Short communication

Deep sequencing of pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses[☆]

Toufic Elbeaino [a,*], Michele Digiaro [a], Mangala Uppala [b], Harikishan Sudini [b]

[a] Istituto Agronomico Mediterraneo, Via Ceglie 9, 70010 Valenzano (BA), Italy
[b] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

ARTICLE INFO

ABSTRACT

The sequences of five viral RNA segments of pigeonpea sterility mosaic virus (PPSMV), the agent of sterility mosaic disease (SMD) of pigeonpea (*Cajanus cajan*, *Fabaceae*), were determined using the deep sequencing technology. Each of the five RNAs encodes a single protein on the negative-sense strand with an open reading frame (ORF) of 6885, 1947, 927, 1086, and 1422 nts, respectively. In order, from RNA1 to RNA5, these ORFs encode the RNA-dependent RNA polymerase (p1, 267.9 kDa), a putative glycoprotein precursor (p2, 74.3 kDa), a putative nucleocapsid protein (p3, 34.6 kDa), a putative movement protein (p4, 40.8 kDa), while p5 (55 kDa) has an unknown function. All RNA segments of PPSMV showed the highest identity with orthologs of fig mosaic virus (FMV) and Rose rosette virus (RRV). In phylogenetic trees constructed with the amino acid sequences of p1, p2 and p3, PPSMV clustered consistently with other emaraviruses, close to clades comprising members of other genera of the family *Bunyaviridae*. Based on the molecular characteristics unveiled in this study and the morphological and epidemiological features similar to other emaraviruses, PPSMV seems to be the seventh species to join the list of emaraviruses known to date and accordingly, its classification in the genus *Emaravirus* seems now legitimate.

© 2014 Elsevier B.V. All rights reserved.

Emaravirus is a recently established viral genus that is unassigned to any family of plant viruses and includes segmented, negative-single stranded RNA viruses with enveloped spherical particles (double membrane-bodies, DMBs), approximately of 80–200 nm in diameter (Muehlbach and Mielke-Ehret, 2011). *European mountain ash ringspot-associated virus* (EMARaV) is the type species of this genus and has a genome composed of four RNA segments (Mielke and Muehlbach, 2007). The other two approved species of the same genus, *Fig mosaic virus* (FMV) and *Rose rosette virus* (RRV), have a genome composed of six (Elbeaino et al., 2012; Ishikawa et al., 2012) and four (Laney et al., 2012) RNAs, respectively. Other tentative emaraviruses have a number of RNA segments differing among species, *i.e.* four in the case of redbud yellow ringspot virus (RYRV, Laney et al., 2011); five for raspberry leaf blotch virus (RLBV, McGavin et al., 2012); from four to six for maize red stripe virus (MRSV, Jensen et al., 1996), and from five to seven for pigeonpea sterility mosaic virus (PPSMV, Kumar et al., 2003). In general, each segment of the genome contains a single ORF that

encodes, in order, the polymerase, the putative glycoprotein, the putative nucleocapsid, the putative movement protein, and other proteins with unknown functions. Intracellular DMBs were identified as particles of the putative causal agents of the fig mosaic (Bradfute et al., 1970; Elbeaino et al., 2009a), the ringspot disease of European mountain ash (Mielke and Muehlbach, 2007) and the dwarfed roses affected by RRV (Laney et al., 2011). Similar DMB structures were also observed in the cells of mosaic-affected plants of pigeonpea (*Cajanus cajan* L.) infected with PPSMV (Kumar et al., 2003), high plain diseased plants of maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) infected with MRSV (Skare et al., 2006; Jensen et al., 1996), plants of *Cercis* spp. affected by RYRV (Laney et al., 2011), and *Rubus* plants affected by RLBV (McGavin et al., 2012). Most of these viruses are vectored in nature by eriophyid mites (Mielke-Ehret and Muehlbach, 2012).

To date, the genomes of most of definitive and putative emaraviruses have been completely sequenced and characterized. One of the few exceptions is represented by PPSMV, the putative causal agent of sterility mosaic disease (SMD), the most devastating disease of pigeonpea in the Indian subcontinent (Kumar et al., 2004; Jones et al., 2004). SMD is characterized by sterility (complete or partial cessation of flowering), severe or mild mosaic symptoms on leaves, reduction in leaf size, excessive vegetative growth, severe stunting, and yield reduction up to 90% (Kannaiyan et al., 1984;

Reddy et al., 1984). The SMD agent is a putative virus transmitted in a semi-persistent manner by the eriophyd mite *Aceria cajani* (Kulkarni et al., 2002), whose the viral nature was only recently hypothesized as belonging to the same group of other mite-borne viruses having DMBs (Kumar et al., 2002, 2003). For PPSMV the only sequence information to date available is limited to a small fragment of 636 nt of RNA1 (Elbeaino et al., 2013) and of another fragment of 764 nt, encoding the nucleocapsid, which is localized in the fifth RNA as magnitude (Kumar et al., 2003) but which was conventionally named RNA3 for similarity with other emaraviruses by Elbeaino et al. (2013). This paper reports the complete sequences of five PPSMV RNA segments obtained by applying the deep sequencing technique, and discusses the molecular and phylogenetic relationships of this virus with definite and tentative members of the genus *Emaravirus* in order to define its taxonomic allocation.

A symptomatic pigeonpea plant (germplasm line ICP 8863) infected with PPSMV isolate "P" (Kumar et al., 2005) was used as a virus source. The plant material was provided by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India.

Tissues of leaves (30 g), exhibiting chlorotic ringspots and severe mosaic symptoms, were used to recover dsRNAs by phenol/chloroform extraction and chromatography through cellulose CF-11 column (Whatman, USA) in the presence of 17% ethanol, according to a protocol described by Dodds (1993) and modified by Elbeaino et al. (2009a). Further purification was by chromatography through a second column of cellulose CC41 (Whatman, USA), followed by digestion with RQ1-DNase (60 µg/ml) (Promega, USA) and RNase A (0.5 µg/ml). Extracts were then passed through Micro Bio-Spin® 30 Columns in RNase-Free Tris (Bio-Rad Laboratories, USA) for the elimination of contaminants, analyzed in 6% PAGE and silver stained.

To determine the genome sequence of PPSMV, a next generation sequencing approach was followed using Illumina technology. A cDNA sequencing library was synthesized from purified viral dsRNAs using a TruSeq RNA Sample Prep Kits v2 (Illumina) and sequenced in a run of 50 bp-single read at University of Bari, Department of Soil Sciences, Plants and Food (http://www.selge.uniba.it) on a HiScan SQ apparatus (Illumina, San Diego, USA). The raw reads were filtered for quality, reduced to unique reads and then were *de novo* assembled into larger contigs using the Velvet Software 1.2.08 (Zerbino and Birney, 2008) with a kmer of 23. Assembled contigs were then screened for sequence homologies using BLASTX (cutoff e-value 10^{-6}), and BLASTP (http://www.ncbi.nlm.nih.gov/). Genomic RNA was reconstructed by manually aligning homologous contigs on the most closely related sequences found in the GenBank database and using the CLC software (CLCbio, Aarhus, Denmark). Once assembled, the five viral segments were re-sequenced through several RT-PCR amplifications, using a battery of sense and antisense specific primers designed on deep sequencing-generated reads, to ensure at least fourfold nucleotide coverage.

All amplified products were ligated in StrataClone™ PCR Cloning vector pSC-A (Stratagene, USA), cloned in *Escherichia coli* DH5α or SoloPACK cells, and custom sequenced (Primm, Milan, Italy). Nucleotide and protein sequences were analyzed with the assistance of the DNA Strider 1.1 program (Marck, 1988). Multiple alignments of nucleotide and amino acid sequences were obtained using the default options of CLUSTALX 1.8 (Thompson et al., 1997). Search for homologies with proteins from the Protein Information Resources database (PIR, release 47.0) was done with the FASTA (Pearson and Lipman, 1988), BlastX and BlastP (Altschul et al., 1997) programs. Protein function analysis was performed using the software for the prediction of glycosylation sites (NetNGlyc 1.0 vers.) (Emanuelsson et al., 2000), transmembrane helices (TMHMM 2.0 vers.), cleavage sites (SignalP 3.0, TragetP vers.) (Emanuelsson

et al., 2007) and the SLP-Local program for the prediction of the subcellular location of proteins (Matsuda et al., 2005). For phylogenetic analysis, amino acid sequences were aligned using "MUSCLE" (Edgar, 2004) and trees were generated by the "Maximum likelihood" method (Jones et al., 1992), applying the JTT matrix and pairwise gap deletion options implemented in MEGA5.05 (Tamura et al., 2011). Branches with less than 50% bootstrap support were collapsed.

Deep sequencing showed a total of 3,541,243 unique reads. After adapters removal and size selection (21–51 nt), a data set of 324,210 short reads was obtained and subjected to *de novo* assembly, generating a total of 87,258 contigs, ranging in size from 33 to 1005 nts with a N50 of 133 nt. BLAST search of the resultant contigs in the NCBI virus database, using a cutoff e-value of 10^{-6} for BLASTX, identified 924 contigs with amino acid sequence similarity to genes encoded by different emaraviruses. Additional 456 contigs with similarity to genes encoded by viruses of genera *Ilarvirus* and *Tombusvirus* had emerged out from the analysis of the cDNA library. However, only sequences relevant to PPSMV were analyzed in the present study. All sequences derived from deep sequencing contigs and PCR clones were compared, gathered and the full genome sequence of PPSMV was constructed.

It consists of five large RNA fragments, with a length of *ca.* 7022; 2223; 1442; 1563; and 1801 nts, that after BLAST analysis were respectively referred to as RNA1 to RNA5 (Fig. 1).

RNA1 (7022 nt) contains a single open reading frame (ORF1) that initiates at position 50–52 and ends with an amber stop codon (UAG) at position 6932–6934 (acc. no. HF568801). It encodes a polypeptide (p1) of 2294 amino acids (aa) with a predicted molecular mass of 267.9 kDa, identified as the RNA dependent RNA polymerase (RdRp) for the presence of conserved motifs common to members of the family *Bunyaviridae* (Fig. 1). PPSMV RdRp has aa sequence similarities to the RdRps of other emaraviruses (RRV, FMV, RYRV, EMARaV, RLBV) ranging from 37 to 54%. Motifs A ($DASKWS_{1125-1130}$) and C ($SDD_{1183-1185}$) are part of the palm domain of the replicase (Bruenn, 2003) and are involved in divalent-metal cation binding. Motif B ($QGNLNHLSS_{1210-1218}$) is thought to be involved in RNA binding with the Gly_{1211} residue allowing for mobility in the peptide backbone. The Lys residues in motif D ($KK_{1276-1277}$) are thought to have catalytic activity as they are reported to be close to the Asp_{1126} residue in motif A through the tertiary structure of the protein. Motif E ($EFLST_{1312-1316}$) is believed to be involved in cap-snatching in bunyaviruses as well as in a possible endonuclease activity (Duijsings et al., 2001).

RNA2 is 2223 nt long and encodes a polypeptide of 648 aa with a predicted molecular mass of 74.3 kDa, whose the translation starts at nt position 42 and ends with an ochre stop codon (UAA) at position 1986–1988 (acc. no. HF568802). This ORF (ORF2) shares 31–45% identity at the aa level with the glycoprotein precursors (GP) of FMV, RRV, RYRV, RLBV and EMARaV. All essential sequence elements of the GP precursor of *Bunyaviridae* species were identified in PPSMV p2 by computer-assisted analysis. Amino acid analysis showed the presence of three putative transmembrane helices located at aa positions 120–142, 178–200 and 593–615, four putative glycosylation sites at aa positions 69, 100, 197 and 446, and a signal peptide sequence with one predicted cleavage site ($NAG_{194-196}\downarrow NPS_{197-199}$), in which the precursor may be cleaved into two single glycoproteins: Gn (22.5 kDa) and Gc (51.8 kDa). Like RRV, RYRSV and EMARaV, PPSMV-p2 contains a phlebovirus glycoprotein motif ($GCYDCQNG_{487-494}$).

RNA3 is 1442 nt long, with protein translation starting at nt position 103 and ending with an ochre stop codon at position 1027–1029 (acc. no. HF568803). This ORF (ORF3) encodes a 34.6 kDa polypeptide (p3) predicted as the nucleocapsid protein (Fig. 1). BlastP analysis of the RNA3 encoded polypeptide showed at the aa level identities with comparable nucleocapsid proteins
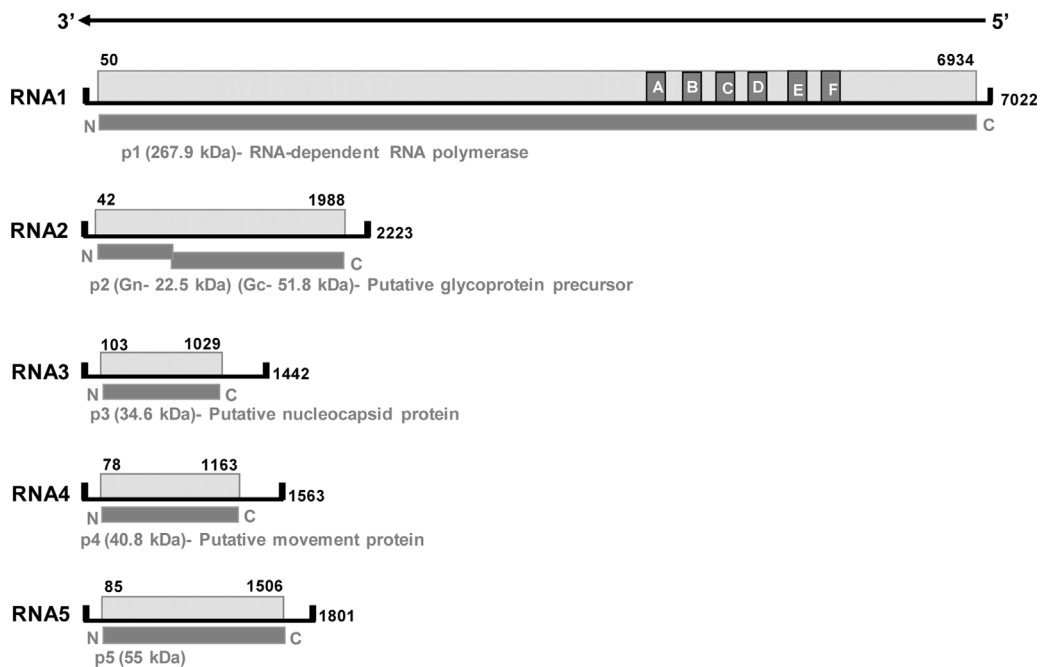
Fig. 1. Schematic representation of the organization of the PPSMV genomic RNA segments. The terminal 13 nucleotides conserved at the 5′ and 3′ termini are indicated as black boxes on each segment. Letters (A–F) represent the conserved motifs on the RdRp (RNA1) gene. Expression products of each RNA (p1–p5) are represented as dark gray boxes. Protein function and estimated mol. wt of each segment are reported below each RNA. Figure not drawn to scale.

of FMV (44%), RRV (43%), RYRSV (37%), EMARaV (35%), RLBV (27%) and MRSV (25%). However, the highest identity was found with 100 aa of PPSMV (99%) that is the unique sequence available for this virus (acc. no. CAD28797) in database (Kumar et al., 2003), together with another segment of 676 nt recently identified by Elbeaino et al. (2013). PPSMV-P3 contains three stretches of aa (NVLSFNK$_{134–140}$, NRLA$_{183–186}$ and GYEF$_{204–207}$) reported to be conserved motifs in other emaraviruses (Elbeaino et al., 2009b) and likely involved in RNA-binding (Laney et al., 2012). Even though smaller than RNA4 and RNA5, this segment has been conventionally referred to as RNA3, by analogy with other emaraviruses, in which notoriously RNA3 encodes the nucleocapsid protein.

RNA4 is 1563 nt long and protein translation initiates with an AUG at position 78–80 and terminates with UAA stop codon at position 1161–1163 (acc. no. HF568804). The molecular mass of the predicted polypeptide is 40.8 kDa (Fig. 1). Database search found similarity of this polypeptide with those encoded by RNA4 of RRV (41%), FMV (40%), RYRSV (30%) and RLBV (24%). The aa Clustal alignment of this segment did not disclose any particular conserved motifs that could be indicative of its viral nature. However, the amino acid analysis using the SLP-Local (Subcellular Location Predictor) program predicted the expression product of PPSMV RNA4 to be a possible membrane-located protein deprived of transmembrane helices that might be involved in virus cell-to-cell movement, similarly to comparable movement proteins coded by RLBV and RRV RNA4 (Yu et al., 2013; McGavin et al., 2012).

RNA5 is 1801 nt long, whose protein translation initiates with an AUG at position 85–87 and terminates with UAA stop codon at position 1504–1506 (acc. no. HF945448). It encodes a polypeptide (p5) of 473 aa with a predicted molecular mass of 55 kDa (Fig. 1). This polypeptide shared with FMV RNA5 encoded protein 33% of sequence identity. The p5 protein has not been described for EMARaV, RRV and RYRSV. No amino acids sequence motif was identified that might suggest its function.

The 5′/3′ termini of all RNA segments possess untranslated regions (UTRs) that extend from 42 to 103 nt (5′ end) and from 88 to 413 nt (3′ end). Sequences of the first 13 nts at both 5′/3′

termini of each segment are similar and exhibit nt complementarities that are unbroken in all RNAs, except for two nucleotides at position 8 and 9 (U^8–U^9). These stretches are conserved in all 5′/3′ genomic RNA ends of all emaraviruses, hence specific primers were designed and used in RT-PCR for the full amplification of RNA segments of many emaraviruses (Laney et al., 2012; Elbeaino et al., 2009b; Mielke and Muehlbach, 2007). Accordingly, these primers were exploited in RT-PCR to amplify additional RNA segments of PPSMV possibly not discovered by the deep-sequencing. These primers were repeatedly applied on the reverse-transcribed dsRNA extract previously utilized for the deep-sequencing, as well as on that of other two PPSMV-infected plants. The sequences of all PCR amplicons of different sizes [with the exception of RNA1, that failed presumably because of the big size of its segment (7022 bp)] were analyzed and showed to coincide with those generated from the deep-sequencing library; while no sequence of any additional segment was detected.

The phylogenetic trees constructed with the complete aa sequences of PPSMV RNA1, RNA2, RNA3 and the homologous polypeptides encoded by the corresponding RNA segments of emaraviruses and of known members of the family *Bunyaviridae* (denoted L, M and S) showed consistent relationships of PPSMV with all emaraviruses, and in particular with FMV and RRV, to form with them a cluster of their own, close to clades comprising members of other genera of the family *Bunyaviridae* (Fig. 2).

The sequencing of five complete RNAs has confirmed the multipartite structure of PPSMV (Kumar et al., 2003) and the viral nature of SMD, which was so far only hypothesized on the basis of similarities in morphology (DMB spherical particles of 80–200 nm and multipartite genome), symptomatology (mosaic and ring spots) and epidemiology (transmission through eriophyd mites) with other viruses now included in the genus *Emaravirus* (Jones et al., 2004; Mielke-Ehret and Muehlbach, 2012). Four of the these genomic segments code for structural proteins [a RNA-dependent RNA polymerase (RNA1), a putative glycoprotein precursor (RNA2), a putative nucleocapsid protein (RNA3), and a putative movement protein (RNA4)] while RNA5 encodes proteins with unknown
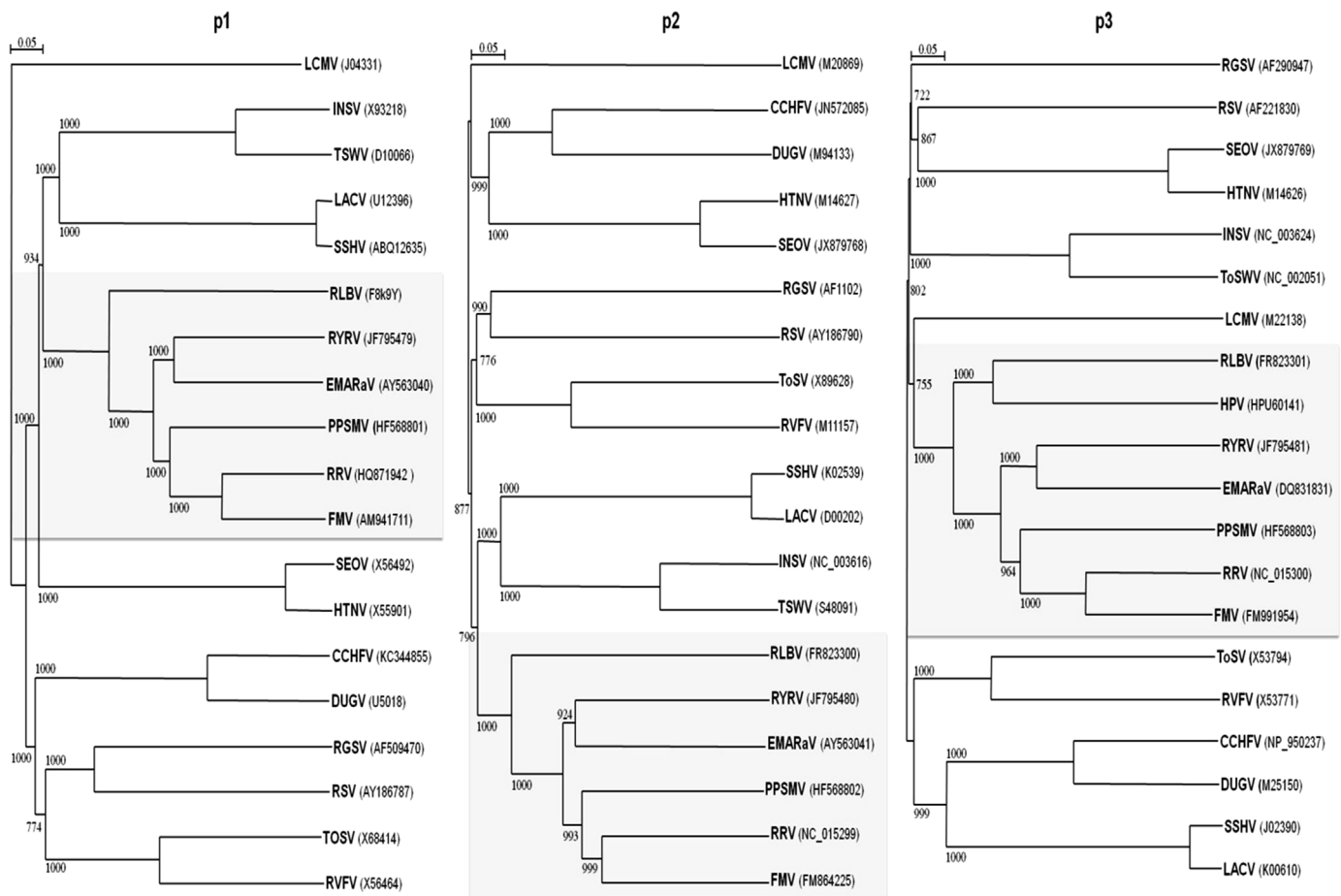
Fig. 2. Phylograms constructed with predicted aa sequences of RNA1 (RNA-dependent RNA polymerase, p1), RNA2 (glycoprotein precursor, p2) and RNA3 (nucleocapsid, p3) of PPSMV together with orthologues (denoted as L, M and S) of members of the family *Bunyaviridae*. The following genera and virus species were used to construct the phylogenetic trees: *Phlebovirus* (TOSV, Toscana virus; RVFV, Rift valley fever virus), *Tenuivirus* (RGSV, rice grassy stunt virus; RSV, rice stripe virus), *Nairovirus* (CCHFV, Crimean Congo hemorrhagic fever virus; DUGV, Dugbe virus), *Hantavirus* (HTNV, Hantaan virus; SEOV, Seoul virus), *Orthobunyavirus* (LACV, California encephalitis virus – La Crosse virus; SSHV, snowshoe hare virus), *Tospovirus* (INSV, impatiens necrotic spot virus; TSWV, tomato spotted wilt virus), *Emaravirus* (EMARaV, European mountain ash ringspot associated virus; FMV, fig mosaic virus; RRV, Rose rosette virus; RYRV, redbud yellow ringspot virus; RLBV, raspberry leaf blotch virus). Lymphocytic choriomeningitis virus (LCMV) was used an outgroup species. Accession numbers of viruses are reported.

functions, but whose the viral nature is highly supported by the similarity with RNAs of other emaraviruses (FMV and RLBV) and by the genomic organization similar to all RNA segments of emaraviruses (Mielke and Muehlbach, 2007; Elbeaino et al., 2009a,b).

The molecular information on PPSMV has remained elusive for decades despite the intensive efforts to determine the sequences and the number of RNA segments composing its genome (Kumar et al., 2002, 2003, 2004). The results obtained in this study from the deep sequencing and the PCR using primers designed on the conserved 5′/3′ termini indicate that the genome of PPSMV contains only five RNA segments. This result is in contrast with a previous indication on the presence of 5–7 RNAs composing the PPSMV genome (Kumar et al., 2002), which was based on a dsRNA analysis likely compromised by the presence of mixed infections with other PPSMV strains or other viruses. Indeed, a similar situation was also encountered in the pigeonpea materials of this study, where multiple infections with other viruses led to a complex dsRNA pattern difficult to be interpreted (data not shown).

Likely, the hypothesis of the existence of different PPSMV strains is substantiated by the well known complex nature of SMD, (i) which occurs with different symptoms (*i.e.* severe systemic mosaic and sterility, systemic mild mosaic and partial sterility, and localized chlorotic ring spots and no sterility) (Kumar et al., 2004); (ii) whose putative agents (in particular P and C isolates) show slight serological variants (Kumar et al., 2004); and (iii) in respect of which there are different responses of susceptibility/resistance by wild lines of different *Cajanus* spp. (Kumar et al., 2005).

Once more, the result of this study seems to confirm the variable number of RNAs segments composing the genome of emaraviral species, *i.e.* six RNAs for FMV, five for RLBV and four for RRV, RYRV, EMARaV.

Based on the molecular characteristics unveiled in this study and its morphological and epidemiological features similar to other emaraviruses, PPSMV seems to be the seventh species to join the list of emaraviruses known to date and accordingly, its classification in the genus *Emaravirus* seems now legitimate.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bradfute, O.R., Whitmoyer, R.E., Nault, R.L., 1970. Ultrastructure of plant leaf tissue infected with mite-borne viral-like particles. Proc. Electron Microsc. Soc. Am. 258, 178–179.

Bruenn, J.A., 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. Nucleic Acids Res. 31, 1821–1829.

Dodds, J.A., 1993. DsRNA in diagnosis. In: Matthews, R.E.F. (Ed.), Diagnosis of Plant Virus Diseases. CRC Press, Boca Raton, USA, pp. 273–294.

Duijsings, D., Kormelink, R., Goldbach, R., 2001. In vivo analysis of the TSWV cap-snatching mechanism: single base complementarity and primer length requirements. EMBO J. 20, 2545–2552.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Elbeaino, T., Digiaro, M., Alabdullah, A., De Stradis, A., Minafra, A., Mielke, N., Castellano, M.A., Martelli, G.P., 2009a. A multipartite single-stranded negative-sense RNA virus is the putative agent of fig mosaic disease. J. Gen. Virol. 90, 1281–1288.

Elbeaino, T., Digiaro, M., Martelli, G.P., 2009b. Complete nucleotide sequence of four RNA segments of Fig mosaic virus. Arch. Virol. 154, 1719–1727.

Elbeaino, T., Digiaro, M., Martelli, G.P., 2012. RNA-5 and -6, two additional negative-sense RNA segments associated with *Fig mosaic virus*. J. Plant Pathol. 94, 421–425.

Elbeaino, T., Whitfield, A., Sharma, M., Digiaro, M., 2013. Emaravirus-specific degenerate PCR primers allowed the identification of partial RNA-dependent RNA polymerase sequences of Maize red stripe virus (MRSV) and Pigeonpea sterility mosaic virus (PPSMV). J. Virol. Methods 118, 37–40.

Emanuelsson, O., Nielsen, H., Brunak, S., Von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.

Emanuelsson, O., Brunak, S., Von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. Nat. Protoc. 2, 953–971.

Ishikawa, K., Maejima, K., Komatsu, K., Kitazawa, Y., Hashimoto, M., Takata, D., Yamaji, Y., Namba, S., 2012. Identification and characterization of two novel genomic RNA segments of fig mosaic virus, RNA5 and RNA6. J. Gen. Virol. 93, 1612–1619.

Jensen, S.G., Lane, L.C., Seifers, D.L., 1996. A new disease of maize and wheat in the High Plains. Plant Dis. 80, 1387–1390.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Jones, A.T., Kumar, P.L., Saxena, K.B., Kulkarni, N.K., Muniyappa, V., Waliyar, F., 2004. Sterility mosaic disease: the 'green plague' of pigeonpea. Plant Dis. 88, 436–445.

Kannaiyan, J., Nene, Y.L., Reddy, M.V., Ryan, J.G., Raju, T.N., 1984. Prevalence of pigeonpea disease and associated crop losses in Asia, Africa and the Americas. Trop. Pest Manage. 30, 62–71.

Kulkarni, N.K., Kumar, P.L., Muniyappa, V., Jones, A.T., Reddy, D.V.R., 2002. Transmission of Pigeon pea sterility mosaic virus by the eriophyid mite, *Aceria cajani* (Acari: Arthropoda). Plant Dis. 86, 1297–1302.

Kumar, P.L., Duncan, G.C., Roberts, I.M., Jones, A.T., Reddy, D.V.R., 2002. Cytopathology of *Pigeonpea sterility mosaic virus* in pigeonpea and *Nicotiana benthamiana*: similarities with those of eriophyid mite-borne agents of undefined aetiology. Ann. Appl. Biol. 140, 87–96.

Kumar, P.L., Jones, A.T., Reddy, D.V.R., 2003. A novel mite-transmitted virus with a divided RNA genome closely associated with pigeonpea sterility mosaic disease. Phytopathology 93, 71–81.

Kumar, P.L., Jones, A.T., Waliyar, F., 2004. Biology, etiology and management of pigeonpea sterility mosaic disease. Annu. Rev. Plant Pathol. 3, 1–24.

Kumar, P.L., Latha, T.K.S., Kulkarni, N.K., Raghavendra, N., Saxena, K.B., Waliyar, F., Rangaswamy, K.T., Muniyappa, V., Doriswamy, S., Jones, A.T., 2005. Broad-based resistance to Pigeonpea sterility mosaic disease in wild relatives of pigeonpea (Cajanus: Phaseolae). Ann. Appl. Biol. 146, 371–379.

Laney, A.G., Gergerich, R.C., Keller, K.E., Martin, R.R., Tzanetakis, I.E., 2011. GenBank Database. Accession No. JF795479–JF795482.

Laney, A.G., Keller, K.E., Martin, R.R., Tzanetakis, I.E., 2012. A discovery 70 years in the making: characterization of the Rose rosette virus. J. Gen. Virol. 92, 1727–1732.

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

Matsuda, S., Vert, J.P., Saigo, H., Ueda, N., Toh, H., Akutsu, T., 2005. A novel representation of protein sequences for prediction of subcellular location using support vector machines. Protein Sci. 14, 2804–2813.

McGavin, W.J., Mitchell, C., Cock, P.J.A., Wright, K.M., MacFarlane, S.A., 2012. Raspberry leaf blotch virus, a putative new member of the genus *Emaravirus*, encodes a novel genomic RNA. J. Gen. Virol. 93, 430–437.

Mielke, N., Muehlbach, H.P., 2007. A novel, multipartite, negative-strand RNA virus is associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). J. Gen. Virol. 88, 1337–1346.

Mielke-Ehret, N., Muehlbach, H.P., 2012. *Emaravirus*: a novel genus of multipartite, negative strand RNA plant viruses. Viruses 4, 1515–1536.

Muehlbach, H.P., Mielke-Ehret, N., 2011. Genus *Emaravirus*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy. Ninth Report of the International Committee on the Taxonomy of Viruses. Elsevier-Academic Press, Amsterdam, The Netherlands, pp. 767–769.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U.S.A. 85, 2444–2448.

Reddy, M.V., Kannaiyan, J., Nene, Y.L., 1984. Increased susceptibility of sterility mosaic infected pigeonpea to powdery mildew. Int. J. Trop. Plant Dis. 2, 35–40.

Skare, J.M., Wijkamp, I., Rezende, J.A.M., Kitajima, E.W., Park, J.W., Desvoyes, B., Rush, C.M., Michels, G., Scholthof, K.B.G., Scholthof, H.B., 2006. A new eriophyid mite-borne membrane-enveloped virus-like complex isolated from plants. Virology 347, 343–353.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Yu, C., Karlin, D.G., Lu, Y., Wright, K., Chen, J., MacFarlane, S., 2013. Experimental and bioinformatic evidence that Raspberry leaf blotch emaravirus p4 is a movement protein of the 30K superfamily. J. Gen. Virol. 94, 2117–2128.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.