Identification and characterization of two novel genomic RNA segments RNA5 and RNA6 in rose rosette virus infecting roses

B. BABU[1], B. K. WASHBURN[2], K. PODUCH[2], G. W. KNOX[1], M. L. PARET[1]

[1]Department of Plant Pathology, North Florida Research and Education Center (NFREC), Institute of Food and Agricultural Sciences, University of Florida, Quincy, Florida 32351, USA; [2]Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA

Summary. – Rose rosette virus (RRV), a negative-strand RNA virus belonging to the genus *Emaravirus*, has recently been characterized to be the causal agent of rose rosette disease. Roses showing typical symptoms of RRV collected from a rose nursery in Florida were subjected to reverse transcription-PCR (RT-PCR) assay using primers corresponding to the conserved inverted 13 nucleotide long stretches found at the termini of the RRV genomic RNA segments. RT-PCR analysis yielded two novel genomic RNA segments, RNA5 and RNA6, in addition to the previously identified four RNA segments. The RNA5 is 1650 bp long and encodes for a polypeptide of 465 amino acids (54.3 K), while RNA6 is 1400 bp long and encodes for a polypeptide of 233 amino acids (27.05 K). RACE analysis showed that, both the RNA segments posses at their 5' and 3' termini, stretches of conserved inverted complementary13 nucleotides long sequence with two nucleotide mismatches as previously identified in other genomic RNA segments. Northern blot analysis as well as RT-PCR using specific primers showed the presence of the novel genomic RNA segments in infected plants, but absent in the non-infected plants. The GenBank Acc. Nos. for the sequences reported in this paper are KT007556 and KT007557.

Keywords: Emaravirus; multipartite RNA virus; rose rosette virus; RNA5; RNA6

Introduction

Roses are one of the most important ornamental flowering shrubs grown worldwide. In the United States, the total wholesale production of roses accounts for 194 million U.S. dollars (United States Department of Agriculture, 2010). Among the diseases of roses, rose rosette disease (RRD) is the most destructive to the cultivated roses, regardless of the cultivars. Symptoms include excessive lateral shoot growth, excessive thorniness, witches' broom, leaf proliferation and malformation, mosaic, red pigmentation, and eventually plant death (Amrine, 2002). Even though RRD was first reported in the early 1940s (Conners, 1941), identification of its causative agent has long been elusive,

E-mails: paret@ufl.edu, binoybabu@ufl.edu; phone: +1-850-875–7100.
Abbreviations: FMV = fig mosaic virus; PPSMV = pigeonpea sterility mosaic virus; RRD = rose rosette disease; RRV = rose rosette virus; vRNA = viral RNA; vcRNA = viral complementary RNA

although an eriophyid mite *Phyllocoptes fructiphilus* was found to be associated with its transmission (Amrine *et al.*, 1988).

Recently, the causal agent of the RRD was found to be a negative-strand RNA virus named rose rosette virus (RRV) belonging to the genus *Emaravirus* (Laney *et al.*, 2011). In November 2013, we first reported the occurrence of RRV infections in Knock Out® roses in Florida (Babu *et al.*, 2014). Like other emaraviruses, RRV has double membrane-bound particles of 120–150 nm in diameter (Gergerich and Kim, 1983; Gergerich *et al.*, 1983), and has multipartite, single stranded negative-sense RNA genome (-ssRNA). Laney *et al.* (2011) reported the complete genome sequence of four RNA segments of RRV using dsRNA-based cDNA cloning and high throughput sequencing. Genome organization and RNA sequences showed high similarities to European mountain ash ringspot associated virus (EMARaV) and fig mosaic virus (FMV) (Laney *et al.*, 2011).

All putative members of the genus *Emaravirus* are thought to consist of four or more negative-strand RNA segments

(Elbeaino *et al.*, 2009; Laney *et al.*, 2011; McGavin *et al.*, 2012; Mielke-Ehret and Muehlbach, 2007; Tantineni *et al.*, 2014), with the presence of an highly conserved, inverted complementary sequences at the terminal ends of the genomic RNA segments. RRV, FMV, EMARaV, and raspberry leaf blotch virus (RLBV) all have a highly conserved 13 nucleotide long sequences at the 5' end (5'-AGUAGUGUUCUCC …3'), which is complementary to the 13 nucleotide long sequences at the 3' end (5'-...-GGAGUUCACUACU-3') (Elbeaino *et al.*, 2009; Laney *et al.*, 2011; McGavin *et al.*, 2012; Mielke-Ehret and Muehlbach, 2007, 2012). These complementary inverted terminal sequences enable viral RNA genomes to form a panhandle structure facilitating the process of viral genomic replication and translation, as also reported for members of other negative strand RNA viruses such as bunyaviruses (Barr *et al.*, 2003; Barr and Wertz, 2005; Kohl *et al.*, 2004; Walter and Barr, 2011).

Rose rosette virus has been reported to contain four genomic RNA segments - RNA1, RNA2, RNA3 and RNA4, which are thought to encode the RNA-dependent RNA polymerase (RdRp), glycoprotein, nucleocapsid and movement protein, respectively, in the negative sense. Among the putative members of the genus *Emaravirus*, FMV possess two additional genomic RNA segments, RNA5 and RNA6, in addition to the RNA1, 2, 3 and 4 (Elbeaino *et al.*, 2012; Ishikawa *et al.*, 2012), which encode complementary sense genes for proteins (known as p5 and p6, respectively) of unknown functions. The genome of pigeonpea sterility mosaic virus (PPSMV) consists of five RNAs (Elbeaino *et al.*, 2014; Kumar *et al.*, 2003; Patil and Kumar, 2015) and, possibly, additional genomic segments, but has no experimental evidence. The RLBV possesses a fifth RNA segment RNA5 (encoding p5 protein) of unknown function (McGavin *et al.*, 2012). Among the members of the genus *Emaravirus*, high plains virus (HPV) (currently known as wheat mosaic virus (WMoV)) contains the largest number of genomic RNAs and comprises of eight genomic RNA segments with subgenomic RNAs for RNA3 (Tantineni *et al.*, 2014). However, in EMARaV and RRV, only four genomic RNA segments have been reported so far.

The objectives of the present study were to identify and characterize novel RRV genomic RNA segments using RT-PCR specific primers based on the 5' and 3' terminal sequences, designed to amplify the full length genomic RNA segments. In this paper we report the discovery of two novel additional genomic RNA segments for RRV.

Materials and Methods

Sample collection, RNA extraction and RT-PCR analysis. Ten infected Knock Out® rose plants showing typical symptoms of RRV were collected from a rose nursery in Florida along with two non symptomatic plants in November 2013. The collected samples were maintained in quarantine facilities at North Florida Research and Education Center (NFREC), University of Florida. Total RNA was extracted from the leaf tissues (100 mg) of symptomatic and healthy rose plants using the Qiagen RNeasy plant mini kit (Qiagen Inc., USA) according to the manufacturers' protocol, and was subjected to one step RT-PCR using the RRV specific diagnostic primers RRV-F and RRV-R (Laney *et al.*, 2011). The amplified products were purified using the Qiagen PCR purification kit (Qiagen Inc., USA) and subjected to sequencing and Basic local alignment search tool (BLAST; http://www.ncbi.nlm.nih.gov/blast/) analysis, in the National Center for Biotechnology information database (NCBI), using Blastn.

RT-PCR analysis of all the genomic RNAs using RRV-5'Ter and RRV-3'Ter specific primers. For simultaneous amplification of the full length genomic RNA segments of RRV, the highly conserved complementary inverted 13 nucleotide long sequences found at the 5' and 3' termini of the genomic RNA segments of RRV were used as specific primers. The sequences of the RRV-5'Ter and RRV-3'Ter were 5'-AGUAGUGUUCUCC-3' and 5'-GGAGUUCACUACU-3', respectively. RT reactions were carried out using Superscript III reverse transcriptase (Invitrogen, Life Technologies, USA) with RRV-3'Ter primer and 1 µg of total RNA in 20 µl reaction mixtures for 30 min at 45°C. The synthesised cDNA was then subjected to PCR amplification using Platinum *Taq* DNA polymerase (Invitrogen, Life Technologies, USA) in a 50 µl reaction mixture containing 4 µl of cDNA, 0.2 µmol/l of each primer RRV-5'Ter and RRV-3-'Ter, 1.5 mmol/l MgCl$_2$ 0.2 mmol/l dNTP, and 1U of Platinum *Taq* DNA polymerase (5 U/µl). The PCR was performed in Eppendorf Mastercycler gradient ES with the following thermal program: 94°C for 3 min, 35 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 5 min, and a final extension at 72°C for 10 min. The amplified product was analysed on 1% agarose gel, stained with ethidium bromide, and photographed under UV-gel documentation system. The amplified PCR products of the two isolates were purified, and cloned into the pCR2.1-TOPO TA vector (Invitrogen, Life Technologies, USA), following the manufacturers' protocol. The positive clones were selected and subjected to colony PCR analysis using the primers M13F and M13R to confirm the size of the insert. Twenty five respective clones of all sizes were then subjected to sequencing using M13F and M13R primers. The sequenced strands were subjected to end-to-end alignment and edited using the BIOEDIT (Hall, 1999). The nucleotide and the deduced amino acid sequences were compared with those in the NCBI databases using Blastn and tBlastx, respectively.

Design of specific primers for the detection of all the six RRV genomic RNAs. The obtained sequence information was used for the design of specific primers (Table 1) for the detection of three RRV genomic RNAs (RNA2, RNA5 and RNA6) using PrimerQuest software (Integrated DNA Technologies) (http://www.idtdna.com/Primerquest/Home/Index). The primers for the RNA1, RNA3 and RNA4 were designed based on the sequences in the NCBI database (Acc. Nos: NC_015298, NC_015300 and NC_015301, respectively),

Table 1. Details of primers used in the RT-PCR analysis of the six RRV genomic RNA segments

Genomic RNA segment	Primers	Target region	Tm (°C)	Amplicon size (bp)
RNA1	RNA1_F 5'- GGTTTGTAGGTTGCTTGCTATTT -3'	1355-2355	53.9	1000
	RNA1_R 5'- ATCCTGAATGGTGTGTGAAGAA -3'		54.2	
RNA2	RNA2_F 5'- GCACTGCCATGCACTTCATT -3'	496-1045	56.7	549
	RNA2_R 5'- GCTATGGTTGTTTTGGCCCG -3'		57.2	
RNA3	RNA3_F 5'- GTTTCATCTGGTGCAGTTCTTG -3'	910-1358	54.4	448
	RNA3_R 5'- CTCTTGCAGCCGATACTGTTA -3'		54.5	
RNA4	RNA4_F 5'- TCTTTGAGCTTGGCCCTAAC -3'	623-1324	55.0	701
	RNA4_R 5'- TGAATCCATCACTGAGCTAACC -3'		54.5	
RNA5	RNA5_F 5'- CCAGGACATCAGCTGGAAAT -3'	805-1130	54.9	325
	RNA5_R 5'- AGTGAAGAGGAATTGGGCTAAC -3'		54.7	
RNA6	RNA6_F 5'- AGTGCCTGTAGGCTATTGT -3'	549-1190	53.3	636
	RNA6_R 5'- GTTCTACATGATGCATATC -3'		53.6	

due to the non availability of the complete sequence information from the screened clones. The synthesised cDNAs from the RRV infected samples as well as healthy controls were tested using the designed primers.

Rapid amplification of cDNA ends (RACE) analysis. To determine the 5' and 3' terminal sequences of genomic RNA segments, 5' and 3' RACE analysis was performed. 5' RACE was performed using the 5' RACE system version 2.0 (Invitrogen, Life Technologies, USA). 3' RACE was performed using the 3' RACE system version E (Invitrogen, Life Technologies, USA), but with modifications. Total RNA extracted from the RRV infected plants were subjected to Poly(A) tail addition at the 3'end using *Escherichia coli* Poly(A) polymerase enzyme (New England Biolabs Inc., USA) according to the manufacturers' protocol. The Poly(A) tailed total RNA was then subjected to 3' RACE analysis. The details of the gene specific primers used for the RT in 5' RACE, and PCR primers for 5' and 3' RACE are listed in Table 2. The respective amplicons from the 5' and 3' RACE analysis were cloned in to pCR2.1-TOPO TA vector (Invitrogen) and sequenced using the M13F and M13R primers.

Bioinformatic analysis and genomic characterization. The obtained sequence information from the RT-PCR analysis using the RRV-5'Ter and RRV-3'Ter, and the RACE techniques were assembled using the BIOEDIT (Hall, 1999). The nucleotide and the deduced amino acid sequences were compared with those in the NCBI databases using Blastn and tBlastx, respectively. Deduced amino acid sequences were analysed using the GCG program MOTIFS (Devereux *et al.*, 1984). The ORF of the novel genomic RNAs have been deduced using the Chemgenome V2 (http://www.scfbio-iitd.res.in/chemgenome/chemgenomenew.jsp) and ORF Finder from the NCBI (http://www.ncbi.nlm.nih.gov/projects/gorf/). A protein sequence match similarity was performed using the Pfam (Finn *et al.*, 2014) accessed at the European Bioinformtics Institute (http://www.ebi.ac.uk/). The protein structure predictions for the genomic RNA segments were performed using Phyre[2] v.2.0 (Kelley and Sternberg, 2009) accessed at the Structural Bioinformatics Group, Imperial College, London (http://www.sbg.bio.ic.ac.uk/). A phylogenetic tree for the deduced amino acids was constructed using the available emaravirus RNA5 and RNA6 sequences in the NCBI database, separately. The tree was constructed using MEGA

Table 2. Details of primers used in the 5' and 3' Rapid amplification of cDNA end (RACE) analysis of the novel genomic RNA segments RNA5 and RNA6

Genomic RNA segment	Primers	Sequence (5'-3')'	Target region
5' RACE			
RNA5	RRV-1130R [a]	AGTGAAGAGGAATTGGGCTAAC	1130
RNA6	RRV-1190R [a]	GTTCTACATGATGCATATC	1190
3' RACE			
RNA5	RRV-805F [b]	CCAGGACATCAGCTGGAAAT	805
	RRV-625F [b]	CGATCTCATCATCACCTGGTTTA	625
RNA6	RRV-767F [b]	TGTTCAATCTGCCAGTGGTATAA	767
	RRV-549F [b]	AGTGCCTGTAGGCTATTGT	549

[a]Gene specific primer used for the reverse transcription of total RNA as well as for PCR amplification along with the abridged anchor primer (AAP) in 5' RACE analysis. [b]Gene specific primer used for the PCR amplification along with the abridged universal amplification primer (AUAP) in 3' RACE analysis.

software (Tamura *et al.*, 2011) using Neighbour-Joining method and 1000 bootstrap replicates.

Development of riboprobes and Northern blot analysis. The digoxigenin labelled riboprobes of the two genomic RNA segments, RNA5 and RNA6, were prepared using the *in vitro* transcription protocol of the DIG RNA labelling kit (Roche Diagnostics Corp., USA). The probes for the respective fragments were prepared in both sense and antisense orientation in order to determine the viral RNA (vRNA) as well as viral complementary RNA (vcRNA). The corresponding amplicons (1.4 kb and 1.2 kb) from the RRV-5'Ter and RRV-3'Ter primers, cloned in the pCR2.1-TOPO TA cloning vector, were subcloned into pSPT 19 vector. Following the linearization of the vector using *SacI* enzyme and subsequent gel purification using Qiagen Gel extraction kit (Qiagen Inc.), the process of *in vitro* transcription was carried out as per the manufacturers' protocol and using SP6 RNA polymerase. The developed riboprobes were quantified using NanoDrop ND-1000 Spectrophotometer (Nanodrop, USA). Northern blot analysis was performed (Sambrook and Russel, 2001) using individual vRNA probes on all the infected and healthy rose sample, while the vcRNA probes were used in combination on two randomly collected infected samples and a healthy control.

Results

Symptomatology and RT-PCR analysis using RRV-specific diagnostic primers

Ten rose plants collected from the nurseries in Florida showed typical symptoms of RRV including witches broom, excessive thorns, deformed leaves and flowers, abnormal growth of leaves from the flower (Fig. 1). The vector of RRV, the eriophyid mite species was not present on any of the infected plants. Analysis of the plant samples using the RRV-specific primers produced an amplicon of 375 bp in all the ten symptomatic samples, but did not produce any amplicon in the non-symptomatic plants (data not shown). Sequencing and BLAST analysis of the amplified product further confirmed the presence of RRV in the infected samples.

Fig. 1

Symptoms of rose rosette virus on Knock Out® roses

(a) Witches' broom, (b) excessive thorn proliferation, (c) abnormal growth of leaves from the flower, (d) excessive stem growth and unusual reddening of leaves, (e) branch proliferation and distorted leaves, (f) distorted leaf and flower.
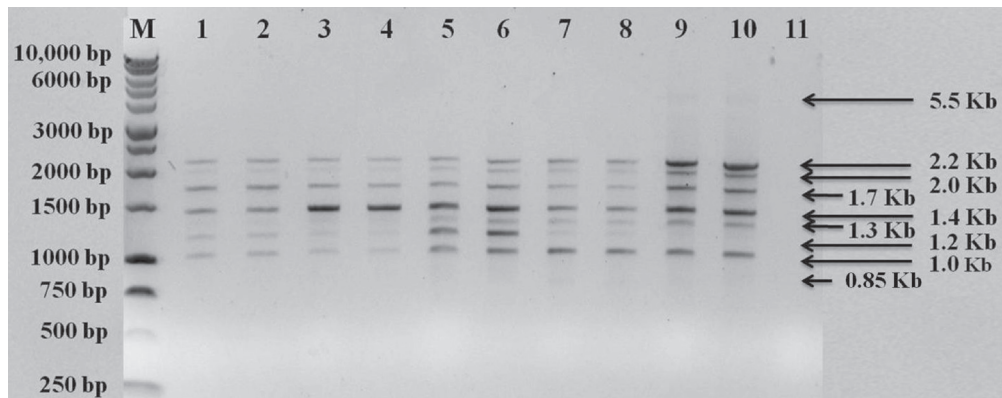
Fig. 2

Reverse transcription-PCR analysis of the rose samples infected by rose rosette virus using RRV-5'Ter and RRV-3'Ter primers
M = 1kb marker; 1–10 = infected rose samples; 11 = healthy rose plant.

RT-PCR analysis

RT-PCR analysis of the infected plants using the primers designed based on the conserved regions of the 5' and 3' end of the genomic RNAs produced eight amplicons of varying sizes (Fig. 2). The sizes of the amplicons were 2.2, 2.0, 1.7, 1.4, 1.3, 1.2, 1.0 kb and 0.85 kb. Sequence analysis of all the cloned amplicons using Blastn indicated that the amplicons 2.2, 2.0, and 1.7 kb corresponded to the RRV genomic RNA segment RNA2. Analysis of the 1.4 kb and the 1.3 kb amplicons showed that they belonged to the same contigs, and did not show any similarity on Blastn analysis. Analysis of the 1.2 kb amplicon also did not show any match with Blastn. Analysis of the 1.0 kb and 0.85 kb amplicons showed that

they corresponded to the RRV RNA4 and RNA3 segments respectively. tBlastx analysis of the 1.2 kb amplicon indicated a low identity of 37 to 40% to the sequence encoding the p6 protein (RNA6) of several FMV isolates (data not shown). tBlastx analysis of the 1.4 kb amplicon indicated an identity of 38% with the coding sequence of p5 protein (RNA5) of PPSMV (Acc. No. CCW28369) and of p5 protein (RNA5) of several FMV isolates, ranging from 32 to 35%.

RT-PCR based detection of six RRV genomic RNA segments

RT-PCR analysis of the rose samples using the designed RRV specific primers for the six genomic RNA segments, re-
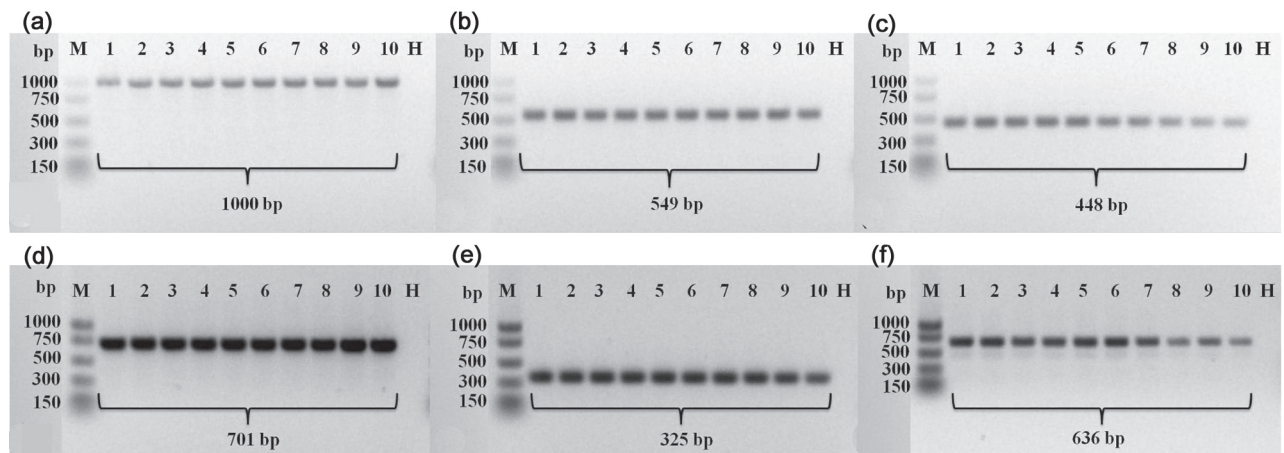

Fig. 3

Reverse transcription-PCR analysis of the rose samples using specific primers designed based on the (a) RNA1, (b) RNA2, (c) RNA3, (d) RNA4, (e) RNA5 and f) RNA6 segments of rose rosette virus
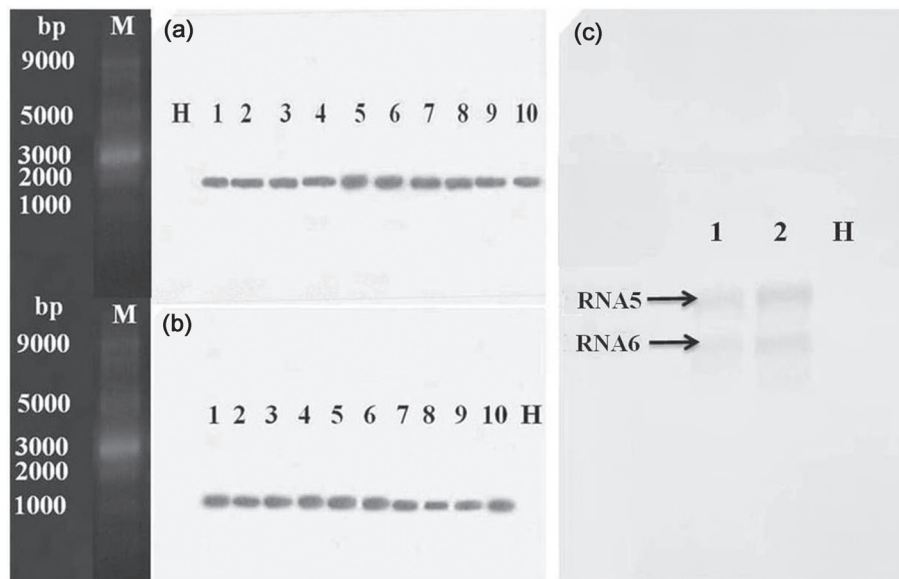M = 100 bp PCR marker; 1–10 = infected rose samples; H = healthy rose plant.

Fig. 4

Northern blot analysis of the rose samples infected by rose rosette virus using (a) vRNA of RNA5 (b) vRNA of RNA6 and (c) vcRNA probes of RNA5 and RNA6

(a, b) M = ssRNA marker analyzed on 1% 3-(N-morpholino) propansulfonic acid (MOPS)-formaldehyde agarose gel; 1–10 = infected rose samples; H = healthy rose plant: (c) 1, 2 = infected samples; H = healthy rose plant.

sulted in specific amplicons in all the RRV infected samples, while the healthy controls did not produce any amplicons (Fig. 3). RT-PCR analysis produced amplicons of 1000 bp (RNA1), 549 bp (RNA2), 448 bp (RNA3), 701 bp (RNA4), 325 bp (RNA5), and 636 bp (RNA6). The amplicons of RNA1, RNA2, RNA3, and RNA4 exhibited 98 to 99% identity with those of the RRV sequences in the NCBI database using Blastn, while the 325 bp of the RNA5 and 636 bp of the RNA6 exhibited very low identity of 38 to 40% with the sequences of RNA5 and RNA6 segments of FMV and PPSMV, in tBlastx analysis.

RACE analysis

In 3' RACE analysis of RNA5, PCR reaction using the two specific primers RRV-805F and RRV-625F in combination with the AUAP primer in the RACE kit, produced amplification products of 896 bp and 1076 bp, respectively; while reactions using primers for RNA6, RRV-767F and RRV-549F with AUAP primer, produced amplification products of 710 bp and 916 bp (data not shown). 5' RACE analysis produced amplification products of 1.1 kb for both RNA5 and RNA6 (data not shown). Sequence analysis of the cloned amplicons using the M13F and M13R primers confirmed that the two genomic RNA segments RNA5 and RNA6 possess the same 13 nucleotide long inverted complementary sequences with two nucleotide mismatches, at their terminal ends. 3' RACE

sequence analysis indicated the presence of an additional 200 bp for the RNA5 and 204 bp for the RNA6 at the 3' end. Thus, RACE analysis indicated the total length of the RNA5 segment to be 1.6 kb and that of RNA6 to be 1.4 kb. Analysis using the 5' RACE did not produce any change in the length and sequence information at the 5' end.

Northern blot analysis

Northern blot analysis with sense riboprobes of the two novel genomic RNA segments yielded signals corresponding to both RNA5 and RNA6 in infected rose samples and not in the healthy controls (Fig. 4 a,b). Northern blot analysis using vcRNA probes of the two novel RNAs also showed a positive reaction only with the infected samples (Fig. 4c).

Characterization of genomic RNA segments RNA5 and RNA6.

RNA5 of the RRV isolates was 1650 bp in length (Acc. No. KT007556), with the vcRNA containing one ORF encoding a predicted protein of 465 amino acids with a molecular mass of 54.3 K. The ORF begins with a start codon at position 1591 and a stop codon at position 191. The ORF of the RNA5 was flanked by a 190 nt 5'UTR and a 59 nt 3'UTR in the viral RNA (Fig. 5a). Database search of the predicted ORF region of the RNA5 using Blastp and Blastx, indicated
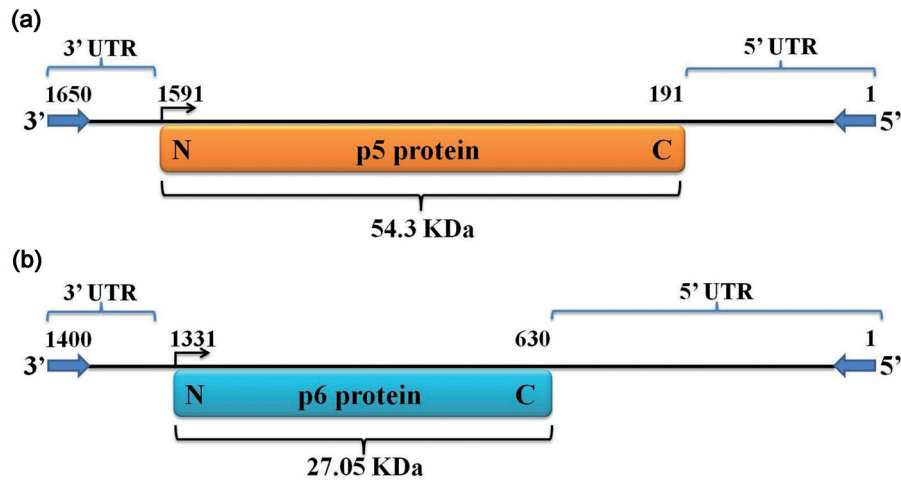
Fig. 5

Structural organization of the two RRV genomic RNA segments (a) RNA5 and (b) RNA6

The genomic RNAs are numbered from the 5'-3' end. The encoded ORFs with their start sites are indicated by open rectangles. The arrows at the 5' and 3' end denote the conserved inverted complementary sequences of the genomic RNA segments.
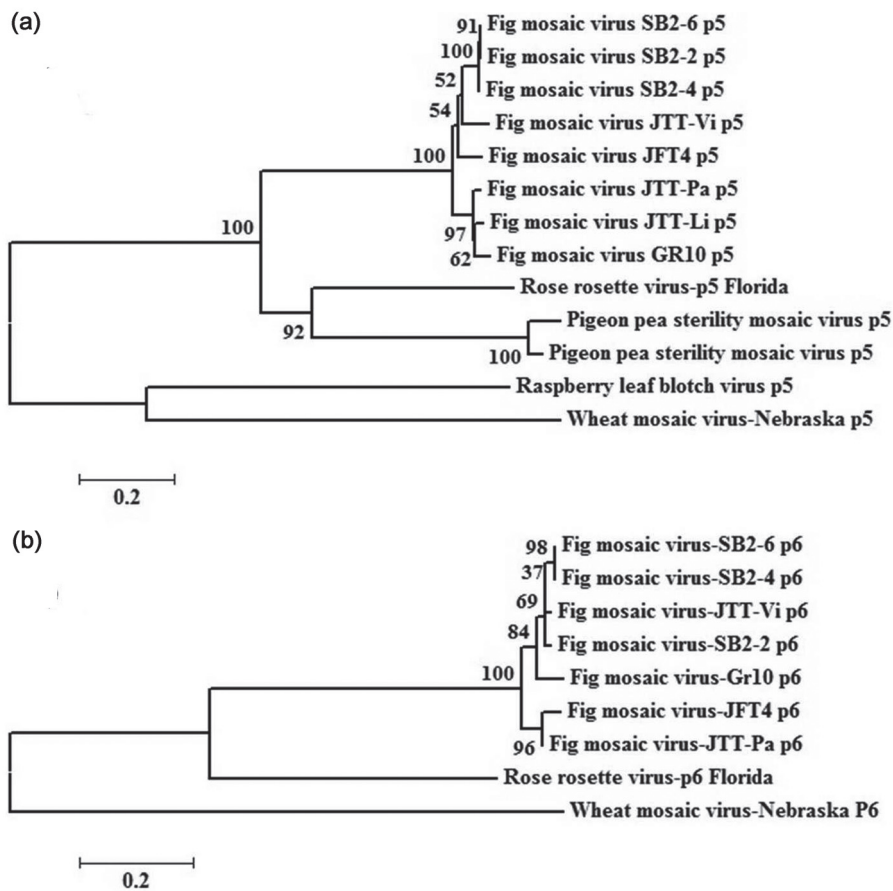

Fig. 6

Phylogenetic Neighbour-Joining tree of the amino acid sequences of (a) RNA5 and (b) RNA6 of the rose rosette virus

Analysis was done with the sequences of emaravirus isolates available in NCBI database; constructed using MEGA 5.2 software using 1000 bootstrap replicates.

a low identity of 38% and 35 to 37% with that of p5 of the PPSMV and FMV, respectively.

RNA6 of the RRV isolates was 1400 bp in length (Acc. No. KT007557), with a vcRNA containing one predicted ORF with a start codon at position 1331 and a stop codon at position 630 (Fig. 5b). The predicted protein constitutes 233 amino acids with a molecular mass of 27.05 K. The ORF is flanked by a 629 nt 5'UTR and 69 nt 3'UTR in the viral RNA. Blastp and Blastx analysis of the predicted amino acid sequences indicated a lower identity of 34 to 37% with that of the p6 protein of the FMV. Pfam analysis as well as protein structure prediction of the RNA5 and RNA6 using Phyre² software did not yield any potential similarity. On phylogenetic analysis, RRV RNA5 showed a distant relation to the RNA5 of the PPSMV, while forming a separate clade with the RNA5 of the FMV isolates. Phylogenetic analysis of the RNA6 also showed the RRV RNA6 to form a separate clade with the RNA6 of the FMV isolates (Fig. 6).

Discussion

Genomic characterization of RRV isolates from Florida indicated the presence of two novel RNAs named RNA5 and RNA6, in addition to the previously described four genomic RNA segments-RNA1, RNA2, RNA3, and RNA4 (Laney *et al.*, 2011). The presence of multiple genomic RNAs for other emaraviruses including FMV, RLBV, WMoV and PPSMV has also been previously reported (Elbeaino *et al.*, 2012, 2014; Ishikawa *et al.*, 2012; Kumar *et al.*, 2003; McGavin *et al.*, 2012; Patil and Kumar, 2015; Tantineni *et al.*, 2014). The presence of the two novel RNAs from RRV was strongly supported by several factors: a) the two RNAs were found in all ten isolates examined and were present only in the RRV infected samples, b) the two genomic RNAs were highly concordant to the other reported RRV RNAs in their structural organization in that they contained a single ORF in the vcRNA, flanked by the 5' and 3' UTR region, c) similar to the other genomic RNAs of RRV, RNA5 and RNA6 also contained conserved 5' and 3' terminal inverted complementary13 nucleotide long stretches with two nucleotide mismatches, as also reported for other emaraviruses (Elbeaino *et al.*, 2009; McGavin *et al.*, 2012; Mielke-Ehret and Muehlbach, 2007, 2012).

RT-PCR analysis using primers complementary to the 5' and 3' termini produced consistent amplicons of varying sizes only in the infected samples, thereby supporting the fact that RRV contains multiple genomic RNAs. The presence of the two amplicons 2.0 and 1.7 kb, in addition to the 2.2 kb amplicon, all of which corresponds to the previously reported RNA2 segment of the RRV, suggests the possibility of internal primer binding sites. Similar observations were also seen in the novel RNA5 with two amplicons 1.4 and 1.3 kb. However genomic RNA segment RNA1 of RRV was not detected in

the RT-PCR analysis using the 5' and 3' terminal primers. This could possibly be due to the limitations in amplifying the large size of the RNA1 (7.0 kb). Similarly, the previously reported RNA4 (1.4 kb) produced only a shorter fragment of 1.0 kb. The RNA3 (1.4 kb) also produced a faint band of 0.85 kb in certain samples (sample 5 to 10), while absent in samples 1 to 4. This could potentially be due to the limitations of multi-target amplifications. The failure to detect the 1.4 kb amplicons of both RNA3 and RNA4 could possibly be due to the size similarity with that of the partial RNA5 (1.4 kb) band, and also due to the difficulty in screening large number of transformed positive colonies. However, RT-PCR analysis using the specifically designed primers for all the RRV RNA segments indicated that the isolates contained all the six RNAs. Additionally, careful observation of the gel indicates the presence of a faint band at approximately 5.5 kb, suggesting the possibility of additional RRV genomic RNA segment RNA7; but lacks scientific evidence at this point. However, this faint band could also be a truncated portion of the RNA1 (7 kb); which needs to be further analysed.

RACE analysis of the two novel genomic RNA segments (RNA5 and RNA6) indicated the presence of conserved 13 nucleotide long stretches at their termini, which further strongly supports the fact that RNA5 and RNA6 share a high similarity in the structural organization as compared to other RRV genomic RNA segments. However, 3' RACE technique detected the presence of additional nucleotide sequences at the 3' end of the RNA5 and RNA6, making the total length of RNA5 and RNA6 to be that of 1.6 kb and 1.4 kb, respectively, which was contradictory to the RT-PCR amplicons obtained by the RRV-5'Ter and RRV-3'Ter primers (1.4 and 1.2 kb). Careful examination of the complete nucleotide sequence of those RNAs indicated the presence of internal RRV-3'Ter primer binding site in both RNA5 and RNA6 segments, which would have resulted in the shorter amplicon (data not shown).

Northern blot analysis using the RNA5 and RNA6 specific probes also proved the presence of two novel genomic RNAs in RRV. Analysis using the vRNA probes indicated that the two RNAs are present in all the infected samples with a size distinction of 1.6 and 1.4 kb. This further supports the fact that the two identified sequences are part of the RRV genomes. However, detection using the vcRNA probes indicated that the vcRNA of the two RNAs are produced during the RRV infection, and hence RNA5 and RNA6 undergo replication within the host. Similar findings were also reported previously in case of FMV in a study that used both vRNA and vcRNA probes (Ishikawa *et al.*, 2012). The two novel RNAs, RNA5 and RNA6, have no detectable similarity to any of the other RRV genomic RNA segments, other than in their highly conserved 13 nucleotide long inverted complementary sequences at the 5' and 3' termini. This also proves the fact that the RNA5 and RNA6 are quite unlikely to

be defective RNAs or truncated or rearranged RNAs derived from helper virus genomes (Simon *et al.*, 2004). The possibility of the RNA5 and RNA6 to be satellite RNAs cannot be completely ruled out. Further studies of the role of the RNA segments in the process of virus infection/movement could provide insights into these aspects. Phylogenetic analysis using the amino acid sequences of RNA5 and RNA6 suggests that these genomic RNA segments of emaraviruses are highly different from each other, and might have different roles in virus life cycles or be involved in specific interactions with certain viral hosts. However in case of RNA6, phylogenetic analysis was performed only with the RNA6 of the FMV and WMoV, due to the unavailability of sequences for other known emaraviruses. Bioinformatic analysis performed in order to assess the structure and function of the proteins encoded by these novel RNA segments did not yield any significant results. Similar to RRV RNA5 and RNA6, description on the potential function of these genomic RNA segments in other emaraviruses are not available; although a predicted potential signal sequence was identified in the N-terminus of the p5 of FMV, suggesting that it could potentially be a membrane-located protein, but lacked scientific evidence (Elbeaino *et al.*, 2012).

In conclusion, we have identified two novel genomic RNA segments – RNA5 and RNA6 – in RRV in addition to the four segments described previously. The number of genomic segments likely reflects the different biological properties of the virus and appears to be an important criterion in the taxonomic classification of segmented viruses. Future studies aiming to decipher the role of the new RNAs would provide more insights into the understanding the biology of RRV, which will be important for the development of effective strategies for the management of the virus induced diseases.

References

Amrine JW, Hindal DF, Stasny TA, Williams RL, Coffman CC (1988): Transmission of the rose rosette disease agent to Rosa multiflora by Phyllocoptes fructiphylus (Acari: Eriophyidae). Entomol. News 99, 239–252.

Amrine J (2002): Multiflora rose. In Van Driesche et al. (Eds): Biological control of invasive plants in the Eastern United States, USDA Forest Service, pp. 265–292.

Barr JN, Elliott RM, Dunn EF, Wertz GW (2003): Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication. Virology 311, 326–338. http://dx.doi.org/10.1016/S0042-6822(03)00130-2

Barr JN, Wertz GW (2005): Role of the conserved nucleotide mismatch within 3'- and 5'-terminal regions of Bunyamwera virus in signaling transcription. J. Virol. 79, 3586–3594. http://dx.doi.org/10.1128/JVI.79.6.3586-3594.2005

Babu B, Dankers H, Newberry E, Baker C, Schubert T, Knox G, Paret ML (2014): First report of Rose rosette virus associated with rose rosette disease infecting knockout roses in Florida. Plant Dis. 98, 1449. http://dx.doi.org/10.1094/PDIS-05-14-0501-PDN

Conners IL (1941): Twentieth annual report of the Canadian plant disease survey, 1940. Department of Agriculture; Ottawa, Canada, 98.

Devereux J, Haeberli P, Smithies O (1984): A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12, 387–395. http://dx.doi.org/10.1093/nar/12.1Part1.387

Elbeaino T, Digiaro M, Martelli GP (2009): Complete nucleotide sequence of four RNA segments of Fig mosaic virus. Arch. Virol. 154, 1719–1727. http://dx.doi.org/10.1007/s00705-009-0509-3

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

Elbeaino T, Digiaro M, Uppala M, Sudini H (2014): Deep sequencing of Pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses. Virus Res. 188, 27–31. http://dx.doi.org/10.1016/j.virusres.2014.03.022

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014): The Pfam protein family database. Nucl. Acids Res. Database Issue 42, D222–D230. http://dx.doi.org/10.1093/nar/gkt1223

Gergerich RC, Kim KS, Kitajima EW (1983): A particle of unique morphology associated with the disease of rose in northwest Arkansas. Phytopathology 73, 500–501.

Gergerich RC, Kim KS (1983): A description of the causal agent of rose rosette disease. Arkansas Farm Res. 32, 7.

Hall TA (1999): Bioedit, A user-friendly biologically sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

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. http://dx.doi.org/10.1099/vir.0.042663-0

Kelley LA, Sternberg MJ (2009): Protein structure prediction on the web: a case study using the Phyre server. Nat. Protocols 4, 363–371. http://dx.doi.org/10.1038/nprot.2009.2

Kohl A, Dunn EF, Lowen AC, Elliott RM (2004): Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength. J. Gen. Virol. 85, 3269–3278. http://dx.doi.org/10.1099/vir.0.80407-0

Kumar PL, Jones AT, Reddy DVR (2003): A novel mite-transmitted virus with a divided RNA genome closely associated with Pigeonpea sterility mosaic disease. Phytopathology 93, 71–81. http://dx.doi.org/10.1094/PHYTO.2003.93.1.71

Laney AG, Keller KE, Martin RR, Tzanetakis IE (2011): A discovery 70 years in the making: characterization of the Rose rosette virus. J. Gen. Virol. 92, 1727–1732. http://dx.doi.org/10.1099/vir.0.031146-0

McGavin WJ, Mitchell C, Cock PJA, Wright KM, MacFarlane SA (2012): Raspberry leaf blotch virus, a putative new member of the genus Emaravirus, encodes a novel genomic RNA. J. Gen. Virol. 93, 430–437. http://dx.doi.org/10.1099/vir.0.037937-0

Mielke-Ehret N, Muehlbach HP (2007): A novel, multipartite, negative-strand RNA virus is associated with ringspot disease of European mountain ash (Sorbus aucuparia L.). J. Gen. Virol. 88, 1337–1346. http://dx.doi.org/10.1099/vir.0.82715-0

Mielke-Ehret N, Muehlbach HP (2012): Emaravirus: A novel genus of multipartite, negative strand RNA plant viruses. Viruses 4, 1515–1536. http://dx.doi.org/10.3390/v4091515

Patil BL, Kumar PL (2015): Pigeonpea sterility mosaic virus: a legume-infecting emaravirus from South asia. Mol. Plant Path. http://dx.doi.org/10.1111/mpp.12238

Sambrook J, Russel DW (2001): Molecular cloning: A Laboratory Manual. 3rd ed. Cold Spring Harbour, New York.

Simon AE, Roossinck MJ, Havelda Z (2004): Plant virus satellite and defective interfering RNAs: new paradigms for a new century. Annu. Rev. Phytopathol. 42, 415–437. http://dx.doi.org/10.1146/annurev.phyto.42.040803.140402

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. http://dx.doi.org/10.1093/molbev/msr121

Tatineni S, McMechan AJ, Wosula EN, Wegulo SN, Graybosch RA, French R, Hein GL (2014): An eriophyid mite-transmitted plant virus contains eight genomic RNA segments with unusual heterogeneity in the nucleocapsid protein. J. Virol. 88, 11834–11845. http://dx.doi.org/10.1128/JVI.01901-14

United States Department of Agriculture (2010): 2007 Census of Agriculture, Washington, DC, USA. 3, 25.

Walter TC, Barr JN (2011): Recent advances in the molecular and cellular biology of bunyaviruses. J. Gen. Virol. 92, 2467–2484. http://dx.doi.org/10.1099/vir.0.035105-0