



Genome sequence and phylogenetic analysis of a novel comovirus from tabasco pepper (*Capsicum frutescens*)

Ricardo Iván Alcalá-Briseño¹ · Pongtharin Lotrakul² · Rodrigo A. Valverde³

Received: 12 June 2019 / Accepted: 28 September 2019 / Published online: 11 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

A virus isolate from tabasco pepper (*Capsicum frutescens*) has been reported as a strain of the comovirus Andean potato mottle virus (APMoV). Using the replicative intermediate viral dsRNA, the pepper virus strain was sequenced by Illumina MiSeq. The viral genome was de novo assembled resulting in two RNAs with lengths of 6028 and 3646 nt. Nucleotide sequence analysis indicated that they corresponded to the RNA-1 and RNA-2 of a novel comovirus which we tentatively named pepper mild mosaic virus (PepMMV). Predictions of the open reading frame (ORF) of RNA-1 resulted in a single ORF of 5871 nt with five cistrons typical of comoviruses, cofactor proteinase, helicase, viral protein genome-linked, 3C-like proteinase (Pro), and RNA-dependent RNA polymerase (RdRP). Similarly, sequence analysis of RNA-2 resulted in a single ORF of 3009 nt with two cistrons typical of comoviruses: movement protein and coat protein (large coat protein and small coat proteins). In pairwise amino acid sequence alignments using the Pro-Pol protein, PepMMV shared the closest identities with broad bean true mosaic virus and cowpea mosaic virus, 56% and 53.9% respectively. In contrast, in alignments of the amino acid sequence of the coat protein (small and large coat proteins) PepMMV shared the closest identities to APMoV and red clover mottle virus, 54% and 40.9% respectively. A phylogenetic tree constructed using the conserved domains for the Pro-Pol from all members of the family *Secoviridae* confirmed the comovirus nature of the virus. Phylogenetic and sequence analyses supports proposing PepMMV as a new species of the genus *Comovirus*.

Keywords *Capsicum annuum* · dsRNA · Next generation sequencing · Replicative intermediate dsRNA · *Secoviridae*

Pepper (*Capsicum* spp.) is an economically important food crop grown worldwide, whose production is severely affected by viral diseases [1]. Plant viruses in the genus *Comovirus* (*Secoviridae*) include viruses with a relatively narrow host range which are mechanically transmitted onto

herbaceous hosts and are vectored by beetles [2]. They cause economically important diseases, particularly in legumes, but some also infect cucurbits and solanaceous crops [3, 4]. In 1995, a putative comovirus was reported naturally infecting tabasco pepper (*Capsicum frutescens*) in Central America [5]. The virus was detected in approximately 25% of tabasco plants with virus symptoms and it was serologically related to Andean potato mottle virus (APMoV) but differed from APMoV in host range, geographical distribution, and vector transmission [5]. Electrophoretic analysis of dsRNA extracted from infected pepper plants resulted in a banding profile similar to two other members of the family *Secoviridae* [5]. We used the dsRNA of the putative comovirus in reverse transcription (RT)-PCR amplifications, with comovirus-degenerate primers for the RNA-dependent RNA polymerase (RdRP) region [6]. Sequence analyses of products obtained by RT-PCR, together with reported unique biological properties, suggested that the virus could be a novel species of the genus *Comovirus*. With the objective of obtaining the complete sequence of the novel virus

Edited by Seung-Kook Choi.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11262-019-01707-6>) contains supplementary material, which is available to authorized users.

✉ Rodrigo A. Valverde
rvalverde@agcenter.lsu.edu

¹ Dept. of Plant Pathology, University of Florida, Gainesville, FL 32611, USA

² Dept. of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

³ Dept. of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

and comparing it to related viruses, we used the replicative intermediate (RI) dsRNA to sequence the viral RNA using Illumina MiSeq. In this investigation, we report the genome sequence, sequence analyses, and phylogenetic relationships of the novel virus which we named pepper mild mosaic virus (PepMMV) and proposed it as a putative species of the genus *Comovirus*.

DsRNA aliquots of the original isolate of PepMMV from El Negrito, Honduras [5] stored at $-70\text{ }^{\circ}\text{C}$ were treated with DNase 1 and S1 nuclease (Promega, Madison, WI, USA) and resolved in 1.2% agarose gels. Two dsRNAs of approximately 6.0 and 4.0 kbp, assumed to be the RI of genomic RNA-1 and RNA-2 of the virus, were gel purified (Qiagen, Valencia, CA, USA) and used for sequencing. DsRNAs were

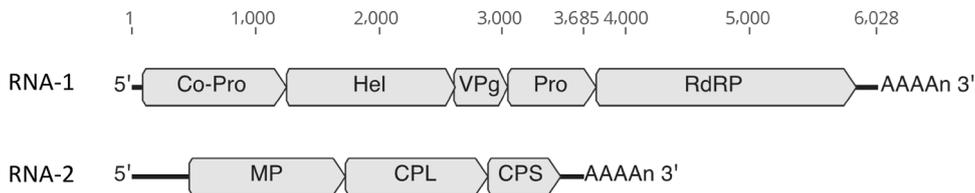


Fig. 1 Genome organization of pepper mild mosaic virus. The boxes represent the cistrons for the cofactor proteinase (Co-Pro), helicase (Hel), viral protein linked to the genome (VPg), the 3C-like protein-

ase (Pro), the RNA-dependent RNA polymerase (RdRP), movement protein (MP), large coat protein (CPL), and small coat protein (CPS)

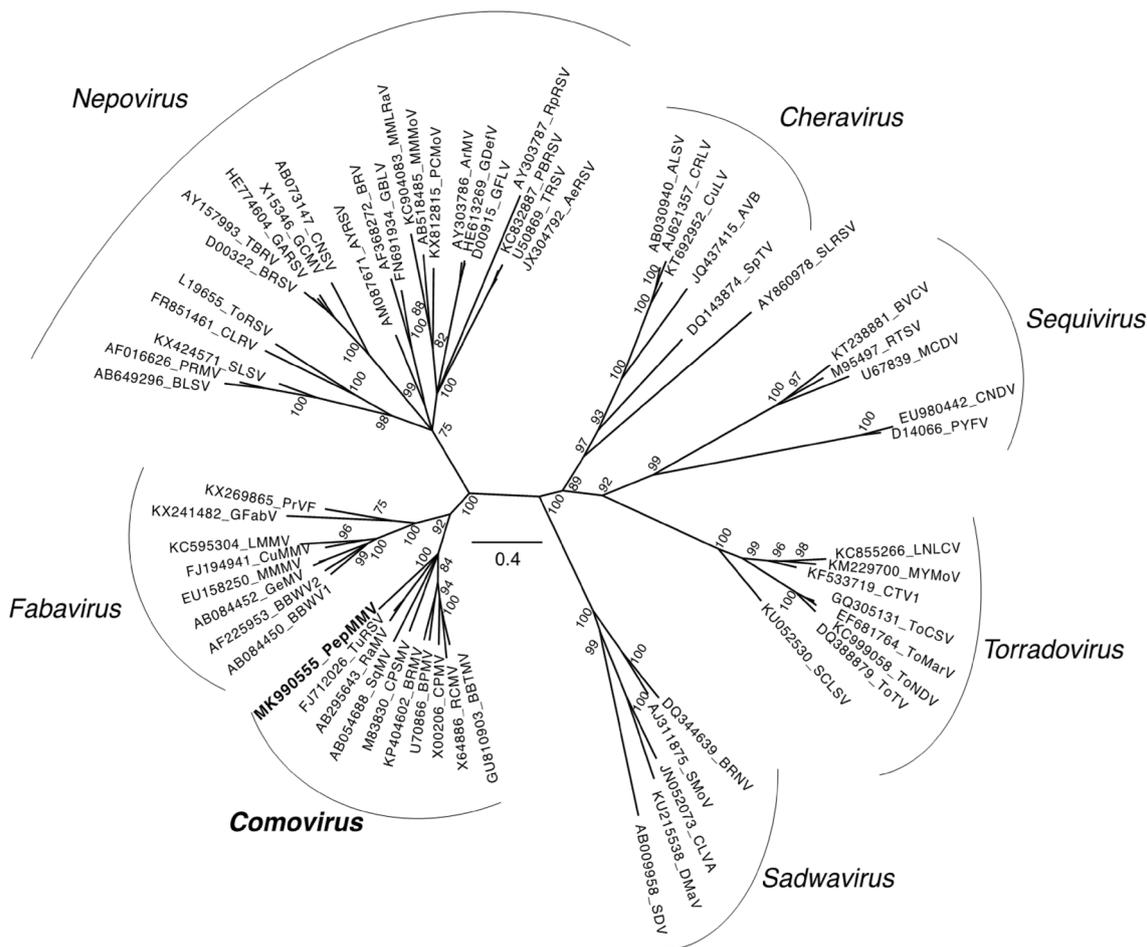


Fig. 2 Phylogenetic tree of the proteinase and RdRP region (Pro–Pol) of members of the *Secoviridae* family. Sixty-two sequences retrieved from NCBI plus PepMMV were aligned with t-coffee [12]. The maximum likelihood tree was performed with IQ-TREE [13], and the best

fitted model according to BIC was LG+F+I+G. A bootstrap of 1000 replications with branches $>70\%$ were collapsed. The bar represents the genetic distance

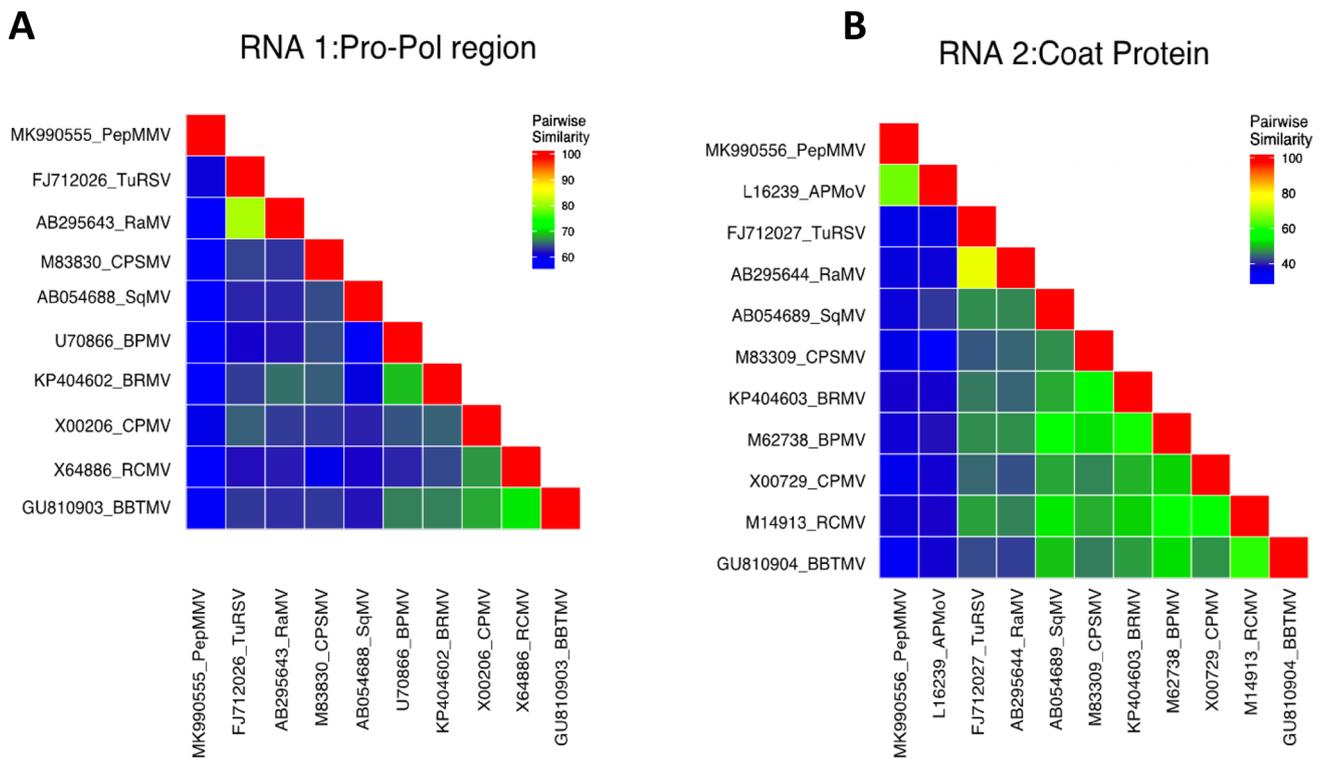


Fig. 3 Pairwise identity matrix of the proteinase and RdRP region (Pro-Pol) (a) and the coat protein (coat protein large and coat protein small) (b) of viruses in the *Comovirus* genus included in Fig. 1. The sequence identity matrices were calculated with the Species Demarcation Tool (SDT) [14] with MAFFT [15]. The plots were generated with ggplot package v.3.5.1 R Core Team [16]. *PepMMV*, pep-

per mild mosaic virus; *TuRSV*, turnip ringspot virus; *RaMV*, radish mosaic virus; *CPSMV*, cowpea severe mosaic virus; *SqMV*, squash mosaic virus; *BPMV*, bean pod mottle virus; *BRMV*, bean rugose mosaic virus; *CPMV*, cowpea mosaic virus; *RCMV*, red clover mottle virus; *BBTMV*, broad bean true mosaic virus; *APMoV*, Andean potato mottle virus

denatured at 95 °C for 5 min and used to prepare RNAseq libraries with the Illumina TruSeq™ RNA Library Prep Kit v2 (Illumina Inc., San Diego, CA, USA). Sequencing was conducted at the Roy J. Carver Biotechnology Center, University of Illinois (Urbana-Champaign, IL, USA) with an Illumina MiSeq next generation sequencer (paired-end 2×250) generating 127,647 reads. The reads were filtered by quality and trimmed, and then de novo assembled with Spades v10.2 [7] using a custom pipeline. The results yielded 392 contigs. Generated contigs were subjected to BLASTx searches using the non-redundant database from NCBI. The sequence analysis indicated the presence of six contigs with homology to comoviruses. Two contigs with lengths of 6028 nt and 3646 nt represented the full length of RNA-1 and RNA-2 respectively of the putative comovirus. Other contigs with sequences corresponding to the families *Endornaviridae* and *Caulimoviridae* were also obtained (Supplementary Fig. 1 and Supplementary Table 2). Read coverage plots for RNA1 and RNA2 are shown in Supplementary Fig. 2A and 2B. The genome sequences of the putative comovirus (*PepMMV*) were submitted to GenBank (Acc. No. MK990555, RNA-1, and MK990556, RNA-2). To

confirm that the assembled sequences corresponded to *PepMMV*, virus-specific primers which amplified the RdRP and the partial CPL and CPS regions were designed and RT-PCR reactions were performed using the dsRNA extracted from the original virus isolate. RT-PCR products covering the RdRP region of RNA1 (1033 nt) and a fragment of the CPL and CPS regions of RNA2 (774 nt) were Sanger sequenced. Alignment of RT-PCR and obtained NGS sequences resulted in 99.4 and 99.2% aa identity with the corresponding regions of the RdRP and CP respectively.

Predictions of the open reading frame (ORF) of RNA-1 resulted in a single ORF of 5871 nt with five cistrons typical of comoviruses, cofactor proteinase (Co-Pro, 1158 nt), helicase (Hel, 1803 nt), viral protein genome-linked (VPg, 84 nt), 3C-like proteinase (Pro, 621 nt), and RdRP (Pol, 2212 nt). The UTRs for the 5' and 3' consisted of 90 and 157 nt, respectively (Fig. 1). Similarly, sequence analysis of RNA-2 resulted in a single ORF of 3009 nt with two cistrons typical of comoviruses: movement protein (MP, 1266 nt), and coat protein (CPL, 1152 nt and CPS, 588 nt). The UTRs for the 5' and 3' consisted of 460 and 177 nt, respectively (Fig. 1). The RNA-1-encoded 3C-like proteinase cleaves both RNA-1 and

RNA-2-encoded polyproteins [2]. The proteinase putative cleavage sites, inferred from sequence alignments, are displayed in Supplementary Table S1.

A maximum likelihood tree was constructed using the amino acid sequence of the conserved domains of the 3C-like proteinase and RdRP (Pro-Pol) region of the RNA-1 of all members of the family *Secoviridae* (Fig. 2). PepMMV clustered with other members of the genus *Comovirus* form a single clade. In pairwise amino acid sequence alignments using the Pro-Pol protein, PepMMV shared the closest identities with broad bean true mosaic virus (GU810903) and cowpea mosaic virus (X00206), 56% and 53.9% respectively. In contrast, in alignments of the amino acid sequence of CPL and CPS, PepMMV shared the closest identities (54%) to APMoV (L16239) and (40.9%) to red clover mottle virus (M14913). These analyses showed an amino acid percentage identity below 73% in the combined CP region and 80% in the Pro-pol region, which are assumed to be the species threshold for comoviruses (Fig. 3) [2].

Surveys for pepper viruses have been conducted using the enzyme-linked immunosorbent assay alone or in combination with PCR as detection tools and have included testing viruses in the genera *Begomovirus*, *Potyvirus*, *Tobamovirus*, *Cucumovirus*, *Polerovirus*, *Crinivirus*, and *Tospovirus*, among others [1, 8–10]. To the best of our knowledge, these surveys have never included testing for virus members of the genus *Comovirus*, and therefore, the true spread of PepMMV in pepper is not known. Nevertheless, the virus was detected in samples from farms at two locations in Honduras (La Flecha and El Negrito) and three locations in Nicaragua (Managua, Palo Verde, and Tipitapa) (5). PepMMV, which was considered a serotype of APMoV, is reported to cause mosaic and yellow mottle symptoms in *C. frutescens* and *C. chinense*, but only very mild mosaic symptoms in six genotypes of *C. annuum* (Durkee Cayenne, Early California Wonder, Yolo Y, Long Red Cayenne, Jalapeño M, and Hungarian Yellow Wax) which is the most widely cultivated *Capsicum* species [5, 11]. Furthermore, PepMMV alone did not cause significant differences in the fresh weight of the four *C. annuum* genotypes. However, in mixed infections with potyviruses, synergistic effects consisting of severe mosaic symptoms and reduced fresh weight were reported in the four *C. annuum* genotypes [11]. Therefore, PepMMV could play an important role in viral diseases of pepper.

The close serological relationship between PepMMV and APMoV reported by Valverde et al. [5] was confirmed by data from the coat protein (CPL and CPS) amino acid sequence alignments shown in Fig. 3b. Nevertheless, differences between APMoV and PepMMV in host range and vector transmission, combined with clear difference in amino acid sequence illustrated in the phylogenetic tree and pairwise identity analysis, support proposing PepMMV as a distinct bipartite comovirus species.

Determining the taxonomic position of PepMMV is important to enable the accurate disease diagnosis.

Acknowledgments We wish to thank Andrea Hebert, Middleton Library, Louisiana State University, for proof reading the manuscript.

Author contributions RIA-B conducted the phylogenetic and sequence analyses. PL conducted preliminary RT-PCR experiments and analyses of sequences of PCR products. RAV planned the experiments, conducted dsRNA purifications, and wrote the manuscript.

Funding Partial support for this investigation was provided by the National Institute of Food and Agriculture, USDA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Kenyon L, Kumar S, Tsai W-S, Hughes JDA (2014) Virus diseases of peppers (*Capsicum* spp.) and their control. *Adv Virus Res* 90:297–354. <https://doi.org/10.1016/B978-0-12-801246-8.00006-8>
2. Thompson JR, Dasgupta I, Fuchs M, Iwanami T, Karasev AV, Petrzik K, Sanfaçon H, Tzanetakis IE, van der Plugt R, Wetzel T, Yoshikawa N, ICTV Report Consortium (2017) ICTV Virus Taxonomy Profile: *Secoviridae*. *J Gen Virol* 98:529–531
3. Maina S, Edwards OR, Jones RAC (2017) Two complete genome sequences of *Squash mosaic virus* from 20-year-old cucurbit leaf samples from Australia. *Genome Announc* 5:e00778–e817. <https://doi.org/10.1128/genomeA.00778-17>
4. Valverde RA, Fulton JP (1996) Comoviruses: Identification and diseases caused. In: Harrison BD (ed) *The plant viruses: polyhedral virions and bipartite rna genomes*. Plenum Publishing, New York, pp 17–33
5. Valverde RA, Black LL, Dufresne DJ (1995) A comovirus affecting tabasco pepper in Central America. *Plant Dis* 79:421–423
6. Maliogka V, Dovas CI, Efthimiou K, Katis NI (2004) Detection and differentiation of Comoviridae using a semi-nested RT-PCR and phylogenetic analysis based on the polymerase protein. *J Phytopathol* 152:404–409
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comp Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
8. Afouda L, Kone D, Zinsou V, Dossou L, Kenyon L, Winter S, Knierim D (2017) Virus surveys of *Capsicum* spp. in the Republic of Benin reveal the prevalence of pepper vein yellows virus and the identification of a previously uncharacterised polerovirus species. *Arch Virol* 162:1599–1607
9. Arli-Sokmen A, Mennan H, Sevik MA, Ecevit O (2005) Occurrence of viruses in field-grown pepper crops and some of their reservoir weed hosts in Samsun, Turkey. *Phytoparasitica* 33:343–358

10. Bolou Bi BA, Moury B, Abo K, Sorho F, Cherif M, Girardot G, Kouassi NP, Kone D (2018) Survey of viruses infecting open-field pepper crops in Côte d'Ivoire and diversity of *Pepper vein mottle virus* and *Cucumber mosaic virus*. *Plant Pathol* 67:1416–1425
11. Dufresne DJ, Valverde RA, Hobbs HA (2000) Effects of coinfections of Andean potato mottle comovirus with two potyviruses in seven *Capsicum* genotypes. *Rev Mex Fitopatol* 17:17–22
12. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tvery ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acid Res*. <https://doi.org/10.1093/nar/gkz268>
13. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
14. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE* 9(9):e108277. <https://doi.org/10.1371/journal.pone.0108277>
15. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acid Res* 30:3059–3066
16. Hornik K (2018) The R FAQ. <https://CRAN.R-project.org/doc/FAQ/R-FAQ.html>. Accessed 1 June 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.