COMPLETE NUCLEOTIDE SEQUENCE OF RADISH MOSAIC VIRUS RNA POLYMERASE GENE AND PHYLOGENETIC RELATIONSHIPS IN THE GENUS *COMOVIRUS*

K. PETRZIK[1*], M. HOLÁ[1,2], J. ŠPAK[1]

[1]Department of Plant Virology, Institute of Plant Molecular Biology, Academy of Sciences of the Czech Republic, Branišovská 31, 370 05 České Budějovice, Czech Republic; [2]University of South Bohemia, Faculty of Biology, České Budějovice, Czech Republic

Received May 23, 2005; accepted August 25, 2005

Summary. – The 3'-terminal part of RNA1 genome segment of Radish mosaic virus (RaMV) including complete RNA polymerase gene was sequenced. The 207 amino acids long polymerase is matured from a polyprotein precursor by cleavage at putative Q/H site by viral protease. The alignment of available amino acid sequences of RNA polymerase genes of comoviruses revealed a closest (55%) identity of RaMV to Red clover mottle virus (RCMV).

Key words: comovirus; sequence; RNA polymerase; phylogeny

Introduction

RaMV is one of the fifteen members of the genus *Comovirus* (the family *Comoviridae*). This family includes nonenveloped, 30 nm in diameter, beetle- and mechanically-transmitted plant viruses with single-stranded RNA (ssRNA) genome of positive polarity in two separately encapsidated segments. RaMV has been originally described in California by Tompkins (1939). Much later it has been found in Japan (1968) and Europe (1972). These and more recent findings from Morocco (Koenig and Fischer, 1981) and Iran (Farzadfar *et al.,* 2004) suggest that the virus is probably distributed worldwide (Brunt *et al.,* 1996).

Typical hosts of comoviruses are *Leguminosae*, with the exception of Andean potato mottle virus (APMoV) infecting *Solanaceae* and RaMV, which is the only comovirus infecting Brassicaceae. Particle structure, composition of the genome, properties of viral proteins and those of the type virus of the genus – Cowpea mosaic virus (CPMV) – have been characterized in detail. The CPMV genome consists of two segments, RNA1 and RNA2 containing 5889 and 3481 nucleotides, respectively. Both contain a VPg protein linked to their 5'-ends and a polyadenylated tail at their 3'-ends. Viral proteins are formed through polyprotein precursors that are cleaved by a virus-coded protease. RNA1 encodes (from 5' to 3') a protease cofactor, a helicase, a VPg, a protease and putative RNA-dependent RNA polymerase. RNA2 encodes (from 5' to 3') a movement protein and large and small capsid proteins (Goldbach and Wellink, 1996).

Complete nucleotide sequences of five comoviruses – Bean pod mottle virus (BPMV), Cowpea mosaic virus (CPMV), Cowpea severe mosaic virus (CPSMV), Red clover mottle virus (RCMV) and Squash mosaic virus (SqMV) – and partial sequence of APMoV have been published so far. In this paper we firstly describe the sequence of the RNA polymerase gene of RaMV and discuss its phylogenetic relationships within the genus *Comovirus*.

*E-mail: petrzik@umbr.cas.cz; fax: +420385-310363.

Abbreviations: APMoV = Andean potato mottle virus, BPMV = Bean pod mottle virus, CPMV = Cowpea mosaic virus, CPSMV = Cowpea severe mosaic virus, RaMV = Radish mosaic virus, RCMV = Red clover mottle virus, SqMV = Squash mosaic virus.

Materials and Methods

Virus. An RaMV1 isolate (Špak, 1992; Špak and Kubelková, 2000), originating from infected winter turnip rape, was propagated by mechanical inoculation of white mustard plants.

RNA isolation and RT-PCR. The virus was precipitated with PEG 6000-NaCl and concentrated and purified by two cycles of differential centrifugation (Klootwijk *et al.,* 1977). RNA was isolated from the purificate with the RNeasy Plant Mini kit (Qiagen). An one-step RT-PCR was performed with the Access RT-PCR kit (Promega). Equimolar mixture of the primers ERIC1 (5'-ATGTA AGCTCCTGGGGATTCAC-3') and ERIC2 (5'-AAGTAAGT GACTGGGGTGAGCG-3') (Versalovic *et al.*, 1994) and an oligo(dT)$_{18}$ primer were used in low stringency annealing conditions (48°C/30 secs). Other reactions were run with the primers 206N9 (5'-TTTAARCCRAARCCNTGT-3') and 206P0 (5'-ACYTGD GTDGACCANGC-3') in identical annealing conditions as above. Combinations of specific primers 209E5 (5'-GTGGTGGTAGT GAAAGTTCTAAACG-3', forward) and 209E4 (5'-TGATGTTG CATGGCAATATG-3', reverse), 209E5 (forward) and 209Z8 (5'-GCACACAAGAACAATAAAAC-3', reverse), and 210A0 (5'-TGGGATCTTTTYTGYTGGGAT-3', forward) and 209Z9 (5'-TGCCTTTGCGCTTTAAGC-3', reverse) were used for amplification of segments covering the complete sequence of the RNA polymerase gene.

Sequencing. The PCR products were cloned in pCR(R)4-TOPO(R) vector (Invitrogen) and sequenced using BigDye™ Terminator Cycle sequencing kit (Applied Biosystems, UK).

Multiple alignments were done by the www service CLUSTALW using http://www2.ebi.ac.uk/clustalw/ and amino acid (aa) sequences translated *in silico* from the nucleotide data on APMoV (Acc. No. M84806), BPMV (NC_003496), CPMV (NC_003549), CPSMV (NC_003545), RCMV (NC_003741), and SqMV (NC_003799).

Phylogenetic analysis was performed using the PROTPARS and PROTDIST programs from the PHYLIP package (Felsenstein, 1993).

Results and Discussion

The RaMV RNA polymerase gene was cloned and sequenced and the obtained sequence, deposited in the GenBank database with the Acc. No. AY96534 and reported for the first time for RaMV, was compared with those of other comoviruses at both nucleotide and amino acid level.

The RaMV RNA polymerase gene is terminated with an UAG followed with a 163 nt long 3'-nontranslated region. Amongst comoviruses, the gene is posttranslationally cleaved from a polyprotein precursor behind one of several glutamines (Q) (Wellink *et al.,* 1986). The exact cleavage site is at present unknown, as there are three Q residues between the protease and polymerase gene and the cleavage site is highly variable among comoviruses: Q/G in RCMV and CPMV, Q/S in BPMV, Q/A in CPSMV (Di *et al.*, 1999) and Q/C in SqMV (Han *et al.,* 2002). In RaMV, a putative cleavage Q/H site corresponds best to the alignment. If it is the correct site, the Q/H should be a new motif among comoviruses and unique for RaMV.

Table 1. Amino acid sequence identity of RNA polymerase genes of comoviruses

	APMoV	CPMV	CPSMV	RCMV	SqMV	BPMV	RaMV
APMoV		48.3	46.3	48.8	48.2	45.9	49.3
CPMV			53.3	61.4	56.2	56.0	53.8
CPSMV				50.3	53.3	53.8	54.7
RCMV					54.2	56.0	55.0
SqMV						50.6	54.4
BPMV							53.6
RaMV							

The polymerase gene is 707 aa long and encodes an about 81 K protein. Only the APMoV polymerase gene is smaller (703 aa). Nevertheless, all polymerase motifs (Ia – VIII), proposed by Koonin *et al.* (1991) for RNA polymerases, are located in the central "core" part of this gene. Outside of the core part, only extremely few conserved motifs are present on this comovirus gene: with the exception of the TSEGFPH motif upstream of the motif Ia, there does not occur any conserved stretch longer than 3 aa (Fig.1). This could be the reason why our attempts to amplify the 3'-end of this gene with degenerate primers, derived from the conserved domains of comoviruses, failed (data not shown). Therefore we had to to use for this purpose unrelated primers in low stringency conditions.

The phylogenetic analysis based on the RNA polymerase gene resulted in a single tree (Fig. 2). This tree grouped BPMV, RCMV and CPMV in one cluster, SqMV and CPSMV in another, and left APMoV and RaMV standing separately. This phylogenetic tree could correlate with diferent hosts of RaMV and APMoV from those of the rest

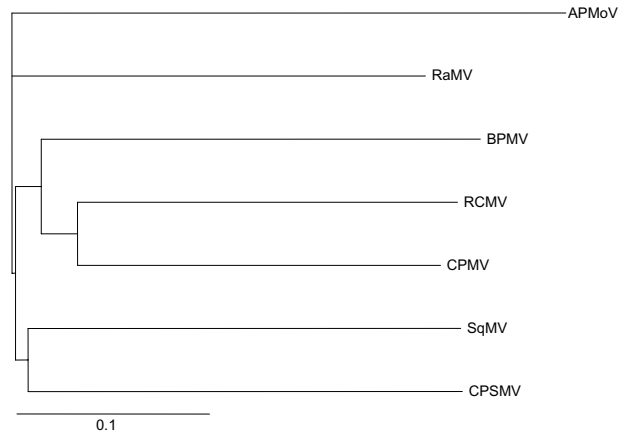

Fig. 2

Phylogenetic tree of comoviruses based on RNA polymerase

The bar represents a genetic distance of 0.1. For the abbreviations of virus names see the front page.

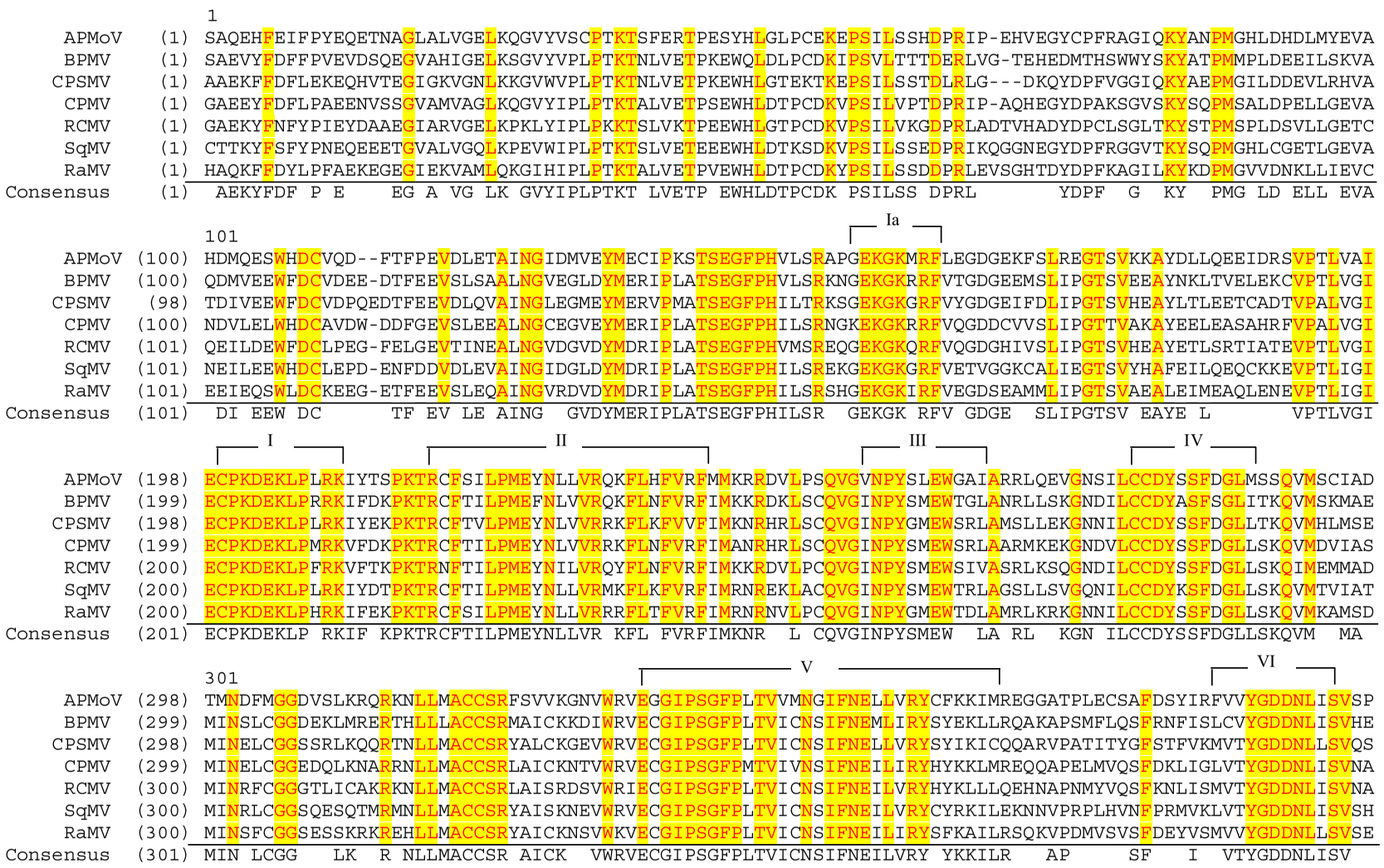

Fig. 1

Alignment of amino acid sequences of RNA polymerase genes of comoviruses

Conserved motifs Ia-VIII on positive RNA strands of RNA polymerase genes are indicated. Identical amino acids are in bold. For the abbreviations of virus names see the front page.

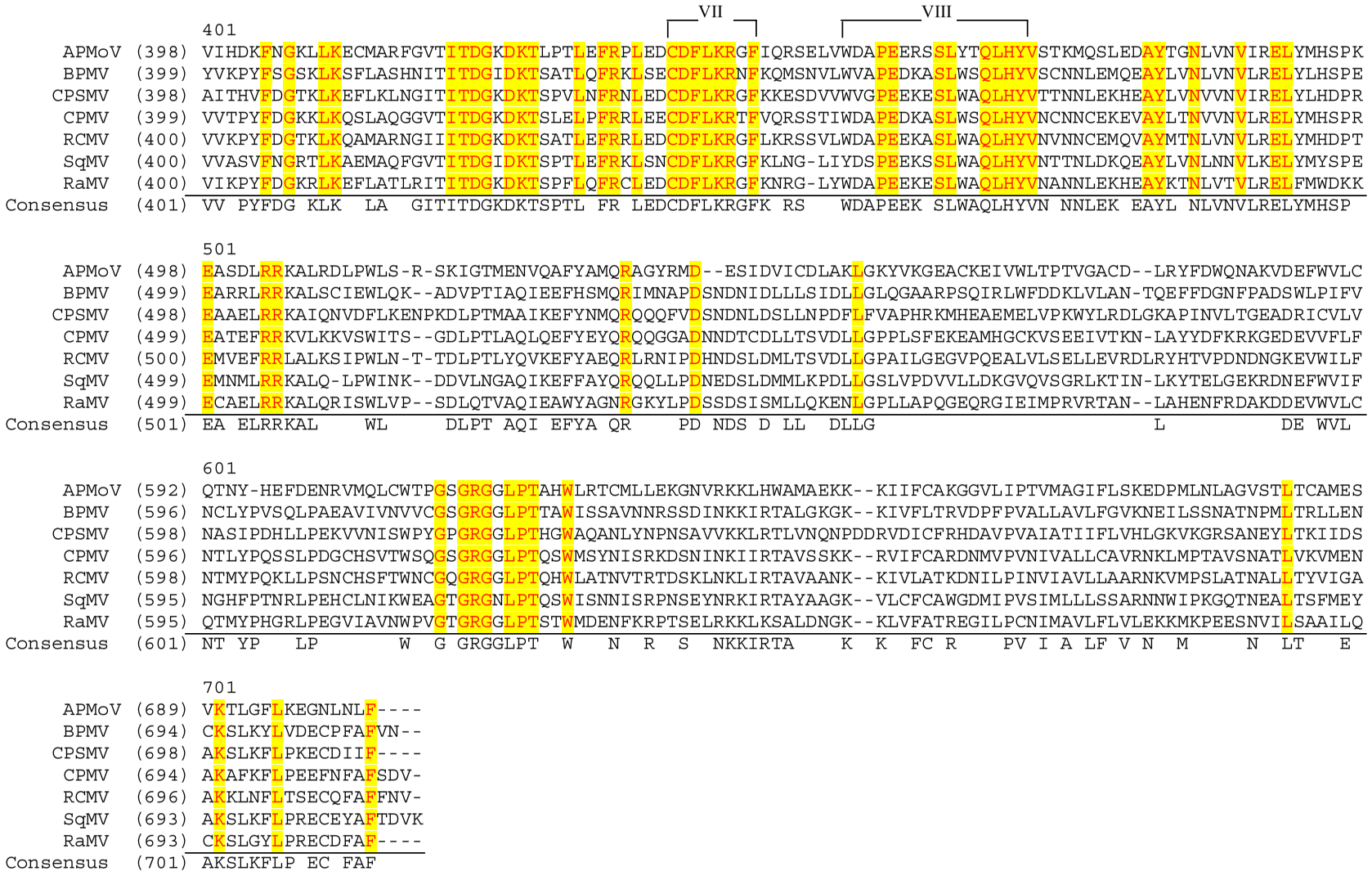
VII
VIII
401
APMoV (398) VIHDKFNGKLLKECMARFGVTITDGKDKTLPTLEFRPLEDCDFLKRGFIQRSELVWDAPEERSSLYTQLHYVSTKMQSLEDAYTGNLVNVIRELYMHSPK
BPMV (399) YVKPYFSGSKLKSFLASHNITITDGIDKTSATLQFRKLSECDFLKRNFKQMSNVLWVAPEDKASLWSQLHYVSCNNLEMQEAYLVNLVNVLRELYLHSPE
CPSMV (398) AITHVFDGTKLKEFLKLNGITITDGKDKTSPVLNFRNLEDCDFLKRGFKKESDVVWVGPEEKESLWAQLHYVTTNNLEKHEAYLVNVVNVIRELYLHDPR
CPMV (399) VVTPYFDGKKLKQSLAQGGVTITDGKDKTSLELPFRRLEECDFLKRTFVQRSSTIWDAPEDKASLWSQLHYVNCNNCEKEVAYLTNVVNVLRELYMHSPR
RCMV (400) VVKPYFDGTKLKQAMARNGIIITDGKDKTSATLEFRRLEDCDFLKRGFLKRSSVLWDAPEEKASLWAQLHYVNVNNCEMQVAYMTNLVNVLRELYMHDPT
SqMV (400) VVASVFNGRTLKAEMAQFGVTITDGIDKTSPTLEFRKLSNCDFLKRGFKLNG-LIYDSPEEKSSLWAQLHYVNTTNLDKQEAYLVNLNNVLKELYMYSPE
RaMV (400) VIKPYFDGKRLKEFLATLRITITDGKDKTSPFLQFRCLEDCDFLKRGFKNRG-LYWDAPEEKESLWAQLHYVNANNLEKHEAYKTNLVTVLRELFMWDKK
Consensus (401) VV PYFDG KLK LA GITITDGKDKTSPTL FR LEDCDFLKRGFK RS WDAPEEK SLWAQLHYVN NNLEK EAYL NLVNVLRELYMHSP
501
APMoV (498) EASDLRRKALRDLPWLS-R-SKIGTMENVQAFYAMQRAGYRMD--ESIDVICDLAKLGKYVKGEACKEIVWLTPTVGACD--LRYFDWQNAKVDEFWVLC
BPMV (499) EARRLRRKALSCIEWLQK--ADVPTIAQIEEFHSMQRIMNAPDSNDNIDLLLSIDLLGLQGAARPSQIRLWFDDKLVLAN-TQEFFDGNFPADSWLPIFV
CPSMV (498) EAAELRRKAIQNVDFLKENPKDLPTMAAIKEFYNMQRQQQFVDSNDNLDSLLNPDFLFVAPHRKMHEAEMELVPKWYLRDLGKAPINVLTGEADRICVLV
CPMV (499) EATEFRRKVLKKVSWITS--GDLPTLAQLQEFYEYQRQQGGADNNDTCDLLTSVDLLGPPLSFEKEAMHGCKVSEEIVTKN-LAYYDFKRKGEDEVVFLF
RCMV (500) EMVEFRRLALKSIPWLN-T-TDLPTLYQVKEFYAEQRLRNIPDHNDSLDMLTSVDLLGPAILGEGVPQEALVLSELLEVRDLRYHTVPDNDNGKEVWILF
SqMV (499) EMNMLRRKALQ-LPWINK--DDVLNGAQIKEFFAYQRQQLLPDNEDSLDMMLKPDLLGSLVPDVVLLDKGVQVSGRLKTIN-LKYTELGEKRDNEFWVIF
RaMV (499) ECAELRRKALQRISWLVP--SDLQTVAQIEAWYAGNRGKYLPDSSDSISMLLQKENLGPLLAPQGEQRGIEIMPRVRTAN--LAHENFRDAKDDEVWVLC
Consensus (501) EA ELRRKAL WL DLPT AQI EFYA QR PD NDS D LL DLLG L DE WVL
601
APMoV (592) QTNY-HEFDENRVMQLCWTPGSGRGGLPTAHWLRTCMLLEKGNVRKKLHWAMAEKK--KIIFCAKGGVLIPTVMAGIFLSKEDPMLNLAGVSTLTCAMES
BPMV (596) NCLYPVSQLPAEAVIVNVVCGSGRGGLPTTAWISSAVNNRSSDINKKIRTALGKGK--KIVFLTRVDPFPVALLAVLFGVKNEILSSNATNPMLTRLLEN
CPSMV (598) NASIPDHLLPEKVVNISWPYGPGRGGLPTHGWAQANLYNPNSAVVKKLRTLVNQNPDDRVDICFRHDAVPVAIATIIFLVHLGKVKGRSANEYLTKIIDS
CPMV (596) NTLYPQSSLPDGCHSVTWSQGSGRGGLPTQSWMSYNISRKDSNINKIIRTAVSSKK--RVIFCARDNMVPVNIVALLCAVRNKLMPTAVSNATLVKVMEN
RCMV (598) NTMYPQKLLPSNCHSFTWNCGQGRGGLPTQHWLATNVTRTDSKLNKLIRTAVAANK--KIVLATKDNILPINVIAVLLAARNKVMPSLATNALLTYVIGA
SqMV (595) NGHFPTNRLPEHCLNIKWEAGTGRGNLPTQSWISNNISRPNSEYNRKIRTAYAAGK--VLCFCAWGDMIPVSIMLLLSSARNNWIPKGQTNEALTSFMEY
RaMV (595) QTMYPHGRLPEGVIAVNWPVGTGRGGLPTSTWMDENFKRPTSELRKKLKSALDNGK--KLVFATREGILPCNIMAVLFLVLEKKMKPEESNVILSAAILQ
Consensus (601) NT YP LP W G GRGGLPT W N R S NKKIRTA K K FC R PV I A LF V N M N LT E
701
APMoV (689) VKTLGFLKEGNLNLF----
BPMV (694) CKSLKYLVDECPFAFVN--
CPSMV (698) AKSLKFLPKECDIIF----
CPMV (694) AKAFKFLPEEFNFAFSDV-
RCMV (696) AKKLNFLTSECQFAFFNV-
SqMV (693) AKSLKFLPRECEYAFTDVK
RaMV (693) CKSLGYLPRECDFAF----
Consensus (701) AKSLKFLP EC FAF

Fig. 1

of comoviruses. On the other hand, the amino acid alignment of RNA polymerases of comoviruses revealed about a 54–55% identity of RaMV with the viruses of both clusters, but a lower one (about 49%) with APMoV (Table 1). Also, our phylogenetic tree did not correlate with known serological relationships: RaMV is serologically related to BPMV, SqMV (Campbell, 1964), RCMV and CPMV (Bruening, 1978). This discrepancy may indicate different evolution history of structural genes and RNA polymerase gene of comoviruses. Only a complete nucleotide sequence and its analysis could solve this discrepancy and reveal a putative recombination event in the RaMV evolution.

Acknowledgements. This study was supported by the grant No. AV0Z50510513 from the Academy of Sciences of the Czech Republic.

References

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L, Zurcher EJ (Eds) (1996 onwards): Radish mosaic comovirus. *Plant Viruses Online: Descriptions and lists from the VIDE database. Version: 20th August 1996. URL http://biology.anu.edu.au/Groups/MES/vide/.*

Bruening G (1978): Comovirus group No **199**. *CMI/AAB Descriptions of Plant Viruses,* England.

Campbell RN (1964): Radish mosaic virus, a crucifer virus serologically related to strains of bean pod mottle virus and squash mosaic virus. *Phytopathology* **54**, 1418–142.

Di R, Hu CH-CH, Ghabrial SA (1999): Complete nucleotide sequence of bean pod mottle virus RNA1: Sequence comparison and evolutionary relationship to other comoviruses. *Virus Genes* **18**, 129–137.

Farzadfar S, Pourrahim R, Golnaraghi AR, Jalali S, Ahoonmanesh A (2004): Occurrence of *Radish mosaic virus* on cauliflower and turnip crops in Iran. *Plant Dis.* **88**, 909.

Felsenstein J (1993): *PHYLIP (Phylogeny Interference Package) version 3.5c.* Department of Genetics, University of Washigton, Seattle.

Goldbach R, Wellink J (1996): Comoviruses: molecular biology and replication. *The Plant Viruses* **5**.

Han SS, Yoshida K, Karasev AV, Iwanami T (2002): Nucleotide sequence of a Japanese isolate of squash mosaic virus. *Arch. Virol.* **147**, 437–443.

Klootwijk J, Klein I, Zabel P, van Kammen A (1977): Cowpea mosaic virus RNAs have neither m7GpppN nor mono-, di- or triphosphates at their 5' ends. *Cell* **11**, 73–82.

Koenig R, Fischer HU (1981): A Moroccan radish mosaic-virus isolate from turnip. *Plant Dis.* **68**, 758–760.

Koonin EV, Choi GH, Nuss DL, Shapira R, Carrington JC (1991): Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a roup of positive-strand RNA plant viruses. *Proc. Natl. Acad. Sci. USA* **88**, 10647–10651.

Špak J (1992): Characterization, purification and serology of the Czechoslovak isolate of radish mosaic virus. *Acta Virol.* **36**, 191–197.

Špak J, Kubelková D (2000): Serological variability among European isolates of radish mosaic virus. *Plant Pathol.* **49**, 295–301.

Tompkins CM (1939): Two mosaic diseases of annual stocks *J. Agric. Res.* **58,** 119–130.

Versalovic J, Schneider M, Lupski JR (1994): Genomic fingerprinting of bacteria using repetitive sequence based PCR (rep-PCR). *Methods Moll. Cell. Biol.* **5**, 25–40.

Wellink J, Rezelman G, Goldbach R, Beyreuther K (1986): Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component RNA of cowpea mosaic virus. *J. Virol.* **59**, 50–58.