bipartite genome comprising 3 kb and 8 kb RNAs. J. Gen. Virol. **71:** 1019–1027.

Rozanov, N.M.N., E.V. Koonin, and A.E. Gorbalenya. 1992. Conservation of the putative methytransferase domain: a hallmark of the sindbis–like supergroup of positive-strand

RNA viruses. J. Gen. Virol. **73:** 2129–2134.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. "Molecular Cloning, a Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Short, M.N. and J.W. Davies. 1983. Narcissus mosaic virus: a potexvirus with an encapsidated subgenomic messanger RNA for coat protein. Biosci. Rep. **3**: 837–846.

Sit, T.L., M.G. AbouHaidar, and S. Holy. 1989. Nucleotide sequence of papaya mosaic virus RNA. J. Gen. Virol. **70:** 2325–2331.

Sit, T.L., D. Leclerc, and M.G. AbouHaidar. 1994. The minimal 5' sequence for in vitro initiation of papaya mosaic potexvirus assembly. Virology **199:** 238–242.

Sit, T.L., K.A. White, S. Holy, U. Padmanabhan, M. Eweida, M. Hiebert, G.A. Mackie, and M.G. Abouhaidar. 1990. Complete nucleotide sequence of clover yellow mosaic virus RNA. J. Gen. Virol. **71:** 1913–1920.

Skryabin, K.G., A.S. Kraev, S.Y. Morozov, M.N. Rozanov, B.K. Chernov, L.I. Lukasheva, and J.G. Atabekov. 1988a. The nucleotide sequence of potato virus X RNA. Nucleic Acids Res. **16:** 10929–10930.

Skryabin, K.G., S.Y. Morozov, A.S. Kraev, M.W. Rozanov, B.K. Chernov, L.I. Lukasheva, and J.G. Atabekov. 1988b. Conserved and variable elements in RNA genomes of potexviruses. FEBS Lett. **240:** 33–40.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalities and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Wang, L.M., D.K. Weber, T. Johnson, and A.Y. Sakaguchi. 1988. Double-stranded DNA sequencing with T7 polymerase. BioTechnique **6:** 843–846.

White, K.A., J.B. Bancroft, and G.A. Mackie. 1992. Mutagenesis of a hexanucleotide sequence conserved in potexvirus RNAs. Virology **189:** 817–820.

Xu, H., D. Leclerc, B. Leung, and M.G. AbouHaidar. 1994. The entire nucleotide sequence and genomic organization of potato aucuba mosaic potexvirus. Arch. Virol. **135:** 461–469.

Zavriev, S.K., K.V. Kanyuka, and K.E. Levay. 1991. The genome organization of potato virus M RNA. J. Gen. Virol. **72:** 9–14.

Zuidema, D., H.J.M. Linthorst, M.J. Huisman, C.J. Asies, and J.F. Bol. 1989. Nucleotide sequence of narcissus mosaic virus RNA. J. Gen. Virol. **70:** 267–276.