Arch Virol (1998) 143: 1489–1500

Archives of Virology

Genomic sequence of physalis mottle virus and its evolutionary relationship with other tymoviruses*

C. T. Ranjith-Kumar[1], **K. Gopinath**[1], **A. N. K. Jacob**[1], **V. Srividhya**[1], **P. Elango**[2], and **H. S. Savithri**[1]

[1]Department of Biochemistry, Indian Institute of Science, Bangalore, India
[2]Bioinformatics Centre, Indian Institute of Science, Bangalore, India

Accepted April 3, 1998

Summary. The genome of physalis mottle tymovirus (PhMV) is 6673 nucleotides long and is rich in cytosine residues (40.58%) like other tymoviruses. The organization of the genes is also similar to that of five other tymoviruses whose sequences are known. However, PhMV has the longest 3′ noncoding region as well as the longest replicase (RP) ORF. The RP sequences are similar to those of other tymoviruses (48–60% identity) whereas the coat proteins (CP) and the overlapping proteins (OP) are conserved to a lesser extent (30–50% and 26–34% respectively). A tetra peptide "GILG" was found to be present in all the tymoviral OPs. The PhMV RP also possesses the methyl transferase, polymerase and the helicase motifs found in all the Sindbis-like super group of plant viruses. A phylogenetic analysis of the six tymoviral sequences revealed that they do not have a rigid hierarchical similarity relationship.

Introduction

Physalis mottle virus, a member of the tymo group of plant viruses, infects solanaceous plants causing systemic mottling symptoms. This virus was isolated by Moline and Fries in Iowa, USA [19]. It was originally known as Physalis strain of belladonna mottle virus (BDMV-I) based on the serological relationships with the European strain, BDMV-E. A phylogenetic tree constructed after pairwise alignment of the tymoviral coat protein (CP) sequences showed that BDMV-I was not a strain of BDMV-E and was therefore renamed physalis mottle virus (PhMV) [14].

*The sequence described in this paper was submitted to GenBank and has been assigned accession no. Y16104.

PhMV, like other tymoviruses, has a single stranded RNA genome of positive polarity. This RNA genome is encapsidated in an icosahedral shell of 180 identical coat protein (CP) subunits of Mr 20,000. The complete nucleotide sequence of five tymoviruses, turnip yellow mosaic virus, TYMV [20], ononis yellow mosaic virus, OYMV [6], kennedya yellow mosaic virus, KYMV [7], eggplant mosaic virus, EMV [18] and erysimum latent virus, ELV (27) are known. The CP gene is 3′-proximal in the genomes of all tymoviruses sequenced thus far. The tymoviral genome also codes for two other proteins, the overlapping protein (OP) and the replicase protein (RP), which are coded by reading frames that overlap [6, 7, 18, 20, 27]. A comparison of tymoviral sequences with that of other plant and animal viruses has shown that these viruses belong to the Sindbis-like super family of plant viruses [1]. Determination of the genomic sequence of the other members of tymovirus group further strengthens the established relationship among the members of the group as well as with various other groups of plant viruses.

Recently, we have demonstrated the role of the pseudoknot structure present at the 3′ terminus of PhMV genomic RNA in virus multiplication using an in vivo protoplast assay system [22]. To locate other regions of PhMV RNA that are important for virus multiplication, it is necessary to determine the full genomic sequence. In this paper we report the complete nucleotide sequence of the PhMV genome and a comparison of this sequence with other tymoviral sequences.

Materials and methods

Viral cDNA synthesis

A synthetic oligodeoxy nucleotide (5′ TGGCAGGCCAATTCGGGGA 3′) complementary to the 5′end of the PhMV sequence determined earlier [14] was used to screen a cDNA library generated using oligo dT as primer. cDNA was also synthesized using oligonucleotides (5′ GAGAGACGAGGGTTGACAAGAGAGGGAGAA3′) and 5′AAGTTCTCTGTTTGGGCAGAGAACAA 3′) complementary to specific sequences of PhMV genome and cloned into appropriately cleaved pBlueScript or pGEM vectors [24]. Eventually four clones TA51 (2091 nts), D88 (1794nts), pG55(1749 nts) and GL18(1390 nts), which spanned the entire length of PhMV genome were obtained.

Sequencing

Initially, the clones were sequenced manually by Sanger's dideoxy chain termination method [25]. Some of the clones were later sequenced using a Perkin-Elmer ABI Prism DNA sequencer. The templates used were the original clones, subclones generated by restriction digestion or the clones generated by exonuclease III deletion. The sequences were determined for both the strands.

Exonuclease III/S1 nuclease deletion analysis

Supercoiled plasmids were prepared by the method of Wang and Rossman [30]. Deletions were generated using exonuclease III/SI nuclease as described in the Promega protocols and applications guide.

Analysis of sequence

Nucleic acid and deduced amino acid sequences were analyzed using the GCG program package at the Bioinformatics center, IISc. The PhMV sequence was compared and aligned with the other tymoviral sequences available in the GenBank and EMBL data base using the "FASTA" and "GAP" programs. The protein sequences of the different tymoviruses were compared using "GAP" program.

Phylogenetic trees

The program CLUSTAL W was used to generate multiple alignment [29]. The aligned sequences were analyzed with the PHYLIP package of programs (10, 11) to infer phylogenesis and to place a confidence limit on the phylogeny. The program SEQBOOT was used to produce 100 bootstrap [9] replacement sequence alignments from the original CLUSTAL W aligned sequences. The bootstrapped samples were analyzed by the maximum parsimony method using the program PROTPARS. The trees produced by PROTPARS were analyzed by the program CONSENSE to deduce a consensus tree. The unrooted phylogenetic tree was drawn using the program DRAWTREE.

Results and discussion

PhMV genomic sequence is 6 673 nucleotides long. The nucleotide sequence was compiled from the four major cDNA clones TA51, D88, pG55 and GL18, their deletion clones and from the sequences of many short overlapping clones. TA51 encompasses the 3′ terminal 2 091 nucleotides and the 5′ 1 390 nucleotides are represented in GL18. pG55 and D88 encompass the region 1 317 nt to 3 066 nt and 2 841 nt to 4 635 nt respectively. A major problem encountered during the determination of genomic sequence of PhMV was the presence of long streches of 'C' residues which lead to termination of the sequence at the same position in many of the clones. Particularly, the nucleotide sequence from 2 154 to 2 177 had 21 'C' residues and was difficult to sequence as it formed a strong secondary structure. An oligonucleotide close to this sequence, when used as primer, finally yielded the sequence corresponding to this region. The genomic sequence of PhMV along with the predicted amino acid sequence of the three open reading frames (ORFs) is shown in Fig. 1.

The PhMV genome has a base composition of A (22.52%), C (40.58%), G (13.1%) and U (23.80%). Like all the other known tymoviral sequences, the C content is very high and G content is low. There are 23 strings of five cytosines (C5), 8 of C6, two each of C8 and C9 and one each of C7 and C12. Such homopolymeric tracts are known to act as potential sites for ribosomal frameshifting which occurs in the translational regulation of several viral RNAs [4]. It is not clear whether such a frameshifting occurs in tymoviruses. The large cytosine content is reflected in both the codon usage and the abundance of proline, leucine and serine residues in the proteins encoded.

5′ and 3′ noncoding regions

Both the 5′ and 3′ noncoding (NC) regions of PhMV are 149 nucleotide residues long. The shortest stretch of 5′ NC sequence is 78 nucleotides in KYMV and

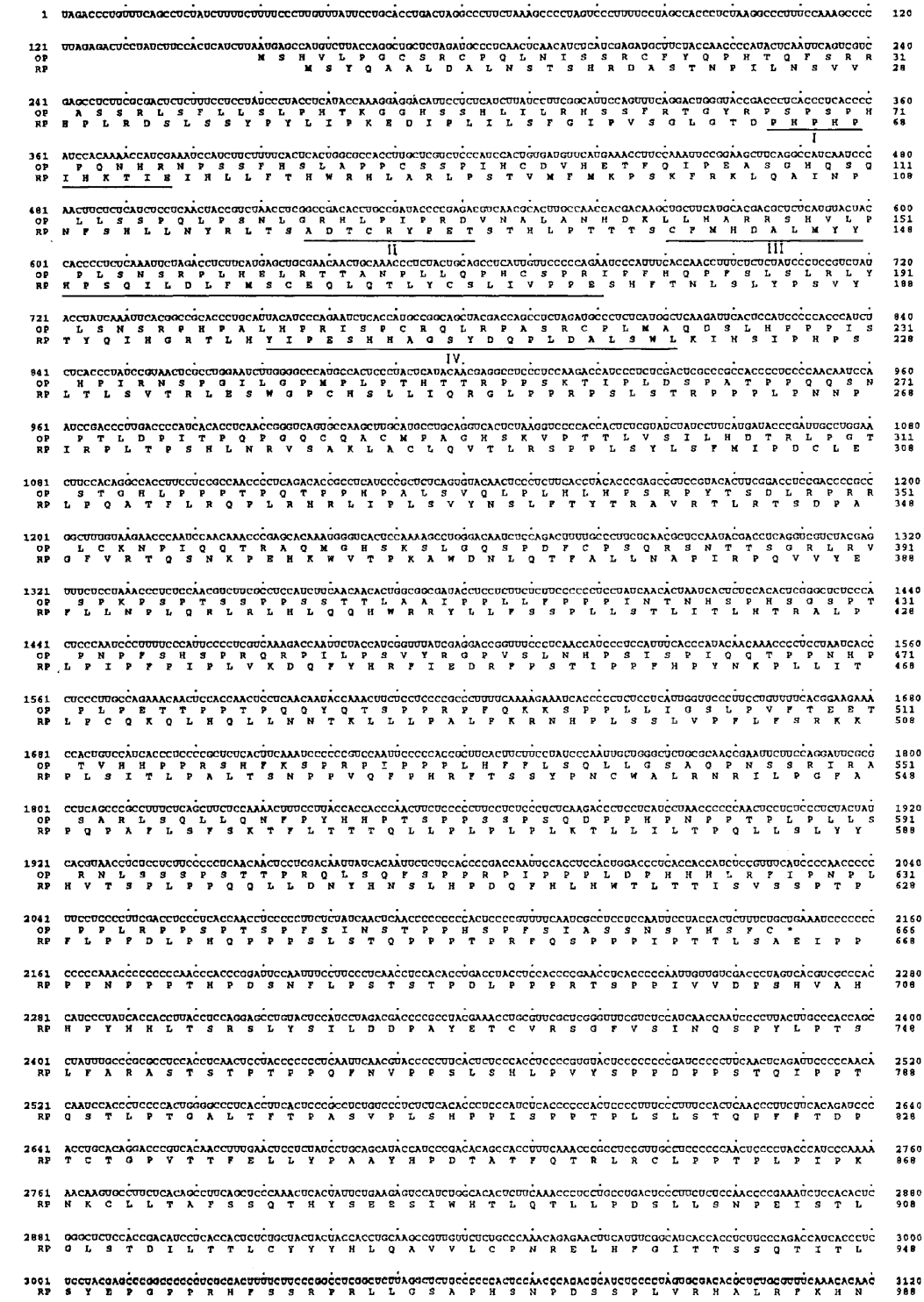
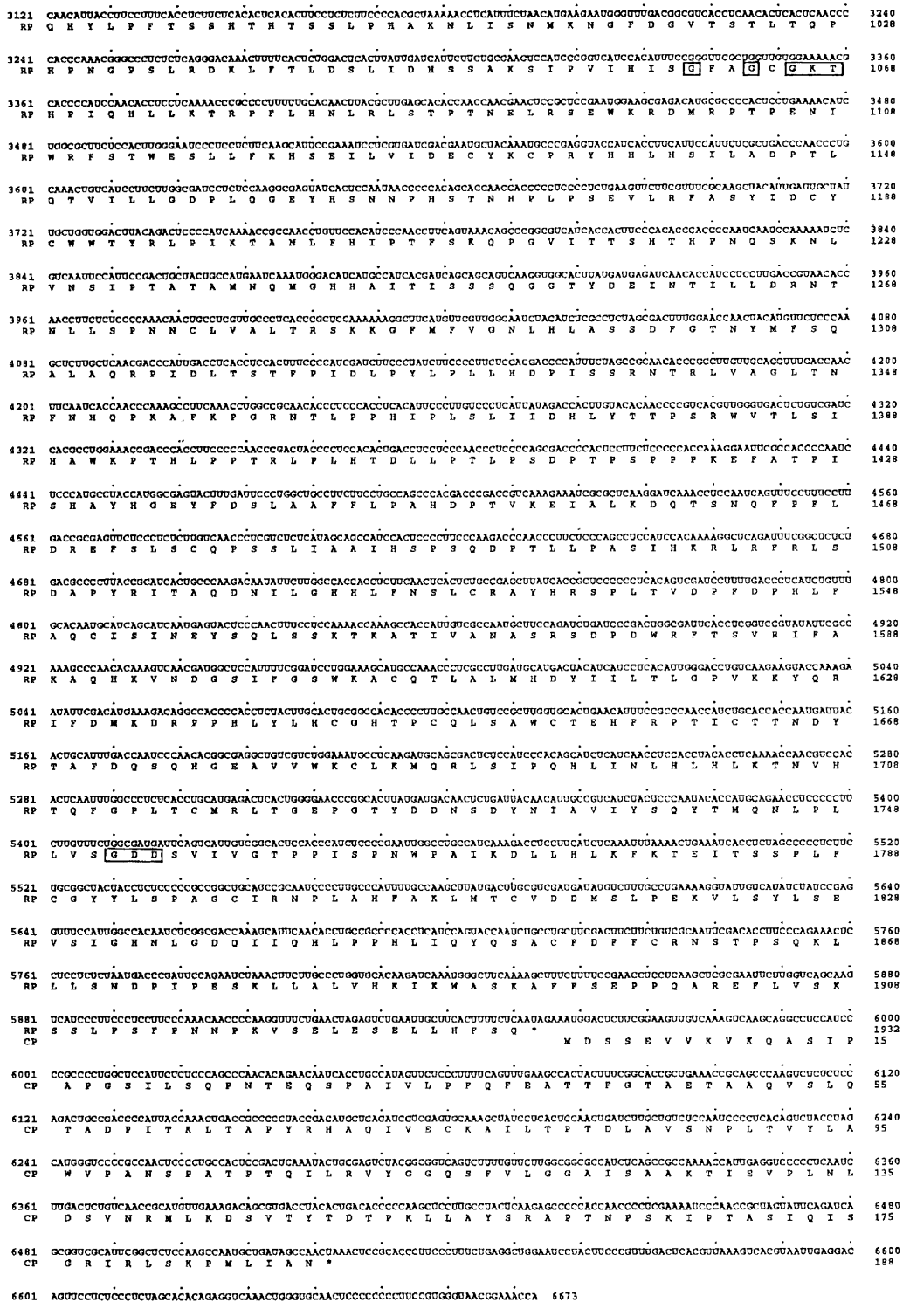


Fig. 1. Nucleotide sequence of 6 673 nt of PhMV genomic RNA. The predicted amino acid sequence of each ORF is shown below the sequence in one-letter code. The nucleotide sequences are numbered on the either side of each line and the amino acid sequence are numbered on the right side of each line. Underlined residues in RP indicate the conserved



motifs within the methyl transferase domain. The motif numbers are indicated as I, II, III and IV. Boxed amino acids correspond to conserved sequences found in other RNA viruses as being a nt binding site (amino acids 1 061–1 068) and a polymerase site (amino acids 1 752–1 754)

ELV. OYMV has the longest 3′ NC sequence (171 nucleotides). PhMV CP has maximum similarity to EMV CP [14]. However, no significant homology was observed between the 5′ NC sequence of EMV and PhMV. Further, the 5′ NC of PhMV is 48 nucleotides longer than that of EMV. The 5′ terminal sequence was determined in three independent clones and begins with 5′-UAGA-3′ whereas in other tymoviruses the 5′ terminus contains the sequence 5′ GGUAA-3′ (5′-GGAAA-3′ for ELV). It is probable that the terminal two G residues were lost when the cDNA clones of PhMV were made and the probable 5′ terminus is 5′-GGUAGA-3′. This however remains to be established. Therefore, the numbering of the nucleotides in the case of PhMV genome is only tentative and it starts from the first nucleotide of the sequence 5′-UAGA3′.

The 5′ NC of tymoviruses contain pyrimidine-rich stretches. In PhMV, this region extends from nucleotide 16–47 interrupted by only three purines. This stretch is followed by shorter stretches of 7–14 residues, also containing pyrimidines. Secondary structure predictions for the 5′ NC region of PhMV showed that it is capable of folding into weak stem-loop structures and the pyrimidine rich stretches occur in the loops. It has been suggested [6] that this pyrimidine rich core sequence base pairs with the 5′-GGAAGG-3′ sequence near the 3′ terminus of wheat 18S RNA and perhaps facilitates translation of the replicase protein. In PhMV, the 5′-CUUCC-3 stretch between nucleotides 135–139, close to the start codon of the OP, shows complementarity to the wheat sequence, similar to that reported for OYMV [6].

It has been shown by Hellendoorn et al. [12] that despite large sequence differences, the 5′ NC regions of tymoviruses have remarkably similar secondary structures. Two or four hairpins containing symmetrical internal loops consisting of adjacent C-C or C-A mismatches, which can be protonated, are found in all tymoviruses. The C-C and C-A mismatches seem to play an important role in RNA-protein interactions and thus in the assembly of the virus [13].

As reported earlier, the 149 nucleotide 3′ NC region of PhMV is the longest observed among tymoviruses thus far and the terminal 81 nucleotides can be folded into a tRNA-like structure (TLS) [14]. The aminoacyl arm of this TLS differs from the canonical tRNA in that it has a pseudoknot structure. Recently, using an in vivo proptoplast assay system, we have shown that mutant transcripts in which the pseudoknot structure was either disrupted or restored failed to compete with genomic RNA and inhibit viral CP synthesis, whereas the wildtype sense NC transcript showed complete inhibition [22]. The pseudoknot structure has also been shown to be important in TYMV replication [5, 26].

ORFs

The deduced amino acid sequences of the three large ORFs in the positive strand of RNA corresponding to OP, RP and CP are shown in Fig. 1. The ORF that initiates closest to the 5′ terminus at nucleotide 150 (the first nucleotide of the triplet) and ends in UGA at position 2 150 (last nucleotide of the stop codon) encodes the overlapping protein of size 666 amino acids (aa). Among the OP of

Table 1. Percentage identity of the amino acid sequence of the RPs (upper triangle) and the OPs (lower triangle) of tymoviruses

RP / OP	PhMV	EMV	OYMV	TYMV	KYMV	ELV
PhMV		59.55	53.52	51.89	51.65	48.10
EMV	34.4		55.84	51.99	53.15	50.09
OYMV	33.97	32.93		50.46	52.92	48.47
TYMV	29.65	30.47	31.26		55.74	49.65
KYMV	32.37	31.62	29.97	31.3		49.43
ELV	26.65	30.68	30.89	30.51	27.85	

the known tymoviruses only KYMV OP (753 aa) is longer than that of PhMV. ELV OP is the shortest and only 440 aa long. Comparisons of the OP of PhMV with five other tymoviruses whose complete genome sequence is available, is shown in Table 1 (lower triangle). As evident, the OPs are very variable in length and exhibit an identity of 26–34%. The OP sequences were compared with the OP sequences of other plant viruses and no sequence similarity could be detected. A tetra peptide, "GILG" (residues 239–242 in the case of PhMV) is found in all the tymoviral OPs known. This tetra peptide may have a structural or functional role. In TYMV, it was shown that the OP is involved in the systemic spread of virus within the plant and possibly also in the cell-to-cell movement [2].

The RP is encoded by the second ORF from the 5′ terminus of PhMV gRNA starting at nucleotide position 157 and terminating with UAG at position 5 955. As in all other tymoviruses, the start codons of the RP ORFs and OP ORFs are 7 nucleotides apart. The PhMV RP is the longest (1932 aa) among all tymoviral RPs known: that of ELV is the shortest 1 747aa in length. The RP is by far the most conserved of all the proteins in tymoviruses. Comparative analysis shows that identities lie between 48–60% (Table 1, upper triangle). The OP ORF is in the -1 reading frame with respect to RP ORF. Therefore, while mutations in the third position of the RP codons (having a preponderance of cytosines) would not affect the amino acid sequence of the RP, it would be reflected in a change in the OP sequence, as these residues would occupy the first positions in the OP codons. Functional constraints operate towards greater conservation of the RP as compared to the OP. The sequence of the RP shows the presence of an RNA methyl transferase domian typically found in the Sindbis-like super group of plus strand RNA viruses [23], the super group to which PhMV belongs. This domain is located within the N-terminal 250 amino acids of the RP and encompasses four distinct conserved motifs (Fig. 1). The motif-GXXGXGK(T/S)- characteristic of many NTP-utilzing enzymes and -GDD-motif typical of viral polymerases are located in the C-terminal half of the PhMV RP (Fig. 1).

In addition to the strategy of using overlapping reading frames, tymoviruses also use the strategy of proteolytic processing for the expression of viral genes [16, 21]. In TYMV, the 206 kDa RP has been shown to undergo proteolytic

processing to p141 and p66 fragments [3, 17]. The residues C783 and H869 are the putative active site residues of a papain-like protease domain present within the RP, which functions to cleave the RP between 1259A–1260T site in the case of TYMV [3]. Although the residues of the putative active site are conserved among all tymoviruses including PhMV, the sequence at the cleavage site is not conserved.

The third and 3′ proximal ORF encoding CP, starts at nucleotide position 5 958 and terminates with the stop codon UAA at nucleotide 6 524 and is 188 amino acids long. The CP ORF is in the same reading frame as the OP ORF and begins two nucleotides after the RP stop codon. The CP is expressed via a subgenomic RNA as in other tymoviruses [15]. The conserved 16 nucleotide stretch "tymobox" (between nucleotides 5 926–5 941) has been proposed to be the promoter element for subgenomic RNA synthesis [8]. The amino acid sequences of both deduced [14] and experimentally determined [28] PhMV coat protein was reported earlier and the similarity of CPs of different tymoviruses lies between 30–50% [14]. Thus the CP is better conserved than the OP but not as much as the RP.

The nucleotide sequence context of the initiation codons of the three ORFs of PhMV was compared with those of other tymoviruses. In all tymoviruses, except ELV the + 4 position with respect to CP ORF is guanine as observed in most plant genes. On the other hand, OP ORFs do not have a G at + 4 position. It was therefore suggested that the RP ORF may be translated in preference to the OP ORF, as RP ORFs have a favorable sequence context for intiation of translation [18]. However, in the case of PhMV RP ORF a 'U' occupies the + 4 position. In vitro translation studies on PhMV and other tymoviruses [16] indicate that both OP and RP ORFs are indeed translated.

Phylogenetic analysis

The phylogenetic analyses of the OP, RP and CP sequences of tymoviruses were done as described in Materials and methods. The cladograms representing the consensus relationships as found from 100 bootstrap samples [9] for OP and CP sequences and 50 bootstrap samples for RP sequence are shown in the Fig. 2A. The number at the branch nodes indicates the number of times that a particular branch appeared in all the trees that were used to determine the consensus tree. The cladogram constructed with the CP sequences is identical to the pattern reported earlier [14, 27]. However, the cladograms obtained with the OP and RP sequences are different from that of the CP in that ELV and OYMV are closer to each other and originate from the same node. These results suggest that the tymoviral sequences are not related by strict hierarchical relationships. Instead, a star phylogeny wherein the sequence differences are accounted for in terms of the taxa diverging from the same time point appears more appropriate. The phylogenetic analysis carried out by Srifah et al. [27] had revealed identical cladograms with all the three tymoviral proteins. It is difficult to comment on the significance of these differences as the methods used are different.

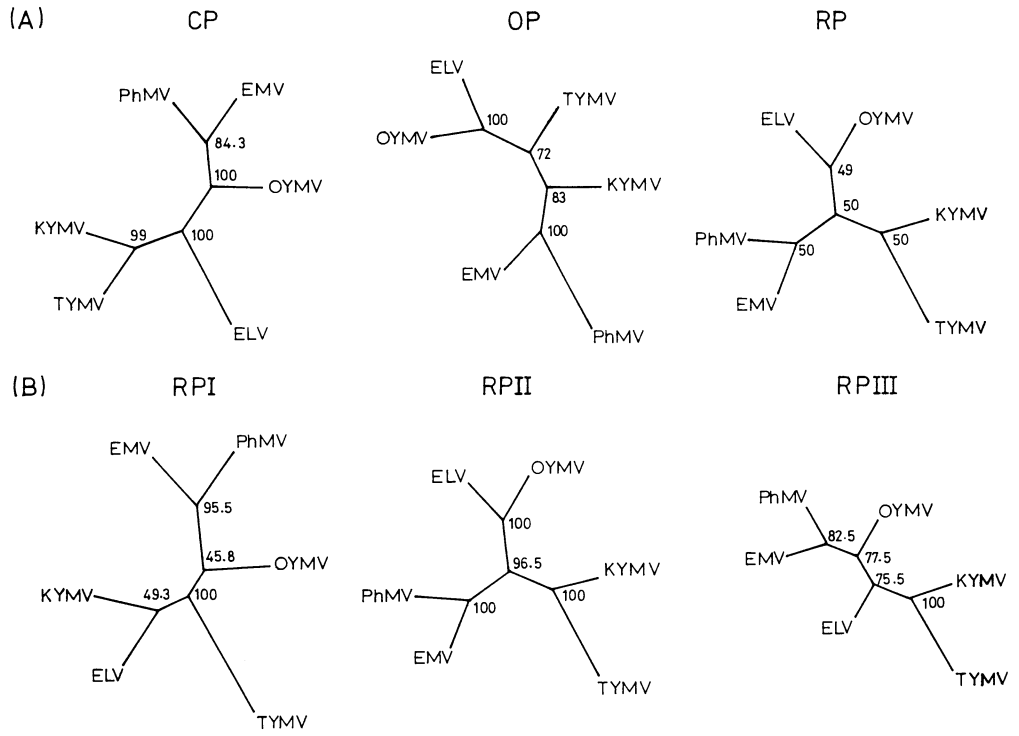


Fig. 2. Phylogenetic analysis of tymoviral proteins. **A** The cladograms representing the consensus relationships obtained from 100 bootstrap samples for OP and CP sequences and 50 bootstrap samples RP sequences. **B** Cladograms constructed using residues 1–252 (RP I), 253–882 (RP II) and 823–1 932 (RP III) in the case of PhMV from 100 bootstrap samples. The number at the branch nodes indicates the number of times that a particular branch appeared in all the trees that were used to determine the consensus tree

The RP ORF is the longest and the most conserved (identity range 48–60%). The CLUSTAL W method used for alignment of the RP ORFs, indicated that N-terminal 1–252 and C-terminal 823–1 932 residues (in the case of PhMV) are highly conserved in all tymoviruses. The region 253–822 (in the case of PhMV) is less conserved (Table 2) and is also variable in length. In fact, PhMV has the maximum length in this segment and therefore has the longest RP sequence. The reason for the variability in segment II is unclear. However, none of the conserved motifs of the RP sequences occur in this segment. Phylogenetic analyses of these segments showed that the cladograms, obtained from 100 bootstrap samples, were different (Fig. 2B); that of segment II being identical to the RP cladogram, segment III to CP whereas the cladogram in segment I had KYMV and ELV originating from the same node, although this relationship occurred rather infrequently (49.3%). These observations suggest that the tymoviruses for which the sequences have been determined so far do not have a rigid hierarchical similarity relationships when examined by parsimony methods.

Table 2. Percentage identity of the amino acid residues 1–252 (I), 253–822 (II) and 823–1932 (III) of PhMV RP with other tymoviral RPs

		PhMV	EMV	OYMV	TYMV	KYMV	ELV
	I						
PhMV	II						
	III						
	I	72.22					
EMV	II	43.80					
	III	63.32					
	I	63.10	65.87				
OYMV	II	40.15	40.61				
	III	56.33	58.42				
	I	59.92	59.92	63.49			
TYMV	II	43.80	42.42	37.53			
	III	55.39	54.93	51.59			
	I	61.11	60.71	61.90	61.91		
KYMV	II	37.58	39.59	39.22	43.42		
	III	55.43	56.83	55.44	60.02		
	I	55.56	55.56	57.14	55.56	55.95	
ELV	II	35.11	37.89	37.22	36.49	38.20	
	III	50.59	52.30	50.54	51.67	51.63	

Acknowledgements

We thank Prof. M. R. N. Murthy for helpful discussions on the analysis of the PhMV genome and Prof. N. Appaji Rao for all the encouragement. The timely gift of oligonucleotides by Dr. Anne-Lise Haenni, Dr. N. Bhaskaran and Dr. Santha Ramakrishnan is gratefully acknowledged. This work was supported by the Indo-French Centre for the Promotion of Advanced Research, Department of Biotechnology and Department of Science and Technology, New Delhi, India. The support of Department of Biotechnology, New Delhi, India in providing Bioinformatics facility and automated DNA sequencing facility is acknowledged. The help of Mrs. Shyamala of the Bioinformatics Centre is gratefully acknowledged. We thank Prof. M. R. S. Rao and Ms Savitha for their help in sequencing some of the cDNA clones described in the paper.

References

1. Ahlquist P, Strauss EG, Rice CM, Strauss JM, Haseloff J, Zimmern D (1985) Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J Gen Virol 53: 536–542
2. Bozarth CS, Weiland JJ, Dreher TW (1992) Expression of ORF-69 of turnip yellow mosaic virus is necessary for viral spread in plants. Virology 187: 124–130
3. Bransom KL, Wallace SE, Dreher DW (1996) Identification of the cleavage site recognised by the turnip yellow mosaic virus protease. Virology 217: 404–406
4. Craigen WJ, Caskey CT (1987) Translational frame shifting: where will it stop? Cell 50: 1–12
5. Deiman BALM, Kortlever RM, Pleij CWA (1997) The role of the pseudoknot at the 3′ end of turnip yellow mosaic virus RNA in minus-strand synthesis by the viral RNA-dependent RNA polymerase. J Virol 71: 5 990–5 996

6. Ding S, Keese P, Gibbs A (1989) Nucleotide sequence of the ononis yellow mosaic tymovirus genome. Virology 172: 555–563
7. Ding S, Keese P, Gibbs A (1990) The nucleotide sequence of the genomic RNA of Kennedya yellow mosaic tymovirus-Jervis Bay isolate: relationships with potex- and carlaviruses. J Gen Virol 71: 925–931
8. Ding S, Home J, Keese P, Mackenzie A, Meek D, Keese MO, Skotnickin M, Srifah P, Torronen M, Gibbs A (1990) The tymobox, a sequence shared by most tymoviruses: its use in molecular studies of tymoviruses. Nucleic Acids Res 18: 1 181–1 187
9. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783
10. Felsentein J (1988) Phylogenies from molecular sequences: inference and reliability. Ann Rev Genet 22: 521–565
11. Felsenstein J (1989) PHYLIP- Phylogeny inference package (version 3.2). Cladistics 5: 164
12. Hellendoorn K, Michiels PJA, Buitenhuis R, Pleij CWA (1996) Protonatable hairpins are conserved in the 5′-untranslated region of tymovirus RNAs. Nucleic Acids Res 24: 4 910–4 917
13. Hellendoorn K, Verlaan PWG, Pleij CWA (1997) A Functional role for the conserved protonatable hairpins in the 5′ untranslated region of turnip yellow mosaic virus RNA. J Virol 71: 8 774: 8 779
14. Jacob ANK, Murthy MRN, Savithri HS (1992) Nucleotide sequence of the 3′ terminal region of belladonna mottle virus-Iowa (renamed Physalis mottle virus) RNA and an analysis of the relationships of tymoviral coat proteins. Arch Virol 123: 367–377
15. Jacob ANK, Savithri HS (1991) In vitro expression of belladonna mottle virus genome. Ind J Biochem Biophys 28: 456–460
16. Kadare G, Drugeon G, Savithri HS, Haenni AL (1992) Comparison of the strategies of expression of five tymovirus RNAs by *in vitro* translation studies. J Gen Virol 73: 493–498
17. Kadare G, Rozanov M, Haenni AL (1995) Expression of the turnip yellow mosaic virus proteinase in Escherichia coli and determination of the cleavage site within the 206 kDa protein. J Gen Virol 76: 2 853–2 857
18. Keese MEO, Keese P, Gibbs A (1989) Nucleotide sequence of the genome of eggplant mosaic tymovirus. Virology 172: 547–554
19. Moline HE, Fries RE (1974) A strain of belladonna mottle virus isolated from *Physalis heterophylla* in Iowa. Phytopathology 64: 44–48
20. Morch MD, Boyer JC, Haenni AL (1988) Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. Nucleic Acids Res 16: 6 157–6 173
21. Morch MD, Drugeon G, Szafranski P, Haenni AL (1989) Proteolytic origin of the 150-kilodalton protein encoded by turnip yellow mosaic virus genomic RNA. J Virol 63: 5 153–5 158
22. Ranjith-Kumar CT, Haenni AL, Savithri HS (1998) Interference with physalis mottle tymovirus replication and coat protein synthesis by transcripts corresponding to the 3′ terminal region of the genomic RNA-role of the pseudoknot structure. J Gen Virol 79: 185–189
23. Rozanov MN, Koonin EV, Gorbalenya AE (1992) Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. J Gen Virol 73: 2 129–2 134
24. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, New York

25. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5 463–5 467
26. Singh RN, Dreher TW (1997) Turnip yellow mosaic virus RNA-dependent RNA polymerase: initiation of minus strand synthesis *in vitro*. Virology 233: 430–439
27. Srifah P, Keese P, Weiller G, Gibbs A (1992) Comparisons of the genomic sequences of erysimum latent virus and other tymoviruses: a search for the molecular basis of their host specificities. J Gen Virol 73: 1 437–1 447
28. Suryanarayana S, Rao NA, Murthy MRN, Savithri HS (1989) Primary structure of belladonna mottle virus coat protein. J Biol Chem 264: 6 273–6 279
29. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4 673–4 680
30. Wang W, Rossman TG (1994) Large-scale supercoiled plasmid preparation by acid phenol extraction. Biotechniques 16: 460–463

Authors' address: Dr. H. S. Savithri Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India.