Host-specific accumulation and temperature effects on the generation of dimeric viral RNA species derived from the S-RNA of members of the Tospovirus genus

André Bertran,1,2† Marina Ciuffo,1† Paolo Margaria,3 Cristina Rosa,3 Renato Oliveira Resende1,2 and Massimo Turina1

1Institute for Sustainable Plant Protection, CNR, Turin, Piemonte, Italy
2Plant Virology Laboratory, Institute of Biological Sciences, University of Brasilia, Brazil
3Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, University Park, PA, USA

Correspondence
Massimo Turina
massimo.turina@ipsp.cnr.it

Received 11 April 2016
Accepted 5 September 2016

INTRODUCTION

Tospovirus, Orthobunyavirus, Hantavirus, Phlebovirus and Nairovirus are negative ssRNA virus genera with tripartite genomes within the family Bunyaviridae (Plyusnin et al., 2011). Tospovirus is the only genus whose members can infect plants and thrips insects, while the other genera infect mammals and arthropod hosts such as mosquitoes, midges, mites and ticks (Plyusnin et al., 2011). Tospoviruses also differ from other bunyaviruses because of their unique ambisense gene expression strategy for both the S-RNA and the M-RNA (de Haan et al., 1989) (Fig. S1, available in the online Supplementary Material).

The Small RNA segment (S-RNA) encodes the nucleocapsid protein (N) (de Haan et al., 1989, 1990), and the NSs protein, which is a non-structural protein associated to suppression of the host-gene-silencing [RNA interference (RNAi)] defence (Takeda et al., 2002) and is necessary for efficient infection of adult insects (Margaria et al., 2014a). The Medium RNA segment (M-RNA) encodes the NSm protein, which is a non-structural protein necessary for cell-to-cell and long-distance movement of the virus in

Polygonum ringspot virus (PolRSV) is a recently characterized Tospovirus reported in Italy. Northern blot analyses of PolRSV infections in Nicotiana benthamiana and tomato plants showed that a viral RNA species with nearly twice the length of the Small genomic RNA (S-RNA) accumulated abundantly in the former host, but was not detected in the latter. Additional assays confirmed that biogenesis of this novel RNA species was common to all PolRSV isolates tested and also to an isolate of Tomato spotted wilt virus (TSWV). Given its size, we hypothesized that the novel RNA species was a dimer molecule and we confirmed this hypothesis by RNA sequencing (RNAseq) analysis and reverse transcription (RT)-PCR of putative predicted dimer junction sites in RNA extracts of N. benthamiana challenged with PolRSV isolates Plg6 and Plg13/2. We also confirmed that these molecules are derived from head-to-tail dimers and often contain deletions at their junction sites. We named these novel molecules imperfect dimer RNAs (IMPD-RNAs). PolRSV IMPD-RNAs systemic accumulation in a range of host plants was restricted to N. benthamiana and Nicotiana occidentalis. Notably, IMPD-RNAs accumulation was modulated by temperature and their generation was restricted to late stages of systemic infection (12 days post-inoculation) in N. benthamiana. Differently from all other PolRSV isolates used in this study, Plg13/2 generated more IMPD-RNAs coupled with low amounts of genomic S-RNA and maintained them even at 18°C, besides having lost the ability to infect tomato plants. This is the first characterization of S-RNA dimers for Tospovirus, and of occurrence of dimers of genomic segments at the whole organism level for Bunyaviridae.

†These authors contributed equally to this work.

The GenBank accession numbers for the genome sequences of PolRSV isolates Plg6 and Plg13/2 are KX468762, KX468763 and KX468764, and KX468765, KX468766 and KX468767 for each genomic segment of each isolate, respectively.

Eight supplementary figures and three supplementary tables are available with the online Supplementary Material.

ã 2016 The Authors

Printed in Great Britain
the plant host (Kormelink et al., 1994), and the glycoprotein precursor that is cleaved by host cell proteases generating Gn and Gc, the mature envelope proteins of tospovirus particles. Gn and Gc are implicated in particle formation (Kikkert et al., 1999; Ribeiro et al., 2008), genome packaging (Ribeiro et al., 2009) and in the capacity to infect thrips (Resende et al., 1991). The Large RNA segment (L-RNA) encodes the RNA-dependent RNA polymerase (RdRp) in the minus-sense orientation (de Haan et al., 1991).

Each of the ORFs belonging to the S- or M-RNA genomic segments is coded in a different strand: the genomic viral RNA strand (the viral RNA present in the mature virus particle) or the antigenic viral RNA strand (the product of the first round of replication by the viral RdRp); this implies that both the viral genomic RNA strand and the antigenic RNA strand (also known as the viral complementary-sense RNA) act as templates for viral sub-genomic mRNA transcription. The ORFs of the non-structural proteins NSs and NSm are coded in the viral genomic strand of S- and M-RNAs, respectively, whereas the ORFs of the N protein and the glycoprotein precursor are coded in the antigenic strand of the S- and M-RNA, respectively (de Haan et al., 1989).

An intergenic region (IGR) rich in A-U is present between the genes in the S- and M-RNAs; IGRs form hairpin-RNA secondary structures associated to transcription termination of the viral mRNAs (Geerts-Dimitriadou et al., 2012). These IGRs and their secondary structure moieties are a feature shared by nearly all the tospovirus species sequenced so far, with the exception of the S-RNA of Polygonum ringspot virus (PolRSV) (Ciuffo et al., 2008; Margaria et al., 2014b).

PolRSV is among the latest tospovirus species officially accepted by ICTV. Although only Polygonum convolvulus and P. dumerilorum were found as natural hosts for PolRSV, its experimental host range is wide and includes solanaceous plants such as tomato and pepper plants (Ciuffo et al., 2008). Nonetheless, the virus may not be a threat to commercial crops because its natural vector Dicythris betae is ecologically restricted to species of the genus Polygonum (Ciuffo et al., 2010).

The full-length genome sequence of PolRSV isolate Plg13 was recently obtained and the presence of a short IGR region in the S segment (2475 nt full-length), not predicted to form a hairpin structure, was confirmed. A short IGR region in the S-RNA segment is a common feature of all the PolRSV isolates sequenced so far (Margaria et al., 2014b). Also recently, the small RNA (sRNA) profiles of Nicotiana benthamiana and Solanum lycopersicum infected with PolRSV have been investigated by a Next Generation Sequencing (NGS) approach revealing that S-RNA-derived sRNAs were more abundant in infected N. benthamiana than in S. lycopersicum, and accumulated preferentially for the ORF of the NSs gene (Margaria et al., 2016).

Defective particles and RNAs have been studied since the late 1970s for the family Bunyaviridae. Typically, defective interfering RNAs (DI-RNAs) of bunyaviruses are derived from the L-RNA and are the product of a single large deletion, where from 30 to 70 % of the L-RNA is lost (Resende et al., 1992; Patel & Elliott, 1992). DI-RNAs are also associated with attenuation of the symptoms of tospovirus infection in plants (Resende et al., 1991). Specifically in the case of the DI-RNAs of tospovirus it has been shown that (i) different plant hosts may influence the generation of DI-RNAs being Nicotiana rustica the most permissive and Emilia sonchifolia the least (Inoue-Nagata et al., 1997); (ii) temperature influences DI-RNA accumulation, being 16 °C more favourable than 22 or 30 °C (Inoue-Nagata et al., 1997); (iii) the presence of DI-RNAs can alter thrips-transmission efficiency (Nagata et al., 2000) and (iv) a dimer of a DI-RNA was found after serial mechanical inoculation of Datura stramonium (Inoue-Nagata et al., 1998).

Canonical DI-RNAs are co-encapsidated with the viral genomic RNAs and propagated through the enveloped virions. Their amount is usually higher than that of the other viral genomic RNAs (Steinbeck et al., 1998). Some exceptions to canonical DI-RNAs for the Bunyaviridae family have been found: uncharacterized RNAs bigger in size than the WT genomic RNAs from which they derived were observed for the S-RNA of Bunyamwera virus (BUNV, genus Orthobunyavirus) (Scallan & Elliott, 1992), and for the L- and S-RNAs of Rift Valley fever virus (RVFV, genus Phlebovirus) (Billecocq et al., 1996) following serial passaging in C6/36 and Vero cells, respectively. The molecular characterization of these RNA species was never performed.

In this paper, molecular characterization of PolRSV through Northern blot analysis revealed the presence of a band with approximately twice the size of the genomic S-RNA in N. benthamiana, but not in tomato plants. Here, we present the first characterization of this unexpected RNA species and the study of its biogenesis. A dimer of the S segment was also found for an Italian isolate of Tomato spotted wilt virus (TSWV) infecting N. benthamiana (TSWV-p105), but not for Iris yellow spot virus (IYSV), which is more closely related to PolRSV. Based on its molecular characterization, we propose for this novel viral RNA species the name IMPD-RNA, and speculate about its origin and regulation.

**RESULTS**

Previously unreported S-RNA related viral RNA species accumulates during PolRSV infection in *N. benthamiana*.

In order to investigate PolRSV gene expression we developed specific RT-PCR primers to amplify, clone and generate riboprobes from fragments of the NSs and the glycoprotein precursor sub-genomic viral mRNAs and for a terminal portion of the L-RNA, based on PolRSV complete nucleotide sequence (Margaria et al., 2014b) (Fig. S1, Table S1). A comparison of the viral S-RNA electrophoretic profiles of PolRSV isolates Plg6 and Plg13/2 inoculated onto N. benthamiana showed the presence of a novel RNA species accumulating in C6/36 and Vero cells, respectively. This unexpected RNA species presented the first characterization of this unexpected RNA species and the study of its biogenesis. A dimer of the S segment was also found for an Italian isolate of Tomato spotted wilt virus (TSWV) infecting N. benthamiana (TSWV-p105), but not for Iris yellow spot virus (IYSV), which is more closely related to PolRSV. Based on its molecular characterization, we propose for this novel viral RNA species the name IMPD-RNA, and speculate about its origin and regulation.
**Fig. 1.** Detection of a novel S-RNA-associated viral RNA species for isolates Plg6 and Plg13/2 of PoIRSV in *N. benthamiana* (a) and their generation at late stages of systemic infection (b). (a) Northern blots were performed with total RNA from systemically infected *N. benthamiana* and tomato leaves (10 dpi) using radioactive riboprobe for the NSs mRNA. (b) IMPD-RNAs are generated at late stages of systemic infection during the first passage of PoIRSV isolate Plg6 from tomato plants (where IMPD-RNAs do not accumulate) to *N. benthamiana*. Samples were taken at two time points after infection for Northern blot: at the beginning of systemic infection (early – 5 dpi) and at a late phase of systemic infection (late – 12 dpi). Plg13/2 was used as positive control for IMPD-RNAs detection. Arrowheads indicate the positions of the viral RNA species of interest. Temperatures are indicated by numerals. Bottom panel, rRNA loading visualized by methylene blue staining. IMPD, imperfect dimer RNAs; g, genomic size S-RNA; sg, sub-genomic size NSs mRNAs; H, healthy *N. benthamiana*; N. bent, *N. benthamiana*; N. benth, *N. benthamiana*.

*Fig. 1*.

**Detection of a novel S-RNA-associated viral RNA species for isolates Plg6 and Plg13/2 of PoIRSV in *N. benthamiana* (a) and their generation at late stages of systemic infection (b). (a) Northern blots were performed with total RNA from systemically infected *N. benthamiana* and tomato leaves (10 dpi) using radioactive riboprobe for the NSs mRNA. (b) IMPD-RNAs are generated at late stages of systemic infection during the first passage of PoIRSV isolate Plg6 from tomato plants (where IMPD-RNAs do not accumulate) to *N. benthamiana*. Samples were taken at two time points after infection for Northern blot: at the beginning of systemic infection (early – 5 dpi) and at a late phase of systemic infection (late – 12 dpi). Plg13/2 was used as positive control for IMPD-RNAs detection. Arrowheads indicate the positions of the viral RNA species of interest. Temperatures are indicated by numerals. Bottom panel, rRNA loading visualized by methylene blue staining. IMPD, imperfect dimer RNAs; g, genomic size S-RNA; sg, sub-genomic size NSs mRNAs; H, healthy *N. benthamiana*; N. bent, *N. benthamiana*; N. benth, *N. benthamiana*.

*Benthamiana* and tomato plants was performed. It showed that isolate Plg6 was able to systemically infect both *N. benthamiana* and tomato plants (Fig. 1a), whereas Plg13/2 did not cause systemic infection in tomato plants (data not shown). The S-RNA profile of isolate Plg6 in tomato plants was as expected with two viral RNA species, the genomic S-RNA and the sub-genomic NSs mRNA. The profile of the S-RNA of the two isolates of PoIRSV in *N. benthamiana* differed from the one of Plg6 in tomato plants by the presence of a third, unexpected, viral RNA species bigger in size than the genomic S-RNA.

Size estimations of the novel S-RNA-associated RNA species based on electrophoretic patterns indicated that this novel RNA molecule was approximately twice the size of the genomic S-RNA, in comparison to an RNA ladder marker and to the electrophoretic migration profile of the genomic M-RNA (data not shown). Both estimates led to the conclusion that the novel RNA species was approximately 4.5–5 kb in size. Taking into consideration the size of the genomic S-RNA (approximately 2.5 kb) and the size range of the novel viral RNA species, our data at this point suggested that the novel viral RNA species could correspond to a dimer-like version of the S-RNA. From here on, this viral RNA species will be referred as IMPD-RNA.

Northern blots analysis for the M- and L-RNA segments using specific riboprobes did not reveal the presence of any other novel RNA species. Putative DI-RNAs derived from the L segment were present in *N. benthamiana* leaves systemically infected by isolates Plg6 and Plg13/2 (Fig. S2).

Additionally, we monitored and compared symptoms of isolates Plg6 and Plg13/2 on the common host *N. benthamiana* at 6 and 14 days post-inoculation (dpi) (Fig. S3), and observed no obvious difference between the two virus isolates in this host in the environmental condition of the experiment.

**Generation of IMPD-RNAs from one single mechanical inoculation occurs after the initial establishment of virus infection**

To further understand when the generation of the IMPD-RNAs took place during the first mechanical passage of PoIRSV to *N. benthamiana*, we inoculated isolate Plg6 using sap from mechanically infected *S. lycopersicum* (therefore, with undetectable presence of IMPD-RNA in the inoculum) onto *N. benthamiana* leaves and assayed systemically infected leaves at the earliest moment of systemic infection (5 dpi) and at a later phase of systemic infection (12 dpi). Only the samples of the late phase of the systemic infection displayed IMPD-RNAs, whereas at the beginning of the systemic infection, IMPD-RNAs could not be detected (Fig. 1b).

**IMPD-RNAs are generated by all PoIRSV isolates tested and their generation is host specific**

We next extended the Northern blot analyses of the S-RNAs to other plant hosts (Fig. S4), and also included other PoIRSV isolates from our collection: two field isolates, Plg9 and Plg14, and isolate Plg3 (described in Methods). Isolates Plg3, Plg6 and Plg13/2 were mechanically inoculated from their original hosts onto a small host range (shown in Table 1, Fig. S4b and listed in Methods) and evaluated for the development of systemic infection and accumulation of IMPD-RNAs both in the mechanically inoculated and systemically infected leaves.

As shown in Fig. 4a, IMPD-RNAs were not present in RNA extracted from leaves of the original host (*P. dumetorum*) systemically infected with the two field isolates, Plg9 and Plg14.

For isolate Plg13/2, the only hosts in which systemic infection developed and IMPD-RNAs accumulated were *N. benthamiana* and *N. occidentalis* (Table 1, Fig S4b). IMPD-
RNAs were also detected in the inoculated leaves of *N. rustica*, *N. tabacum*, *N. clevelandii* and *N. glutinosa*, even though PolRSV infection was not systemic in these hosts (Table 1, Fig. S4b).

Similar to Plg13/2, isolate Plg6 was able to systemically infect *N. benthamiana* and *N. occidentalis*, and additionally *N. clevelandii* and *S. lycopersicum* (Table 1, Fig. S4b). For this isolate, the IMPD-RNA species was detected only in systemically infected *N. benthamiana* leaves (Table 1, Fig. S4c).

Isolate Plg3 behaved similarly to isolate Plg6, in terms of host range and accumulation of IMPD-RNAs in local and systemically infected leaves (Table 1).

**Temperature modulates IMPD-RNA biogenesis and persistence**

It has been shown that low temperatures (i.e. 16°C) favour the accumulation of defective interfering molecules derived from the L-RNA (Inoue-Nagata et al., 1997). For this reason, we decided to test what would be the effect of temperature on IMPD-RNAs biogenesis and accumulation. First, we assayed the presence and accumulation of the IMPD-RNAs in *N. benthamiana* plants mechanically inoculated with isolate Plg3 and kept at either 30 or 18°C. Northern blots showed that Plg3-infected *N. benthamiana* kept at either 30 or 18°C. Northern blots showed that Plg3-infected *N. benthamiana* kept at 30°C accumulated IMPD-RNAs species, whereas plants kept at 18°C did not (Fig. 2a). We also tested if high temperatures could stimulate IMPD-RNA biogenesis in *S. lycopersicum* plants mechanically inoculated for at least two successive passages with isolate Plg3 and kept at 30°C. Our results indicated that even at high temperatures, IMPD-RNA biogenesis was host-specific, with no accumulation of IMPD-RNAs in *S. lycopersicum* (Fig. 2b). It is worth pointing out that the *N. benthamiana* plant extract used to inoculate tomato plants was rich in IMPD-RNAs, and therefore, lack of accumulation in tomato implies that in this plant species the IMPD-RNAs not only cannot be generated, but also it cannot be replicated to detectable levels.

For isolates Plg6 and Plg13/2 we tested IMPD-RNAs accumulation at two different temperatures, 25 and 18°C. Plg6 behaved similarly to Plg3, showing no accumulation of IMPD-RNAs at the lower temperature, whereas Plg13/2 maintained a high accumulation of IMPD-RNAs also at 18°C (although the overall level of virus accumulation was lower than that occurring at 25°C). Furthermore, it is worth mentioning that isolate Plg13/2 had undetectable levels of S segment genomic RNA, yet the NSs sgRNA (subgenomic RNA) was detected (Fig. 2c).

In addition, we tested if high temperatures could induce the formation of IMPD-RNAs in *N. benthamiana* for tospovirus species other than PolRSV, such as TSWV and IYSV. To our surprise, Northern blots showed accumulation of putative IMPD-RNAs for an Italian TSWV isolate (p105) at 25°C (Fig. 3a), but the same did not happen for IYSV, even at an incubation temperature of 30°C and after long exposure of the radioactively probed membranes to X-ray films (Fig. 3b).

**IMPD-RNAs are not covalently closed (circular) viral RNA molecules**

Dimers of viral genomes may arise by the same strategy that allows intense multiplication of viroids or of viruses that replicate by generating concatemers, i.e. templates that are or become covalently closed circular molecules. In order to evaluate if this was the case for the IMPD-RNAs, we performed an RNase R exonuclease assay for isolate Plg13/2 infecting *N. benthamiana* and used isolate Plg6 infecting tomato plants as a negative control for IMPD-RNA formation. As shown in Fig. 4, RNase R digested both Plg13/2 (displaying mostly IMPD-RNAs and sub-genomic viral RNAs) and Plg6 (displaying genomic and sub-genomic viral RNAs) and Plg6 (displaying genomic and sub-genomic viral RNAs).

---

**Table 1. Summary of PolRSV host-range experiments for isolates Plg13/2, Plg6 and Plg3 in relation to the development of systemic infection and IMPD-RNAs accumulation in inoculated and systemically infected leaves**

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Plg13/2</th>
<th>Plg6 and Plg3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systemic inf.</td>
<td>IMPD-RNA local</td>
</tr>
<tr>
<td></td>
<td>Systemic inf.</td>
<td>IMPD-RNA local</td>
</tr>
</tbody>
</table>

+ Positive in the host-range experiment; -, negative in the host-range experiment; NA, data not available; inf., infection; local, mechanically inoculated leaves; syst., systemically infected leaves.

N. benthamiana

IMPD-RNAs in RNA biogenesis was host-specific, with no accumulation of results indicated that even at high temperatures, IMPD-RNAs at the lower temperature, whereas Plg13/2 maintained a high accumulation of IMPD-RNAs also at 18°C (although the overall level of virus accumulation was lower than that occurring at 25°C). Furthermore, it is worth mentioning that isolate Plg13/2 had undetectable levels of S segment genomic RNA, yet the NSs sgRNA (subgenomic RNA) was detected (Fig. 2c).

In addition, we tested if high temperatures could induce the formation of IMPD-RNAs in *N. benthamiana* for tospovirus species other than PolRSV, such as TSWV and IYSV. To our surprise, Northern blots showed accumulation of putative IMPD-RNAs for an Italian TSWV isolate (p105) at 25°C (Fig. 3a), but the same did not happen for IYSV, even at an incubation temperature of 30°C and after long exposure of the radioactively probed membranes to X-ray films (Fig. 3b).

**IMPD-RNAs are not covalently closed (circular) viral RNA molecules**

Dimers of viral genomes may arise by the same strategy that allows intense multiplication of viroids or of viruses that replicate by generating concatemers, i.e. templates that are or become covalently closed circular molecules. In order to evaluate if this was the case for the IMPD-RNAs, we performed an RNase R exonuclease assay for isolate Plg13/2 infecting *N. benthamiana* and used isolate Plg6 infecting tomato plants as a negative control for IMPD-RNA formation. As shown in Fig. 4, RNase R digested both Plg13/2 (displaying mostly IMPD-RNAs and sub-genomic viral RNAs) and Plg6 (displaying genomic and sub-genomic viral RNAs).
IMPD-RNAs are dimers of the S-RNA segment often carrying internal deletions

RNA sequencing (RNAseq) has been shown to be a very good approach to describe the complex of viral populations inside a sample, including the possibility to reveal recombination hotspots (Kutnjak et al., 2015; Routh et al., 2012). For this reason, we performed an RNAseq analysis of rRNA-depleted total RNA extracted from systemically infected \textit{N. benthamiana} leaves inoculated with Plg13/2 and Plg6, kept at 25 °C. Northern blot analysis of the RNA used for RNAseq showed the presence of a full-length S-RNA segment and, in a much lower relative concentration, of IMPD-RNAs for isolate Plg6, while the IMPD-RNAs from Plg13/2 accumulated to higher abundance, both in relation to the full-length S-RNA segment and to isolate Plg6 (Fig. S2).

The tripartite genome sequences of both Plg13/2 and Plg6 isolates were reconstructed \textit{de novo} and some minor differences were observed compared to the full genome of an isolate previously completely sequenced, Plg13 (Margaria et al., 2014b), from which Plg13/2 was derived (Table S2). The L protein of isolates Plg6 and Plg13/2 carries a deletion from amino acid positions 408 to 410 according to the reference sequence deposited in GenBank (accession no. KJ541746.1) (Table S2).

The relative percentage of reads mapping to the host genome/transcriptome or the viral genome showed the same trend in Plg13/2- and Plg6-infected \textit{N. benthamiana}, with the highest percentage of reads aligning against the host genome/transcriptome, followed by the viral genome, and ~11% of unmapped reads (Fig. S5a). Allocation of reads against each of the viral genomic segments was consistent in the two samples, with the lowest percentage of reads aligning against the L-RNA segment, followed by the S- and M-RNA segments, respectively (Fig. S5b).
Prediction of recombinant reads showed that 1.6% of the reads mapping the S-RNA of isolate Plg13/2 were composed of recombinant fragments aligning to distant regions on the segment, while the percentage of recombinant reads mapping the S-RNA of Plg6 was one order of magnitude lower (Fig. S5c, Table 2). Recombinant reads from the M- and L-RNAs were below 0.1% in both isolates (Fig. S5c, Table 2). Given the higher percentage of recombinant reads from the S-RNA segment, and their numerical correlation with IMPD-RNAs abundance, we next analysed in closer detail the nature of this specific subset of reads.

Alignment of the recombinant reads against the corresponding S-RNA reference sequence showed the presence of four major peaks in sample Plg13/2 (S-RNA segment 2476 nt in length) in positions 1–150, 355–500, 1850–1911 and 2300–2476, and two major peaks in sample Plg6 (S-RNA segment 2475 nt in length) in positions 1–150 and 2300–2475 (Fig. S6a, b). On the contrary, in the case of the M- and L-RNAs, recombinant reads were uniformly distributed along the genomic segments (Fig. S6a). To further determine the position of the junction sites, we aligned three subsets of reads (two subsets for Plg13/2 and one from Plg6) against the reference sequence, and we observed the presence of two abundant recombination events in the S-RNA of isolate Plg13/2 (junction 1911-1, termed A and 2476-355, termed B), and one event in isolate Plg6 (junction 2475-1) (Fig. S6b). As a further confirmation, alignment of the same subsets of reads against the corresponding simulated recombinant sequences showed that reads were mapping continuously at the predicted junction site (Fig. S7).

Analysis with the VIREMA software further confirmed the position of the recombination sites on the S segment of both isolates (Table 3). All the recombinant reads from Plg13/2 were associated to the 1911-1 and 2476-355 events, and the recombinant reads from Plg6 were mostly (46%) associated to two head-to-tail recombination events (positions 2474-1 and 2475-1, respectively) with most of the other events with frequency below 1% (Table 3). Recombination events on the M- and L-RNA segments showed low frequencies, confirming the previous analysis shown in Fig. S5c: reads were fairly dispersed in the genome without any noticeable hotspot and no events with a junction including the 3' or 5' end were predicted (Table S3).

Sanger sequencing analysis of RT-PCR fragments obtained using primers across the different predicted junction sites confirmed the RNAseq results (Figs 5 and S8). Specifically, PolRSV isolate Plg13/2 presented two concurrent IMPD-RNA populations showing a deletion of 565 nt for Plg13/2...
A and 354 nt for Plg13/2 B. The four Plg6 clones we sequenced showed no deletion, a single nucleotide deletion or a small rearrangement at the junction sites (Fig. S8).

Isolate Plg3 was not subjected to RNAseq analysis, but RT-PCR across a putative dimer junction showed the presence of a 321 bp deletion in all the clones sequenced (Fig. S8). The same RT-PCR approach across the junction site of the putative TSWV IMPD-RNAs resulted in a collection of clones with small deletions, from 3 to 230 nt, across the junction (Fig. S8).

**DISCUSSION**

**IMPD-RNA characterization and relation to existing models of viral dimer RNA formation**

In this paper, a novel viral RNA species derived from the S-RNA of the tospovirus PolRSV was discovered and characterized from systemically infected *N. benthamiana* plants. The novel RNA species, IMPD-RNA, is the product of a duplication event (dimerization) of the S-RNA; evidence of actual replication of the IMPD-RNA comes from the fact that both strands accumulate abundantly when using strand-specific probes in Northern blot analysis (data not shown).

**Table 3.** Junction-site positions and abundance of unique reads associated to the recombination events on the S-RNA genomic segments of Plg13/2 and Plg6 predicted using ViReMa

<table>
<thead>
<tr>
<th>Segment</th>
<th>Recombination event (nt position)*</th>
<th>No. of unique reads</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_Plg13/2</td>
<td>1911</td>
<td>15 258</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>2476</td>
<td>3 015</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Other events†</td>
<td>569</td>
<td>3</td>
</tr>
<tr>
<td>S_Plg6</td>
<td>2474</td>
<td>2 721</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>2475</td>
<td>966</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>2463</td>
<td>180</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1090</td>
<td>179</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2408</td>
<td>139</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1701</td>
<td>139</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1945</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2475</td>
<td>88</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1251</td>
<td>88</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2475</td>
<td>121</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1690</td>
<td>104</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Other events‡</td>
<td>3 251</td>
<td>40</td>
</tr>
</tbody>
</table>

*Junction position is reported only for recombination events with accumulation percentage above 1%.
†Other events corresponded to 24 recombination events, each of one with percentage below 1%.
‡Other events corresponded to 122 recombination events, each of one with percentage below 1%.
Fig. 5. Schematic view of the IMPD viral RNA populations characterized for PolRSV by RNAseq. Different primer regions used for RT-PCR and Sanger sequencing are indicated by black arrows (also shown in Table S1). (a) IMPD-RNA characterized for isolate Plg6. (b and c) Two different IMPD-RNA populations (named A and B) characterized for isolate Plg13/2. The deleted regions found for each IMPD-RNA population are also indicated in each figure. UTR, viral untranslated region; IGR, S-RNA intergenic region; NSs, non structural protein encoded by the small genomic segment; N, nucleocapsid protein.

Maintenance of PolRSV through mechanical inoculation of N. benthamiana brought to the serendipitous selection of specific isolates containing distinct dimer forms, in some cases containing the full-length head-to-tail dimer of the S genomic segment, and in other cases having discrete deletions, either at the 5′ or at the 3′ ends of one of the two genomic molecules, but always leaving unaltered at least one full-length molecule of the genomic RNA inside the dimer. While dimers or multimers of genomic sequences are obligatory replication intermediates for viroids and some RNA satellites (virusoids) including Hepatitis delta virus (Flores et al., 2011), to our knowledge, their presence during the replication of (+) or (−) stranded RNA viruses has been fully characterized only in the case of Flock house virus (FHV). Interestingly, in the case of FHV, dimers of the viral RNAs were shown during replication in yeast, mammal and insect cells (Albariño et al., 2001).

For members of the Bunyaviridae, accumulation of RNA molecules larger than the corresponding RNA genomic segments has been observed for BUNV and RVFV (Scallan & Elliott, 1992), but their exact molecular nature or biological significance remained unexplored. Here, in the case of PolRSV, we show an unprecedentedly high accumulation of these dimers, which in one specific isolate (Plg13/2) accumulated more than the WT S-RNA genomic RNA; this is in contrast to the FHV dimers for which accumulation is minimal when compared to WT genomic segment. The presence of dimer forms is more common in the case of DI-RNAs for Cucumber necrosis virus (CNV) different DI-RNA dimers were observed concurrently to DI-RNA monomers in an experiment where wild-type transcripts of CNV were co-inoculated with different CNV DI-RNA monomer molecules. The dimer to monomer ratios for the different CNV DI-RNA varied but in one specific case, the DI dimer outcompeted the monomer (Finnen & Rochon, 1993); nevertheless, all the DI dimer forms so far studied are much smaller than their corresponding genomic RNA. The best-studied system of dimer-like molecules of satellite RNA or DI-RNA is arguably Turnip crinkle virus (TCV) and its dimer satellite and DI molecules (Altenbach & Howell, 1981; Cascone et al., 1990; Carpenter et al., 1991; Simon et al., 1988). The molecular mechanism hypothesized to give origin to dimer molecules is described by Carpenter et al. (1991), and relies on the assumption that after synthesis of one (+) strand molecule of RNA from (−) strand RNA template, the replicase complex will initiate a second molecule synthesis switching template without releasing the first molecule, therefore originating a head-to-tail dimer. A careful analysis of dimer junction sequences for TCV dimer satellites revealed that specific deletions at the 3′ end of the first molecule in the dimer influenced TCV infectivity (Simon et al., 1988).
Our system is the first that shows a high dimer/monomer ratio in the case of full-length genomic segments. In the case of both FHV and CNV DIs, available reverse genetic tools demonstrated that dimer sequences are indeed templates for monomer accumulation during replication, raising the interesting hypothesis that they are a necessary common replication intermediate and not simply a dead-end product of aberrant replication (Finnen & Rochon, 1993).

Factors that regulate IMPD-RNA accumulation and a comparison to regulation of DI-RNA accumulation

We have shown that temperature is an important factor regulating IMPD-RNAs biogenesis and that IMPD-RNAs accumulation positively correlates with temperature. Taken together, our results are different from those found for DI-RNAs, for which low temperatures (16°C) were shown to favour accumulation (Inoue-Nagata et al., 1997). The positive effect of high temperatures on IMPD-RNA levels seen for N. benthamiana seems to be effective for this host only, because when tomato plants infected with PolRSV were kept at 30°C, no IMPD-RNAs were generated despite the high temperature setting (Fig. S2b).

Important observations presented in this study about the host specificity observed for PolRSV IMPD-RNA accumulation led us to speculate on the association between a faulty RNAi pathway (as is the case for N. benthamiana, that lacks one of the RdRps, RDR1, involved in RNA silencing) and IMPD-RNAs formation. Current efforts are made to obtain a different N. benthamiana ecotype or transgenic lines that recover the RDR1-analogue function.

N. benthamiana is a PolRSV host where virus titre is always very high compared to some other hosts, and one could be tempted to assume that viral overload results in generation of IMPD-RNAs. On the contrary, it should be noted that IMPD-RNA formation is not always linked to viral overload in plants. It is instead specific to N. benthamiana and N. occidentalis infections only; in fact, Fig. S4b shows that another solanaceous host (D. stramonium) supports high virus accumulation without producing IMPD-RNAs.

It is also of notice that the polymerase sequences of isolates Plg6 and Plg13/2 are not completely identical to the reference sequence for PolRSV. They carry a deletion of three amino acids in the corresponding nucleotide positions 7632–7640 of the L-RNA genomic sequence of isolate Plg13. This deletion, even though not directly associated to any know functional feature of the L protein of tospoviruses, could be related to the formation of IMPD-RNAs for PolRSV.

Northern blot analysis of isolate Plg13/2 raised the possibility that transcription of viral sgRNAs can occur from an IMPD-RNA template; in fact, despite the apparent absence of genomic RNA in many of our blots for this PolRSV isolate, abundant sgRNAs were detected (cf. Fig. 2c), implying that the IMPD-RNAs themselves could be templates for the transcription of sgRNAs.

The host ranges of the PolRSV isolates considered in this work show that isolate Plg13/2 was never able to infect S. lycopersicum and was the only isolate that kept accumulating IMPD-RNAs even at 18°C, suggesting that in the same way as DI-RNAs, IMPD-RNAs may also interfere with viral infection, changing viral host range and inhibiting systemic infection. Nevertheless, we cannot rule out that a specific mutation in other genomic regions has restricted the host range of Plg13/2.

In contrast to isolate Plg6, Plg13/2 had low levels of WT genomic S-RNA, even when kept at 18°C, as shown in Fig. 2c: this specific isolate, seems to have lost the ability to fine tune the dimer/monomer ratio in different experimental conditions and fixed a high IMPD-RNAs accumulation through maintenance of two distinct populations of dimer molecules indistinguishable by size in Northern blots, with single deletions of 565 and 354 nt in different positions in respect to the S sequence: one 3’ centred and the other 5′ centred, respectively. The two populations of Plg13/2 IMPD-RNAs accumulated to different relative frequencies in the recombinant S-RNA population of molecules, with one being significantly more frequent than the other (the one with the biggest deletion), by more than sixfold in total number of copies. This scenario indicates that some sort of selection process, be it size related or sequence content related, governs the dynamics between these two populations of IMPD-RNAs in the same isolate.

We showed that besides PolRSV, an Italian isolate of TSWV (p105 – main TSWV isolate in Northern Italy) infecting N. benthamiana also generated IMPD-RNAs at 25°C (Fig. 3a) and at 30°C (data not shown). The IYSV isolate we worked with, however, even when kept at 30°C, did not generate IMPD-RNAs (Fig. 3b). Even though we were not able to induce IMPD-RNA biogenesis for IYSV, the fact that TSWV produced IMPD-RNAs indicates that IMPD-RNA formation may be a common feature for tospoviruses in specific virus–host combinations or environmental conditions. This hypothesis is corroborated by the occurrence of dimer-like S- and L-RNAs for BUNV and RVFV (Scallan & Elliott, 1992; Langmead & Salzberg, 2012). In fact, it is probable that dimer-like formation of genome-associated viral molecules is a common feature of all Bunyaviridae, depending on abnormal cellular conditions (i.e. stressed cellular environments such as immortalized cell culture lines and heavily infected cells).

IMPD-RNA characterization stimulates re-interpretation of previous results

The discovery of the IMPD-RNAs in PolRSV-infected N. benthamiana gives us a new perspective to interpret recent results on the analysis of viral small-RNA (vsRNA) profiles of PolRSV-infected S. lycopersicum and N. benthamiana (Margaria et al., 2016). In fact, the particularly
high number of vsRNA derived from the S segment found in *N. benthamiana* in comparison to *S. lycopersicum* can now be attributed to the abundant accumulation of IMPD-RNA species in the former host. These results strongly suggest that any sRNA analysis should always be associated with a comprehensive Northern blot analysis that allows the detection of previously unreported RNA species that might be origin and/or target of the silencing machinery.

The bioinformatics approach, Northern blot analyses and RNase R assay, demonstrated that in the case of PolRSV, the abundant quantity of reads mapping across a putative dimer junction is not due to circularization of the genomic segment, but to the presence of a true dimer-like molecule particularly abundant for the S segment. In a recent paper, Li *et al.* (2015) based on sequence read distributions across 3’–5’ junction sites, suggested that *Chuviridae*, a newly characterized clade of multi-segmented, negative-stranded viruses, are the first example of RNA viruses possessing a circular genome. As exemplified by our case study, the same distribution of reads across putative junction sites is indeed due to the presence of dimers, and not circular genomic sequences. In this sense, we think that the proposed circular nature of the genomes in the *Chuviridae* family should be confirmed by an RNase R assay and Northern blot analysis before its acceptance by the scientific community.

**METHODS**

**Source and maintenance of virus isolates and host-range experiments.** PolRSV field isolates from *P. dumerouire* (Plg9 and Plg14 from Piedmont) were collected and mechanically inoculated onto *N. benthamiana* by leaf homogenization in inoculation buffer (50 mM phosphate buffer, pH 7, containing 1 mM Na-EDTA, 5 mM sodium diethyldithiocarbamate and 5 mM sodium thioglycolate). Liquid nitrogen-stored, infected *N. benthamiana* plant material was used as starting inoculum for isolates Plg3 and Plg6, previously described (Ciuffo *et al.*, 2008). Isolate Plg13/2 was derived from mechanically passaged isolate Plg13 described previously (Margaria *et al.*, 2014b, 2016) at 30 °C, and lost the ability to infect *S. lycopersicum*. TSWV isolates p105, p105RBMar and p105-803RB were described previously (Margaria *et al.*, 2007, 2014a). The IYSV isolate (Cip3) used in this work was part of the PLAVIT (Plant Virus of Italy) collection, stored in liquid nitrogen. Temperature experiments were carried out in growth chambers at constant 18, 25 or 30 °C (±2 °C), with a 14 h/10 h light/dark cycle. Isolates Plg6 and Plg13/2 from greenhouse-maintained *N. benthamiana* were used for a host-range experiment at 25 °C that included a small number of solanaceous species, and *Gomphrena globosa*.

**RNA extraction, cloning and sequencing of cDNA fragments.** Total RNA was extracted from infected leaves using Spectrum Total RNA Plant kit (Sigma) according to the manufacturer’s instructions. Amplicons made with specific primers designed based on consensus S segment sequences were used to obtain clones spanning the IGR, and across the putative dimer junction using the combination of oligonucleotides displayed in Table S1. Reverse transcription was carried out with Thermoscript RT-PCR System (Invitrogen) according to the manufacturer’s instructions using random hexamers. The list of primers used to generate cDNA to be used for riboprobe synthesis and junction-site analysis is shown in Table S1. Selected PCR fragments were gel-purified (Qiagen), cloned in pGEM-T Easy Vector (Promega) and sequenced. Briefly, primer pair 661F and 324R was used to obtain the sequence of the IMPD-RNA A of Plg13/2 and primer pair 2123F and 695R was used to obtain the sequence of the IMPD-RNA B of Plg13/2. Primer pair 2123F (or 1930F) and 324R was used to sequence the junction site of the IMPD-RNA of isolate Plg6 and Plg3. Clones spanning the IGR region for Plg3, Plg6 and Plg13/2 were obtained using oligonucleotides Plg-1375F and Plg-1801Rev, and used to confirm the *in silico* assembly data from the RNAsseq experiment.

**Northern blot analysis.** Total RNA from virus-infected plants was separated under denaturing conditions in 1% agarose gels containing glyoxal/DMSO as detailed in Sambrook (2001), using HEPES buffer (20 mM HEPES, 1 mM EDTA, pH 7) instead of sodium phosphate buffer. RNAs from the gel were transferred to nylon membranes (Immobilon-Ny; Merck Biochemicals). Hybridization and detection were performed following protocols previously described (Rossi *et al.*, 2015). Radioactively labelled RNA probes using [32P]dUTP in both orientations were obtained using the Maxiscript SP6/T7 kit (Ambion) from two plasmids (pGEM-T Easy Vector; Promega) linearized with EcoRI containing, respectively, a fragment each of the NSs- and N-coding regions. A fragment of the glycoprotein precursor coding region and another from the 3’ end of the genomic sense of the L segment were used for specific M- and L-RNA detection. Northern blot for IYSV was carried out using a clone corresponding to the full-length N gene (pCip3) obtained with oligonucleotides described previously (Tomassoli *et al.*, 2009). For the NSs gene of TSWV, we used a clone spanning the region between nucleotide positions 490 and 760 of the NSs gene of isolate p105. All primer sequences and positions are shown in Table S1.

**RNase R assay.** In order to check the possible existence of circular forms of viral RNA, we performed RNase R digestion (Epipentric) of total RNA of both Plg6- and Plg13/2-infected *N. benthamiana* immediately after denaturing the RNA for 5 min at 70 °C. We followed the manufacturer’s protocol using a total of 2.5 µg of RNA with or without enzyme, for each sample. Aliquots of each reaction collected 10 and 60 min after adding the enzymes. In order to exclude the possibility of non-specific RNase activity, we repeated the assay reducing the amount of enzyme to one-fifth of what was suggested by the manufacturer based on declared activity.

**RNAsseq of plant samples, genome assembly and bioinformatics analysis.** Total RNA extracted from virus-infected plants (Plg6- and Plg13/2-infected *N. benthamiana*) was assayed by Northern blots, and 10 and 20 µg of RNA from each sample, respectively, were sent to BMR Genomics for construction of the libraries. The Ribozero™ Plant Leaf kit (Epipentric) was used to remove rRNA. The mRNA-depleted total RNA was then used to generate paired-end libraries with the TrueSeq Stranded Total RNA Library Prep kit (Illumina). Sequencing was performed on an Illumina NextSeq 500 platform. Prior to assembly and mapping, adapters were filtered out by Trimmomatic version 0.32 (Bolger *et al.*, 2014) and read quality was checked using the FastQC suite (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The complete genomic sequences of isolates Plg13/2 and Plg6 were assembled using Trinity version 2.06 (Grabherr *et al.*, 2011), and used as reference viral genomes for further analysis. Accession numbers are reported in Table S2. Reads were mapped against the host genome (*N. benthamiana* genome version 0.5) (Fernandez-Pozo *et al.*, 2015) and transcriptome (*N. benthamiana* transcriptome version 5) (Fernandez-Pozo *et al.*, 2015), or against the viral genomes using butter version 0.3.3 (Axtell, 2013), allowing one mismatch. Reads mapping only a single fragment of the corresponding viral genomic segment or recombinant reads (i.e. reads composed of at least two fragments mapping distant regions of the same genomic segment) were identified using Blastn (Johnson *et al.*, 2008) set at an E value of 1e-07, which in preliminary analyses provided optimal stringency. The subset of recombinant reads was extracted using in-house Perl scripts and further analysed by alignment against the corresponding viral genomic segment sequence using Bowtie2 version 2.2.4 (Langmead & Salzberg, 2012). Output BAM
format files were visualized in IGV version 2.3.63 (Robinson et al., 2011; Thorvaldsdóttir et al., 2013), to inspect the distribution of the recombination events along the genomic segments. Depth of coverage values along the genomic segments were exported using IGVTools version 2.3.66 and imaged using Microsoft Excel® v. 10. Robustness of the results for the most abundant recombination events was further assessed by alignment of the reads against simulated recombination reference sequences, spanning the region before and after the predicted recombination event junction (200 nt on each side), to ensure that recombinant reads were covering each predicted recombinant site. Exact position of the recombination events along the genomic segments was further confirmed using ViReMa version 0.6 (Routh et al., 2012; Routh & Johnson, 2014), set at a seed length of 25 and considering as input-only de-duplicated multi-mapping reads. Using this approach, all detected recombination events were described by unique reads, meaning that highly populated recombination events were identified by multiple unique reads. Recombination events were further considered only if supported by at least 10 non-identical reads.

Results from RNAseq analyses were confirmed through RT-PCR using primer pairs designed across the predicted putative dimer junction (Table S1). For isolate Plg3 and TSWV-p105S, we used primer pairs across the putative junction sufficiently distant to include possible deletions (Table S1).

ACKNOWLEDGEMENTS

Particular thanks to Ricardo Lenzi and Caterina Perrone (IPSP-CNR, Italy) for technical assistance. Bioinformatic analysis was performed with Advanced CyberInfrastructure computational resources provided by The Institute for CyberScience at The Pennsylvania State University, University Park, PA, USA (http://ics.psu.edu). The authors also acknowledge Consiglio Nazionale delle Ricerche (CNR) for a grant (Progetto bilaterale Italia-Brasile 2011) for M.T., Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processo 490529/2011 – Chamada Cooperação Internacional Itália – CNPq/CNR) for a grant for A. B. and R. O. R. and Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) for funding (R. O. R.) and a scholarship in the Science without borders programme of the Brazilian government (A. B.) (Processo 99999.001310/2014-03).

REFERENCES


