Complete genomic sequence of a *Rubus yellow net virus* isolate and detection of genome-wide pararetrovirus-derived small RNAs

CrossMark

Melanie L. Kalischuk [a,c], Adriana F. Fusaro [b], Peter M. Waterhouse [b],
Hanu R. Pappu [a], Lawrence M. Kawchuk [c,*]

[a] *Department of Plant Pathology, Washington State University, Pullman, WA 99164-106, United States*
[b] *School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006, Australia*
[c] *Agriculture and Agri-Food Canada, P.O. Box 3000, Lethbridge, Alberta T1J 4B1, Canada*

ARTICLE INFO

ABSTRACT

Rubus yellow net virus (RYNV) was cloned and sequenced from a red raspberry (*Rubus idaeus* L.) plant exhibiting symptoms of mosaic and mottling in the leaves. Its genomic sequence indicates that it is a distinct member of the genus *Badnavirus*, with 7932 bp and seven ORFs, the first three corresponding in size and location to the ORFs found in the type member *Commelina yellow mottle virus*. Bioinformatic analysis of the genomic sequence detected several features including nucleic acid binding motifs, multiple zinc finger-like sequences and domains associated with cellular signaling. Subsequent sequencing of the small RNAs (sRNAs) from RYNV-infected *R. idaeus* leaf tissue was used to determine any RYNV sequences targeted by RNA silencing and identified abundant virus-derived small RNAs (vsRNAs). The majority of the vsRNAs were 22-nt in length. We observed a highly uneven genome-wide distribution of vsRNAs with strong clustering to small defined regions distributed over both strands of the RYNV genome. Together, our data show that sequences of the aphid-transmitted pararetrovirus RYNV are targeted in red raspberry by the interfering RNA pathway, a predominant antiviral defense mechanism in plants.

1. Introduction

Red raspberry (*Rubus idaeus* L.) is grown in temperate regions worldwide for fresh market, jam, juice, purée or individually quick frozen fruit. Raspberry is a host to many different viruses and mixed virus infections appear to be common (Quito-Avila and Martin, 2012). In North America and Europe, *Rubus yellow net virus* isolates cause raspberry mosaic disease (RMD) while in association with members of *Black raspberry necrosis virus* and *Raspberry leaf mottle virus* (McGavin and MacFarlane, 2010). Although complete genomic sequence is available for some of the viruses involved in RMD, the complete genomic sequence of a *Rubus yellow net virus* isolate is currently unavailable, limiting the development of immunological and nucleic acid diagnostics for the molecular characterization of this disease that reduces berry growth and yield. Rubus yellow net virus (RYNV) produces a bacillifom particle and partial sequence from a highly conserved portion of the virus genome, previously

identified RYNV as a putative member of the family *Caulimoviridae*, genus *Badnavirus* (Jones et al., 2002).

Badnaviruses include some of the most destructive plant viruses, often causing devastating losses to infected crops (Harper et al., 2004; Huang and Hartung, 2001; Thresh and Owusu, 1986). Initially, it was reported that the badnaviruses were found infecting hosts growing exclusively in tropical and sub-tropical climates. However, badnaviruses have also been found infecting host plants occupying North American and other temperate climates. These include host shrub-like species including red raspberry, gooseberry and ornamental spiraea infected with strains of *Rubus yellow net virus*, *Gooseberry vein banding virus* and *Spiraea yellow leaf spot virus*, respectively.

Badnaviruses have a $120–150 \times 30$ nm non-enveloped bacilli-form capsid containing a singular circular dsDNA genome that is 7.3–8.0 kb in size (Lockhart, 1990). Virions enter a host in a non-circulative semi-persistent manner by a mealybug or an aphid vector. In the case of raspberry, RYNV is vectored by *Amphorophora agathonica* Hottes also known as the large raspberry aphid. The typical genomic structure for badnaviruses consists of three open reading frames (ORFs), and all of the genes are encoded on the same discontinuous strand (Xu et al., 2011). The first two ORFs encode proteins that are 17 and 15 kDa, respectively and little is known about their function. Most information regarding the badnaviruses is derived from the characterization,

* Corresponding author. Tel.: +1 403 317 2271; fax: +1 403 382 3156.
E-mail address: lawrence.kawchuk@agr.gc.ca (L.M. Kawchuk).

organization and highly homologous nature of ORF 3 which encodes a 216 kDa polyprotein that is cleaved post-translationally by the viral-encoded aspartic protease to produce a movement protein, a coat protein and a replicase comprised of a reverse transcriptase and ribonuclease H (Laney et al., 2012; Sether et al., 2012).

High throughput small RNA sequencing has increasingly been used to characterize plant viruses and host–pathogen interactions (Hwang et al., 2013; Kreuze et al., 2009). RNA silencing is a universal system among most eukaryotes to developmentally and temporally regulate gene expression through the production of small RNAs (Huntzinger and Izaurralde, 2011). RNA silencing also plays an important role in antiviral defense whereby, small RNAs (sRNAs) accumulate to abundant levels in infected organisms (Ratcliff et al., 1997; Waterhouse et al., 2001). In plants, RNA silencing is orchestrated by a diverse family of endonucleases, known as Dicers-like proteins (DCLs). In *Arabidopsis thaliana*, four Dicers (DCL1-4) have been identified (Margis et al., 2006). The dicer-like proteins involved in antiviral defense include DCL2 and DCL4 that generate 22–24-nt small interfering RNAs, from double-stranded replicative intermediates of plant RNA viruses (Voinnet, 2001).

To examine the molecular structure and characteristics of RYNV, nucleic acid from the virus was cloned, sequenced and compared to other members of the genus *Badnavirus*, family *Caulimoviridae*. The RYNV genomic sequence most closely resembled that of GVBV, the only other member of the genus *Badnavirus* from North America with a fully sequenced genome. Small RNA sequence profiling of RYNV infected red raspberry leaf tissue yielded a highly uneven genome-wide distribution of virus-derived RNAs (vsRNAs) with strong clustering to small defined regions distributed over both strands of the RYNV genome. This shows that the raspberry interfering RNA pathway targets specific regions of a plant dsDNA virus sequence, possibly as an antiviral defense mechanism.

2. Materials and methods

2.1. Source of infected plant material

To obtain RYNV genomic sequence, leaf tissue was collected from one red raspberry (*R. idaeus* L.) plant located near Lethbridge, Alberta, Canada expressing symptoms characteristic of RYNV infection (Kalischuk et al., 2008). Tissue was ground in liquid nitrogen and total DNA extracted using hexadecyltrimethylammonium bromide (CTAB) (Doyle and Doyle, 1990). Total RNA for the small RNA sequencing was extracted using the miRNeasy spin column kit (Qiagen, Toronto, Canada) from three RYNV infected plants showing the symptoms similar to those described above and located at the same location. The quantity and quality of the RNA were checked using spectrophotometry and agarose gel electrophoresis.

2.2. Polyclonal antibodies to the coat protein of RYNV

A DNA fragment containing a portion of the putative coat protein was prepared by PCR with the oligonucleotides 5′CTGGAATATCTACAAAACGTCTGCAAG3′ and 5′GATGCTACAGATTTCATCGTCATCTGA3′ (Kalischuk et al., 2008). The fragment was cloned and expressed in the glutathione S-transferase gene fusion system pGEX (GE Healthcare, USA) following the manufacturer's protocol. Amplified products inserted into the recombinant clones were confirmed by DNA sequencing. Protein from inclusion bodies was purified using glutathione sepharose 4B (GE Healthcare, USA). The purified fusion protein was used to immunize rabbits for the production of polyclonal antibodies.

2.3. Development of a full length RYNV infectious clone

A full length clone was constructed by ligating two PCR products to produce a greater than unit length fragment of the RYNV genome based on preliminary sequence. *Nhe*I was identified as a unique endonuclease restriction site in initial RYNV genomic sequence. Oligonucleotides 5′ATATAGGA**GCTAGC**CGCAGCTGTGGA3′ and 5′AGC<u>GAATTC</u>GCCTGATTCCGAGCTGCTTGTTG3′ were used to PCR amplify 4501 bp from the *Nhe*I area of the genome (bold nucleotides represent the *Nhe*I site and the underlined nucleotides represent an added *Eco*RI site). A second PCR fragment was amplified 5871 bp from *Nhe*I in the opposite direction of the RYNV genome using oligonucleotides 5′AGC<u>GTCGAC</u>CTGGTGAACAAGGAGGTACAGAGC3′ and 5′TCCACAGCTGCG**GCTAGC**TCCTATAT3′ (bold nucleotides represent the *Nhe*I site and underlined nucleotides represent a *Sal*I site). Phusion high-fidelity DNA polymerase (New England Biolabs, Canada) was used for PCR and procedures followed the manufacturer's recommendations. PCR products were examined using agrose gel electrophoresis and DNA was excised, purified using a Qiaex II gel extraction kit (Qiagen, Toronto) and ligated using T4-DNA ligase into the binary vector pBINPLUS. Ligation reactions were directly transformed into *Agrobacterium tumefaciens* EHA105 by electroporation. Cells were incubated for 3 h and then 20 μl of the inoculum was injected using a Hamilton syringe into the leaf veins (5 spots per leaf) of young vegetatively propagated woodland strawberry (*Fragaria vesca* L.) or *R. idaeus*. Another construct containing empty pBINPLUS was prepared and tested in a similar manner. Transformed and healthy plants were evaluated for RYNV infection with an immunocapture and subsequent PCR amplification of a 451 bp fragment that was negative with blank and preserum controls (Kalischuk et al., 2008).

2.4. Genomic sequencing of RYNV

The complete genomic sequence of RYNV was determined from the RYNV full length infectious clone. The 4501 bp *Nhe*I/*Eco*RI restricted product was subcloned into the *Spe*I/*Eco*RI site of pBluescript II SK(−) (Stratagene, Santa Clara, CA, USA). However, due to instability in *Escherichia coli*, direct cloning of the 5871 bp fragment was unsuccessful. Instead, the 5871 bp fragment was digested with *Bam*HI and each of the two DNA fragments subcloned separately. The *Sal*I/*Bam*HI fragment was cloned into the corresponding sites of pBluescript II SK and the *Nhe*I/*Bam*HI fragment was cloned into the corresponding sites of pLitmus 38i (New England Biolabs, Ipswich, MA, USA). Standard procedures were used for the remaining cloning and Sanger sequencing reactions. Primer walking was used to obtain sequence from the entire length of each insert. Sequencing was performed on both DNA strands to produce at least three copies of overlapping sequence for the entire genome (primers are listed in Supplemental Material).

2.5. RYNV genomic sequence assembly and analysis

Assembly of the RYNV genome was completed using Sequencher V4.7 (Gene Codes Inc., Ann Arbor, MI, USA) and the alignments used only high quality value bases as determined by Sequence Scanner Software v1.0 (Applied Biosystems, Carlsbad, CA, USA). Nucleotide and the deduced amino acid sequence were compared to existing sequences in GenBank and in other databases such as conserved protein domain (CDD) (http://www.ncbi.nlm.nih.gov/structure/cdd.shtml), Pfam (http://pfam.sanger.ac.uk/), SMART (http://smart.embl-heidelberg.de/) and Prosite (http://prosite.expasy.org/) using the basic local alignment search tool (BLAST). Putative ORFs were identified using ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) and

theoretical molecular weights were predicted using Compute pI/Mw (http://www.expasy.org/tools/pi_tool.html). Potential promoter elements and non-coding RNA's were identified using PLACE (http://www.dna.affrc.go.jp/PLACE/), PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and Signal Scan (http://www.dna.affrc.go.jp/sigscan/signal.html/). Secondary structure predictions were made using Jpred (www.compbio.dundee.ac.uk/www-jpred/) and Paircoil (http://groups.csail.mit.edu/cb/paircoil/cgi-bin/). Patterns and profiles were predicted using ELM (http://elm.eu.org/search/). Direct and inverted symmetry repeats were identified using Radar (http://www.ebi.ac.uk/Tools/Radar/index.html).

2.6. Phylogenetic analysis

Phylogenetic relationships among the badnaviruses were estimated using CLUSTAL X v2 multiple alignment with ORF 3 amino acid sequence. The neighbor-joining clustering algorithm was used to estimate the phylogeny and a rooted tree constructed using CLC Main Workbench software (CLC Biosciences, Cambridge, MA, USA). Confidence estimates were based on bootstrap sets of re-sampling alignments with 1000 replicates and these values were added to the nodes of the trees. GenBank accessions used in the analysis are listed in Supplemental Material.

2.7. Small RNA sequencing and mapping of RYNV infected leaf tissue

The RNA was sequenced on an Illumina Genome Analyzer IIx platform (Illumina, Inc., Madison, WI). Briefly, the processing by Illumina consisted of the following successive steps: purification of 20–30 nt RNAs by filtering through a polyacrylamide gel, ligation of 3′ and 5′ adapters to the 20–30 nt RNAs, c-DNA synthesis by reverse transcription followed by acrylamide gel purification and a final steps of bridge amplification and paired-end sequencing.

RYNV-derived sRNA sequences were identified by mapping the reads to the RYNV genome. Raw reads were filtered for quality and adaptor sequences trimmed using FastX Test Kit (http://hannonlab.cshl.edu/fastx_toolkit). A mappable read had a Phred score greater than 30, a 3′ADT and greater than 15 bases after the 3′ADT cut. Reads that were 18–25 nt and that met the filtering criteria were mapped to the RYNV genome using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml). The seed length was set to 16 nt and one mismatch was allowed in the seed region during mapping. The command line for Bowtie on a Mac Unix system was as follows: bowtie -q -l 16 -n 1 -e 40 -k 1 raspberry_virus Rubida.fq. Output data were exported to a Microsoft Excel spreadsheet and summarized using standard graphics software (http://www.corel.com).

3. Results

The complete genomic sequence of RYNV was determined and consisted of 7932 bp (Genbank Accession: KF241951).

3.1. Non-coding regions of RYNV genomic DNA

The intergenic region (IR) of RYNV was 969 bp and contained many of the conserved nucleic acid sequences previously described for plant dsDNA viruses (Benfey and Chua, 1990; Medberry and Olszewski, 1993). A plant tRNA[Met] sequence was predicted with 67 percent nucleotide identity to the complementary sequence of 3′ACCAUAGUCUCGGUCCAA5′, and this region has been previously described as one of the priming site for reverse transcription (Medberry et al., 1990). The 5′ end of the tRNA[Met] was designated as the first nucleotide for numbering the RYNV genome as done

with previously characterized badnaviruses. Nucleic acid motifs previously identified and located in the intergenic region of other members of the family *Caulimoviridae* and identified along the RYNV intergenic region included a TATA box (nt 7523–7536), a cap signal associated with the TATA box (nt 7561–7564), a hexamer motif (nt 7633–7638), an as-1-like sequence (nt 7363–7379) and a GATA box motif (nt 7486–7490) (see Supplemental Material for details on conserved sequences).

3.2. Coding regions of RYNV genomic DNA

The nucleic acid sequence of RYNV ORF 1 potentially encodes a 210 amino acid protein of 24 kDa (Table 1) with two nuclear export signals (NES), two coiled-coil regions and a proline rich C-terminus. Homology was detected between sequences of RYNV ORF 1 and both P24 of *Rice tungro bacilliform virus* and ORF 1 of *Gooseberry vein banding virus* isolates.

The nucleic acid sequence of RYNV ORF 2 encodes a predicted 17 kDa protein (153 aa) with homology to ORF 2 from other badnaviruses including citrus mosaic virus (CMBV), dracaena mottle virus (DMV), dioscorea bacilliform virus (DBV), kalanchoe top-spotting virus (KTSV) and sugarcane bacilliform virus (SCBV). RYNV ORF 2 contained one NES, one 14 amino acid direct repeat, one coiled-coil domain and a proline rich C-terminus.

Nucleotide sequence of RYNV ORF 3 was predicted to produce a 1971 amino acid protein with a molecular mass of 210 kDa (Table 1). RYNV ORF 3 resembled that of other badnaviruses encoding a polyprotein containing the movement protein (MP), coat protein (CP), aspartic protease (PR), reverse transcriptase (RT) and ribonuclease H (RNaseH) (Laco et al., 1995; Marmey et al., 1999; Medberry et al., 1990; Tzafrir et al., 1997). The active site $HXGX_9HRX_3GX_8D...EXDX_3G$ (where X represents any amino acid) for the MP was located at RYNV ORF 3 amino acid position 137–182. In addition to similarities to the badnaviruses (Bouhida et al., 1993; Tzafrir et al., 1997), the active site for the MP region of the RYNV genome also showed homology to the 30 K superfamily movement protein from members of the families *Caulimoviridae* and *Flexiviridae* and the genus *Tobamovirus* (Melcher, 2000). A second motif $GX_2SXRFXNYX_7P$ (where X represents any amino acid) recognized to be involved in systemic infection was also found within the RYNV amino acids 299–315 (Tzafrir et al., 1997). A total of five proline, glutamic acid, serine and threonine (PEST) sequences found in proteins with short cell half-lives, were located within the putative movement protein region of RYNV. PEST sequences have been described previously (Kawchuk et al., 2001).

The badnavirus coat proteins were previously identified in ORF 3 as containing two cysteine–histidine rich motifs (Cys–His motif), with sequences $CXCX_2CX_4HX_4C$ and downstream $CX_2CX_5HX_5CX_2CX_4CX_2C$ (where X represents any amino acid) (Medberry et al., 1990). These two Cys–His motifs within ORF 3 of RYNV were located at amino acids 864–879 and 998–1024, respectively. At amino acids 1209–1216, 3′ to the RYNV putative CP, was an aspartic protease (PR) active site motif AX_2DXGXT (where X represents any amino acid). The consensus sequences for RT and RNaseH were identified at the 3′ end of ORF 3 at amino acids 1553–1556 and 1553–1823, respectively.

Open reading frame 4 (136 aa) encodes a predicted protein of 15 kDa (Table 1). RYNV ORF 4 overlapped the 3′ region of ORF 3 but in a different reading frame. Homology was absent between RYNV ORF 4 and other known proteins. Comparative genomics between ORF 4 and a partial sequence for another RYNV isolate (Jones et al., 2002) showed 76% amino acid similarity.

The nucleotide sequence of RYNV ORF 5 was located 305 bp downstream of ORF 3, in the same reading frame as ORF 3 and potentially encoded a 17 kDa protein of 152 amino acids (Table 1). This ORF 5 sequence overlapped the putative consensus sequence

Table 1
Coding capacity of Rubus yellow net virus open reading frames.

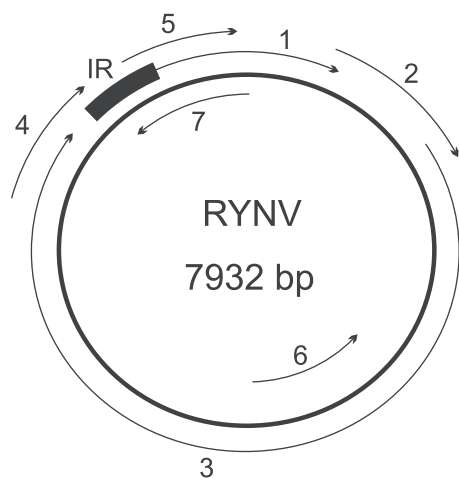| ORF	1st nucleotide	Last nucleotide	Size (bp)	No. amino acids	M_r (kDa)
1	386	1018 (TGA)	633	210	24
2	1015	1476 (TGA)	462	153	17
3	1433	7348 (TGA)	5916	1971	210
4	6940	7356 (TAG)	417	136	15
5	7654	180 (TGA)	459	152	17
6	3330	3767 (TGA)	438	145	16
7	7906	405 (TAA)	432	143	17

Fig. 1. Rubus yellow net virus (RYNV) genomic organization. A schematic diagram showing the dsDNA genomic organization of RYNV. The rectangle represents the intergenic region containing the tRNA^{Met} consensus sequence. Lines with arrows represent the open reading frames 1–7. ORF 3 encodes a polyprotein containing the putative movement protein, coat protein, protease, reverse transcriptase and ribonuclease H.

complementary to the plant tRNA^{Met} and other predicted promoter elements. Features predicted for RYNV ORF 5 included a signal peptide for localization within the secretory pathway near the N-terminus (1–21 aa), a transmembrane helix near the central portion of the sequence (66–85 aa) and an endocytosis signal (YxxF) towards the C-terminus, with all of these features being previously described (Kawchuk et al., 2001).

3.3. Open reading frames along the antisense strand

There were two ORFs predicted at greater than 10 kDa along the antisense strand of the RYNV genome and they were designated as ORFs 6 and 7 (Fig. 1). ORF 6 would produce a 145 amino acid 16.2 kDa protein. Similarly, ORF 7 would encode a 143 amino acids 16.7 kDa protein that contained a zinc finger-like motif with 86% amino acid similarity to the second Cys–His motif of the RYNV putative coat protein.

3.4. Infectivity of the full length clone

To confirm that the RYNV genomic DNA sequence was infectious, a full length clone was developed and agro-inoculated into raspberry and strawberry plants. At 14 days post-inoculation, a PCR product corresponding to the RYNV genomic sequence was detected in the immunocapture assay of newly emerging leaves of all plants inoculated with *A. tumefaciens* carrying the full length clone. The PCR product was not detected in any plants injected with the RYNV full length clone nucleic acid or agro-inoculated with the empty pBINPLUS vector. Leaves agro-inoculated with the RYNV full length clone produced necrotic lesions in *F. vesca* and *R. idaeus* (data not shown). Transmission electron microscopy (TEM)

of RYNV agro-inoculated plants detected 140 × 30 nm bacilliform particles (data not shown).

3.5. Phylogenetic analysis

The phylogeny derived from the badnavirus ORF 3 amino acid sequence (Fig. 2 & Table 2) showed that RYNV was most closely related to GVBV but they remain as two distinct species as they have differences in host range and differences in polymerase nucleotide sequences greater than 20%.

3.6. Virus derived small RNA profiling

A total of 14 million sRNA reads between 15 and 30 nt in length were obtained from high throughput sequencing of RYNV-infected raspberry. After filtering the data and using a greater than 99% accuracy cut-off value for base calls, 6.7 million reads were mapped against the RYNV genomic sequence. Of the mappable reads, a total of 0.4% showed sequence homology with RYNV, with a greater number of sRNAs represented by the sense strand than by the antisense strand. The genomic coverage of RYNV by sRNAs was 84% and the size classes of the RNAs were mainly 21–24 nt with 22-nt being the most abundant size class (Fig. 3). Mapping the viral small RNAs (vsRNA) to the RYNV genome indicated several prominent areas targeted for RNA silencing. ORF 2 exhibited concentrated high levels of RYNV vsRNAs and these regions corresponded to predicted secondary structures (Fig. 3). There were forty-four regions dispersed across the RYNV genome that were devoid of vsRNAs. Although 82% of these regions were less than 100 nt in length, one identified sRNA desert comprised 6.5% of the RYNV genome or 514 nt (Fig. 3). This 514 nt region corresponded to the 5′ portion of the large polyprotein sequence and contained the sequence for the active site for the MP (Fig. 3).

4. Discussion

The genome of RYNV, a pararetrovirus that infects red raspberry and is transmitted by an aphid vector, was sequenced and found to be 7932 base pairs in size. All badnaviruses, including RYNV, share ORFs 1–3 which have approximately the same size and location within the genome (Lockhart, 1990; Medberry et al., 1990). Nucleotide sequence of ORF 3 encodes a large polycistronic transcript critical to the dsDNA viral lifecycle, as it produces the movement protein, coat protein and reverse transcriptase. Phylogenetic analysis of the amino acid sequence for ORF 3 suggests that RYNV is a member of genus *Badnavirus* and is most related to GVBV, another member belonging to genus *Badnavirus* found infecting a temperate climate host.

Analysis of ORF 3 indicated that RYNV closely resembled other members of the *Badnavirus* genus regardless of the many biological and molecular differences. Unlike ORF 3, ORF 1 and 2 had low homology with the badnaviruses and other known proteins. RYNV ORF 1 and 2 contained a proline rich C-terminus indicating possible non-sequence specific nucleic acid binding potential. Non-specific nucleic acid binding was demonstrated at the proline rich
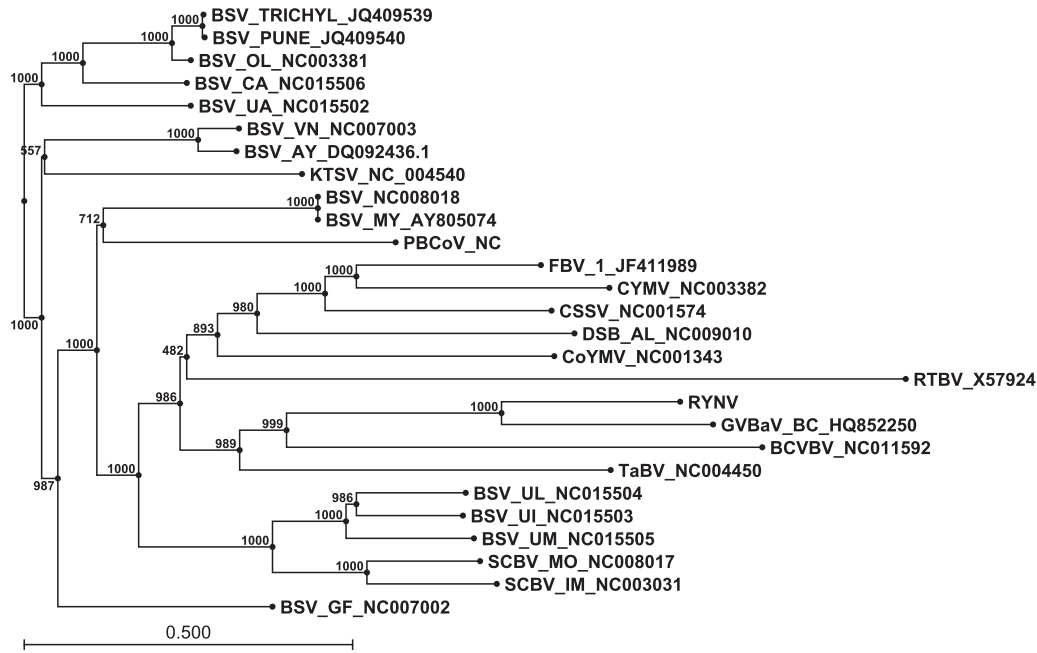
1000 BSV_TRICHYL_JQ409539
1000 BSV_PUNE_JQ409540
1000 BSV_OL_NC003381
BSV_CA_NC015506
1000 BSV_UA_NC015502
1000 BSV_VN_NC007003
BSV_AY_DQ092436.1
557 KTSV_NC_004540
1000 BSV_NC008018
712 BSV_MY_AY805074
PBCoV_NC
1000 FBV_1_JF411989
1000 CYMV_NC003382
980 CSSV_NC001574
893 DSB_AL_NC009010
482 CoYMV_NC001343
RTBV_X57924
986 1000 RYNV
999 GVBaV_BC_HQ852250
989 BCVBV_NC011592
1000 TaBV_NC004450
986 BSV_UL_NC015504
1000 BSV_UI_NC015503
BSV_UM_NC015505
1000 SCBV_MO_NC008017
SCBV_IM_NC003031
987 BSV_GF_NC007002

0.500

Fig. 2. Neighbor-joining dendrogram of sequence relationships determined using the amino acid sequence alignment of the reverse transcriptase among species within genus *Badnavirus*. The tree was rooted to rice tungro bacilliform virus (RTBV). Details of the accessions used in the analysis are in the Supplemental Material. Alignments were produced by a CLUSTAL algorithm and the dendrogram was produced by CDC Main Workbench software. Horizontal distances were proportional to sequence distances and vertical distances were arbitrary. The dendrograms was bootstrapped 1000 times (shown at nodes).

terminus region for ORF 2 of cacao swollen shoot virus (CSSV) and RTBV (Jacquot et al., 1996, 1997). RYNV ORF 1 and 2 also contained coiled-coil regions that are associated with protein–protein interactions. Coiled-coil domains located in these ORFs are widespread across plant pararetroviruses (Leclerc et al., 1998) and protein–protein interactions were detected through

tetramerization (Stavolone et al., 2001). The amino acid sequence of RYNV ORF 2 was especially interesting because in addition to possible protein–protein interactions and nucleic acid binding potential, the DXG motif necessary for aphid transmission of CaMV (Schmidt et al., 1994) was present at the C-terminus of the sequence and may function in aphid transmission. Evidence supporting

Table 2
Pairwise sequence alignments for comparing Rubus yellow net virus with the other badnaviruses.

Virus	Genome size (bp)	Nucleotide identity[a] (%)	Amino acid identity (similarity) (%)[a]				
			ORF 1	ORF 2	ORF 3	ORF 4[b]	ORF5[c]
BSV_OLV	7389	53	16 (36)	17 (40)	37 (51)		
BSV_Trichyl	6950	51	19 (39)	17 (41)	36 (50)		
BSV_Pune	6950	51	15 (35)	17 (40)	36 (50)		
BSV_CA	7408	54	19 (36)	22 (43)	37 (51)		
BSV_UA	7519	53	20 (39)	20 (39)	36 (51)		
BSV_VN	7801	54	15 (29)	19 (40)	36 (51)		
BSV_AY	7722	54	19 (39)	22 (42)	36 (50)		
BSV_UL	7401	53	21 (38)	21 (37)	34 (50)		
BSV_UI	7458	53	23 (40)	23 (39)	34 (51)		
BSV_UM	7532	54	19 (39)	20 (33)	34 (50)		
BSV_MYV	7650	54	16 (35)	23 (41)	39 (54)		
SCBV_MV	7568	53	19 (34)	24 (38)	34 (50)		
SCBV_IMV	7687	53	17 (34)	22 (39)	33 (49)		
KTSV	7591	53	20 (40)	27 (45)	35 (50)		
CSSV	7161	54	18 (33)	23 (36)	40 (54)	21 (35)	
CMBV	7559	53	18 (32)	31 (48)	39 (53)	22 (37)	
ComYMV	7489	52	18 (34)	20 (38)	36 (51)		
FBV_1	7140	54	18 (34)	29 (46)	40 (54)		
DBV	7261	52	15 (30)	26 (45)	36 (52)		
DMV	7531	54	19 (34)	21 (47)	39 (55)	19 (33)	18 (26)
BCVBV	8759	53	17 (28)	20 (35)	35 (50)		
GVBAV	7649	61	49 (70)	38 (64)	56 (68)		
TaBV	7458	52	18 (35)	23 (40)	38 (53)		
PBV_Co	7451	54	17 (35)	24 (41)	35 (51)		
RTBV	8002	50	19 (43)	18 (29)	24 (39)		9 (16)

[a] Needleman–Wunsch algorithm was used for alignments. Identity and similarity were calculated as # matches divided by longest total sequence length of either query or subject multiplied by 100.

[b] Based on same position along the genome, RYNV ORF 4 was compared to CSSV ORF Y, CYMV ORF 6, and DMV ORF 6.

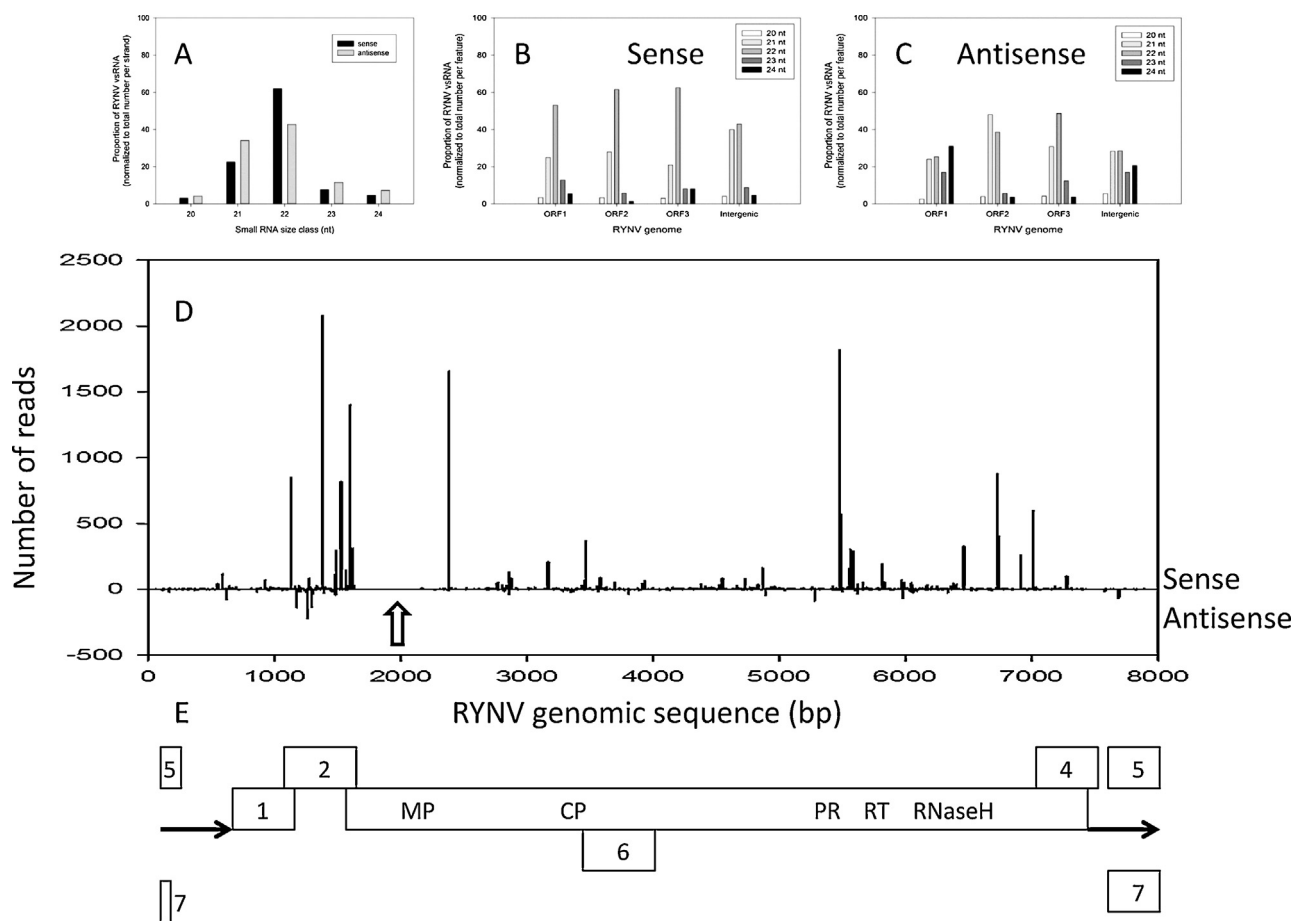[c] RYNV ORF 5 was compared to DMV ORF 7 and RTBV P46.

Fig. 3. Illumina deep-sequencing analysis of the Rubus yellow net virus (RYNV) vsRNA from infected *Rubus idaeus* L. leaf tissue. (A) Bar graph showing the proportion of mapped sense and antisense RYNV vsRNA classified into 20–24 nt distribution. Bar graphs showing the proportions of 20–24 nt RYNV vsRNA mapped to ORF 1–3 and the intergenic region on the sense (B) and antisense (C) strand of the RYNV genomic sequence. (D) Genome-wide map of RYNV vsRNA at single-nucleotide resolution. The diagram plots the number of 20–24 nt vsRNA at each nucleotide position of the 7932 bp RYNV genome. Base numbering for the RYNV genome begins at the 5′ nucleotide position of the tRNA[Met] consensus sequence, on the sense strand. Vertical lines above the axis represent sense reads starting at each respective position and those below the axis represent the antisense reads. The arrow indicates the 514 bp region devoid of RYNV-derived sRNAs. (E) In scale, linear genomic map of RYNV sense and antisense strands. Lines with arrows represent the intergenic region and rectangles represent ORFs. ORF 3 is the polyprotein consisting of the movement protein (MP), coat protein (CP), aspartic protease (PR), reverse transcriptase (RT) and the ribonuclease H (RNaseH).

additional function includes the prediction of nuclear export signals (NES) contained in ORFs 1 and 2, suggesting a nucleocytoplasmic function.

Unique to the RYNV genome were two small ORFs located on the sense strand. CSSV, *Citrus mosaic virus* (CMBV) and *Dracaena mottle virus* (DMV) also have an ORF located in relatively the same position as RYNV ORF 4 (Su et al., 2007; Huang and Hartung, 2001). Although homology was not detected between RYNV ORF 5 and any other proteins, DMV, CaMV and RTBV have an ORF of a similar size near the same vicinity. Another feature unique to the RYNV genome was that ORF 6 and ORF 7 occur in the antisense sequence and may encode several zinc finger-like motifs. ORF 7 encoded a zinc finger-like motif with 86% amino acid similarity to the second Cys–His motif identified in the RYNV putative coat protein. Zinc finger motifs are often involved in forming finger-like protrusions involved in binding DNA, RNA, protein and/or lipid substrates. There are many occurrences of zinc finger signatures in viral genomes and functional studies have suggested diverse roles such as in structure and regulation (Tanchou et al., 1998). Since RYNV ORFs 6 and 7 are located on the antisense strand and are not preceded by sequences associated with promoter activity, the transcription of RYNV is likely asymmetric like that of the other members of genera *Badnavirus* and *Caulimovirus* (Medberry et al., 1990).

An increasing number of DNA viruses are being reported that are targeted by the interfering RNA pathway and produce RNA silencing suppressors (Blevins et al., 2006; Moissiard and Voinnet, 2006; Bronkhorst et al., 2012). In our study, approximately 0.4% of total sRNAs generated by raspberry were RYNV-derived vsRNAs which spanned 84% of the RYNV genome. Involvement of all known DCLs in the production of a diverse pool of 21–24 nt vsRNAs has been demonstrated as a general host plant response to pararetrovirus infection (Blevins et al., 2006; Moissiard and Voinnet, 2006). Our sequencing of sRNAs from RYNV-infected raspberry leaf tissue revealed a diverse pool of 21–24-nt vsRNAs, supporting the aforementioned interaction between pararetrovirus and host. The 22-nt vsRNAs were the predominant size class of vsRNAs spanning the RYNV coding regions and concentrated in areas such as ORF 2 that could form considerable secondary structure (Stavolone et al., 2001). Another study that partially examined size and distribution of vsRNAs in badnavirus-infected tissue showed that the majority of vsRNAs were also 22-nt in size and distributed along the genome (Kreuze et al., 2009).

Mapping the vsRNAs to the RYNV intergenic region revealed that they were predominately 21 and 22-nts in size (Fig. 3). Previously, it was demonstrated that in *Arabidopsis thaliana* and *Brassica rapa* exposed to CaMV, the 24-nt size class of vsRNAs was predominant and that most of the vsRNAs clustered within the long

pararetrovirus leader sequence (Blevins et al., 2011; Moissiard and Voinnet, 2006), rather than being dispersed throughout the genome, as with RYNV. Mapping of the RYNV vsRNAs against the entire viral genome revealed a 514 nt region that appears to have escaped RNA silencing. This sequence lacking sRNAs, spanned the region of the RYNV genome associated with the binding domain in the badnavirus movement protein. Deep sequencing to identify novel viruses in sweet potato reported partial vsRNA profiles for two badnaviruses that also lacked vsRNA in a similar location (Kreuze et al., 2009).

Genomes of RNA viruses produce a double-stranded RNA intermediate during replication (Weber et al., 2006). These dsRNA products are recognized by the RNAi mechanism and cleaved into virus-derived siRNAs that target homologous ssRNA for further degradation. Although DNA viruses do not require dsRNA intermediates, they are known to produce dsRNA during infection (Weber et al., 2006). Recently it was shown that infection of *Drosophila melanogaster* with invertebrate iridescent virus 6 (IIV-6) elicited the RNAi pathway involving *Dcr-2* and *Argonaute-2* (Bronkhorst et al., 2012). The highly uneven genome-wide distribution with clustering of vsRNAs to defined regions designated hotspots in both strands of IIV-6 were also observed with RYNV. Together, our results show that RYNV induces the siRNA antiviral defense pathway in red raspberry, producing an uneven genome-wide distribution of vsRNAs with strong clustering to small defined regions distributed over both strands of the genome. These results increase the complexity of the pararetroviruses and provide evidence of intriguing host–badnavirus interactions in economically important plants.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2013.09.026.

References

Benfey, P.N., Chua, N., 1990. The *cauliflower mosaic virus* 35S promoter: combinational regulation of transcription in plants. Science 250, 959–966.

Blevins, T., Rajeswaran, R., Shivaprasad, P.V., Beknazariants, D., Si-Ammour, A., Park, H., Vazquex, F., Robertson, D., Meins Jr., F., Hohn, T., Pooggin, M.M., 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res. 34, 6233–6246.

Blevins, T., Rajeswaran, R., Aregger, M., Borah, B.K., Schepetilnikov, M., Baerlocher, L., Farinelli, L., Meins Jr., F., Hohn, T., Pooggin, M.M., 2011. Massive production of small RNAs from a non-coding region of *Cauliflower mosaic virus* in plant defense and viral counter-defense. Nucleic Acids Res. 39, 5003–5014.

Bouhida, M., Lockhart, B.E.L., Olszewski, N.E., 1993. An analysis of the complete sequence of a *sugarcane bacilliform virus* genome infectious to banana and rice. J. Gen. Virol. 74, 15–22.

Bronkhorst, A., van Cleef, K.W.R., Vodovar, N., Ince, I.A., Blanc, H., Vlak, J.M., Saleh, M.C., van Rij, R.P., 2012. The DNA virus invertebrate iridescent virus 6 is a target of the Drosophila RNAi machinery. Proc. Natl. Acad. Sci. U.S.A. 109, E3604–E3613.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Harper, G., Hart, D., Moult, S., Hull, R., 2004. Banana streak virus is very diverse in Uganda. Virus Res. 100, 51–56.

Huang, Q., Hartung, J.S., 2001. Cloning and sequence analysis of an infectious clone of *Citrus yellow mosaic virus* that can infect sweet orange via *Agrobacterium*-mediated inoculation. J. Gen. Virol. 82, 2549–2558.

Huntzinger, E., Izaurralde, E., 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. 12, 99–110.

Hwang, Y.T., Kalischuk, M., Fusaro, A., Waterhouse, P.M., Kawchuk, L., 2013. Small RNA sequencing of *Potato leafroll virus* infected plants reveals an additional subgenomic RNA encoding a sequence-specific RNA binding protein. Virology 438, 61–69.

Jacquot, E., Hagen, L.S., Jacquemond, M., Yot, P., 1996. The open reading frame 2 product of *cacao swollen shoot bandavirus* is a nucleic acid-binding protein. Virology 225, 191–195.

Jacquot, E., Keller, M., Yot, P., 1997. A short domain supports a nucleic acid-binding activity in the *rice tungro bacilliform virus* open reading frame 2 product. Virology 239, 352–359.

Jones, A.T., McGavin, W.J., Geering, A.D.W., Lockhart, B.E.L., 2002. Identification of *Rubus yellow net virus* as a distinct badnavirus and its detection by PCR in *Rubus* species and in aphids. Ann. Appl. Biol. 141, 1–10.

Kalischuk, M.L., Kawchuk, L.M., Leggett, F., 2008. First report of *Rubus yellow net virus* on *Rubus idaeus* in Alberta, Canada. Plant Disease 92, 974.

Kawchuk, L.M., Hachey, J., Lynch, D.R., Kulcsar, F., van Rooijen, G., Waterer, D.R., Robertson, A., Kokko, E., Byers, R., Howard, R.J., Fischer, R., Prüfer, D., 2001. Tomato *Ve* disease resistance genes encode cell surface-like receptors. Proc. Natl. Acad. Sci. U.S.A. 98, 6511–6515.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simons, R., 2009. Complete viral genomic sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnostics, discovery and sequencing of viruses. Virology 388, 1–7.

Laco, G.S., Kent, S.B.H., Beachy, R., 1995. Analysis of the proteolytic processing and activation of the *rice tungro bacilliform virus* reverse transcriptase. Virology 208, 207–214.

Leclerc, D., Burri, L., Kajava, A.V., Mougeot, J., Hess, D., Lustig, A., Kleemann, G., Hohn, T., 1998. The open reading frame III product of *cauliflower mosaic virus* forms a tetramer through a N-terminal coiled-coil. J. Biol. Chem. 273, 29015–29021.

Laney, A.G., Hassan, M., Tzanetakin, I.E., 2012. An intergrated badnavirus is prevalent in fig germplasm. Phytopathology 102, 1182–1189.

Lockhart, B.E.L., 1990. Evidence for a double-stranded circular DNA genome in a second group of plant viruses. Phytopathology 80, 127–131.

Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J., Waterhouse, P.M., 2006. The evolution and diversification of Dicers in plants. FEBS Lett. 580, 2442–2450.

Marmey, P., Bothner, B., Jacquot, E., de Kochko, A., Ong, C.A., Yot, P., Siuzdak, G., Beachy, R.N., Fauquet, C.M., 1999. Rice tungro bacilliform virus open reading frame 3 encodes a single 37-kDa coat protein. Virology 253, 319–326.

McGavin, W.J., MacFarlane, S.A., 2010. Sequences similarities between Raspberry leaf mottle virus Raspberry leaf spot virus and the closterovirus Raspberry mottle virus. Ann. Appl. Biol. 156, 439–448.

Medberry, S.L., Lockhart, B.E., Olszewski, N.E., 1990. Properties of *Commelina yellow mottle virus's* complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. Nucleic Acids Res. 18, 5505–5513.

Medberry, S., Olszewski, N.E., 1993. Identification of *cis* elements involved in Commelina yellow mottle virus promoter activity. Plant J. 3 (4), 619–626.

Melcher, U., 2000. The '30K' superfamily of viral movement proteins. J. Gen. Virol. 81, 257–266.

Moissiard, G., Voinnet, O., 2006. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four *Arabidopsis* Dicer-like proteins. Proc. Natl. Acad. Sci. U.S.A. 103, 19593–19598.

Quito-Avila, D.F., Martin, R.R., 2012. Real-time RT-PCR for detection of *Raspberry bushy dwarf virus*, *Raspberry leaf mottle virus* and characterizing synergistic interactions in mixed infections. J. Virol. Methods 179, 38–44.

Ratcliff, F., Harrison, B.D., Baulcombe, D.C., 1997. A similarity between viral defense and gene silencing in plants. Science 276, 1558–1560.

Schmidt, I., Blanc, S., Esperandieu, P., Kuhl, G., Devauchelle, G., Louis, C., Cerutti, M., 1994. Interaction between the aphid transmission factor and virus particles is a part of the molecular mechanism of cauliflower mosaic virus aphid transmission. Proc. Natl. Acad. Sci. U.S.A. 91, 8885–8889.

Sether, D.M., Melzer, M.J., Borth, W.B., Hu, J.S., 2012. Pineapple bacilliform CO virus: diversity, detection, distribution and transmission. Plant Disease 96, 1798–1804.

Stavolone, L., Etienne, H., Leclerc, D., Hohn, T., 2001. Tetramerization is a conserved feature of the virion-associated protein in plant pararetroviruses. J. Virol. 75, 7739–7743.

Su, L., Gae, S., Huang, Y., Ji, C., Wang, D., Ma, Y., Fang, R., Chen, X., 2007. Complete genomic sequence of *Dracaena mottle virus*, a distinct badnavirus. Virus Genes 35, 423–429.

Tanchou, V., Decimo, D., Péchoux, C., Lener, D., Rogemond, V., Berthoux, L., Ottmann, M., Darlix, J.L., 1998. Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid in virus structure and replication. J. Virol. 72, 4442–4447.

Thresh, J.M., Owusu, G.K., 1986. The control of cocoa swollen shoot disease in Ghana: an evaluation of eradication procedures. Crop Prot. 5 (1), 41–52.

Tzafrir, I., Ayala-Navarrete, L., Lockhart, B.E.L., Olszewski, N.E., 1997. The N-terminal portion of the 216-kDa polyprotein of Commelina yellow mottle *badnavirus* is required for virus movement but not for replication. Virology 232, 359–368.

Voinnet, O., 2001. RNA silencing as a plant immune system against viruses. Trends Genet. 17, 449–459.

Waterhouse, P.M., Wang, M.B., Lough, T., 2001. Gene silencing as an adaptive defence against viruses. Nature 411, 834–842.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., Paludan, S.R., 2006. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. 80, 5059–5064.

Xu, D., Mock, R., Kinard, G., 2011. Molecular analysis of the complete genomic sequences of four isolates of Goosebery vein banding associated virus. Virus Genes 43, 130–137.